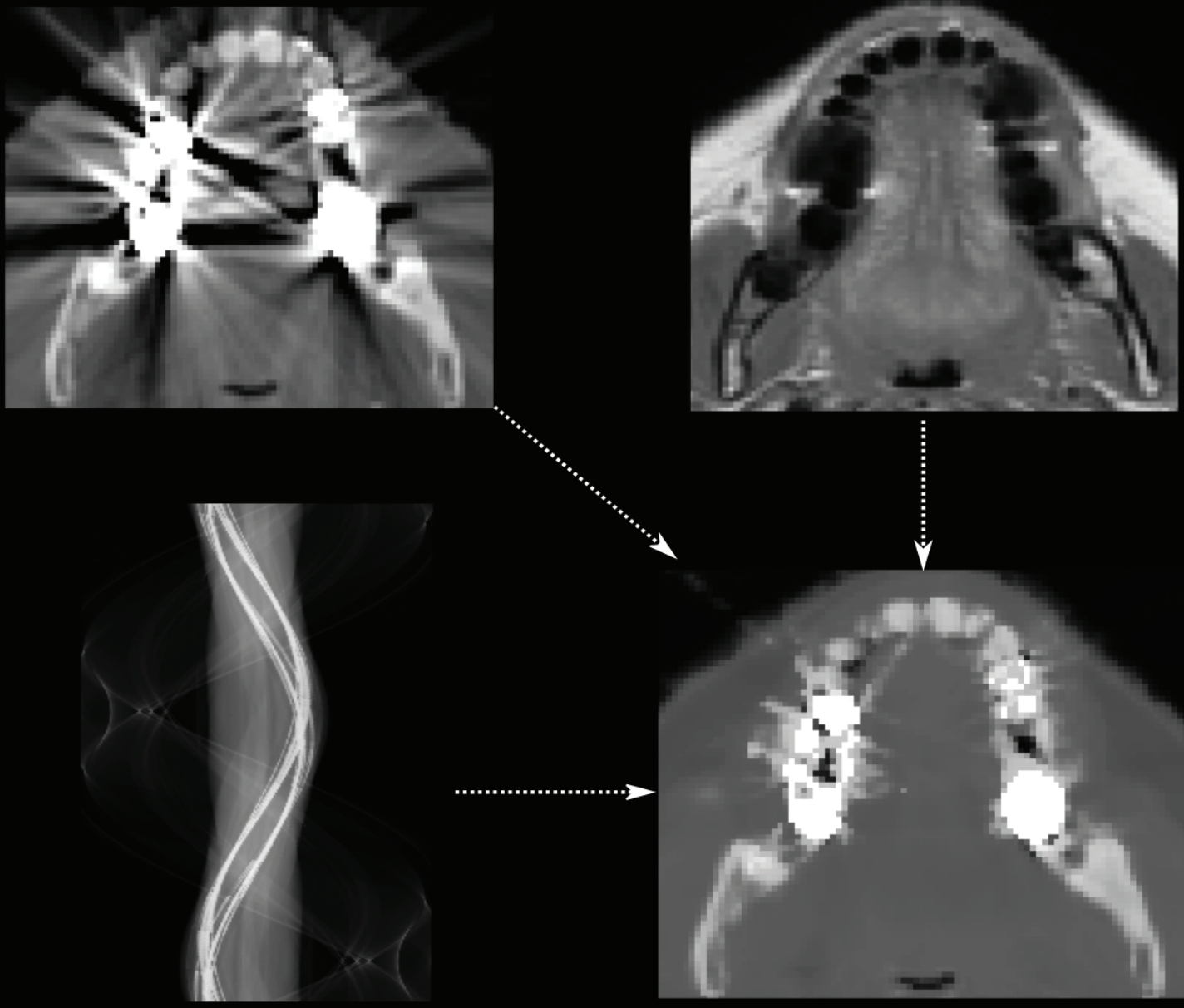
DTU Health Tech
Department of Health Technology

Computed tomography metal artifact reduction for radiotherapy using magnetic resonance imaging

Jonathan Scharff Nielsen, PhD Thesis May 2019



TECHNICAL UNIVERSITY OF DENMARK

Computed tomography metal artifact reduction for radiotherapy using magnetic resonance imaging

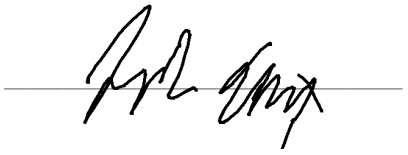
by

Jonathan Scharff Nielsen
Email: jonathanscharff2@gmail.com
Supervised by: *Koen Van Leemput* (PhD, Asc. Prof.) and
Jens Edmund (PhD, Hospital Physicist

Technical University of Denmark
Department of Applied Mathematics and Computer Science
Image analysis and computer graphics section
Richard Petersens Plads, building 324
2800 Kongens Lyngby, Denmark
Phone +45 4525 3031
compute@compute.dtu.dk

Preface

This thesis was prepared at the Department of Applied Mathematics and Computer Science at the Technical University of Denmark (DTU) in fulfillment of the requirements for acquiring a PhD degree. The thesis was prepared with funding from DTU Compute, Herlev and Gentofte hospital. It was co-supervised by PhD, DABR Jens M. Edmund and associate professor Koen Van Leemput. The thesis deals with the reduction of Computed Tomography metal artifacts using coacquired Magnetic Resonance images for improved radiotherapy treatment planning.



Jonathan Scharff Nielsen, 05/05/2019.

Abstract

English: When metal implants are scanned during medical x-ray computed tomography (CT), the conventional model that relates the image coefficients to the x-ray measurements (the sinogram) breaks down. The resulting streak and cupping artifacts affect the accuracy of external beam radiotherapy (RT) treatment plans that rely on CT images for electron density and particle relative stopping power estimates. Particularly in the highly sensitive proton RT this may lead to notable errors the estimated beam penetration depth, and consequent dosimetric errors that may exclude some patients from proton treatment that would otherwise benefit from it. Metal artifact reduction (MAR) algorithms are therefore employed that however despite many years of development remain inadequate. This thesis investigates a promising way to increase the efficacy of MAR by using the superior anatomical information in artifact corrupted regions that is available in a coregistered Magnetic Resonance Image (MRI) that in RT may have been acquired for tumor delineation. This is challenging due to the inherently weak correlation between conventional-sequence MRI intensities and CT values, as well as the potentially imperfect coregistration of the images. We develop a novel, Bayesian generative model of MRIs and CTs that is suitable to handle these challenges, and use it with promising results for MAR by image inpainting in the corrupted regions; as a prior for sinogram inpainting of metal-affected projections; and for prior modelling in model-based iterative reconstruction (MBIR). We finally validate our methods in the context of head-and-neck RT, finding dosimetric as well as image space improvements compared to a standard MAR algorithm in clinical use.

Dansk: Når metal-implantater skannes med røntgen-baseret "computed tomography" (CT), brydes modellen der relaterer billedekoefficienterne til røntgenmålingerne (sinogrammet). Resultatet er intense striber og systematiske variationer i CT-værdierne som påvirker nøjagtigheden af ekstern stråleterapi, som bruger CT værdierne til at estimere elektrondensiteter samt "stopping power" versus partikler. I specielt den meget sensitive protonterapi kan dette føre til anseelige fejl i parametre relateret til penetrationsdybden af strålerne, og deraf afledte fejl i dosisbestemmelsen som kan ekskludere patienter fra protonterapi some ellers kunne have haft nytte af det. Derfor anvendes Metal Artefakt Reduktions (MAR) algoritmer, som dog trods mange års udvikling ikke er tilstrækkelige. Denne afhandling undersøger en lovende metode til at højne effektiviteten af MAR ved at anvende den tilgængelige anatomiske information i de korrumperede regioner der findes i et Magnetisk Resonans (MR) billede for den samme patient. Dette er et udfordrende problem på grund af den svage korrelation imellem MR intensiteter og CT værdier samt den potentielt uperfekte coregistrering af billederne. Vi udvikler en ny, Bayesians generative model af MR og CT billeder som håndterer disse problemer, og bruger den med lovende resultater til MAR ved direkte estimering af CT-værdier i de korrumperede regioner; som a priori information i en sinogram-behandlingsalgoritme; og som en a priori statistisk model for billedet i model-baseret iterativ rekonstruktion (MBIR). Vi validerer vores metoder i hoved/hals RT og finder dosimetriske samt billedekvalitetsforbedringer sammenlignet med en standard algoritme i klinisk brug.

Acknowledgements

First of all, I would like to thank both of my supervisors, Jens M. Edmund and Koen Van Leemput for their guidance and company throughout the years. It has been a pleasure working with you.
Also thanks to the radiotherapy research unit at Herlev Hospital for finding much of the funding that made this project a reality and the wonderful, yearly "Øresund" meetings.
Many thanks to professor Johan Nuyts at Leuven University Hospital for valuable discussions with him and his (then) PhD student Tao Sun, and for accommodating me during my external stays in his medical imaging research group.
Thank you to Laura Ann Rechner at Righshospitalet for taking the time to teach me about proton radiotherapy treatment planning.
Thanks to all my colleagues in the image analysis and computer graphics section at DTU and in the radiotherapy research unit for the occasional internal review of my work and the consistently good company.
Finally, I would like to thank my friends and family for their support and sense of perspective, especially in stressful times.

Contents

Abbreviations

Acronym	What (it) Stands For
CT	**C**omputed **T**omography
MR	**M**agnetic **R**esonance
MRI	**MR i**mage
NEC	**N**oise **E**quivalent **C**ount
RT	**R**adio **t**herapy
FBP	**F**iltered **B**ack **P**rojection
MAR	**M**etal **A**rtifact **R**eduction
nMAR	**N**ormalized **MAR**
MBIR	**M**odel **B**ased **I**terative **R**econstruction
MLTR	**M**aximum **L**ikelihood **T**ransmission **Reconstruction**
IMPACT	**I**terative **M**ethod for **P**olychrom **a**tic **CT**
KDE	**K**ernel **D**ensity **E**stimation
pCT	**P**seudo **CT**
kerMAR	**KER**nel regression **MAR**
oMAR	**MAR** for **o**rthopedic implants
WET	**W**ater **E**quivalent **T**hickness

Symbols

Symbol	Description
$\mathcal{T}$	Set of voxels in the patient volume
$\mathcal{T}_u$	Set of uncorrupted voxels ($\mathcal{T}_u \subseteq \mathcal{T}$)
$\mathcal{T}_c$	Set of corrupted voxels ($\mathcal{T}_c \subseteq \mathcal{T}$)
$\mathcal{S}$	Set of detector indices in the sinogram
n_j	The x-ray intensity measurement at detector $j \in \mathcal{S}$
Γ_j	The emitted x-ray intensity toward detector j
λ_j	The simulated x-ray intensity measurement given a CT image at detector j
C	The estimated Noise Equivalent Count (NEC) scaling coefficient.
$\mathcal{N}\{x\|\mu, \sigma^2\}$	A Gaussian (normal) distribution with mean μ and variance σ^2
$\mathcal{N}\{\mathbf{x}\|\boldsymbol{\mu}, \boldsymbol{\Sigma}\}$	A multivariate Gaussian (normal) distribution with mean $\boldsymbol{\mu}$ and covariance matrix$\boldsymbol{\Sigma}$
$\mathbf{x}_i$	The position of voxel $i \in \mathcal{T}$ (mm)
y_i	True CT value in voxel i
t_i	Observed (possibly corrupted) CT value in voxel i
$\mathbf{m}_i$	The M-dimensional, cuboidal MRI patch centered on voxel i
$\mathcal{A}_i$	Indices to the set of regression points ($\{y_n, \mathbf{m}_n\}_{n \in \mathcal{A}_i}$)
β_m	The MRI kernel precision
β_y	The CT kernel precision
β_t^*	The minimum artifact noise precision in the image (achieved near the metal implants)
$f(\mathbf{x}_i)$	Sigmoidal function to scale the artifact noise precision with the distance to the metal
β_t	The artifact noise precision in voxel i ($\beta_t^{-1} = f(\mathbf{x}_i)\beta_t^{*-1}$)
$\boldsymbol{\beta}$	The set of hyperparameters $\boldsymbol{\beta} = \{\beta_t^*, \beta_y, \beta_m\}$
κ	A scale parameter in $f(\mathbf{x}_i)$

Chapter 1

Introduction

1.1 Computed Tomography metal artifacts in radiotherapy

The attenuation coefficients reconstructed by medical x-ray Computed Tomography (CT) are closely related to the electron density distribution and particle stopping power of the scanned tissue [1–3]. External-beam radiotherapy (RT) planners therefore use them for patient-specific simulation of the deposited dose under radiation by photons, electrons or protons. For this process to provide accurate dose distributions, the quality of the CT is of critical importance[4–6].

Unfortunately, many cancer patients, in particular head-and-neck cases, may contain metal hardware near the tumor in the form of dental inserts, crowns or implants. This is a problem for the CT image quality due to inaccuracies in the image reconstruction model used by the ubiqitous Filtered Back Projection (FBP) algorithm. This model breaks down through various physical effects for measurements through the highly dense metal, to a degree unsalvagable by ordinary calibration methods; this leads to characteristic streak and cupping artifacts such as those in the three head-and-neck FBPs in the top row of fig. 1.1[7].

1.2 Reduction of metal artifacts

Metal artifact reduction (MAR) algorithms are therefore an important part of error-management in RT. Since the metal artifacts arise from a variety of sources, no algorithm has however yet solved the problem in general applications, despite at least 40 years of development. This holds

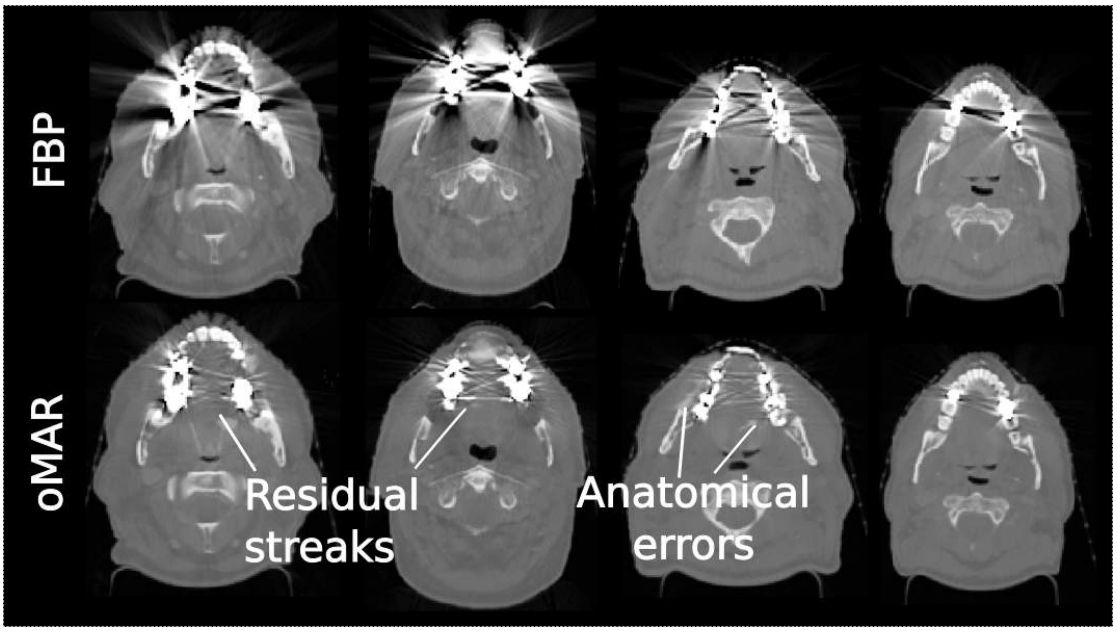


FIGURE 1.1: FBP (Filtered Back Projection) CTs and oMAR (MAR for orthopedic implants) artifact reduced CTs of four head-and-neck patients.

true in particular in the case of RT[8]; fig. 1.1 shows exemplary images using a state-of-the-art algorithm in clinical use, the "MAR for orthopedic implants" (*oMAR*) algorithm (Philips Healthcare). While the artifacts left behind by oMAR may seem small, they may have an impact on the dose distribution calculations, in particular in sensitive proton RT treatments. As a consequence, the artifacts must be planned around when designing treatment plans, decreasing the degrees of freedom available to the planner and potentially disqualifying patients from proton RT who would otherwise benefit from it.

The driving principle behind oMAR[9] is to override artifacts in the CT using a combination of segmentation and raw data processing, which generates a *prior image* for x-ray measurement simulation. The residual streaks in fig. 1.1 are owed to difficulties in generating this prior image due to severe image corruption. This difficulty is not peculiar to oMAR, as a similar use of prior image modelling is a common feature in many MAR algorithms: *Sinogram inpainting* algorithms use prior images to calculate estimates of the erroneous x-ray measurements through the metal, while the more complex model-based iterative reconstruction (*MBIR*) algorithms that improve the image reconstruction itself to account for the artifact sources may include image prior distributions. In the absence of more accurate prior information, such prior distributions may be simple functionals that impose regularities in the images.

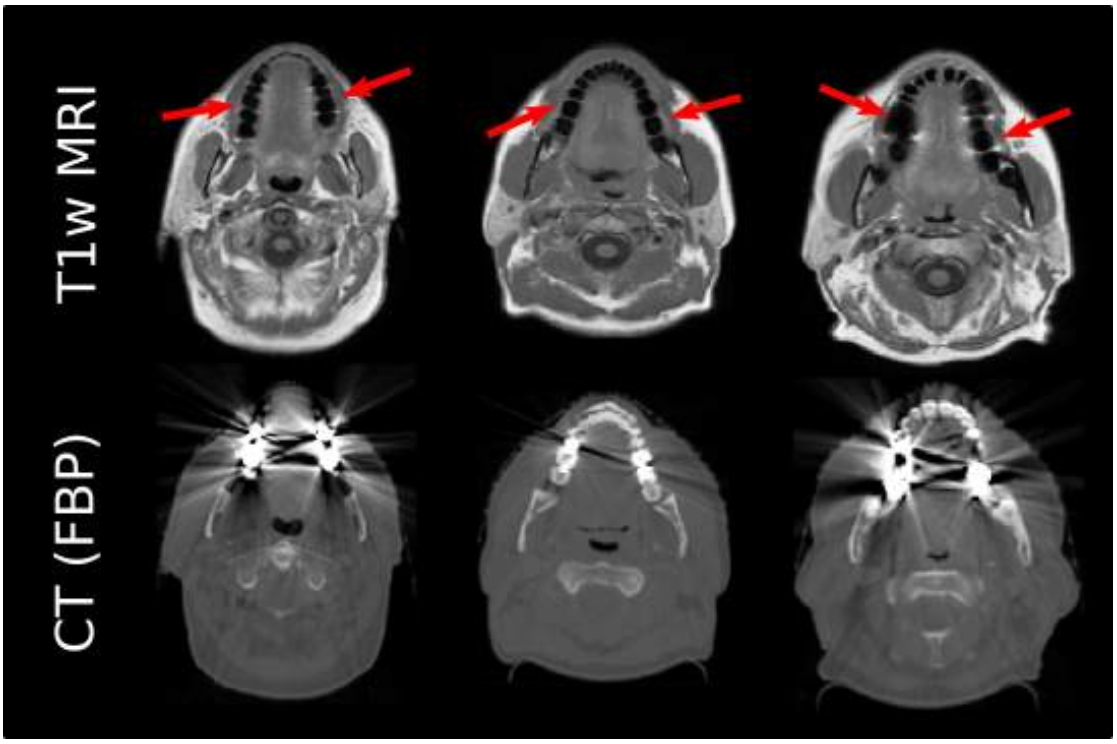


FIGURE 1.2: Axial FBP CT slices versus MRIs acquired with a T1-weighted sequence. The CT artifacts are far more extensive than the subtle signal voids in the MRIs near the implants.

1.3 MRI-based prior modelling for MAR

In the context of RT, a promising source of image information is available that may aid the prior modelling task, since it is a common practice to acquire an MRI in the same configuration as the CT for organ and tumor delineation, due to its superior soft-tissue contrast. Since, as shown in fig. 1.2, the metal artifacts in MR images may be far more localized than the streak and cupping artifacts in the CT[1], this provides a valuable source of anatomical information in corrupted regions[10].

The most obvious way to now use this MRI information for MAR is to insert the observed MRI intensities in a functional map to CT values. There are two fundamental challenges with this: first, there is the potentially imperfect coregistration of the CT and MRI, which directly limits the accuracy of any voxel-based prediction. Second, plotting the CT values for each voxel in a head-and-neck RT patient versus their corresponding T1w MRI intensities (fig. 1.3(a)) shows that such a map is not easily established; the observation of especially a low MRI intensity may be associated with a range of CT values. This is due to an inherent difference in the physical quantities that lead to the CT and MRI contrasts[11–13], and the correlation between their values is therefore weak. An extreme example of this occurs in bone and air regions, which are

[1] In the experience of the author, this tends to be the case at least in the head and neck, which is more in focus in this thesis; with larger implants, such as dual hip implants, the MR signal voids may be more of a concern.

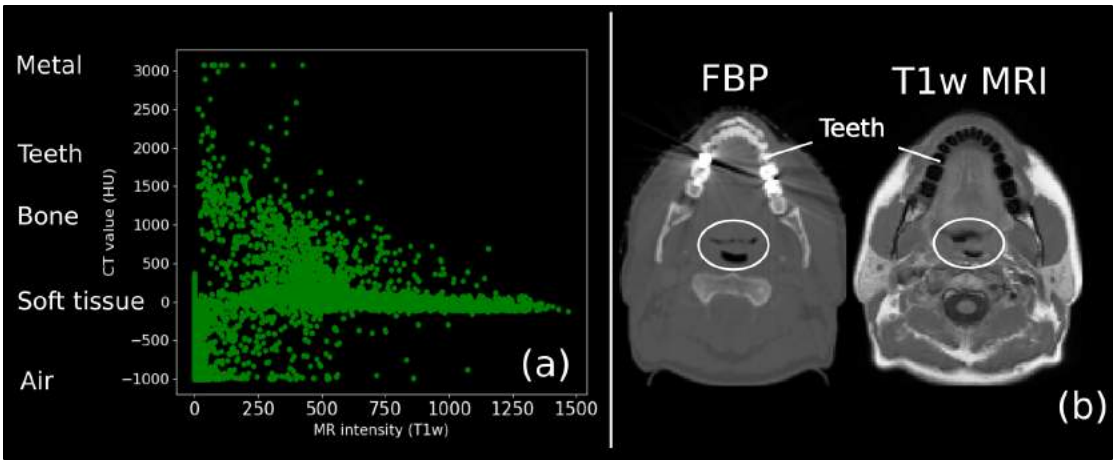


FIGURE 1.3: (a): A scatter plot of single voxel CT values and MRI intensities from a coregistered pair of images. While large MR intensities are good predictors of soft tissue CT values, lower intensities have little predictive power due to a contrast disparity. (b): The contrast disparity is especially clear between bone and air, which have opposite contrasts with CT and MR. The ring additionally shows anatomical differences between the MR and CT acquisitions that also leads to uncertainty.

indistinguishable in the MRI but have excellent CT contrast, as shown in fig. 1.3(b). In such regions, the MRI is therefore a poor CT value predictor.[10–12, 14].

1.4 Goals of the project

The main focus of this project was to develop succesful methods to perform MRI-based CT metal artifact reduction, in particular using probabilistic (Bayesian) generative modelling combined with CT reconstruction from raw CT data. Additionally, the goal was to benchmark our methods in the setting of external beam RT to the clinical state-of-the-art, especially in the head and neck where metal implants occur frequenctly and lead to artifacts in small, complex regions containing both bone and air.

1.5 Scientific contributions

Papers included in this thesis:

Paper A: Jonathan Scharff Nielsen, Jens M. Edmund and Koen Van Leemput: "CT metal artifact reduction using MR image patches", in proceedings of SPIE. SPIE - International Society for Optical Engineering, 2018, vol. 10573. doi:10.1117/12.2293815.

Paper B: Jonathan Scharff Nielsen, Jens M. Edmund and Koen Van Leemput: "MRI-based CT metal artifact reduction using Bayesian modelling", *Unpublished*

Paper C: Jonathan Scharff Nielsen, Koen Van Leemput and Jens M. Edmund: "MR based CT

metal artifact reduction for head-and-neck photon, electron and proton radiotherapy", *Under journal review.*

As of the date of thesis submission (31/01/19), paper A has been peer-reviewed, while paper B, which is a more mature journal manuscript expanding on the model we introduced in A, is in the final stages of writing. Paper C is undergoing review at the International Journal of Medical Physics Research and Practice.

Peer-reviewed conference contributions (not included in this thesis):

Oral presentation: Jonathan Scharff Nielsen, Koen Van Leemput and Jens M. Edmund: "CT metal artifact reduction by means of MRI", Biology-guided Adaptive Radiotherapy (BigART) conference, Århus 2017.
Oral presentation: Jonathan Scharff Nielsen, Jens M. Edmund and Koen Van Leemput: "CT metal artifact reduction using MR image patches", International Society for Optical Engineering (the medical imaging conference), Houston 2018).
Poster: Jonathan Scharff Nielsen, Koen Van Leemput and Jens M. Edmund: "Impact on proton range estimates of a novel magnetic resonance based metal artifact reduction algorithm", European Congress for Medical Physics (ECMP), Copenhagen 2018.

The second oral presentation at SPIE covered the contents of paper A. While the other two contributions include experimental methods and results that we do not cover in this thesis, the presented results constituted initial experiments that led up to the more refined final investigations in the included papers. We therefore chose to not reference them in this thesis.

1.6 Thesis overview

This thesis is structured as follows:

- Chapter 2 provides a basic overview of CT and the sources of the metal artifacts. It also considers reconstruction in practice from the medical CT raw data acquired in this project. The chapter contains simulated experiments that highlight the effects of the different artifact sources as well as examples from head-and-neck RT patients.
- Chapter 3 is an introduction to MAR algorithms, with focus on the methods that we include in the experiments of papers A-C.
- Chapter 4 is an overview of the material in papers A and B. It in particular develops our MRI-based generative model and our experiments with applying it to artifact reduction

by image inpainting, as a prior image for sinogram inpainting and for prior modelling in MBIR.

- Chapter 5 is an overview of the material in paper C, showing the potential improvements to the clinical state of the art in RT dose planning of using our MRI-based predictive model. Before this, it provides some RT background theory to aid in interpreting our results.
- Chapter 6 is a final discussion that includes suggestions for future work.

Chapter 2

Computed tomography basics

CT metal artifacts arise from multiple physical sources that may only be understood with some basic knowledge of the image acquisition process. It is also a non-trivial engineering problem to accurately perform CT reconstruction from raw, medical CT data. In this chapter, we therefore provide an overview of the basics in CT reconstruction, how to reconstruct from real medical data, and finally arrive at the origins of the artifacts.

2.1 The CT problem and its usual solution

The goal of CT is to obtain a quantitative map of a patient's insides, which we describe by the position dependent function $y(\mathbf{x})$. In x-ray CT, this is achieved by acquiring a set of x-ray measurements from a set of x-ray source positions around the patient, indexed by $j \in \mathcal{S}$. The x-rays traverse the image in a set of *beam lines* that we define as curves in 3D space denoted by $\{\mathcal{L}_j\}_{j\in\mathcal{S}}$; given the physics of x-ray attenuation, the intensities $\{n_j\}_{j\in\mathcal{S}}$ measured on the other side of the patient, stacked in the vector $\mathbf{n}$, may then be functionally related to $y(\mathbf{x})$. This setup is illustrated for a certain scanner geometry in fig 2.1, the *first generation scanner*. Here, the x-ray source is rotated to a set of N_v *view* angles, at each of which it performs N_d measurements while moving to positions ψ along a line orthogonal to a beamline through the isocenter of the rotation[7].

Given the geometry, the image distribution $y(\mathbf{x})$ is related to the measurements $\mathbf{n}$ by a forward model. The problem of CT is to invert an appropriate formulation of this model, thus inferring $y(\mathbf{x})$ from $\mathbf{n}$.

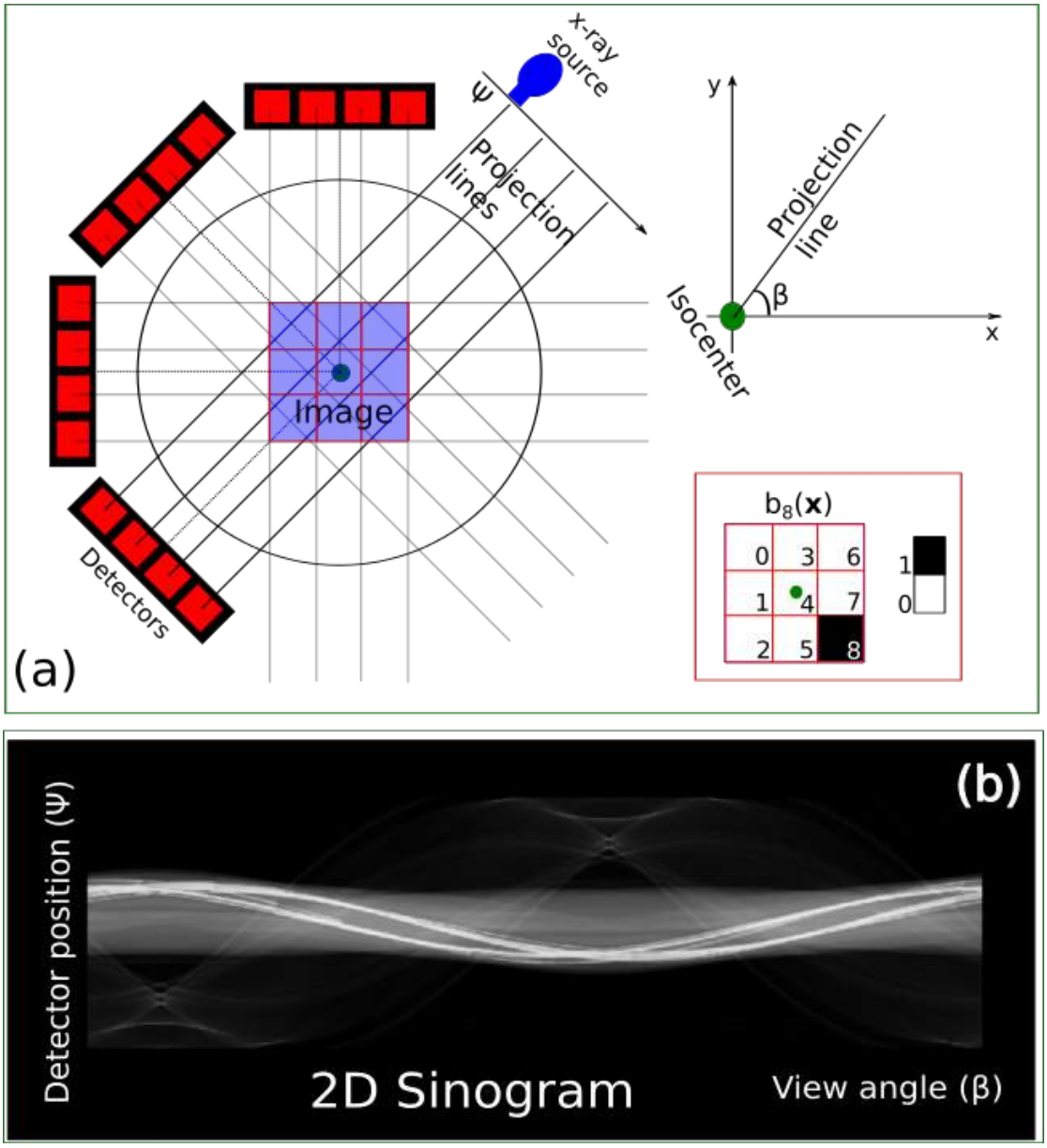


FIGURE 2.1: (a) The first generation scanner geometry: the x-ray source and linear detector array are stationed on each side of the patient and rotated to view angles β from 0 to at least π (half a full rotation). At each position, the source is translated to positions ψ perpendicular to a beamline through the isocenter. Shown is also the definition of one of the voxel basis functions used to discretize an image, a function that is non-zero only in a specific cuboidal region. (b) A 2D *sinogram* acquired with the first generation scanner, which shows the measured projections along the y-axis, varying the view angle along the x-axis. The contrast in each pixel quantifies the measured projection at the corresponding beamline, a brighter value signifying more attenuating material along the beamline.

2.1.1 The Lambert-Beer law

Conventionally, the *Lambert-Beer*[7, 14] law of mono-energetic x-ray attenuation is used to define the CT forward model:

$$n_j = \Gamma_j \mathsf{e}^{-\int_{\mathcal{L}_j} y(\mathbf{x}) d\mathbf{x}}, \tag{2.1}$$

where Γ_j is the x-ray intensity projected towards detector j. $y(\mathbf{x})$ is here identified as the *attenuation coefficient*, the probability per unit travelled length of a photon being attenuated in the material at $\mathbf{x}$. The *projection* in the exponent is a line integral through the image, and is directly related to the measurements:

$$p_j = -\log\left(\frac{n_j}{\Gamma_j}\right) = \int_{\mathcal{L}_j} y(\mathbf{x}) d\mathbf{x}. \tag{2.2}$$

Eqn. (2.2) is a linear relationship between the image and the projections, which is easier to handle than the non-linear eqn. (2.1). This linear model is therefore generally used for CT.

2.1.2 Discrete CT and the System Matrix

To perform the CT reconstruction, we must first discretize $y(\mathbf{x})$. We therefore cover the volume with $|\mathcal{T}|$ cuboidal *voxels* centered at locations $\{\mathbf{x}_j\}_{j\in\mathcal{T}}$, where $\mathcal{T}$ contains the set of voxel indices and $|\cdot|$ denotes the number of elements in the set. Letting the voxels have halved side-lengths of $\mathbf{d} = \{d_x/2, d_y/2, d_z/2\}$, we write the image as a sum of local basis functions as follows[7, 15]:

$$y(\mathbf{x}) = \sum_{i\in\mathcal{T}} y_i b_i(\mathbf{x}) \quad \text{where} \quad b_i(\mathbf{x}) = \begin{cases} 1 & |\mathbf{x} - \mathbf{x}_i| \preceq \mathbf{d} \\ 0 & \text{otherwise} \end{cases}$$

Here, $|\mathbf{x}|$ denotes the elementwise absolute value of the vector $\mathbf{x}$ while $\preceq$ is the element-wise joint "less-than-or-equal-to" operator. An exemplary basis function $b_8(\mathbf{x})$ corresponding to voxel 8 in a 3x3 image is illustrated in the lower right corner of fig. 2.1. It is 1 only in the support of voxel 8, 0 elsewhere. We insert this definition in eqn. (2.2):

$$p_j = \sum_{i\in\mathcal{T}} y_i l_{j,i} \quad \text{where} \quad l_{j,i} = \int_{\mathcal{L}_j} b_i(\mathbf{x}) d\mathbf{x}, \tag{2.3}$$

where we defined the *system matrix*[7, 15] $\mathbf{L}$ with entries $l_{j,i}$. This matrix depends on the scanner geometry and defines which voxels are traversed by which beam lines, as it corresponds to line integrals through the basis functions. The entry $l_{j,i}$ is thus non-zero only if voxel i is traversed by line j.

We stack the coefficients of the basis function expansion, $\{y_i\}_{i\in\mathcal{T}}$, in the vector $\mathbf{y}$, which provides our discrete representation of the image. Also stacking the projections in vector $\mathbf{p}$ and allowing

for random projection noise $\boldsymbol{\epsilon}$, the CT forward model becomes a matrix equation:

$$\mathbf{p} = \mathbf{L}\mathbf{y} + \boldsymbol{\epsilon}. \tag{2.4}$$

The vector $\mathbf{p}$ may be visualized as a 2D *sinogram* with the N_v view angles β along one axis and the N_d detector positions along the other (fig. 2.1(b)). The *forward projection* operation $\mathbf{L}[\cdot]$ that multiplies by the system matrix calculates a set of projections, and thus a sinogram, from an image.

The inverse problem of estimating $\mathbf{y}$ from $\mathbf{p}$ is overdetermined: with a modern CT scanner, i.e. the one that provided the CT material for this project, a single image contains $|\mathcal{T}| = 512 \times 512 \approx 3 \cdot 10^5$ voxels reconstructed from $|\mathcal{S}| = 816 \times 2640 \approx 2 \cdot 10^6$ measurements, and so the number of data points is an order of magnitude larger than the number of parameters.

2.1.3 The pseudo-inverse and Filtered Back Projection

A standard way to solve such overdetermined, linear problems is to find the solution that minimizes the squared error:

$$\mathbf{y}^* = \arg\min_{\mathbf{y}} ||\mathbf{p} - \mathbf{L}\mathbf{y}||^2, \tag{2.5}$$

where $||\mathbf{x}|| = \sqrt{\sum_k |x_k|^2}$ denotes the l^2 vector norm. Applying the Moore-Penrose pseudo-inverse[16, 17] $\mathbf{L}^+$ solves this minimization problem, and:

$$\mathbf{y}^* = \mathbf{L}^+\mathbf{p} \quad \text{with} \quad \mathbf{L}^+ \equiv (\mathbf{L}^T\mathbf{L})^{-1}\mathbf{L}^T, \tag{2.6}$$

where $\mathbf{y}^*$ is the least-squares solution. CT may thus in principle be performed by simply calculating $\mathbf{L}^+$ and applying it to $\mathbf{p}$, which however is complicated by the fact that $\mathbf{L}$ is huge: it contains $|\mathcal{T}| \times |\mathcal{S}| \approx 10^{11}$ entries and must instead be used through implicit calculation of its operations. While, using e.g. raytracing[18–20], this may be done with relative ease for both the *forward projection* $\mathbf{L}$ and the *back projection* $\mathbf{L}^T$, the remainder of $\mathbf{L}^+$, $(\mathbf{L}^T\mathbf{L})^{-1}$, is less straightforward. It however has a simple interpretation that lets it be applied implicitly when the scanner geometry follows the mentioned first generation geometry, and the forward model becomes the *Radon transform*; in this case, the fast *Filtered Back Projection* (FBP) algorithm may be applied. FBP may be understood by considering the following form of the pseudo-inverse:

$$\mathbf{L}^+ = (\mathbf{L}^T\mathbf{L})^{-1}\mathbf{L}^T = (\mathbf{L}^T\mathbf{L})^{-1}\mathbf{L}^T(\mathbf{L}\mathbf{L}^T)(\mathbf{L}\mathbf{L}^T)^{-1} = \mathbf{L}^T(\mathbf{L}\mathbf{L}^T)^{-1}, \tag{2.7}$$

where we used the following universal relationship: $\mathbf{L}^T(\mathbf{L}\mathbf{L}^T) = (\mathbf{L}^T\mathbf{L})\mathbf{L}^T$. In the first generation scanner geometry, the matrix $(\mathbf{L}\mathbf{L}^T)^{-1}$ becomes a linear frequency filter applied along

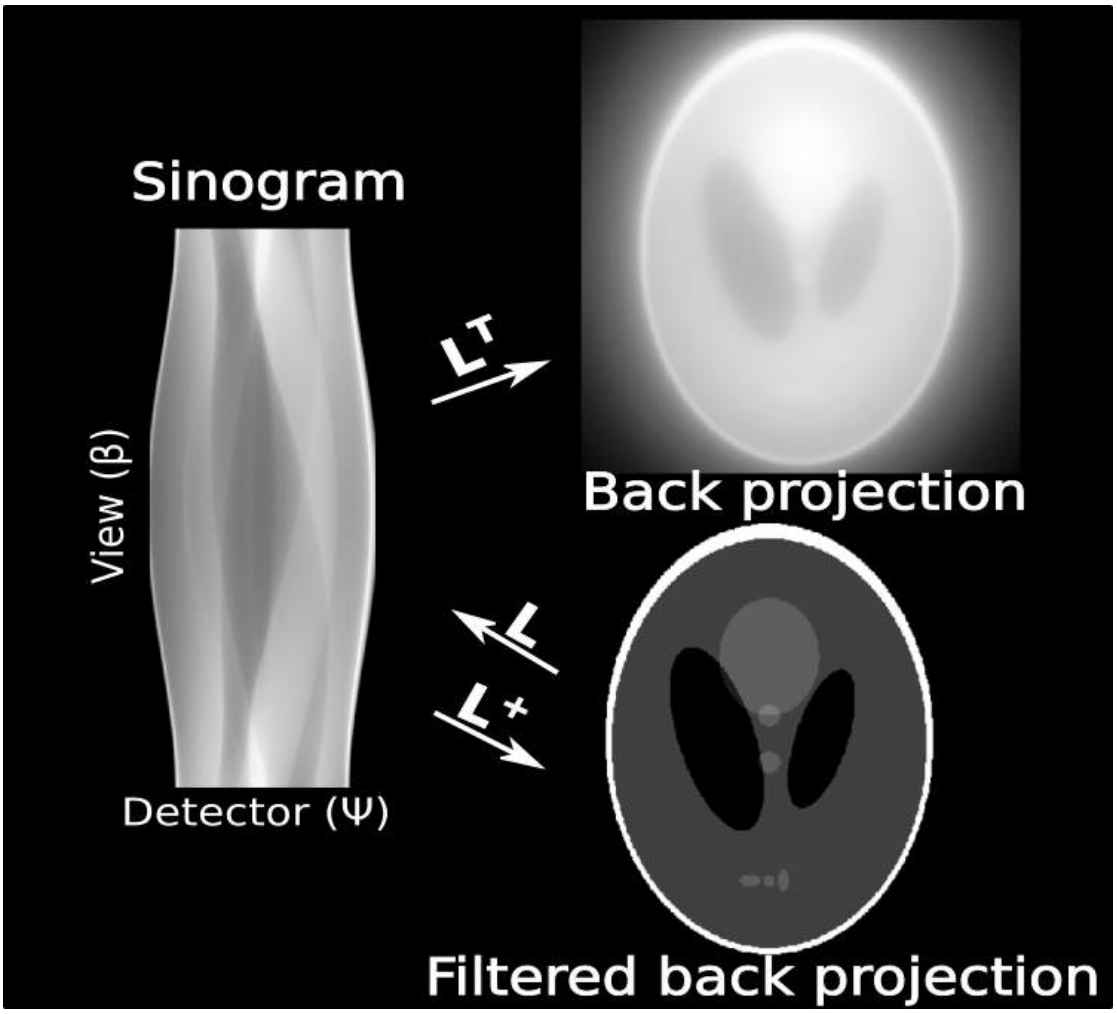


FIGURE 2.2: Definition of the three basic operations in CT, illustrated for the Shepp-Logan phantom: Multiplying with the system matrix $\mathbf{L}$ is a forward projection that calculates a sinogram from an image; multiplying with its transpose $\mathbf{L}^T$ is a back projection, which smears the projections over the image; and multiplying with the pseudo-inverse $\mathbf{L}^+$ calculates a least-squares solution to the CT problem, which in an idealized first generation scanner geometry may be achieved using Filtered Back Projection.

the detector arrays[7]. This filter is first applied to the projections, upon which the result is backprojected by applying $\mathbf{L}^T$.

2.1.4 CT in practice

We summarize the three main CT operations in fig. 2.2, using the digital *Shepp-Logan* phantom as an example: multiplying by $\mathbf{L}$ forward projects, creating a set of projections from an image by calculating line integrals; $\mathbf{L}^T$ back projects, "smearing" the projections over the images; and applying the pseudo-inverse $\mathbf{L}^+$ reconstructs the image as a least-squares solution to the CT problem, which in the first generation scanner geometry may be achieved using FBP.

Open source CT packages exist that provide implementations of both the primitive operations $\mathbf{L}$ and $\mathbf{L}^T$ and reconstruction algorithms such as FBP. In this project, we predominantly used the ASTRA package[18, 19], which provides accelerated versions of these operations on Graphical Processing Units (GPUs). Since both the forward and back projections are parallelizable[20], this GPU acceleration provided important speedups: testing the GPU and CPU versions of the forward and back projections in ASTRA on the Shepp-Logan phantom with a 2640×816 first generation sinogram, we found a 60 times speed-up for the forward projection (6s vs. 0.1 s) and a 20 times speed-up for the back projection (6.4s vs. 0.33s) on the following system: CPU: Intel Core i7-4712HQ @ 2.30GHz; GPU: NVIDIA GK107GLM (Quadro K1100M).

2.2 Reconstruction from 3D spiral CT data

Modern (medical) CT scanners do not follow the first generation scanner geometry that we discussed in the previous section. A more widespread model is the *third generation scanner*, which is the kind that we had access to for this project. In this section, we consider how to perform CT using data from such a scanner[7].

2.2.1 Scanner specifications

Our data came from a Philips Brilliance Big Bore medical CT scanner, which is third generation in the sense that the source and detector array are fixed relative to each other during acquisition and rotate continuously in the xy-plane around an isocenter, while the patient table is moved along the z-axis. This leads to a *spiral*, or helical source trajectory, as illustrated in fig. 2.3(a). Additionally, instead of the flat detector of the first generation scanner, it uses a curved, 2-dimensional cylindrical detector array illuminated by a cone-beam, illustrated in fig. 2.3(b). With $N_s = 16$ detector rows in the z-direction, an acquisition in this geometry leads to 16 spiral sinograms, one of which is shown in fig. 2.3(b).

The relevant geometric quantities of our scanner are shown in table 2.1.[21]. The focus-isocenter distance (FI) is the radius of the source rotation around the patient isocenter; the focus-detector distance is the radius of the circle in the xy-plane upon which each row of detectors lies, while the fan angle defines the angular distance spanned by the detector row; the z-collimation[1] defines the (effective) number of detectors in the z-direction (here 16) and their widths (here 0.75).

Dynamic focal spot While the source is moved in a spiral trajectory, the possible image resolution at a given detector sampling rate is increased by using a *flying focal spot*[23], where

[1] This may be varied between scans; these were the settings for all our patients

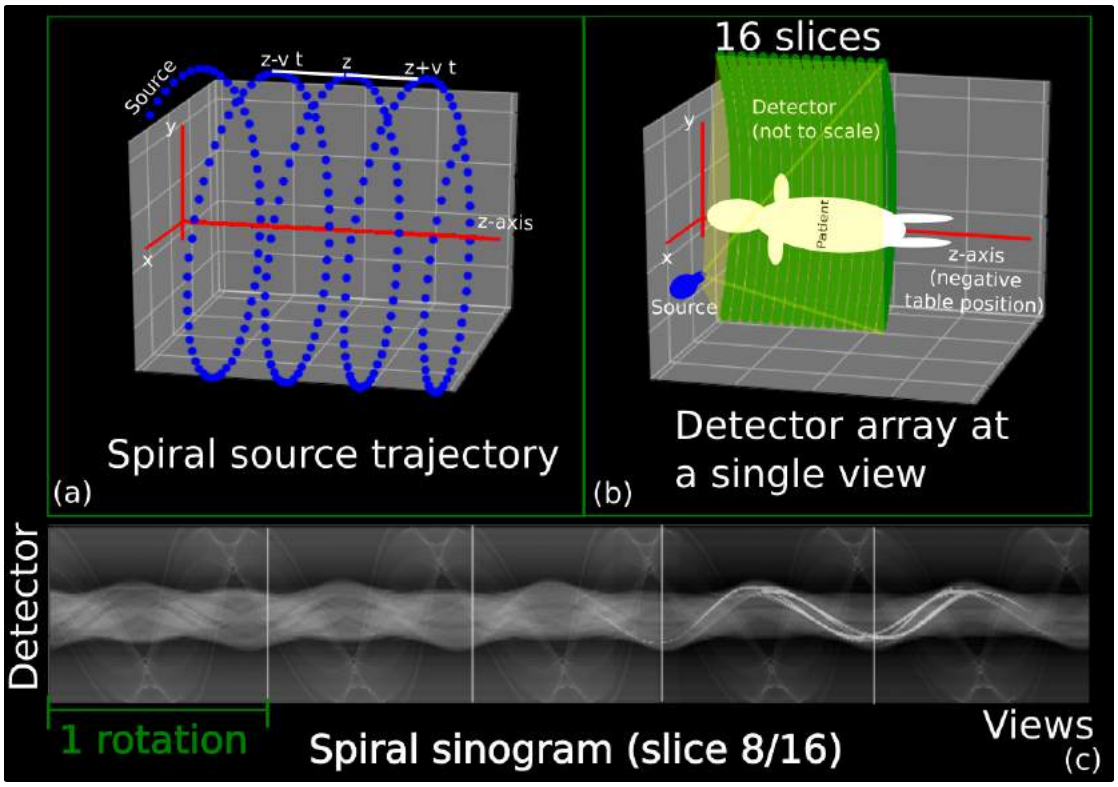


FIGURE 2.3: Illustration of third generation, spiral/helical CT acquisition. (a): As the source rotates, the table is moved along the z-axis at velocity v, leading to a spiral trajectory allowing for rapid 3D scanning. (b): At each view, measurements are performed simultaneously in a cylindrical detector using a conebeam x-ray source. (c): Using a detector array with 16 detector rows in the z-direction of 816 detector elements in the xy-plane, we get 16 spiral sinograms. Shown is number 8/16 from a head-and-neck acquisition.

Source (focus) - isocenter distance (FI)	**Source (focus) - detector distance (FD)**	**Z-collimation**	**Number of detectors (N_d)**	**Fan angle (in the xy-plane)**
$645mm$	$1183.45mm$	$16 \times 0.75mm$	816	55.6364

TABLE 2.1: Some scanner specifications of the Philips Brilliance Big Bore scanner. We acquired these data from the headers of the sinogram data files, but similar values are reported in the comparison of big-bore scanners in ref. [22].

the focal spot[2] in the x-ray source is dynamically deflected back and forth between subsequent views.

In our scanner, this is achieved using the *dynamic focal spot* technology. In particular, denote by $\Delta\alpha$ the angular separation of two detectors in the detector array and the current view angle at view number k by β_k. As we found indicated in the header section of the raw data files, the effective view angle of the x-ray beam is then deflected to an angular position of $\beta_k - \frac{\Delta\alpha}{8}$, $\beta_{k+1} + \frac{3\Delta\alpha}{8}$, etc.

[2]The small metal area that emits the x-rays

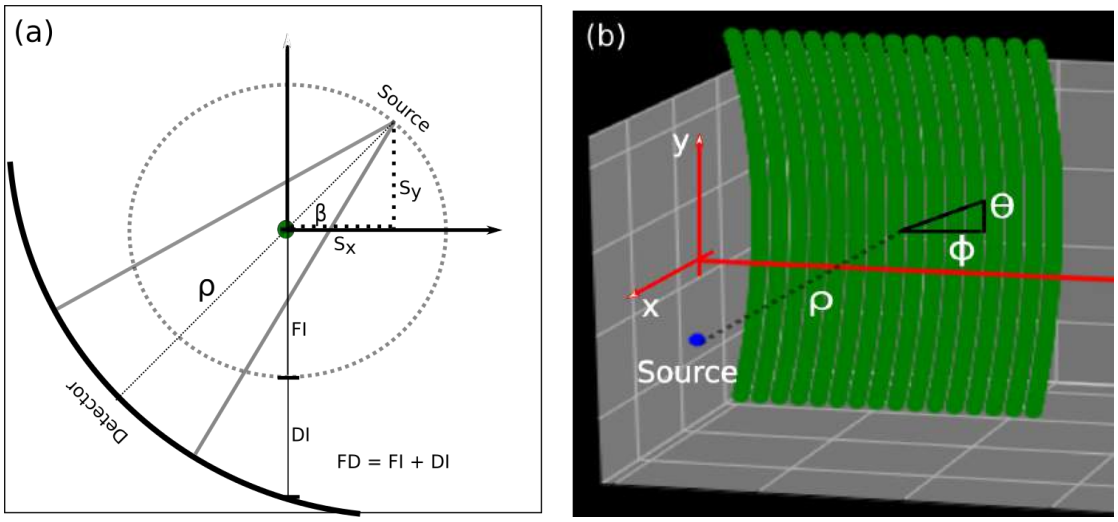


FIGURE 2.4: Diagram of the source and detector position. (a): Diagram of a single detector row. The source rotates on a circle around the isocenter with radius FI. The position of the source in the isocenter coordinate system is given by (S_x, S_y). The detector array lies on a cylinder centered on the source position with radius FD. The detector position in the isocenter coordinate system may thus be calculated as the source position added by a cylindrical coordinate on the detector array, as illustrated in (b).

2.2.2 3D reconstruction

In 3D, the forward- and back-projection operations $\mathbf{L}[\cdot]$ and $\mathbf{L}^T[\cdot]$ may be applied using 3D analogues to the 2D raytracing operations combined with a simulation of the detector and source movement during acquisition. We implemented such a simulation and describe it in the following.

2.2.2.1 Source and detector position during helical acquisition

Assume that the source and detector array rotate at an angular frequency of $\omega = \frac{2\pi}{\tau}$, and that the table moves at a speed of v. Suppose further that the x-ray tube acquires a total of N_v measurements at equal time intervals as it rotates in the xy-plane defined in fig. 2.3.

As illustrated for the xy-plane in fig. 2.4(a), without the flying focal spot the source position during rotation is now given by:

$$S_x = FI\cos(\omega t),$$
$$S_y = FI\sin(\omega t),$$
$$S_z = vt.$$

To get the detector positions in the cylindrical detector array, suppose a detector element is positioned at a rotation away from the central detector element by an angle of θ in the direction of detector rotation, and displaced by ϕ in the z-direction, as illustrated in fig. 2.4(b). The position of the detector element becomes the source position added by a cylindrical coordinate

in a system centered on the source position[24].

$$
\begin{aligned}
D_x &= FI\cos(\omega t) + FD\cos(\omega t + \pi + \theta),\\
D_y &= FI\sin(\omega t) + FD\sin(\omega t + \pi + \theta),\\
D_z &= vt + \phi.
\end{aligned}
$$

Here, π was added to θ since the central detector is found at this rotation.

Including the flying focal spot now depends on how the source deflections are technically realized. In our case, we realized the dynamic focal spot deflections defined in subsection 2.2.1 by using a quarter-detector periodic deflection along with a constant displacement of the detector array by an eighth of a detector width; this leads to the mentioned periodic deflections. We achieved this by first defining the frequency of the periodic modulation as half the detector sampling frequency, $\nu = \omega\frac{N_v}{2}$. We then modelled the dynamic focal spot deflections defined in subsection 2.2.1 by rotating the detector array by $(1/8)\Delta\alpha$, where $\Delta\alpha$ is the angular width of a detector, and modulating the view angle as $\omega t \to \omega t - \frac{\Delta\alpha}{4}\cos(\nu t)$.

2.2.2.2 FBP in 3D

Having modelled the source and detector movements, iterative reconstruction may now be performed, but FBP cannot be applied since the forward model no longer follows the first generation geometry. The sinogram must therefore either first be interpolated and rebinned to emulate data from an appropriate geometry, or, for more accurate reconstruction, the algorithm must be adapted for 3D reconstruction. FBP-based algorithms that achieve this exist, such as the Feldkamp-David-Kreuz (FDK) algorithm for cone-beam CT reconstruction[7, 25, 26].

We however found during the project that 3D reconstruction was too slow for our purposes, since the primitives we had access to[3] were not GPU-accelerated: a single 3D back projection would therefore take about 10 minutes on our system. We therefore chose to interpolate and rebin to a 2D geometry compatible with the ASTRA toolbox.

As illustrated in fig. 2.5, we split this task into two steps: first, we interpolated the spiral sinogram into a set of parallel, axial slices for subsequent sequental 2D reconstruction; and second, we rebinned the 2D sinograms to the first generation geometry, which is supported by ASTRA.

[3] Borrowed with gratitude from Johan Nuyts and his group at UZ Leuven, Belgium.

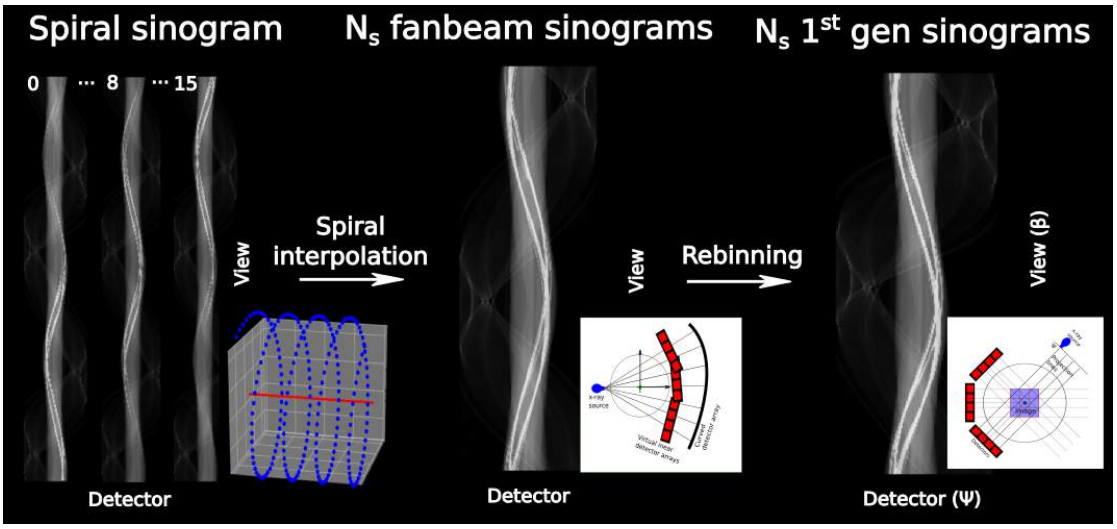


FIGURE 2.5: Illustration of the two steps in our sinogram interpolation approach. First, the spiral sinogram is interpolated to a set of 2D, planar sinograms. Next, the 2D, planar sinograms, which follow a fanbeam geometry with a dynamic focal spot, are rebinned and interpolated to a first generation scanner geometry with a flat detector array.

2.2.2.3 Spiral interpolation

We used the 360LI (360 degrees Linear Interpolation) algorithm[25] for interpolation of the 3D spiral sinogram to a set of 2D, planar sinograms: the sinogram value at the axial position z was calculated as the linear interpolation of the values at position $z - v \cdot \tau$ and $z + v \cdot \tau$, illustrated in fig. 2.3(a), where v is the table speed and τ is the source rotation period; they were thus replaced by linear combinations of measurements displaced by a 360 degrees rotation.

While in principle any number of 2D sinograms corresponding to slices at any location may be created through this interpolation method, it is not possible to achieve a resolution that is larger than the collimation of the detector array, i.e. the effective size of the detector elements in the z-direction. As seen in table 2.1, this was in our case $0.75mm$, and so we rebinned to $0.75mm$ slices. The CT volumes were reconstructed sequentially in these slices.

2.2.2.4 Rebinning to the first generation geometry

The 2D sinograms that result from the spiral interpolation correspond to fan-beam sinograms with a curved detector array and a dynamic focal spot. ASTRA currently supports both fan-beam and parallel beam geometries, but currently only with flat detector arrays. We chose to rebin our sinograms to a flat detector, parallel beam geometry, corresponding to the first generation scanner: as illustrated in fig. 2.6, all beam-lines in the fanbeam can be associated with a virtual detector at a certain view angle β and detector displacement ψ. Using our source/detector movement simulation we implemented for use in 3D reconstruction, we associated each projection with a source and detector position in polar coordinates relative to the isocenter.

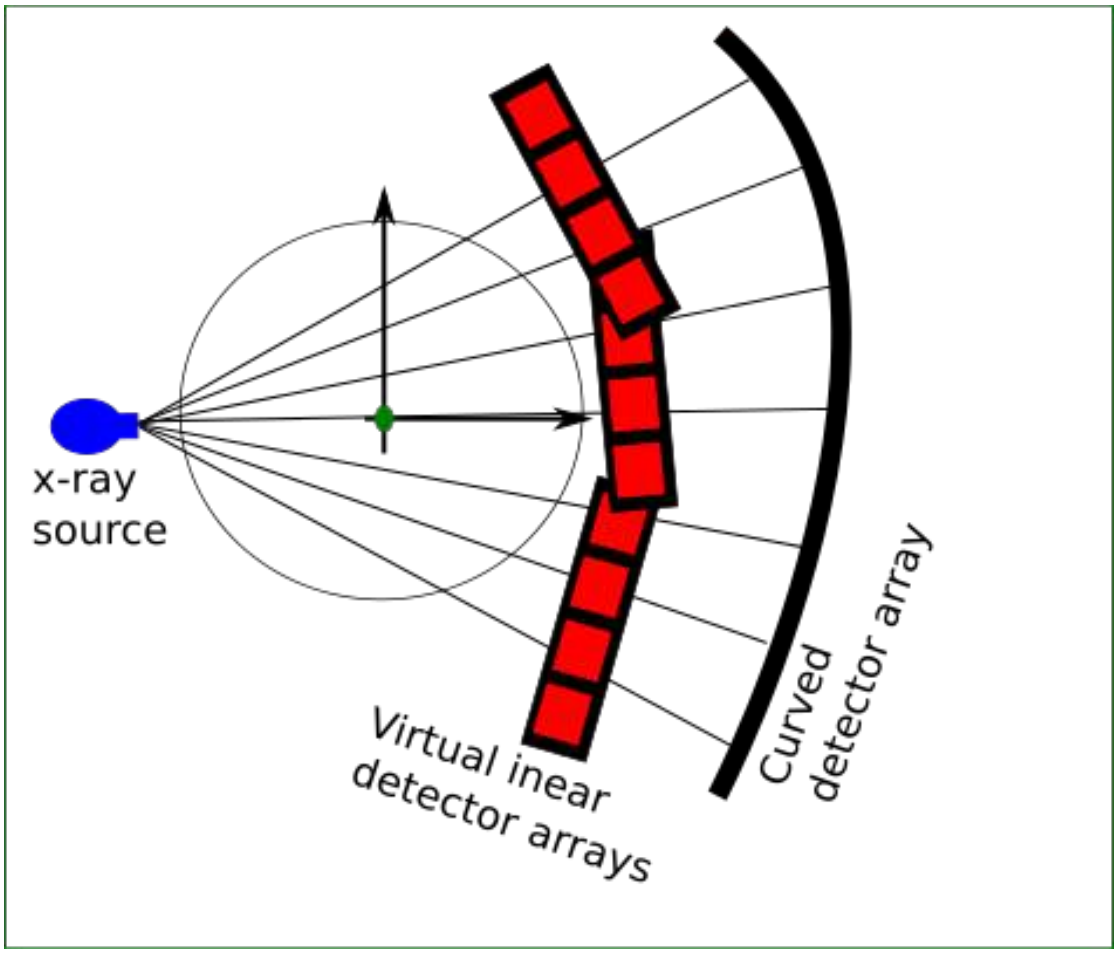


FIGURE 2.6: Illustration of the relationship between the fanbeam and first generation geometries: The beamlines in the fan may each be assigned to a virtual detector array.

For each position, we then calculated the angle β and orthogonal displacement ψ along the detector array. We finally binned the data by the view angles and used cubic interpolation along the detector arrays to put them on a regular grid, thus realising a first generation scanner geometry.

2.3 Inaccuracies of the linear model

In addition to diverging in geometry from the canonical first generation scanner model, as we considered in the last section, data from a real medical CT scanner do not conform to the linear model we presented in section 2.1. In this section, we consider such inaccuracies, the problems they lead to and how they are addressed in practice. As we will see, they are exacerbated when metal implants are involved, leading to the metal artifacts that are the focus of this thesis.

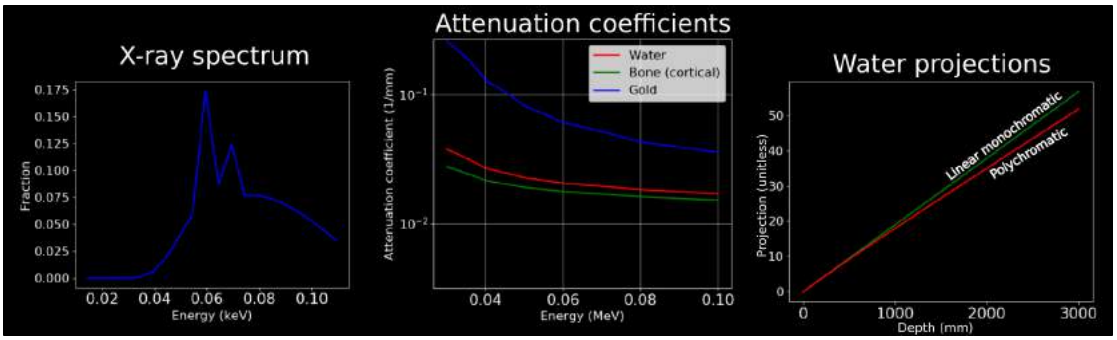


FIGURE 2.7: Left: Simulated energy spectrum of x-rays emitted from the source (using SpekCalc[27]). The spectrum assumed a tungsten source with a kVp of 120kV, a tube anode angle of 7 deg and a source filter containing 3mm aluminium and 0.1mm copper. Center: Energy dependent attenuation coefficients (without K-edge) (source: NIST[28]). Right: The calculated projections at different depths assuming 1) mono-energetic x-rays (blue, linear curve) and 2) the full spectrum in (a) (red, sub-linear curve).

2.3.1 The x-ray energy spectrum

The Lambert-Beer model in eqn. (2.1) assumed that the attenuation of the x-rays may be described by a single attenuation map $y(\mathbf{x})$. For real CT scanners, the source emits x-rays at a spectrum of energies that are attenuated at an energy-dependent rate, and so in reality a higher dimensional map that provides the attenuation coefficient at each possible energy E, $y(\mathbf{x}, E)$, should be reconstructed.

Fig. 2.7(a) shows an exemplary, simulated x-ray spectrum with a maximum energy of $E_{max} = 120keV$[4]. The polyenergetic source spectrum means that eqn. (2.1) does not hold; a more appropriate model of the transmitted x-ray count is[29]:

$$n_j = \int_0^{E_{max}} \Gamma_j(E) \mathsf{e}^{-\int_{\mathcal{L}_j} y(\mathbf{x},E)d\mathbf{x}} dE, \tag{2.8}$$

where the emitted x-ray intensities have been made energy-dependent. The energy-dependent attenuation coefficients of a few tissue types (water, cortical bone, gold) are shown in fig. 2.7(b) over the range of energies in the spectrum; the energy-dependence is notable, especially for lower energies and especially for the gold. We calculate the projections using eqn. (2.2):

$$p_j^{poly} = \log(\Gamma_j) - \log \int_0^{E_{max}} \Gamma_j(E) \mathsf{e}^{-\int_{\mathcal{L}_j} y(\mathbf{x},E)d\mathbf{x}} dE, \tag{2.9}$$

where $\Gamma_j = \int_0^\infty \Gamma_j(E)dE$ is the total background count, summed over energies. Suppose now that the projections were acquired through a fixed depth x of water with the constant (energy dependent) attenuation coefficient $y_{water}(E)$ using 1) the full spectrum in fig. 2.7(a) and 2) mono-energetic x-rays at the effective mean energy of the spectrum, $E_0 = \frac{\int_0^{E_{max}} E\Gamma_j(E)dE}{\int_0^{E_{max}} \Gamma_j dE}$, $\forall j \in \mathcal{S}$. The polychromatic and monochromatic projections p_j^{poly} and p_j^{mono} (eqn. (2.2)) then

[4]This is controlled by the voltage used to accelerate the electrons that release the x-rays, and may be varied between scans

become:

$$p_j^{poly} = \log(\Gamma_j) - \log \int_0^{E_{max}} \Gamma_j(E) \mathsf{e}^{-y_{water}(E)\cdot x} dE$$
$$\text{and} \quad p_j^{mono} = \log(\Gamma_j) + y_{water}(E_0) \cdot x.$$

p_j^{mono} is here a linear function of a single, effective attenuation coefficient $y_{water}(E_0)$, while p_j^{poly} is non-linear. In particular, it is generally sub-linear, as illustrated in fig. 2.7(c) that show how p_j^{poly} calculated in water using the spectrum in fig. 2.7(a) diverges from the linear p_j^{mono} as x increases. The sublinearity is minor for water, but the larger the atom number of the material, the more pronounced it becomes; this is due both to a larger energy-dependence of the attenuation coefficient and to its larger values.

The Hounsfield Unit When assuming the linear model of the projections, and thus an effective attenuation coefficient of the sort introduced in the previous subsection, the interpretation of reconstructed CT values depends on the particular acquisition. This leads to varying contrasts between scanners, and so for ease of comparison the attenuation coefficient is linearly transformed to *Hounsfield Units*:

$$HU_i = 1000 \frac{y_i - y_w}{y_w - y_a},$$

where y_w and y_a are respectively the effective attenuation coefficients of air and water reconstructed using the linear model. In these units, $HU_i = 0$ corresponds to water, $HU_i = -1000$ to air. The conversion is done in practice by a phantom calibration method that both accounts for the x-ray source spectrum and other potentially unknown scanner-specific factors[7].

Beam hardening The sublinearity of eqn. (2.9) is due to the *beam hardening* effect: as the polyenergetic x-ray beam traverses the material, the less energetic "softer" x-rays are preferentially attenuated such that only the more energetic "harder" x-rays are left. Accordingly, the beam becomes more penetrating as x increases, leading to a decrease in attenuation compared to the linear approximation[23, 30, 31].

The effect of this beam hardening may be seen in fig. 2.8, where we used eqn. (2.9) to simulate a polychromatic measurement of a cylindrical water phantom with and without iron inserts. In the pure water phantom, the main effect of beam hardening is the *cupping artifact*, where the reconstructed attenuation coefficient decreases artificially towards the center. This is explained in fig. (b): a projection acquired along a line through the image is lowered by beam hardening to a degree depending on its length, and so the reconstructed attenuation coefficients decrease toward the center to compensate[8].

In the phantom with metal inserts, the sharp transition between degrees of beam hardening between the soft tissue and metal additionally leads to periodically occuring dark and bright streaks in the reconstruction. This is caused by the beam hardening as follows: the value of the measured projection controls the values of the reconstructed attenuation coefficient along the beam line. If the projection values are smaller than expected by the model, as happens with beam hardening through the metal, the attenuation coefficients become smaller than the truth, leading to the dark areas between the implants. At the same time, the projection values are as expected along beamlines passing only the water between the implants, which can only be reconciled with the dark areas by inserting compensatory light streaks[8, 32].

Fig. 2.7 suggests a way to account for this beam hardening effect, as there seems to exist a simple transform between the linear and sublinear curves. The sublinear curve may be acquired from phantom measurements, e.g. a water phantom, from which a map to a linear curve may be calculated. Such a map is in practice applied to the projections as a standard sinogram preprocessing step known as *empirical beam hardening correction*. This is however not always sufficient: to illustrate this, we applied it to the sinograms using simulated water measurements for the phantoms, both with and without metal, yielding the beam hardening corrected FBP images in fig. 2.8(a). The correction was succesful for the pure soft tissue phantom, but did very little with the metal implants[23, 33, 34].

2.3.2 Noise and other intensity variations

2.3.2.1 Intensity modulations

Fig. 2.9(a) shows the intensity (x-ray count per area) measured by a detector near the edge (number 815) at a number of view angles in a spiral head-and-neck scan. The profile has a period of 2640 measurements, corresponding to a rotation of the source. The measurements suddenly increase in value at a certain rotation that coincides approximately with when the head enters the scan. This is due to an automatic increase in the current to the source that increases the emitted x-ray intensities $\mathbf{\Gamma}$ to 1) provide better measurements in the diagnostically interesting region and 2) provide sufficient x-rays to accurately scan the more attenuating regions in the oral regions and skull. This modulation must of course be accounted for when calibrating the data, which, if details on the modulation are not available (as in our case), may be done by considering measurements such as those in 2.9(a) that directly provide values for the background intensity at each view angle.

The intensities measured at a single view from the central row (number 8) of detectors in our scanner are shown in fig. 2.9(b), both for an empty-bore (air) scan (blue) and a head-and-neck patient (red). The shape of the curve shows a systematic modulation of the signal that increases

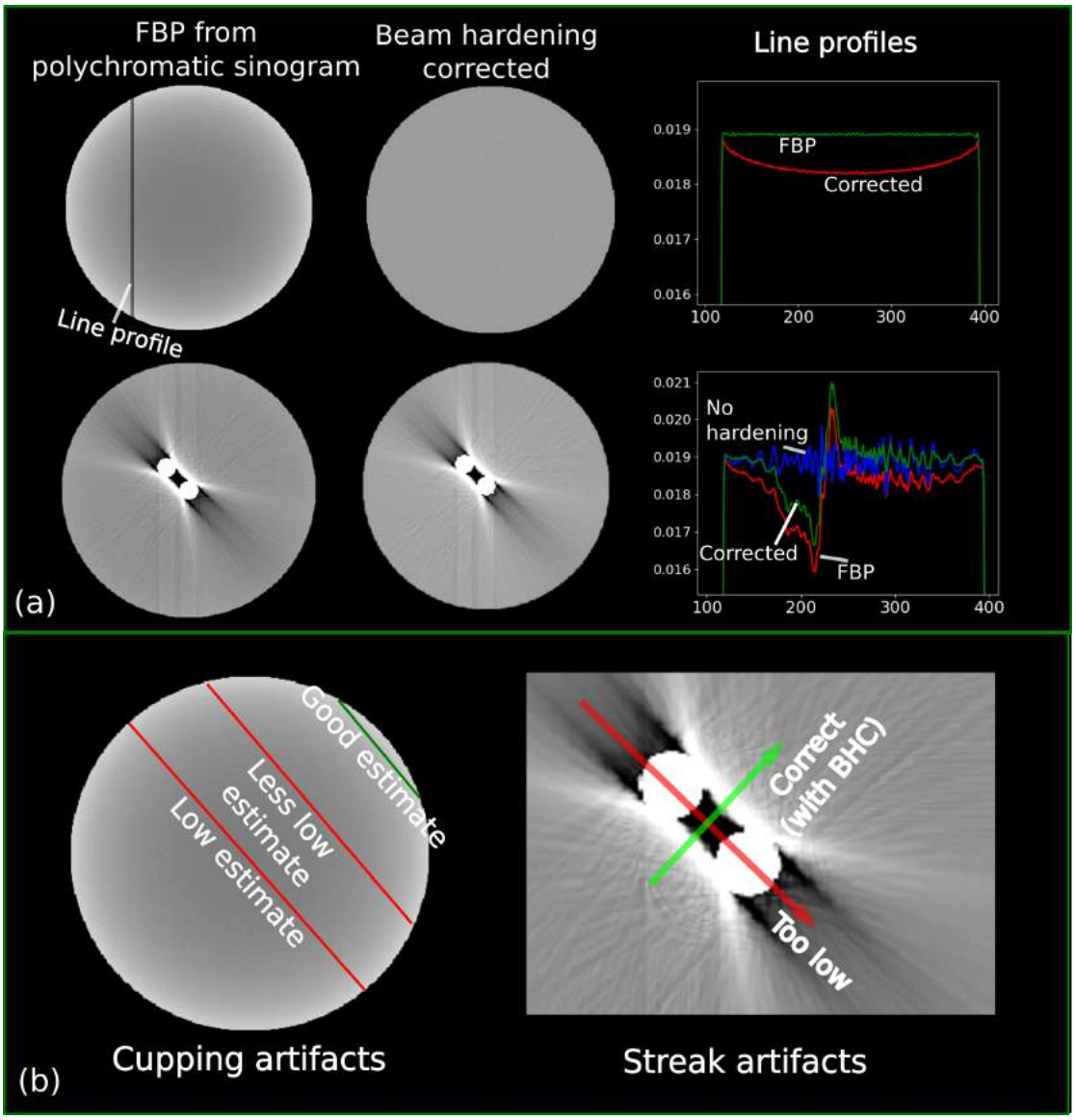


FIGURE 2.8: (a): A digital phantom with and without iron inserts reconstructed using FBP from a sinogram created with the polychromatic eqn. (2.9). Shown is also the reconstruction after empirical water correction as well as profiles of the attenuation coefficients along a line through the phantoms. For the phantom with metal inserts we additionally included a (blue) line profile for an FBP reconstruction with a monochromatic spectrum, and thus no beam hardening. (b): Illustration of cupping and streak artifacts.

the intensity toward the center of the conebeam. This modulation is realized by the use of two types of filter: a pre-hardening filter that absorbs low-energy x-rays from the beam to diminish the beam hardening effect; and a *bowtie* filter that serves to focus the intensity towards the center of the beam where the attenuation tends to be larger and so more x-rays are needed.[7] There is additionally a periodic modulation with a period of 16 detectors, which is an artifact of the detector hardware.

To account for these effects, the co-acquired air measurements are divided on the patient measurements to yield the ratio n_j/Γ_j in eqn. (2.2) prior to reconstruction using eqn. (2.4).

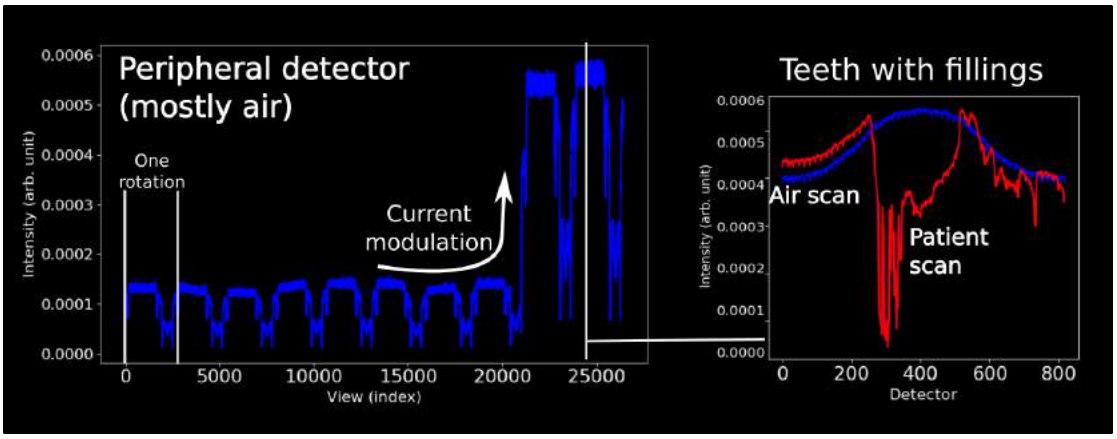


FIGURE 2.9: Intensity measurements from a head-and-neck patient. (a): The signal from a detector on the edge of the cylindrical array, which mostly measures through empty air but sometimes through an edge of the table. The signal repeats after each rotation and the intensity is automatically increased through current modulation as the head enters the field of view. (b): The measurements from a row of detectors at a single view taken from both a pure air scan and the patient scan. Due to source filters there is an intensity increase towards the center.

2.3.2.2 Measurement noise

On close inspection of the curves in fig. 2.9, the measurements also show the presence of noise whose magnitude depends on the magnitude of the intensities; in particular, due to the statistics of the x-ray generation process in the tube, it may be modelled by a Poisson distribution.

Noise equivalent count scaling Due to potential unknown scalings of the data during storage and the sinogram pre-processing steps, the measurements may however not be truly Poisson distributed. One may attempt to undo such scalings by *Noise Equivalent Count* (NEC) scaling[7, 35]: assume a set of Poisson measurements n_j, $j \in \mathcal{S}$, are available at the same mean value Γ, but which have been scaled somehow. Then, supposing that we apply a certain scaling factor C, the variance is:

$$\mathsf{E}[(Cn_j - \mathsf{E}[Cn_j])^2] = C^2\mathsf{E}[(n_j - \mathsf{E}[n_j])^2],$$

where $\mathsf{E}[\cdot]$ here denotes an empirical average. For the data to be Poisson distributed, we require the mean and variance to be equal:

$$\mathsf{E}[Cn_j] = C\Gamma = C^2\mathsf{E}[(n_j - \mathsf{E}[n_j])^2],$$

and so, since $\Gamma = \mathsf{E}[n_j]$:

$$C = \frac{\mathsf{E}[n_j]}{\mathsf{E}[(n_j - \mathsf{E}[n_j)^2]]}.$$

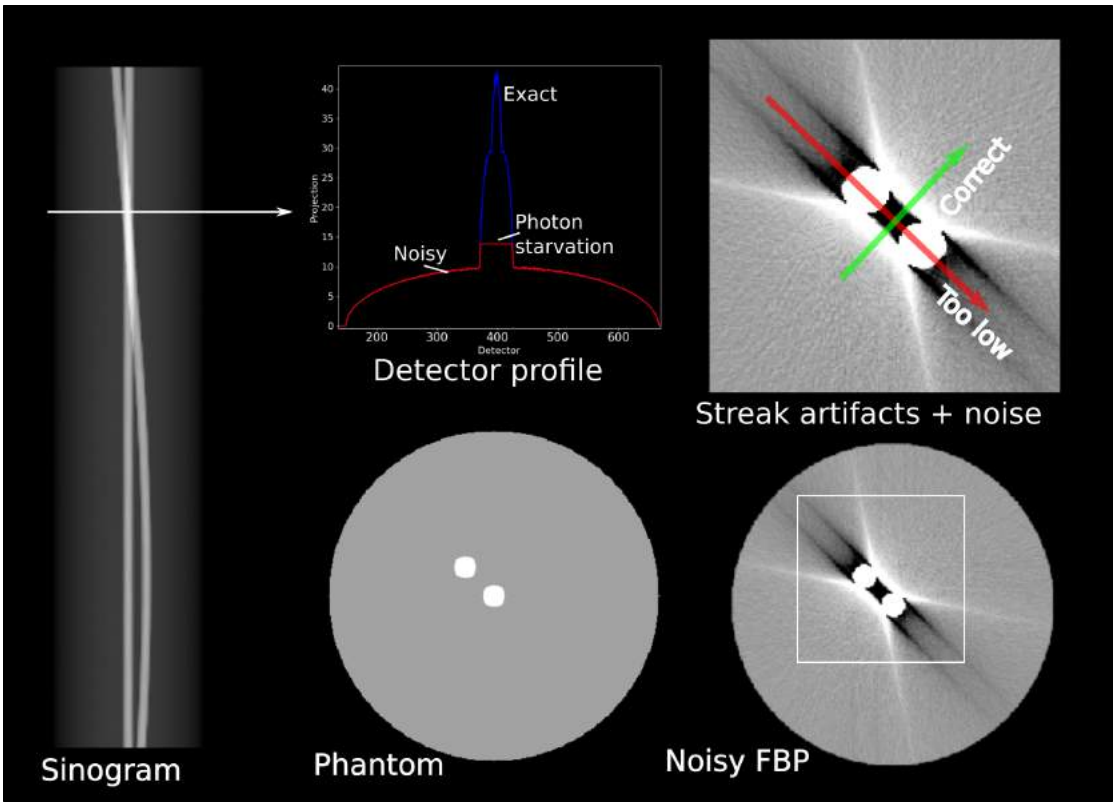


FIGURE 2.10: Noise and photon starvation artifacts in CT. The detector profile shows the simulated projections for the detector array at a single view angle through the simulated phantom with iron inserts. The photon starvation leads to artificially lowered projection measurements through the metal, which leads to crmpensatory dark and bright streaks.

We in practice calculated this factor from an air scan, but it may be estimated from any known measurements in the patient acquisition with the same mean value, e.g. those of the peripheral detectors.

Noise and photon starvation artifacts The lowest-valued, most noisy measurements in fig. 2.9(b) were acquired through metal fillings in the teeth. Here, both the noise and the consequently small Signal-to-Noise Ratio (SNR) as well as the potentially undetectably low counts may have consequences for the CT reconstruction. The resulting *photon starvation* that occurs as a potentially insufficient number of photons are measured leads to artifacts by two mechanisms: first, the SNR becomes small, and so the measurements are highly uncertain, leading to noise[29, 31, 36]. Second, the mean count may become too small for the detector to measure[37], which leads to a systematic error. To illustrate these effects, fig. 2.10 shows the results of an experiment where we emulated a noisy CT acquisition on the digital phantom with iron inserts by exponentiating the sinogram and applying Poisson noise at a constant $\Gamma_j = 10^9$; measurements with an integer value of 0, which would lead to a divergence in eqn. (2.2), were assigned the value of 1, corresponding to the minimum value measurable by the detector.

As may be seen in fig. 2.10, the noise in the measurements added noise to the image, but also led to more dramatic streak artifacts due to the photon starvation: Fig. (b) shows the projections at a single view, which illustrates how the minimum intensity due to photon starvation translated to a maximum possible projection value. In the reconstruction, this led to severe streak artifacts, by much the same mechanism as the beam hardening[37].

In less severe cases of photon starvation, such as those we see with small dental implants in the head and neck, this extreme case where photon starvation occurs for all metal projections is unlikely; however, due to the measurement noise it may happen to a subset of metal projections.

2.3.3 Scatter

As mentioned in section 2.3.1, the attenuation of x-rays occurs both by absorption and scatter of the x-rays, in the diagnostic CT window in particular using photoelectric absorption and Compton scatter[29, 32].

In addition to adding noise, the effect of scatter on the image may be similar to the photon starvation and beam hardening effects: while scattered photons lead to an artifically increased x-ray count in all detectors, the smaller-valued, noisier measurements through the metal are particularly affected, which leads to a relative inflation in the metal measurements that is similar to the effects of beam hardening[7, 14, 32].

In modern medical CT systems, the scatter is mitigated by the use of a physical anti-scatter grid on the detectors, and so while scatter may be expected to contribute some noise its contribution to the artifacts should be somewhat smaller than the photon starvation and beam hardening effects[7, 29, 38].

2.3.4 Non-linear partial volume effect

The beam hardening artifacts we discussed in section 2.3.1 occured due to the modelling of a truly non-linear forward model as linear. This effect was here due to the dependence of the attenuation on a distribution over energies rather than an attenuation coefficient at a single energy.

Analogous errors may arise from any effect that leads to a similar non-linear form of the forward model. In particular, the *Non-linear partial volume* (NLPV) effect arises when the attenuation coefficient varies perpendicular to the projection line $\mathcal{L}_j$ within the view of a single detector. Since the measurement n_j of detector j is in reality the integral over the x-rays passing a cross-section of the volume, n_j should follow an integrated Lambert-Beer equation similar to eqn.

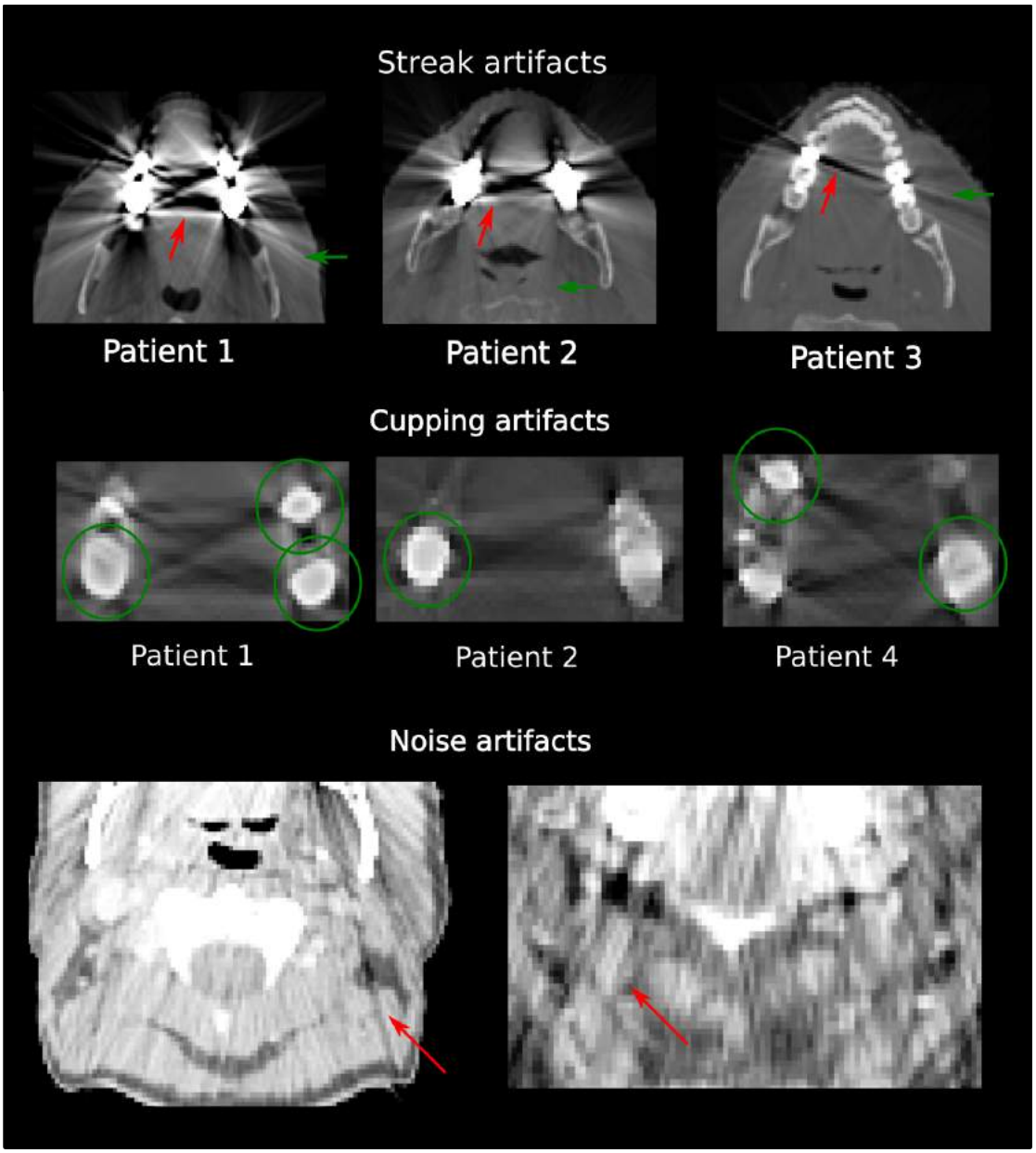


FIGURE 2.11: Examples from three head-and-neck patients of metal artifacts: Streaks, cupping and noise artifacts from top to bottom.

(2.9), but with the integral over a sub-voxel surface distribution rather than a spectrum of energies[29, 32, 38].

2.3.5 Summary

We summarize the four main artifact sources in table 2.2, while fig. 2.11 shows the FBPs of three patients with dental implants and the different types of metal artifacts. As seen from the table, the notable streak artifacts (fig. a) arise from a combination of four sources. Cupping is generally not seen in the patient soft tissue due to the empirical beam hardening correction, but

	Detector noise /photon starvation	**Beam hardening**	**Scatter**	**NLPV**
Source	Truncation, random signal	Linear approximation	Attenuated photons detected	Linear approximation
Mechanism	Effective minimum measurement, random error	Over-measurements	Over-measurements, random error	Wrong measurements.
Artifact	Noise and streaks	Streaks and cupping	Noise and streaks	Streaks

Table 2.2: Metal artifact sources

may be seen in fig. (b) as a decrease in the CT values toward the center of the metal implants. Noise is finally often seen as thin streaks that propagate far into the image (fig. (c)).

Chapter 3

Metal Artifact Reduction

Chapter 2 provided an overview of CT and considered the issues with the conventional linear forward model used for CT reconstruction, and how these issues were exacerbated in the presence of metal so as to cause metal artifacts. This chapter in turn considers how to address these metal artifacts using Metal Artifact Reduction (MAR) algorithms, which fall into three categories: Image inpainting, sinogram inpainting and model-based iterative reconstruction (MBIR) algorithms.

3.1 Manual image inpainting

The most direct MAR approach is to directly paint over the corrupted CT values in image space. Since such methods do not directly address the artifact sources and are prone to introducing new artifacts to the image, automatic, algorithmic versions of such methods tend to not be used as-is, but rather as part of a hybrid method including sinogram inpainting. We return to some such methods when considering prior generation for sinogram inpainting in sec. 3.2[8, 32].

Some applications however call for manual manipulation of the CT before use; in radiotherapy, the CT is used for computer-guided treatment planning that is itself time consuming, and so it is feasible to manually address the artifacts in image space. The technique for this task depends on the practitioner, but a common technique is water override, where the planner replaces soft tissue regions with a bulk CT value of HU=0[39, 40]. This method has obvious downsides: 1) it is time consuming; 2) it is prone to human errors in the detection of the artifacts; and 3) the assigned bulk CT value may be systematically wrong.

3.2 Sinogram inpainting algorithms

Sinogram inpainting algorithms generally consist of two steps. First, identify the projections that pass through metal (the *metal projections*) by a process such as the one illustrated in fig. 3.1(a): acquire the FBP and segment the metal implants by some means (usually simple thresholding). Then, forward project through the resulting metal mask, leading to a *metal trace* that identifies the metal projections. Finally, inpaint the assumedly erroneous metal projections with new estimates, apply FBP for a corrected image and in post-processing reintroduce the metal CT values by inpainting using the metal mask.

3.2.1 Linear Interpolation MAR

Linear Interpolation MAR (liMAR)[41] is the arguably simplest sinogram inpainting algorithm. As illustrated in fig. 3.1(b), it linearly interpolates the metal projections from the adjacent, non-metal projections. In particular for a 2D scan, it considers the detector profile for each view separately and linearly interpolates over the metal projections[41, 42].

3.2.2 Normalized MAR

A related but more succesful algorithm is the normalized MAR (nMAR), which improves upon liMAR by including information in a prior image[41–43].

3.2.2.1 Prior generation

The prior image may be generated by any means, but the standard way is using some combination of thresholding and bulk CT value assignation, filtering and liMAR, which we implemented within our ASTRA-based framework[42, 43]. The original nMAR publication suggested using K-means clustering[44] to automatically create the segmentation thresholds[41]; fig. 3.2 shows image priors generated in that way, using varying numbers of clusters (tissue classes) in the segmentation and different image processing strategies.

When we implemented nMAR within our reconstruction framework, we settled on the method indicated by the yellow box, i.e. K-means clustering with three clusters on a liMAR image; compared to not using liMAR, this led to better accuracy in the oral cavity, and using a Gaussian filter before K-means led to unnecessary smoothing.

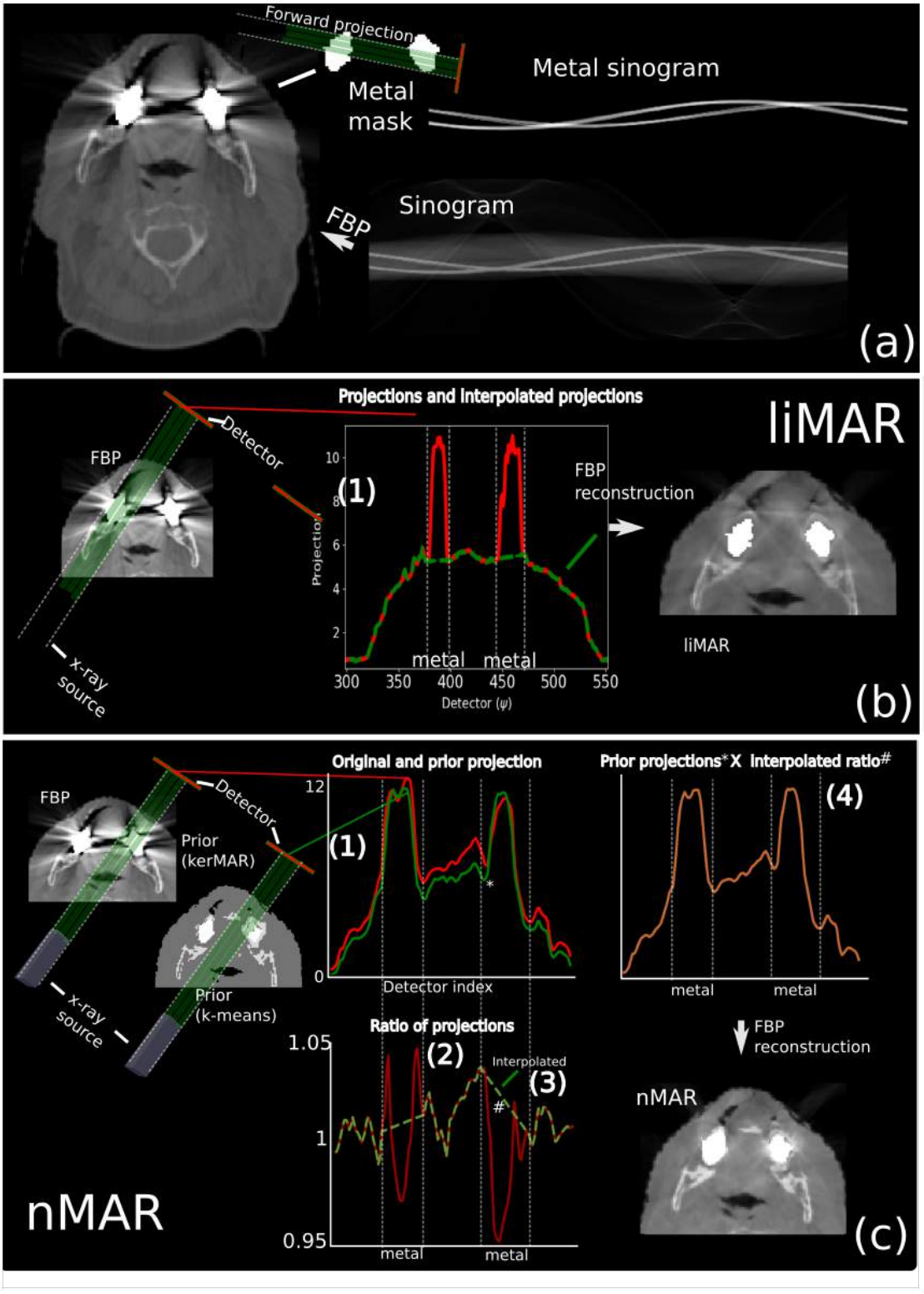


FIGURE 3.1: (a): In sinogram inpainting algorithms, the metal projections are often detected by forward projecting through a metal segmentation. (b): Linear interpolation MAR (liMAR) considers the measurements at each view angle and linearly interpolates over the detected metal projections. (c): Normalized MAR (nMAR) works as follows: 1) A prior image is generated, and a prior sinogram created by forward projecting through it; 2) The original sinogram is divided by the prior sinogram, leading to a ratio sinogram; (3) The ratios are interpolated over the metal projections, as in liMAR; (4) The prior sinogram is multiplied by the interpolated ratios, leading to a corrected sinogram that is finally reconstructed.

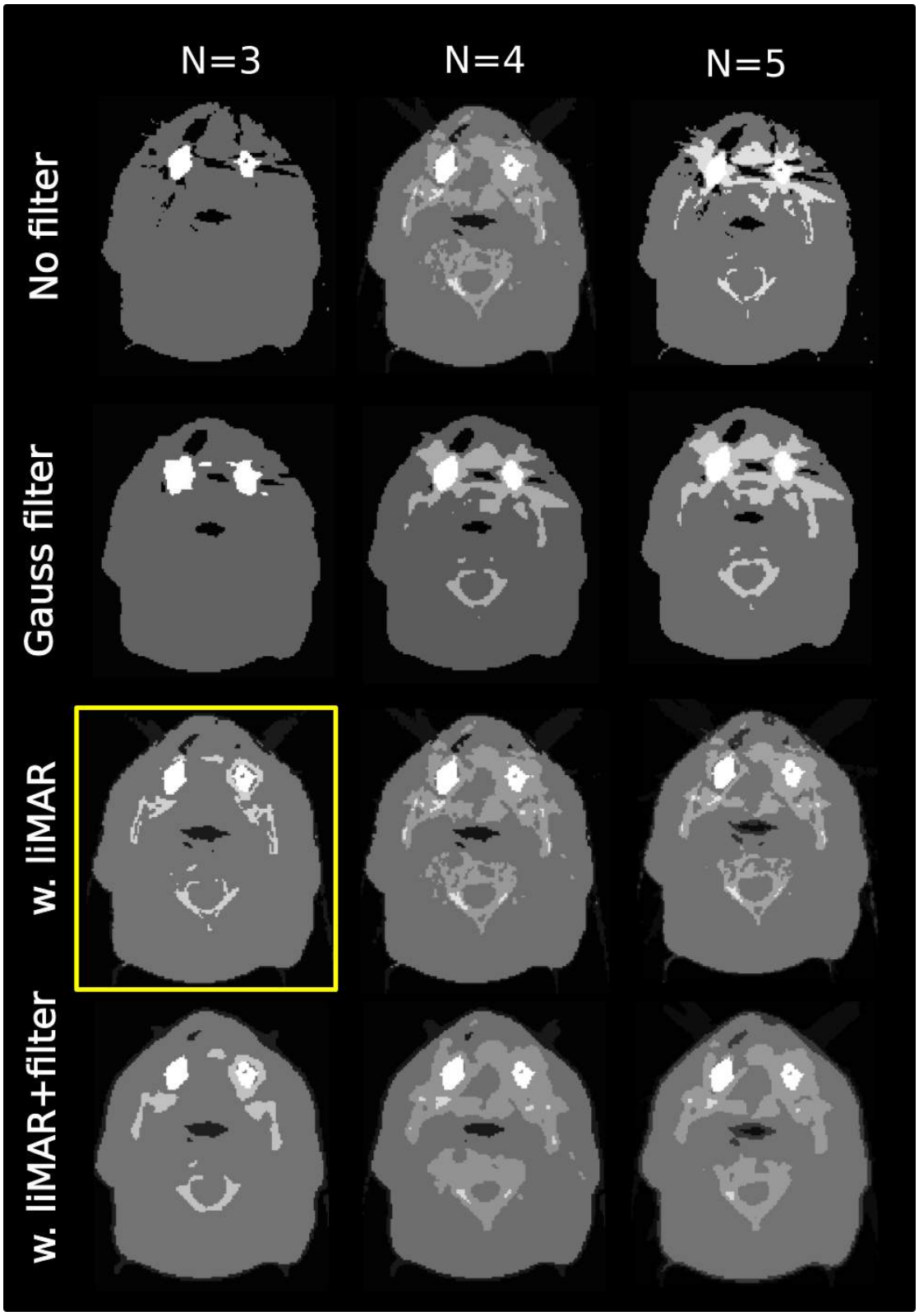


FIGURE 3.2: Prior images generated using K-means clustering[44] combined with different image processing steps. The metal implants were here segmented and kept out of the clustering, but reintroduced in post-processing. Top-bottom: No filter, a Gaussian filter, liMAR, liMAR with Gaussian filter. Left-right: Number of clusters in the K-means clustering. The box highlights our final choice for our implementations.

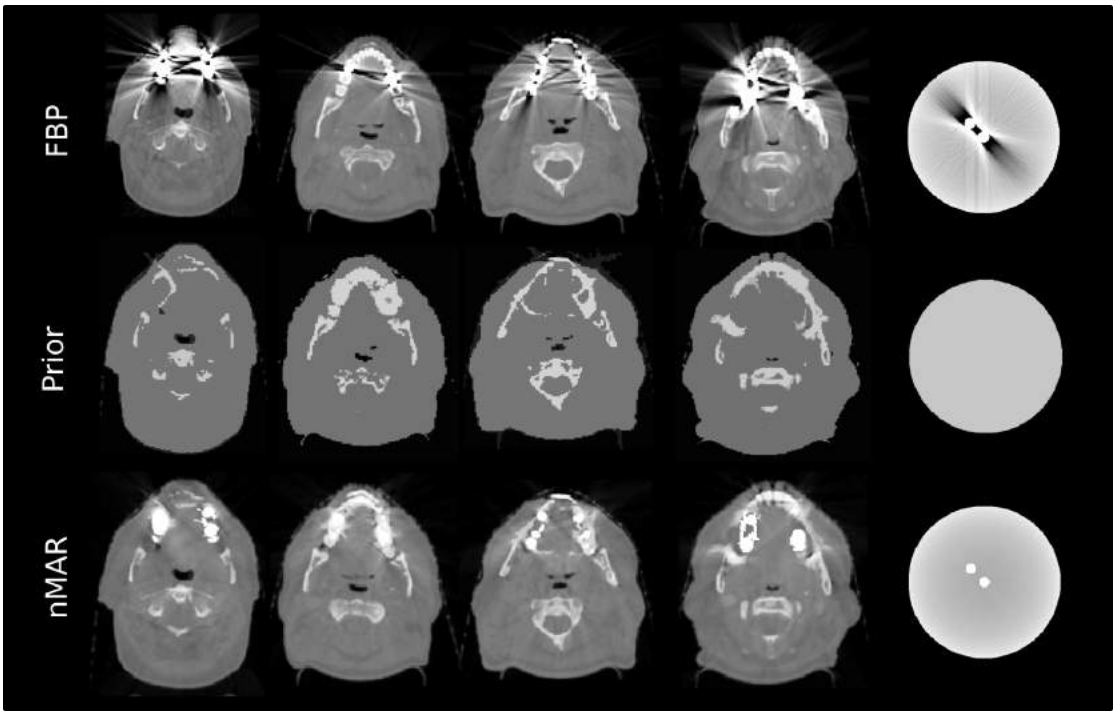


FIGURE 3.3: Normalized (nMAR) results for 4 head-and-neck patients and a simulated water phantom with iron inserts. The priors for nMAR were generated using liMAR and K-means clustering with 3 clusters. Note that the metal implants were not included in the priors, but were reintroduced in the post-processing of the nMAR images.

3.2.2.2 Sinogram inpainting

As illustrated in fig. 3.1(c), the second step of nMAR simulates a prior sinogram by forward projection through the prior image, which for a single view leads to the two detector profiles marked by (1). Dividing the prior sinogram by the original sinogram provides a set of ratios (2) that are interpolated over the metal projections as in liMAR (3). These ratios are multiplied on the prior projections, leading to the inpainted projections projections (4). Effectively, nMAR thus replaces the metal projections by scaled versions of the prior projections.

Fig. 3.3 shows the results of using nMAR on a few head-and-neck patients. The quality of the prior varies much between patients, which carries over to the corrected images. The performance of nMAR in conclusion depends sensitively on the prior, which in turn is not trivial to create based on the corrupted CT.

3.2.3 MAR for orthopedic implants (oMAR)

MAR for orthopedic implants (oMAR), which we considered briefly in the introduction (chapter 1), is a hybrid image processing and sinogram inpainting algorithm that works on the same basic principle as nMAR in that it creates a prior image by segmenting the image. It is important

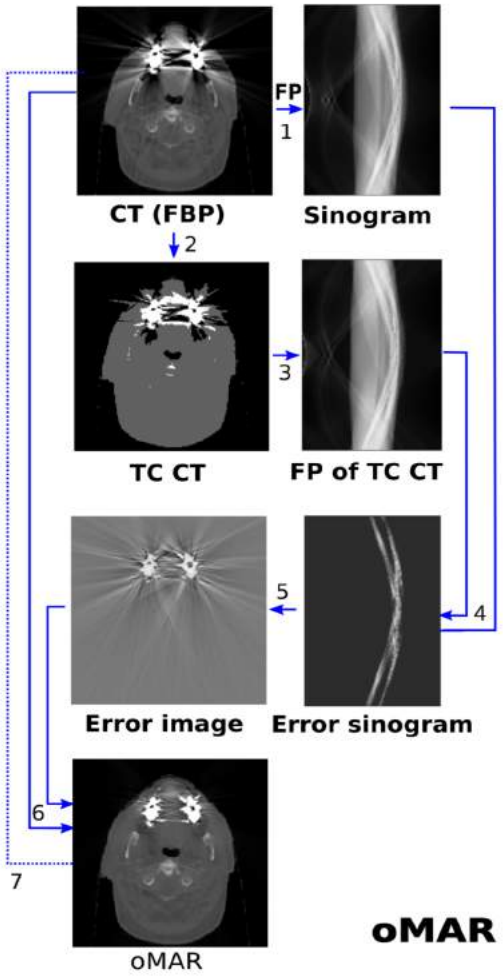


FIGURE 3.4: *Source:* Paper C. Schematic illustration of the Philips oMAR algorithm. oMAR proceeds as follows: 1) Forward project (FP) the input CT image (reconstructed by FBP) to simulate the sinogram; 2) segment the input CT to get a tissue classified prior image (TC CT); 3) forward project through the prior; 4) subtract the prior projection from the original sinogram; 5) reconstruct the error sinogram for an error image; 6) subtract the error image from the FBP; 7) replace the FBP with the updated image for iteration, or stop here.

for this thesis because it comes as a commercial add-on for use in, among other applications, radiotherapy[9].

oMAR differs from nMAR by working in an iterative fashion, as shown in fig. 3.4(a). First, it creates a prior image from the initial FBP by tissue classification, in particular by using an interpolation-based sinogram inpainting method similar to the nMAR prior generation we considered in the previous subsection. The prior provides a set of projection estimates, again similar to in nMAR, that are subtracted from the original sinogram to yield a difference sinogram. The difference sinogram is masked by a metal trace to remove non-metal datapoints, and then reconstructed, leading to a difference image. Adding this difference image to the original image finally removes some of the artifacts. As the result of this procedure is fed back into the algorithm over a number of iterations, the prior image is expected to improve[9].

We did not implement oMAR ourselves for this project, but show example results of the oMAR algorithm from the vendor-provided scanner software in the introduction, paper C and chapter 5 (fig. 5.3).

3.3 Model based iterative reconstruction

A common feature of the inpainting algorithms is their focus on pre- or post-processing the data while being compatible with the linear forward model, and thus fast reconstruction by FBP. This makes them versatile, easy to implement and, importantly, compatible with vendor-provided processing steps and optimized reconstruction software.

However, they may lack in accuracy as they do not address the ultimate cause of the artifacts. The *Model-based iterative reconstruction* (MBIR)[38, 45, 46] algorithms seek to do this by more accurately defining the reconstruction model. As we discussed in subsection 2.3.2, the measured x-ray intensities are inherently noisy observations due to the physics behind the x-ray generation and detection. MBIR methods therefore view the CT reconstruction as a parameter inference task from the conditional probability of the image given the data; using the vector representations of the x-ray measurements $\mathbf{n}$ and image $\mathbf{y}$ that we introduced in chapter 2:

$$p(\mathbf{y}|\mathbf{n}) \propto p(\mathbf{n}|\mathbf{y})p(\mathbf{y}), \tag{3.1}$$

which follows from Bayes theorem[47, 48]. The probability of the measurements given the unkown image, $p(\mathbf{n}|\mathbf{y})$, is the *likelihood* function while $p(\mathbf{y})$ is the *image prior distribution*. MBIR performs *Maximum a Posteriori* (MAP) estimation by maximizing the posterior distribution for the image $\mathbf{y}$. For algebraic and computational convenience, the log is usually taken prior to maximization as this does not change the maximum, leading to the following maximization problem:

$$\arg\max_{\mathbf{y}} \mathcal{C}(\mathbf{y}) \quad \text{with} \quad \mathcal{C}(\mathbf{y}) = \ln p(\mathbf{n}|\mathbf{y}) + \ln p(\mathbf{y}). \tag{3.2}$$

3.3.1 Maximum Likelihood Transmission Reconstruction

As discussed in subsection 2.3.2, a simple but accurate noise model, assuming a monochromatic x-ray source, is the Poisson distribution, which looks as follows[38, 49, 50]:

$$p(n_j|\mathbf{y}) = \frac{\lambda_j^{n_j}}{n_j!}\mathsf{e}^{-\lambda_j} \quad \text{with} \quad \lambda_j = C \cdot \mathsf{e}^{-l_{j,*}\mathbf{y}}. \tag{3.3}$$

Here, $l_{j,*}$ denotes a row of the system matrix, and its application to the image $\mathbf{y}$ calculates the projection at detector j. Many so-called *pre-log*[46] MBIR methods use this likelihood, in

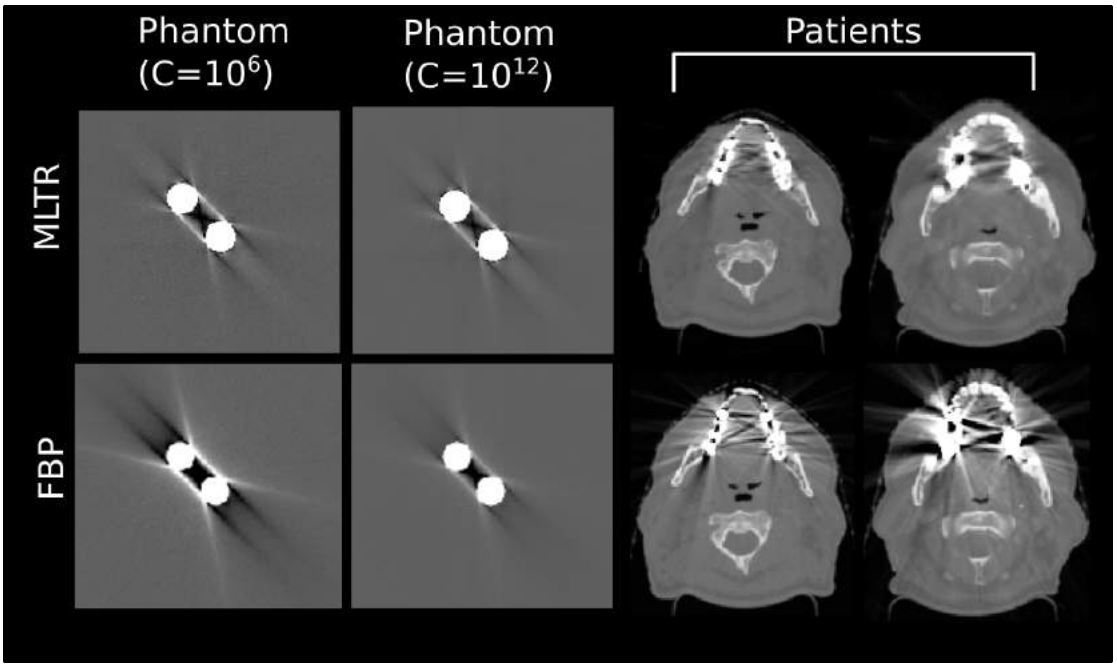


FIGURE 3.5: MBIR results for the digital phantom with iron inserts and two head-and-neck patients. Top: MLTR, with a noise model but no beam hardening model, leading to similar artifacts for both values of the background intensity C. Bottom: FBP, with no noise model and thus far more artifacts for $C = 10^6$ than 10^{12}

particular the *Maximum Likelihood Transmission Reconstruction* (MLTR) algorithm[38, 49, 50]. In its basic conception, MLTR assumes a flat prior ($p(\mathbf{y}) = 1$) and simply maximizes this likelihood, which explains its name; we will however use the version with a general prior and yet keep the name.

Efficiently maximizing the likelihood is not trivial using e.g. simple gradient ascent or Newton's Second order method[51]; MLTR therefore uses an Expectation-Maximization[52, 53]-like approach that leads to the additive update equation in algorithm 1. We here used the index convention and set definitions in chapter 2, stacking the image coefficients $\{y_i\}_{i\in\mathcal{T}}$, x-ray intensities $\{n\}_{j\in\mathcal{S}}$ and intensity estimates $\{\lambda_j\}_{j\in\mathcal{S}}$ in column vectors $\mathbf{y}$, $\mathbf{n}$ and $\boldsymbol{\lambda}$. $\mathsf{e}^{\{\cdot\}}$ is applied

Algorithm 1 Maximum Likelihood Transmission Reconstruction (MLTR) with an image prior

1: Transform the sinogram to intensities: $\mathbf{n} = C\mathsf{e}^{-\mathbf{p}}$, estimating C by NEC scaling.
2: Initialize $\mathbf{y}$ as a small-valued, homogeneous image; $\delta \leftarrow 1$.
3: Calculate $\boldsymbol{\alpha} = [\mathbf{L1}]$.
4: **while** $\delta > 10^{-6}$ **do**

$$\mathbf{y}_0 \leftarrow \mathbf{y} \tag{3.4}$$

$$y_i \leftarrow y_i + \frac{l_{*,i}^T[\boldsymbol{\lambda} - \boldsymbol{n}] + \frac{\partial \ln p(y_i')}{\partial y_i'}\big|_{y_i}}{l_{*,i}^T[\boldsymbol{\alpha} \odot \boldsymbol{\lambda}] + \frac{\partial^2 \ln p(y_i')}{\partial {y_i'}^2}\big|_{y_i}}, \quad \forall i \in \mathcal{T} \quad \text{with} \quad \lambda_j = C\mathsf{e}^{-\mathbf{l}_{j,*}\mathbf{y}}$$

$$\delta \leftarrow \frac{\sqrt{(\mathbf{y} - \mathbf{y}_0)^T(\mathbf{y} - \mathbf{y}_0)}}{|\mathcal{T}|} \tag{3.5}$$

5: **end while**

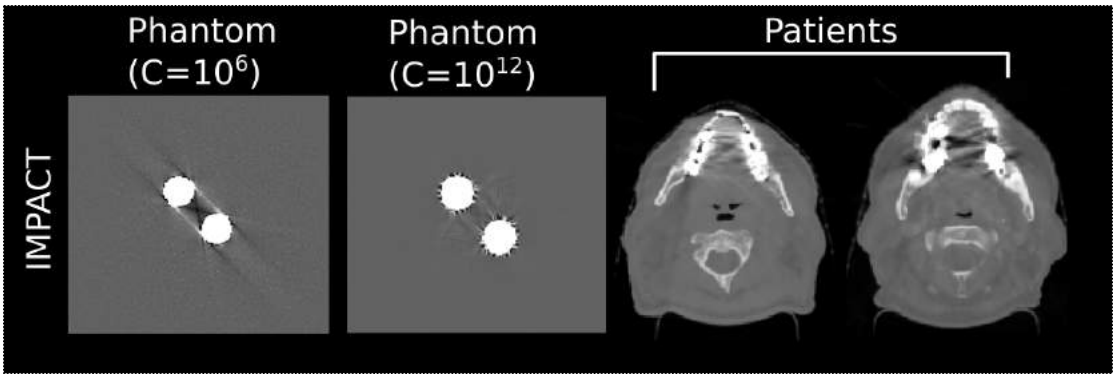


FIGURE 3.6: IMPACT, which includes a beam hardening model, leading to artifacts for $C = 10^6$ due to photon starvation but none for $C = 10^{12}$. The patients, however, are similar to the MLTR results in fig. 3.5.

element-wise, $\mathbf{1}$ is an image of ones and $l_{*,i}$ denotes a column of the system matrix $\mathbf{L}$. We used NEC scaling to pre-treat the data and make it closer to Poisson distributed, as inferior results have been reported with MLTR on non-Poisson data[49], and used the voxel-averaged magnitude of the iterative step as a stop-criterion.

The algorithm runs in parallel over the voxels $i \in \mathcal{T}$, and so calculates the full back projection $\mathbf{L}^T[\cdot]$ in each iteration through the applications of $l^T_{*,i}$. The likelihood step thus matches the back projection of the simulated intensities $\boldsymbol{\lambda}$ to that of the measured $\mathbf{n}$, i.e. the *summed* intensities. This is important for MAR applications because this linearly weighs the contribution of a measurement to the update equation according to its magnitude, such that the low SNR measurements through e.g. the highly attenuating metal get a relatively low weight.

Fig. 3.5 shows the results of applying MLTR for the noisy, simulated phantom with beam hardening. We considered both a low and a high noise level with emitted x-ray counts of $C = 10^6$ and $C = 10^{12}$, leading to different degrees of noise and photon starvation. MLTR achieved similar results for both C, contrary to FBP which shows a noticeable increase in streak artifacts due to photon starvation for $C = 10^6$. The remaining artifacts for both C are due to the beam hardening effect.

The figure additionally shows results for head-and-neck patients, for which we see a clear reduction in the metal artifacts compared to the FBPs (bottom row), which is similar to what we saw with the noisy phantom for $C = 10^6$ and is thus likely owed to better handling of the noise and photon starvation.

3.3.2 IMPACT (including a beam hardening model)

To include the beam hardening effect in the model, one may use the more general expression in eqn. (2.8) given knowledge of the x-ray source energy spectrum. This requires the definition of

some voxel-specific parameter that allows for the calculation of the energy-dependent attenuation coefficient $y_j(E)$ over the spectrum energies, which may be done in various ways; one method that builds upon the MLTR algorithm uses the following basis function expansion[29]:

$$y_j(E) = \phi_j \Phi(E) + \theta_j \Theta(E),$$

where $\Phi(E)$ and $\Theta(E)$ are proportional to the energy-dependent cross sections of the photoelectric absorption and Compton scatter that are the main x-ray attenuation events. This expansion is well-known from *dual-energy CT*, which acquires images with spectra at different modal energies to sensitize the acquisition to respectively the photoelectric and Compton scatter effects. ϕ_j and θ_j are voxel-specific functions that, depending on the tissue in the voxel, define the relative probabilities of the processes occuring given an attenuation event[29, 38].

Using empirical data for the energy dependent attenuation of different tissue types that are assumed present in the patient, ϕ_j and θ_j may be calculated given the attenuation coefficient $y_j(E_0)$ at in principle any energy E_0, upon which $y_j(E)$ may be calculated and used in the calculations of the MLTR update equation. This leads to the IMPACT (Iterative Maximum-likelihood Polychromatic Algorithm for CT) algorithm, which reconstructs the attenuation coefficients $y_j(E_0)$ at the chosen E_0; for details on this algorithm, we point to references [29, 38, 50].

We implemented IMPACT and ran it for the same cases as MLTR, using the simulated spectrum in fig. 2.7(a) and setting the effective energy to the mean energy, $E_0 = 75.2kEv$. From the results in fig. 3.6, we observe that the beam hardening model provided noticeable improvements compared to what we saw for MLTR with $C = 10^{12}$, but displayed only similar results for $C = 10^6$. The photon starvation thus dominated for $C = 10^6$.

Importantly, we observe the same for the patients, where IMPACT appears indistinguishable from MLTR in terms of the artifact reduction. This implies the importance of other effects that caused the sophisticated beam hardening model to provide limited benefits, such as the following:

1. **Inaccurate spectrum and implant knowledge:** Since we did not know the exact energy spectrum of the source, we ran the algorithm using a variety of simulated spectra, varying the thickness of the aluminium filter, the copper filter and the kVp. Since this did not lead to positive results, we additionally varied the implant material, experimenting with dental amalgam, composite resin, gold, titanium and porcelain. This was however to little avail.

2. **Convergence issues:** IMPACT can be slow to converge, which we addressed by implementing local model reconstruction as in ref. [50]. While this did have an effect on the results, they were not obviously positive for the patients.

3. **Photon starvation and other artifacts:** As discussed in chapter 2, other artifact sources, in particular photon starvation, may have been a factor, in which case, as we saw with the simulated phantom, the beam hardening model in IMPACT does not help.

Due to these factors, we chose in our investigations of MBIR using an MRI-based prior in papers A and B to focus on the faster and simpler MLTR algorithm.

Chapter 4

MRI-based Metal Artifact Reduction

4.1 The context of our model

While MRI-based MAR is a mostly untouched field, with only a handful existing algorithms[54–56] that we briefly discuss in paper B, MRI-based CT value prediction is the subject matter of pseudo-CT (pCT) generation for MRI-only RT, where entire CT images are simulated from MRIs[12, 57–60]; here, the difficulty of estimating CT values from MR intensities is a known issue: first, the so-called *voxel-based* pCT algorithms that use regression models or other intensity-based techniques for CT value prediction, such as segmentations and bulk CT value assignment, sometimes rely on specialized *Ultra-Short Echo Time* (*UTE*) sequences to increase the MRI bone/air contrast[59]. The *atlas-based* algorithms do not require such specialized sequences, and instead use a database of coregistered CTs and MRIs in which one may look up CT values for the different anatomical regions. Such regions may be encoded using image features in the MRI rather than single-voxel intensities, e.g. MR image patches (local spatial contexts)[12, 57, 58].

In the field of MRI-based MAR, we are fortunate as we not only have access to the MRI for the target patient, but also a CT. We may therefore use the patient volume itself to look for anatomical correspondences and associated CT-values. We thus, like the atlas-based pCT algorithms, avoid the need for a specialized MRI sequence but, like the voxel-based algorithms, do not need an actual atlas. We even have potentially exploitable CT information in the corrupted regions. The model that we present in this chapter capitalizes on these features by 1) using data from only the target patient and 2) incorporating the corrupted CT values in the regression model.

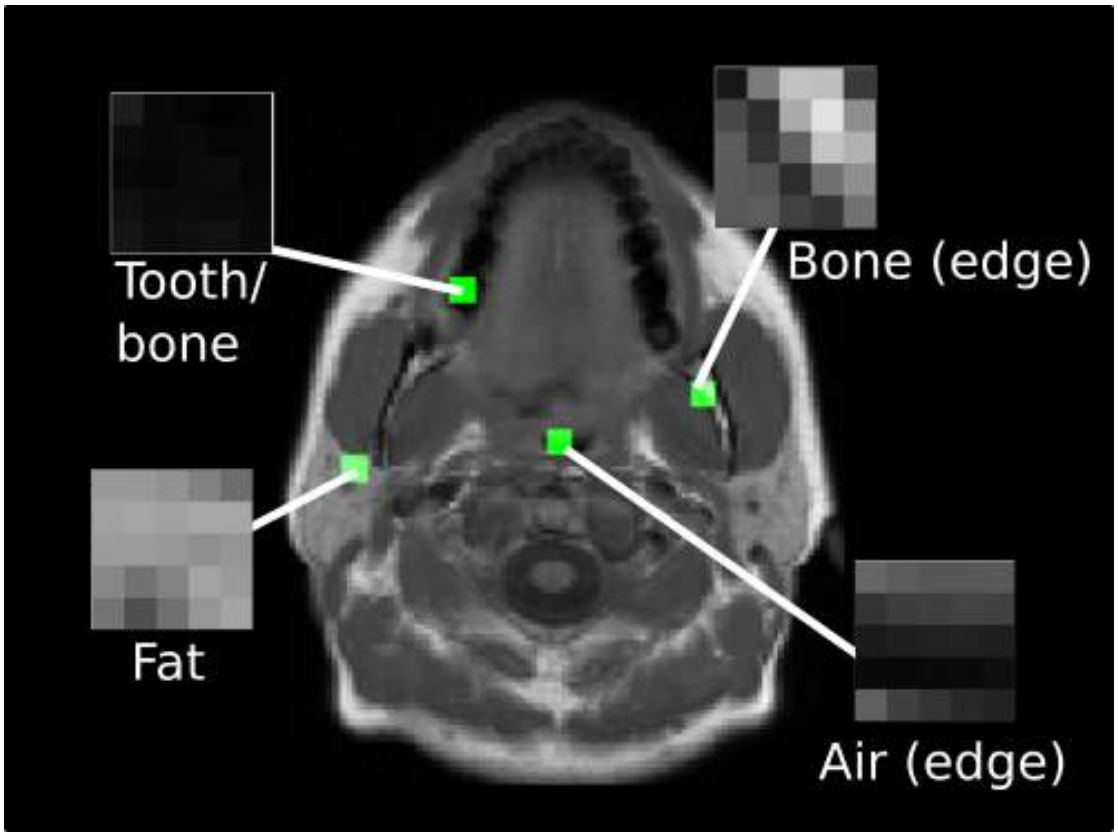


FIGURE 4.1: Illustration of feature-encoding with patches. While the central voxels in the patch do not contain much information, the wider patch may encode specific anatomical features.

4.1.1 Patches as feature encoders

We adopt the use of patches instead of single voxel intensities from certain atlas-based pCT algorithms[57, 58]. Patches address the contrast disparity issue by implicitly encoding anatomical features in the image, which are shared between modalities. Fig. 4.1 illustrates this concept: while the central voxel in each patch does not contain good information about the tissue type at their location, as evidenced by the similar contrast in e.g. the bone and air, the collections of voxels in the patches do. Using such patches may therefore provide a better CT value prediction model.

We thus extract patches from the MRI in the form of vectors $\{\mathbf{m}_i\}_{i\in\mathcal{T}}$ of MR intensities from local cubodial windows centered on the voxels, whose indices are stored in the set $\mathcal{T}$.

4.2 Bayesian generative modelling of CT and MRI images

A generative model defines the probability distribution assumed to have generated the observed data as well as the potentially unknown quantities[61]. From the MRI we have access to MR

image patches centered on every voxel in the image, i.e., using the index convention of chapter 2 where $\mathcal{T}$ contains voxel indices, $\{\mathbf{m}_i\}_{i\in\mathcal{T}}$. We additionally acquire an FBP CT and thus have access to the, potentially corrupted, FBP CT values $\{t_i\}_{i\in\mathcal{T}}$. The unknown parameters of the model are in turn the underlying true CT values, $\{y_i\}_{i\in\mathcal{T}}$.

For notational convenience, we will in the following let $\{\cdot\} \equiv \{\cdot\}_{i\in\mathcal{T}}$, i.e. let the absence of an index imply an index over all voxels in the volume.

4.2.1 The joint distribution

We now model the full joint distribution:

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

where we assumed the measurements for different voxels to be statistically independent, and $\boldsymbol{\beta} = \{\beta_t^*, \beta_y, \beta_m\}$ is the set of *hyperparameters* in the model, which we define later. We factorize the distribution as follows, using the product rule of probabilities[47, 48, 61]:

$$p(t_i, \mathbf{m}_i, y_i|\boldsymbol{\beta}) = p(t_i|y_i, \boldsymbol{\beta})p(\mathbf{m}_i, y_i|\boldsymbol{\beta}), \tag{4.1}$$

where we assumed t_i and $\mathbf{m}_i$ to be conditionally independent given y_i, i.e. that knowing the true CT value y_i entirely defines the distribution of the corrupted measurement t_i; thus, $p(t_i|\mathbf{m}_i, y_i, \boldsymbol{\beta}) = p(t_i|y_i, \boldsymbol{\beta})$. This assumption is reasonable since t_i arises from y_i in some distortion process that is independent of the MRI acquisition.

This factorization splits the generative model into an artifact noise model $p(t_i|y_i, \boldsymbol{\beta})$ and a lower-level generative model of CT and MRI patches, $p(\mathbf{m}_i, y_i|\boldsymbol{\beta})$. We model these in the next subsections.

4.2.2 Kernel density estimation

We learn $p(y_i, \mathbf{m}_i|\beta)$ empirically from samples of assumedly uncorrupted CT values and MRI patches, $\{y_n, \mathbf{m}_n\}_{n\in\mathcal{A}_i}$, with indices $\mathcal{A}_i \subseteq \mathcal{T}$. Fig. 4.2(a) illustrates how we pick this set by looking up patches in the MRI volume at locations where the FBP is uncorrupted; we return to this later. On this regression point set, we use a specific kind of *kernel density estimation* (KDE)[61]:

$$p(y_i, \mathbf{m}_i|\boldsymbol{\beta}) = \frac{1}{|\mathcal{A}_i|} \sum_{n\in\mathcal{A}_i} K(y_i, y_n; \mathbf{m}_i, \mathbf{m}_n|\boldsymbol{\beta}), \tag{4.2}$$

with the *kernel function* K.

For eqn. (4.2) to define a probability distribution, K must integrate to 1 over y_i and $\mathbf{m}_i$ and becomes itself a probability distribution. To interpret eqn. (4.2), we thus use the product rule to identify it as identical in shape to a mixture of $|\mathcal{A}_i|$ distributions, where $|\cdot|$ denotes set cardinality, that are conditional on the hidden variable n:

$$p(y_i, \mathbf{m}_i|\boldsymbol{\beta}) = \sum_{n \in \mathcal{A}_i} p(y_i, \mathbf{m}_i|n, \boldsymbol{\beta}) p(n|\boldsymbol{\beta}) \tag{4.3}$$

$$\text{with} \quad p(y_i, \mathbf{m}_i|n, \boldsymbol{\beta}) = K(y_i, y_n; \mathbf{m}_i, \mathbf{m}_n|\boldsymbol{\beta}) \quad \text{and} \quad p(n|\boldsymbol{\beta}) = \frac{1}{|\mathcal{A}_i|},$$

where we for notational convenience let the conditional on n contain the dependence on both $\mathbf{m}_n$ and y_n. This expression helps us decide how to model the kernel function: $p(y_i, \mathbf{m}_i|n, \boldsymbol{\beta})$ should provide the probability of the observation $(y_i, \mathbf{m}_i)$ given knowledge of n. Now suppose, as an example, that $\{y_n, \mathbf{m}_n\}_{n \in \mathcal{A}_i}$ only contains three voxels with three ideal types of tissue; e.g. one air voxel ($n = 1$), one pure water voxel ($n = 2$) and one cortical bone voxel ($n = 3$). In this situation, n directly encodes a tissue type, and $p(y_i, \mathbf{m}_i|n, \boldsymbol{\beta})$ should define the distribution of CT values and MRI patches in tissue n. Due to the poor CT value/MR intensity correlation[1] it is appropriate to assume that CT values and MRI patches are independent within a given tissue type, and we thus factorize this distribution into a separate CT and MRI component as $p(y_i, \mathbf{m}_i|n, \boldsymbol{\beta}) = p(y_i|y_n, \boldsymbol{\beta}) p(\mathbf{m}_i|\mathbf{m}_n, \boldsymbol{\beta})$.

Within the tissue type, we assume that the observations follow a Gaussian distribution for the CT and a multivariate Gaussian with a diagonal covariance for the MRI patches. The distribution defined by the KDE then becomes the following mixture of Gaussians:

$$p(y_i, \mathbf{m}_i|\boldsymbol{\beta}) = \frac{1}{|\mathcal{A}_i|} \sum_{n \in \mathcal{A}_i} \mathcal{N}\{y_i|y_n, \beta_y^{-1}\} \mathcal{N}\{\mathbf{m}_i|\mathbf{m}_n, \beta_m^{-1}\mathbf{I}_M\}. \tag{4.4}$$

$\mathcal{N}\{x|\mu, \sigma^2\}$ here denotes a Gaussian with mean μ and variance σ^2 while $\mathcal{N}\{\mathbf{x}|\boldsymbol{\mu}, \boldsymbol{\Sigma}\}$ is a multivariate Gaussian with mean $\boldsymbol{\mu}$ and covariance $\boldsymbol{\Sigma}$. β_y and β_m are the precisions (reciprocal variances) of the Gaussian kernels, and $\mathbf{I}_M$ is the M-dimensional identity vector where $M = |\mathbf{m}_i|$ is the number of voxels in a patch.

Fig. 4.2(b) plots eqn. 4.4 for $1 \times 1 \times 1$ patches ($M = 1$), in which case it becomes a 2D plane; in general, it would be an $M+1$ dimensional hyperplane. The red curve traced on the surface of the plane at a certain $\mathbf{m}_i$ corresponds to the conditional distribution $p(y_i|\mathbf{m}_i, \boldsymbol{\beta})$, i.e. the distribution of the CT value in voxel i given the observation of the patch $\mathbf{m}_i$. This distribution has multiple peaks that reflect the ambiguous relationship between CT values and MRI intensities.

[1] At least for low MR intensities, see fig. 1.3 in chapter 1.

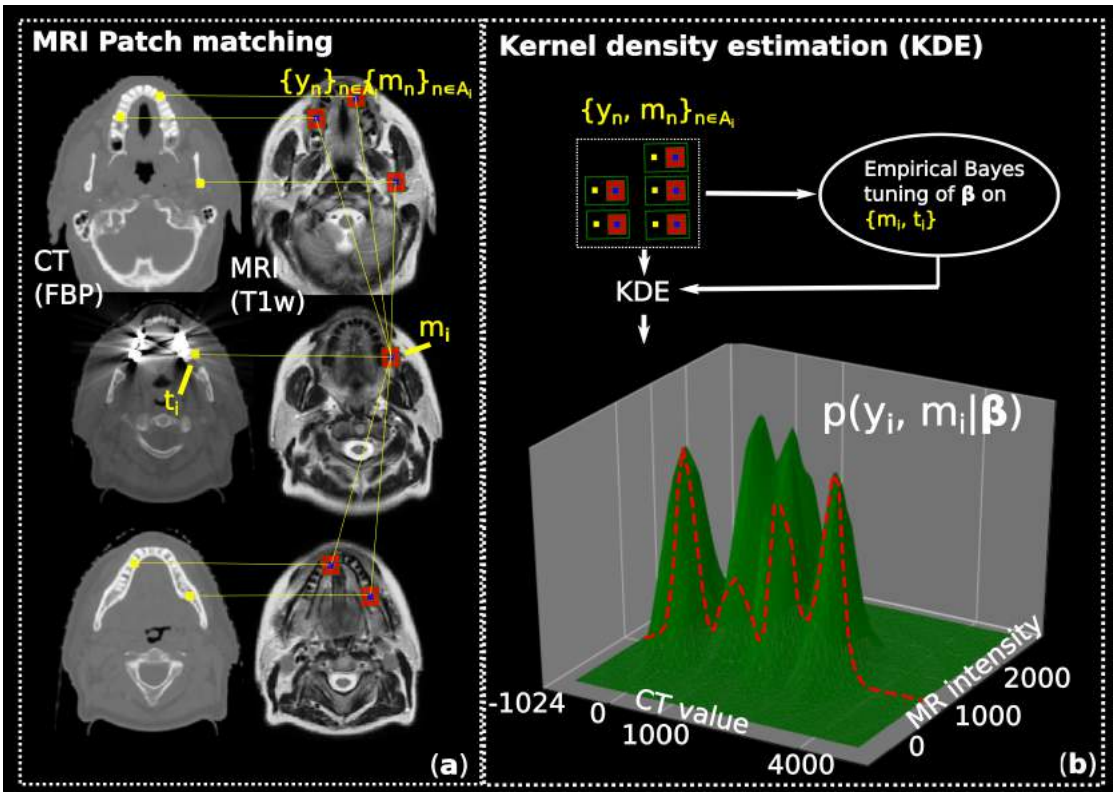


FIGURE 4.2: *Source:* Paper B.(a): For each corrupted voxel, a regression point set of CT value / MRI patch pairs, $\{y_n, \mathbf{m}_n\}$, is found by matching the observed MRI patch $\mathbf{m}_i$ to patches at uncorrupted locations (far from the implants). (b): On the regression point set, kernel density estimation (KDE) is used to estimate the joint distribution $p(y_i, \mathbf{m}_i|\boldsymbol{\beta})$ (showas a surface for $1 \times 1 \times 1$ patches). The KDE surface depends directly on the hyperparameters $\boldsymbol{\beta}$, which are tuned on the data using empirical Bayes estimation. The red curve is a trace on the surface at a specific $\mathbf{m}_i$, whose relevance is explained in fig. 4.4.

4.2.2.1 The kernel precisions

The kernel precisions β_m and β_y affect the shape of $p(y_i, \mathbf{m}_i|\boldsymbol{\beta})$ as illustrated in fig. 4.3, where we performed kernel density estimation on CT and MRI data simulated from a Gaussian mixture model similar to eqn. (4.3). We here used 1x1x1 patches and 4 tissue classes ($|\mathcal{A}_i| = 4$): $n = 1$: Air; $n = 2$: Water; $n = 3$: Bone; $n = 4$: Teeth (enamel). We chose the values of $\{y_n, m_n\}_{n \in \mathcal{A}_i}$ by considering an FBP CT and T1w MRI, and let $(\beta_y, \beta_m) = (10^{-6}, 10^{-6})$.

Using small precisions in the KDE here led to a smooth surface that does not capture the details of the distribution, while large precisions led to a too jagged surface with too many peaks and thus a highly ambiguous model. Choosing the hyperparameters used to generate the data led to something in between that appears to be a better description. This illustrates how, while the hyperparameters may be viewed as free parameters of the model, their choice affects the data fit and may as such be chosen non-arbitrarily to fit the model. This is an important point that we return to when we later consider their settings.

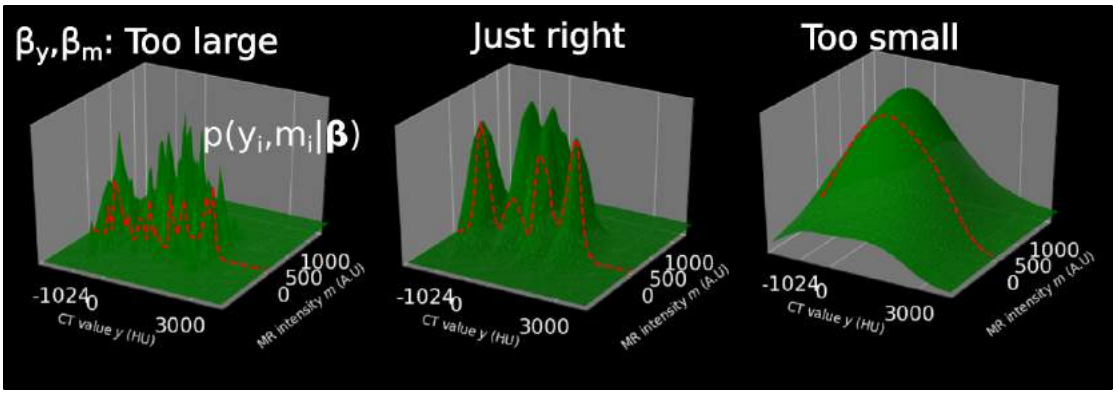


FIGURE 4.3: KDE surfaces for different settings of the kernel precisions β_y and β_m, calculated on simulated data from a Gaussian mixture model. Left: 10 times larger than the parameters used to generate the data; Center: equal; Right: 10 times smaller.

4.2.3 Artifact noise model

The artifact noise model for voxel i, $p(t_i|y_i, \boldsymbol{\beta})$, models the process by which the artifact corrupted measurement t_i arises from the true CT values y_i. In reality, the artifacts come from several processes in conjunction and vary in complicated ways over the volume in a manner that depends on extended parts of the true image, in particular the implant position and geometry[8, 29, 32]. Since taking such features into account is a difficult task, we model the features of the artifact noise in a phenomenological fashion: first, since both light and dark streaks arise from the artifacts in roughly equal proportions in the image, we choose a symmetric Gaussian model. Second, the artifact noise level decreases heavily with distance to the metal implants, suggesting a variation with voxel position $\mathbf{x}_i$ of the variance. We therefore define the artifact noise model as follows:

$$p(t_i|y_i, \boldsymbol{\beta}) = \mathcal{N}\{t_i|y_i, \beta_t^{-1}\} \quad \text{with} \quad \beta_t^{-1} = f(\mathbf{x}_i)\beta_t^{*-1}, \tag{4.5}$$

where $f(\mathbf{x}_i)$ is a function that modulates the variance of the artifact noise with the distance to the metal (which we segment using Otsu's method[62]), while β_t^* is an introduced hyperparameter of the model. We define $f(\mathbf{x}_i)$ as the following, decreasing sigmoidal:

$$f(\mathbf{x}_i) = 1 + \tanh\left(\frac{-D_\perp^2}{\kappa}\right), \tag{4.6}$$

where $D_\perp$ is the perpendicular distance to the set of metal voxels.

The noise model is illustrated in fig. 4.4(b) and (c). $f(\mathbf{x}_i)$ decreases from 1 to 0 with the distance to the metal, imposing a decrease of the variance from a maximal value of β_t^{*-1} in the highly corrupted regions near the metal. Far from the metal, the variance approaches 0, modelling the effect that the artifact noise is non-existent far from the metal.

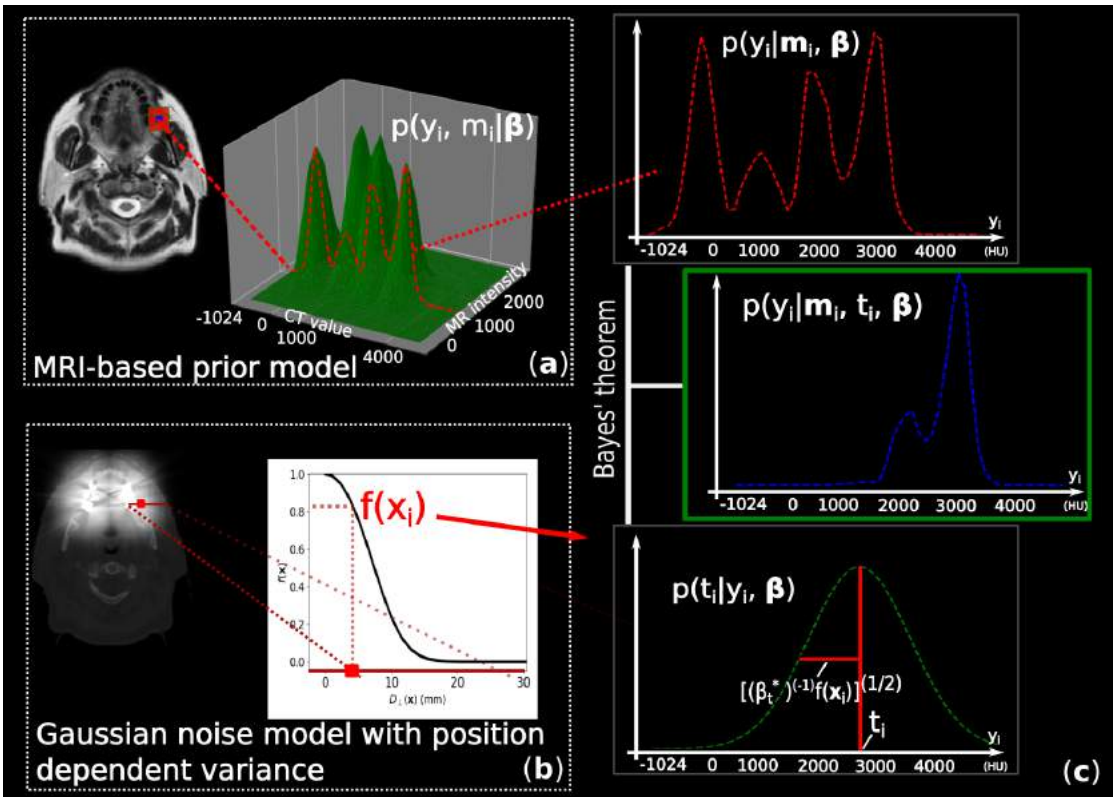


FIGURE 4.4: *Source*: Paper B (a): Given an observed MRI patch, the prior distribution $p(y_i|\mathbf{m}_i, \boldsymbol{\beta})$ corresponds to tracing a curve on the KDE hypersurface (fig. 4.2 (b)) at the observation. (b): The function $f(\mathbf{x}_i)$ decreases sigmoidally from the metal implants. Its value at the position of the voxel i, $\mathbf{x}_i$, is used to scale the variance of artifact noise model. (c): The artifact noise model $p(t_i|y_i, \boldsymbol{\beta})$ is Gaussian with a variance that decreases sigmoidally with the distance to the metal implants as $\beta_t^{-1} = \beta_t^{*-1} f(\mathbf{x}_i)$; the hyperparameter β_t^* is tuned on the data. The noise model acts as a likelihood function in y_i centered at t_i together with $p(y_i|\mathbf{m}_i, \boldsymbol{\beta})$ to define an improved posterior predictive model, $p(y_i|\mathbf{m}_i, t_i, \boldsymbol{\beta})$.

4.2.4 Posterior predictive distribution

Plugging eqns. (4.5) and (4.4) into eqn. (4.1), we now have a generative model describing our data. To use this distribution for CT value prediction, we calculate the *posterior predictive distribution* of the true CT value y_i given the rest of the observations. The distribution may be calculated in closed form using the basic relations of probability via the steps in paper B:

$$p(y_i|\mathbf{m}_i, t_i, \boldsymbol{\beta}) = \sum_{n \in \mathcal{A}_i} v_n^i \mathcal{N}\{y_i|\mu_n^i, (\beta_y + \beta_t)^{-1}\} \tag{4.7}$$

$$\text{with} \quad v_n^i = \frac{\mathcal{N}\{t_i|y_n, \beta_y^{-1} + \beta_t^{-1}\}\mathcal{N}\{\mathbf{m}_i|\mathbf{m}_n, \beta_m^{-1}\mathbf{I}_m\}}{\sum_{n' \in \mathcal{A}_i} \mathcal{N}\{t_i|y_{n'}, \beta_y^{-1} + \beta_t^{-1}\}\mathcal{N}\{\mathbf{m}_i|\mathbf{m}_{n'}, \beta_m^{-1}\mathbf{I}_m\}} \quad \text{and} \quad \mu_n^i = \frac{\beta_t}{\beta_t + \beta_y} t_i + \frac{\beta_y}{\beta_t + \beta_y} y_n. \tag{4.8}$$

Eqn. (4.7) depends on both t_i and the regression points $\{\mathbf{m}_n, y_n\}_{n \in \mathcal{A}_i}$: It is a mixture of Gaussians centered at $\{\mu_n^i\}_{n \in \mathcal{A}_i}$. It is perhaps easier interpreted by considering it in the following

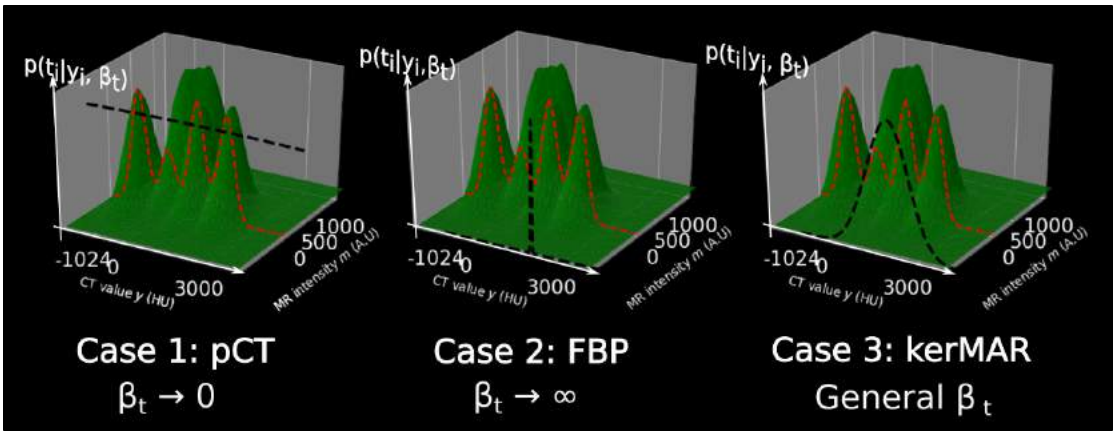


FIGURE 4.5: The artifact noise model $p(t_i|y_i, \boldsymbol{\beta})$ along the y_i axis of the KDE surface $p(y_i, \mathbf{m}_i|\boldsymbol{\beta})$ (supposing 1x1x1 patches), upon which $p(y_i|\mathbf{m}_i, \boldsymbol{\beta})$ is drawn. We show three case settings for β_t, which lead to respectively pCT (a purely MRI-based prediction), FBP (accepting the reconstructed CT values) and, in the general case, kerMAR.

shape, derived from Bayes' theorem:

$$p(y_i|\mathbf{m}_i, t_i, \boldsymbol{\beta}) \propto p(t_i|y_i, \boldsymbol{\beta})p(y_i|\mathbf{m}_i, \boldsymbol{\beta}).$$

Fig. 4.4(c) ilustrates how this expression corresponds to multiplying an effective MRI-based prior distribution $p(y_i|\mathbf{m}_i, \boldsymbol{\beta})$(fig. a), found by tracing a curve on the KDE hypersurface, and a likelihood function defined by the artifact noise model (fig. b+c). This shows how the observation of the corrupted CT value t_i "picks out" certain peaks from the ambiguous MRI-based prior distribution, leading to a predictive distribution with fewer, more accurate modes.

4.2.4.1 Image inpainting: Kernel regression MAR (kerMAR)

A Bayesian parameter estimate may be calculated from eqn. (4.7) by calculating the conditional expectation[48]:

$$\bar{y}_i = \int_{-\infty}^{\infty} y_i p(y_i|\mathbf{m}_i, t_i, \boldsymbol{\beta}) dy_i = \sum_{n \in \mathcal{A}_i} v_n^i \mu_n^i \quad \forall i \in \mathcal{T}. \tag{4.9}$$

This estimate defines an image inpainting MAR algorithm that we name *kerMAR* (kernel regression MAR). kerMAR bases its estimate on both the corrupted CT value t_i and the MRI patch $\mathbf{m}_i$, in a mixture that is directly determined by the choice of the hyperparameters $\boldsymbol{\beta} = \{\beta_t^*, \beta_y, \beta_m\}$.

4.2.4.2 kerMAR and pCT

The spatially varying artifact noise precision $\beta_t = f(\mathbf{x}_i)^{-1}\beta_t^*$ in particular defines the degree to which the CT measurement is included in the model; this is illustrated in fig. 4.5, where we show the noise model along the y_i-axis of the KDE surface for the following three special cases:

$\beta_t \to 0$ **(Case 1: CT measurement t_i fully corrupted)**

For this setting, the noise model is flat and eqn. (4.9) is calculated from the MRI-based prior distribution, leading to a purely MRI-based estimate:

$$\bar{y}_i^{\mathsf{pCT}} = \int_{-\infty}^{\infty} y_i p(y_i|\mathbf{m}_i, \boldsymbol{\beta}) dy_i = \sum_{n \in \mathcal{A}_i} w_n^i y_n^i \quad \text{with} \quad w_n^i = \frac{\mathcal{N}\{\mathbf{m}_i|\mathbf{m}_n, \beta_m^{-1}\mathbf{I}_M\}}{\sum_{n' \in \mathcal{A}_i} \mathcal{N}\{\mathbf{m}_i|\mathbf{m}_{n'}, \beta_m^{-1}\mathbf{I}_M\}}. \tag{4.10}$$

$\beta_t \to \infty$ **(Case 2: CT measurement t_i not corrupted)**

For this setting, the noise model is a δ-function at the FBP value t_i, which is therefore directly accepted without reference to the MRI:

$$\bar{y}_i = t_i.$$

$0 < \beta_t < \infty$ **(General case)**

For general hyperparameter settings, the estimate is calculated from the kerMAR estimate in eqn. (4.9) and is constructed in reference to both t_i and $\mathbf{m}_i$.

β_t thus determines where the model lies between pure MRI-based prediction and simply accepting all the observed CT values as reconstructed by FBP. Case 1 corresponds directly to MRI patch-based *kernel regression*, which has previously been used for atlas-based pseudo-CT (pCT) generation in MRI-only radiotherapy[57]. Case 2, on the other hand, corresponds to not touching the CT values reconstructed by the FBP. By scaling β_t with the distance to the metal, we impose an automatic transition from case 3 in the corrupted regions with $\beta_t \approx \beta_t^*$, to case 2 in the uncorrupted regions far from the artifacts.

4.3 Automatic hyperparameter choice and other decisions

We now have a predictive model suitable for MRI-based MAR in eqns. (4.7) and (4.9). Before they may be applied, we must however first define how to choose the following quantities (automatically or otherwise): the regression point sets $\{\mathcal{A}_i\}_{i \in \mathcal{T}}$; the dimension M of the MRI patches; and of course the hyperparameters $\boldsymbol{\beta}$.

4.3.1 Regression point sets and patch size

The voxel subset $\mathcal{A}_i \subseteq \mathcal{T}$ contains the indices of the KDE regression points $\{y_n, \mathbf{m}_n\}_{i \in \mathcal{A}_i}$. A possible choice is $\mathcal{A}_i = \mathcal{T}_u$, where $\mathcal{T}_u$ is an assumed uncorrupted part of the patient volume; we

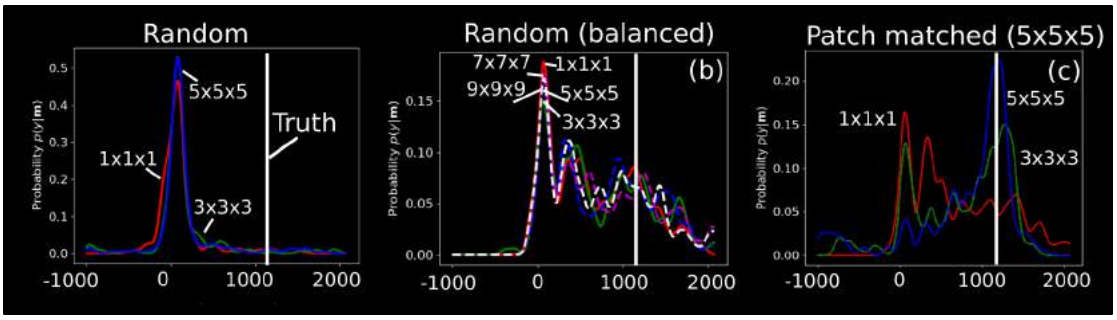


FIGURE 4.6: The conditional distribution $p(y_i|\mathbf{m}_i, \boldsymbol{\beta})$ for a tooth voxel ($HU \sim 1200$) shown for varying patch size and three different choices of regression point set clA_i. The CT value of the voxel is shown as a white bar. (a): Uniformly random sampling. (b): Balanced sampling from clusters determined by K-means clustering; (c): Using MRI patch matching (Fast PatchMatch)[63].

in practice find this set by thresholding $f(\mathbf{x}_i)$ so that $\mathcal{T}_u \equiv \{i \in \mathcal{T} | f(\mathbf{x_i}) < 0.5\}$, and thus choose voxels that are far from the metal. This set is however prohibitively large and time consuming to handle, and a sub-sampling strategy for picking $\mathcal{A}_i$ is necessary for the algorithm to be useful in practice.

4.3.1.1 Random sub-sampling strategies

One strategy is to pick a random subset $\mathcal{A}_i \subset \mathcal{T}_u$, which leads to a poor prior model $p(y_i|\mathbf{m}_i, \boldsymbol{\beta})$ for instructive reasons. Figs. 4.6(a) and (b) show the prior model for a voxel in the teeth ($HU \sim 1200$), varying the patch size while using two randomization strategies. β_m was chosen by the methods we describe later.

In (a), we picked the regression points uniformly in $3D$. This led to a preponderance of soft tissue indices with $HU \approx 0$ in $\mathcal{A}_i$, and thus a strong, but wrongly positioned, peak. The model does not improve with increased patch size as no patches in the regression point set are good correspondences.

In (b), we balanced the regression point set by segmenting the CT using K-means clustering and picking an equal, random number of points from soft tissue and bone. Here, the increase in patch size did not concentrate the wrong peak at $HU = 0$, as we saw with the unbalanced set, but also did not clearly help. This is because the weights w_n^i are determined by the relative magnitude of the squared l^2-norm between patches $(\mathbf{m}_i - \mathbf{m}_n)^T(\mathbf{m}_i - \mathbf{m}_i)$ that figures in the Gaussian kernels; the patches with particularly small values of this quantiy, which would get a large weight w_n^i, are rare and so were not discovered by the random search.

To find such patches with large weights, we therefore opted to find a set of patches that were similar to $\mathbf{m}_i$ in the said l^2 norm using a *patch matching algorithm*.

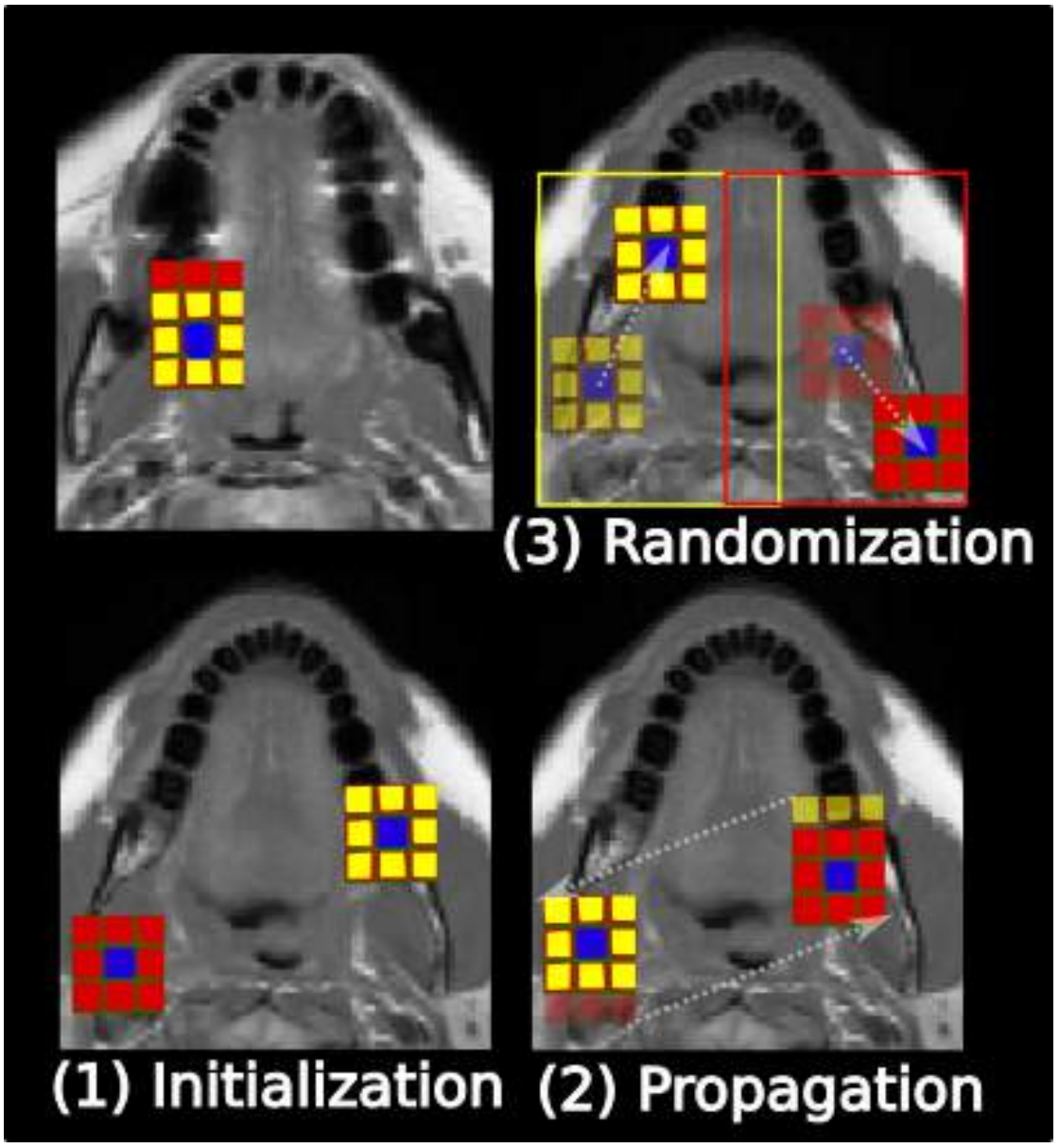


FIGURE 4.7: Illustration of the principle behind the patch matching algorithm (Fast PatchMatch) for one patch and its neighbor, assuming the discovered patches are the so-far best correspondences (1): In the initialization phase, a random patch is chosen for the target patch and its neighbors (one neighbor shown). (2): In the propagation phase, appropriately displaced versions of the patches found for the neighbors are considered; (3): In the randomization phase, a random patch is considered in a local window near the currently best patch.

4.3.1.2 Fast PatchMatch

We implemented an adaptation of the patch matching algorithm by Ta et al. in reference [63], which we refer to as *Fast PatchMatch*. The goal of the algorithm is to find patches for each point in the volume that are similar in the sense that the mentioned l^2 patch norm is small.

As illustrated in fig. 4.7 for two neighboring patches, the algorithm has three stages: (1) Initialization, (2) propagation and (3) randomization. In the initialization phase, a randomly chosen

regression point n is picked for each voxel i, and the l^2 patch norm is calculated between $\mathbf{m}_i$ and $\mathbf{m}_n$. The remaining two phases occur alternately during iteration: in the propagation phase, the current best regression points j for the 6 neighbors of each voxel i are evaluated by calculating the l^2 norm; fig. 4.7 assumes a better regression point was discovered for the two neighbors in this way. In the randomization phase, for each patch a random point is evaluated in the vicinity of the currently best regression point, which helps avoids local optima in the search.

The propagation step in particular lets the algorithm rapidly progress to better patch correspondences by using the fact that a good correspondence for a given patch is likely also good for its neighbor. In our final experiments, we used $T = 10$ iterations and ran the algorithm 20 times, saving the best point in each iteration to end up with a total of $\mathcal{A}_i = 200$ points. Such a regression point set led to the $p(y_i|\mathbf{m}_i, \boldsymbol{\beta})$ in fig. 4.7 (c), where the model noticeably improves with increasing patch size.

This also motivates us to use larger patch sizes. This however happens in a tradeoff with computation time; Ta et al. in ref. [63] found the best results (fastest and most accurate) for $5 \times 5 \times 5$ patches (with images of comparable resolution to ours), which is also the choice we made for our final experiments.

4.3.2 Empirical Bayes hyperparameter estimation

In theory, the hyperparameters $\boldsymbol{\beta}$ are unknown quantities and should therefore not be assigned a definite value but instead be modelled by a prior distribution $p(\boldsymbol{\beta})$, upon which it may be marginalized out of the predictive model as[48]:

$$p(\{y_i\}|\{\mathbf{m}_i, t_i\}) = \int_{\boldsymbol{\beta}} p(\{y_i\}|\{\mathbf{m}_i, t_i\}, \boldsymbol{\beta}) p(\boldsymbol{\beta}|\{\mathbf{m}_i, t_i\}) d\boldsymbol{\beta} \quad \text{with} \quad p(\boldsymbol{\beta}|\{\mathbf{m}_i, t_i\}) \propto p(\{\mathbf{m}_i, t_i\}|\boldsymbol{\beta}) p(\boldsymbol{\beta}).$$

It is however convenient to find a point estimate of $\boldsymbol{\beta}$, which also under certain assumptions may not be much inferior to the full treatment: *Empirical Bayes*[47, 48] hyperparameter estimation assumes that $p(\boldsymbol{\beta}|\{\mathbf{m}_i, t_i\})$ is a sharply peaked function, such that it approximately becomes a δ-function:

$$p(\{\boldsymbol{\beta}\}|\{\mathbf{m}_i, t_i\}) \approx \delta(\boldsymbol{\beta} - \boldsymbol{\beta}_M)$$

around its maximum, $\boldsymbol{\beta}_M$. In this case, we get:

$$p(\{y_i\}|\{\mathbf{m}_i, t_i\}) \approx \int_{\boldsymbol{\beta}} p(\{y_i\}|\{\mathbf{m}_i, t_i\}, \boldsymbol{\beta}) \delta(\boldsymbol{\beta} - \boldsymbol{\beta}_M) d\boldsymbol{\beta} = p(\{y_i\}|\{\mathbf{m}_i, t_i\}, \boldsymbol{\beta}_M).$$

With a large amount of data, i.e. a reasonably large number of voxels $|\mathcal{T}|$, the marginal data likelihood $p(\{\mathbf{m}_i, t_i\}|\boldsymbol{\beta})$ becomes sharp as it is the product of a corresponding number of distributions, and this assumption holds well.

Assuming a flat (constant-valued) prior $p(\boldsymbol{\beta})$, Empirical Bayes calculates $\boldsymbol{\beta}_M$ by maximizing the logarithm of the marginal data likelihood given the hyperparameters ($\log p(\{t_i, \mathbf{m}_i\}|\boldsymbol{\beta})$) [47, 48]. Using the expression for this marginal data likelihood shown in paper B (eqn. 9), which is arrived at by marginalizing the joint distribution eqn. (4.1), this leads to the following maximization problem for $\boldsymbol{\beta} = \{\beta_t^*, \beta_y, \beta_m\}$:

$$\begin{aligned} &\arg\max_{\boldsymbol{\beta}} \Phi(\boldsymbol{\beta}) \quad \text{with} \quad \Phi(\boldsymbol{\beta}) = \sum_{i\in\mathcal{T}} \sum_{n\in\mathcal{A}_i} \log \phi_n^i(\boldsymbol{\beta}) \quad \text{and} \\ &\phi_n^i(\boldsymbol{\beta}) = \mathcal{N}\{t_i|y_n, f(\mathbf{x}_i)\beta_t^{*-1} + \beta_y^{-1}\}\mathcal{N}\{\mathbf{m}_i|\mathbf{m}_n, \beta_m^{-1}\mathbf{I}_M\}], \end{aligned} \tag{4.11}$$

In defining $\Phi(\boldsymbol{\beta})$, we here assumed a constant $|\mathcal{A}_i|$ that was accordingly dropped as a constant term.

4.3.2.1 Expectation Maximization for hyperparameter estimation

The objective function cannot be maximized in closed form, and so we seek an iterative algorithm that defines the mapping $\boldsymbol{\beta}_{k+1} \leftarrow \boldsymbol{\beta}_k$ for a sequence of hyperparameter estimates $\{\boldsymbol{\beta}_k\}$ that converge to the optimal solution ($\boldsymbol{\beta}_M$). We achieve this using *Expectation Maximization* (EM)[52, 53], which in each iteration maximizes a lower-bound $l(\boldsymbol{\beta}|\boldsymbol{\beta}_k)$ to the objective that is equal to it at the current estimate ($l(\boldsymbol{\beta}_k|\boldsymbol{\beta}_k) = \Phi(\boldsymbol{\beta}_k)$). Maximizing it is thus guaranteed to increase the value of the objective function[53].

We therefore apply *Jensen's inequality*[52, 53]:

$$\log \frac{\sum_{n\in\mathcal{A}_i} v_n^{i\,(k)} \phi_n^i}{\sum_{n\in\mathcal{A}_i} v_n^{i\,(k)}} \geq \frac{\sum_{n\in\mathcal{A}_i} v_n^{i\,(k)} \log \phi_n^i}{\sum_{n\in\mathcal{A}_i} v_n^{i\,(k)}},$$

which holds for any sequence $\{\phi_n^i\}_{n\in\mathcal{A}_i}$ (i.e. the summand in eqn. (4.11)). We here defined $v_n^{i\,(k)} = v_n^i(\boldsymbol{\beta}_k)$ as the weights in eqn. (4.7) evaluated at the hyperparameter estimate, $\boldsymbol{\beta}_k$. Introducing $v_n^{i\,(k)}$ into the objective by simultaneously multiplying and dividing it, using the fact that $\sum_{n\in\mathcal{A}_i} v_n^i = 1$ and dropping some constants terms, we end up with the following lower bound $l(\boldsymbol{\beta}|\boldsymbol{\beta}_k)$:

$$\Phi(\boldsymbol{\beta}) = \sum_{i\in\mathcal{T}} \log\left(\sum_{n\in\mathcal{A}_i} \frac{v_n^{i\,(k)} \phi_n^i}{v_n^{i\,(k)}}\right) \geq \sum_{i\in\mathcal{T}} \sum_{n\in\mathcal{A}_i} v_n^{i\,(k)} \log \phi_n^i \equiv l(\boldsymbol{\beta}|\boldsymbol{\beta}_k).$$

The Expectation (E)-step of our EM algorithm evaluates the weights $v_n^{i\,(k))}$ to define the lower bound, while the Maximization (M)-step maximizes it.

The potential benefit of EM over conventional gradient-based optimization such as Newton's second order method arises when the M-step becomes fast and simple, preferably by having

a closed-form solution[52]. Due to the function $f(\mathbf{x}_i)$, this is unfortunately not the case for our problem, but, as we show in appendix A.4 of paper B where we derive the algorithm, we got around this issue by using the following approximation: defining the corrupted set as $\mathcal{T}_c \equiv \{i \in \mathcal{T} : f(\mathbf{x}_i) > 0.5\}$ and the uncorrupted set as $\mathcal{T}_u \equiv \{i \in \mathcal{T} : f(\mathbf{x}_i) \leq 0.5\}$, we truncate $f(\mathbf{x}_i)$ such that $\forall i \in \mathcal{T}_c$, $f(\mathbf{x}_i) = 1$, and $\forall i \in \mathcal{T}_u$, $f(\mathbf{x}_i) = 0$. Since $f(\mathbf{x}_i)$ is a relatively sharp sigmoid, this has only a small impact on the results[2].

Our final hyperparameter estimation algorithm is shown in alg. 2, which has been reformulated without the iteration index k for clarity of expression. Its function may easily be interpreted: The kernel variance of the uncorrupted CT values β_y^{-1} is estimated by a weighted variance over the uncorrupted volume; next, the artifact noise variance in the corrupted region β_t^{*-1} is estimated as the added variance in the corrupted region over β_y^{-1}; and finally the MRI variance is determined through a weighted variance over the full volume.

As an example of how this scheme tunes the model to the individual patient case, consider the case where the FBP image in the corrupted region, $\{t_i\}_{i\in\mathcal{T}_c}$, has severe artifacts. Then, $(t_i - y_n)^2$ will tend to be large as t_i varies due to artifacts, and β_t^* becomes small. This widens the Gaussian distortion model, which, referring to fig. 4.5, brings the final predictive model closer to the purely MRI-based one in case 1.

Algorithm 2 Empirical Bayes Hyperparameter estimation

1: Choose an initial estimate of the hyperparameters (e.g. $\boldsymbol{\beta} \leftarrow \{0, 0, 0\}$, and set $\delta \leftarrow 0$
2: **while** $\delta > 10^{-3}$ **do**
3: $\boldsymbol{\beta}_0 \leftarrow \boldsymbol{\beta}$
4: **E-step**: Calculate v_n^i using eqn. (4.8).
5: **M-step**: Update the hyperparameter estimates:

$$[\beta_y]^{-1} \leftarrow \frac{1}{|\mathcal{T}_u|} \sum_{i\in\mathcal{T}_u} \sum_{n\in\mathcal{A}_i} v_n^i (t_i - y_n)^2$$

$$[\beta_t{}^*]^{-1} \leftarrow \frac{1}{|\mathcal{T}_c|} \sum_{i\in\mathcal{T}_c} \sum_{n\in\mathcal{A}_i} v_n^i (t_i - y_n)^2 - [\beta_y]^{-1}$$

$$[\beta_m]^{-1} \leftarrow \frac{1}{|\mathcal{T}|} \sum_{i\in\mathcal{T}} \sum_{n\in\mathcal{A}_i} v_n^i (\mathbf{m}_i - \mathbf{m}_n)^T (\mathbf{m}_i - \mathbf{m}_n).$$

$\delta \leftarrow \sqrt{(\boldsymbol{\beta} - \boldsymbol{\beta}_0)^T (\boldsymbol{\beta} - \boldsymbol{\beta}_0)}/3$
6: **end while**

4.4 Summary of the predictive model

We have now defined our MRI-based generative model and how to choose its parameters. We summarize how to calculate the posterior predictive distribution in algorithm 3 (copied from

[2]For general (even non-analytical) $f(\mathbf{x}_i)$, the M-step may be performed iteratively using e.g. Newton's Second order method.

paper B). After arriving at point 8, the posterior is available for use in the CT-value prediction

Algorithm 3 Calculation of the posterior predictive distribution

1: Calculate the FBP.
2: Threshold the FBP using Otsu's heuristic to define a metal segmentation.
3: Calculate $f(\mathbf{x}_i) = 1 + \tanh(-\frac{D_\perp(\mathbf{x}_i)^2}{\kappa})$ ($\kappa = (10mm)^2$) using the metal segmentation.
4: Threshold $f(\mathbf{x}_i)$ (≤ 0.5) to define the set of uncorrupted voxel indices, $\mathcal{T}_u$.
5: Find $\mathcal{A}_i \subseteq \mathcal{T}_u$, $\forall i \in \mathcal{T}$, using Fast PatchMatch.
6: Store $\mathcal{A}_i$ along with $(\mathbf{m}_i - \mathbf{m}_n)^T(\mathbf{m}_i - \mathbf{m}_n)$ ($\forall n \in \mathcal{A}_i$).
7: Estimate $\boldsymbol{\beta}$ using algorithm 2.
8: The posterior may now be evaluated using the analytical expression in eqn. (4.7).

task, which we consider in the next section.

4.5 Experiments with MAR using the MRI-based predictive model

We experimented with three different metal artifact reduction algorithms we derived from our model: an image inpainting algorithm (kerMAR, as defined in eqn. (4.9)); a sinogram inpainting algorithm that uses kerMAR as a prior for nMAR; and an MBIR algorithm that uses our predictive model (eqn. (4.7)) as an image prior distribution for MLTR. In this section we define these algorithms, show results for nine head-and-neck patients and discuss their performance.

4.5.1 Technical details

We report the experimental results from paper B. We used regression sets of size $|\mathcal{A}_i| = 200, \forall i \in \mathcal{T}$, resulting from $T = 10$ iterations of Fast PatchMatch repeated 20 times, saving the 10 best patch correspondences in each repetition. We used cuboidal patches with 5 voxels on each side on image sets with a resolution of $1.2 \times 1.2 \times 2.0mm$ (CT) and $0.5 \times 0.5 \times 5.5mm$ (T1w MRI). The CT values were stored in HU + 1024, and the MRIs were resampled to the CT resolution after coregistration using mutual information. The spatial dimensions of the patches were thus $6 \times 6 \times 10mm$, comparable to the size of a tooth. The parameter κ in $f(\mathbf{x}_i)$ was chosen as $(10mm)^2$ for all patients. For additional details on the image set (scanner models, sequence parameters, etc.), refer to appendix A.1 of paper B.

4.5.2 Image inpainting by Bayesian estimation

Our first method, kerMAR, uses pure image inpainting to replace corrupted CT values with Bayesian estimates calculated from eqn. (4.9). The main novelties of kerMAR are that it 1)

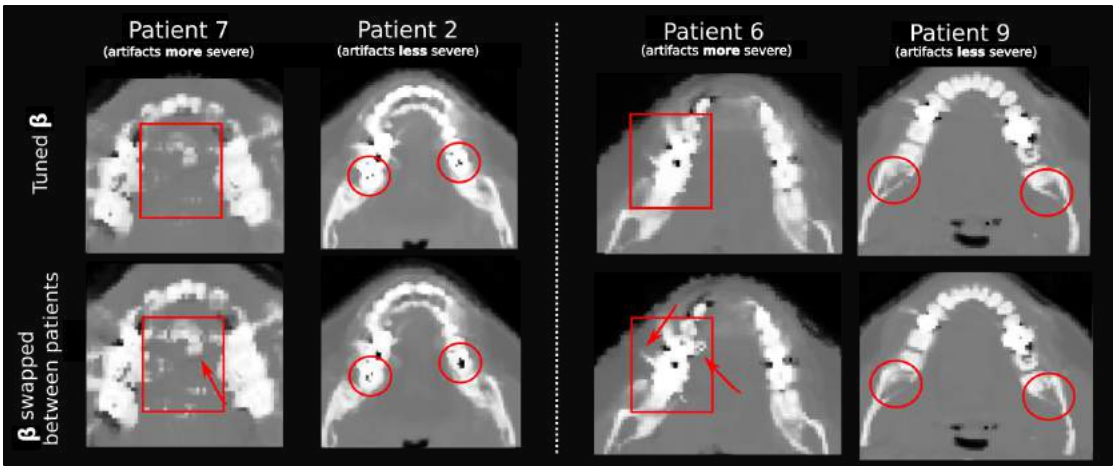


FIGURE 4.8: *Source*: Paper B. kerMAR calculated for 4 head-and-neck patients using (top) the tuned hyper-parameters $\boldsymbol{\beta}$ and (bottom) hyper-parameters swapped (7 with 2, 6 with 9). Patients 7 and 6 showed more severe artifacts than respectively 2 and 9. The arrows point to increases in anatomical errors and artifacts, which occurred for patients 7 and 6 due to a relative under-estimation of the artifact noise level. The circles show bone/air/soft tissue disambiguation issues, which occured for patients 2 and 9 due to a relative *over*-estimation of the artifact noise level.

combines both the corrupted CT measurements t_i and the regression points $\{y_n\}_{i\in\mathcal{A}_i}$ in its estimate in a mixture determined by the hyperparameters, and 2) tunes said hyperparameters on the data, thus adapting the model to each specific patient.

We compared kerMAR to purely MRI-based inference (pCT), i.e. eqn. (4.10), with the following modification: rather than forcibly letting $\beta_t \to 0$ for all voxels, we simply downscaled β_t^* by a large factor (1000), thus restricting the artifact reduction to the corrupted region. This would be achieved by some means in a serious application to MAR, and thus creates a fairer basis for evaluating the inclusion of a distortion model, which is the main methodological novelty of kerMAR.

4.5.2.1 Influence of the patient-specific hyperparameter tuning

The results of applying kerMAR and pCT are shown in figs. 4.9 and 4.10. The Empirical Bayes hyperparameter estimation provided the following average values ± standard deviations over the patients of $\langle\beta_y\rangle = (5.8 \pm 5)\cdot 10^{-4}$, $\langle\beta_t^*\rangle = (7.6 \pm 4.3)\cdot 10^{-6}$ and $\langle\beta_m\rangle = (14.7 \pm 12)\cdot 10^{-7}$. While the relative values of the three parameters differ by orders of magnitude, there is thus a rather large amount of variation between the patients, showing how the algorithm adapted them to each patient case. To investigate the influence of this patient-specific tuning, fig. 4.8 shows the kerMARs for two patients with their hyperparameters swapped (fig. 4.8). One of the patients here displayed more artifact corruption than the other, and thus had a small β_t^*, while the other displayed less corruption and thus had a larger β_t^*.

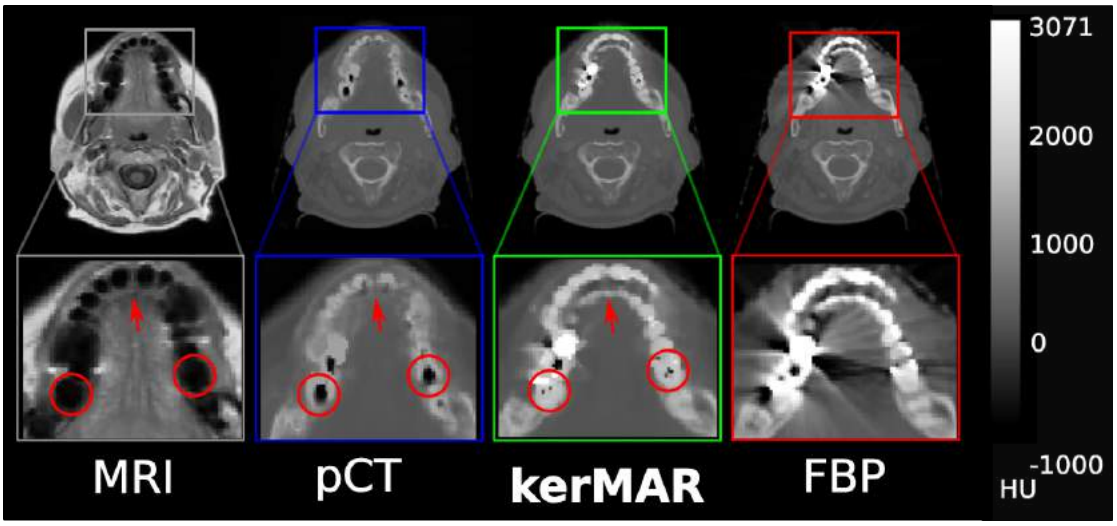


FIGURE 4.9: *Source:* Paper B. Axial slices of kerMAR and pCT (kerMAR with infinite artifact noise variance) shown together with the FBP for a head-and-neck patient. The tendency to miss-classify bone with pCT is much decreased through the use of the CT information using kerMAR (blue arrows), as are anatomical errors due to an imperfect co-registration (red arrow). The latter may be seen by comparing the MRI to the FBP, noting how the pCT imposes the MRI anatomy to a larger degree than kerMAR.

For the more corrupted patient, swapping the parameters led to decreased artifact reduction as the precision β_t^*, and thus the effective trustworthiness of the corrupted FBP CT values, was overestimated. For the less corrupted patient, on the other hand, swapping led to worse bone/air/soft tissue disambiguation, and filling in of small air pockets and solid bone regions, as the trustworthiness of the FBP was underestimated.

An additional benefit of the patient-specific tuning is the automatic handling of systematic variations in, especially, the MRI intensities due to variations in the MR gradient sequence parameters.

4.5.2.2 Handling of the contrast disparity

The main benefit of including the corrupted FBP CT value in the estimation may be seen by comparing the results for purely MRI-based pCT and kerMAR in fig. 4.9. The regions marked by red circles are tooth regions of a thickness comparable to the patch size, and so the pCT introduced air voxels. Including the CT value in the prediction with kerMAR, on the other hand, drastically reduced this effect. The same feature may be seen for multiple patients in fig. 4.10, again shown by red rings.

4.5.2.3 Solving coregistration issues

The case in fig. 4.9 also shows clear alignment/coregistration issues between MRI and FBP. This was mostly due to a difference in the resolution between the MRI and CT, which was respectively

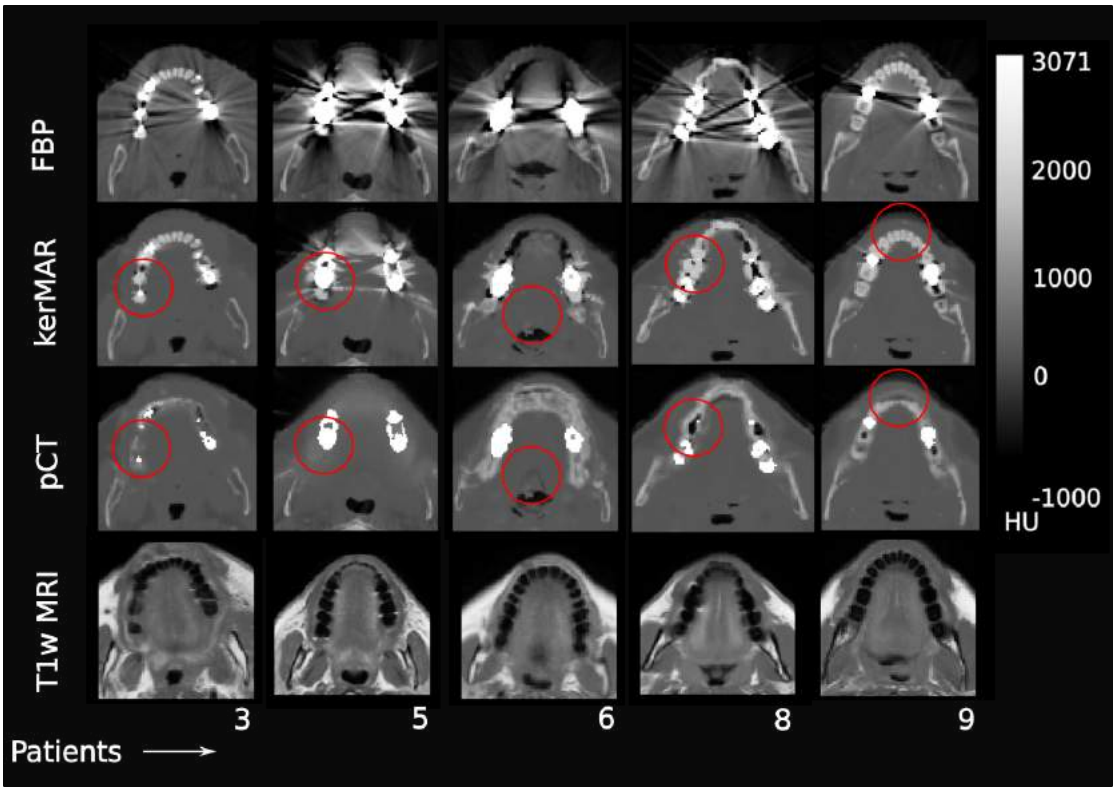


FIGURE 4.10: *Source:* Paper B. Additional axial slices for 5 patients, showing results for FBP, kerMAR and oMAR.

5.5mm and 2.0mm in the direction from head to toe in the patient. As shown by the red arrow, pCT imposes the MRI geometry, while kerMAR succesfully uses the MRI for artifact reduction while preserving the CT anatomy. This recurs in the additional images in fig. 4.10, in particular for patient 6.

4.5.3 Sinogram inpainting with MRI-based prior

As mentioned in chapter 3, image inpainting algorithms may be combined with sinogram inpainting algorithms for improved results. We therefore used our image inpainting method kerMAR to create a prior for the normalized MAR (nMAR)[41–43] algorithm.

4.5.3.1 Improvement over the CT-based prior

We compared using our MRI-based prior with the nMAR algorithm to using conventionally generated priors. Fig. 4.11 illustrates the potential benefit of this method, which we call nMAR-k: flaws in the prior for conventional nMAR, which generated from a corrupted CT, introduced artifacts to the image that were not present when using the MRI-based prior. This occured for several patients, as may be seen in the additional images in fig. 4.12.

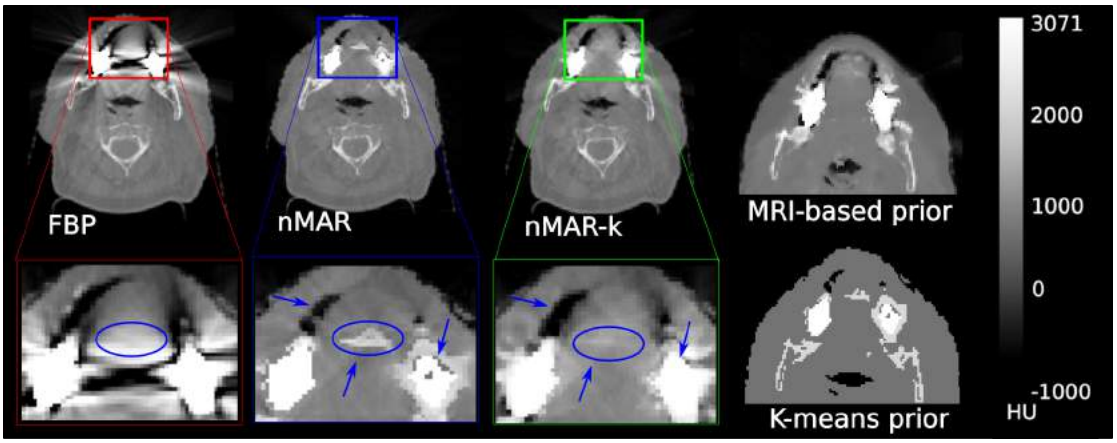


FIGURE 4.11: *Source:* Paper B. Results using the nMAR sinogram inpainting algorithm with (middle) a K-means clustering[42, 61] based prior and (center-right) with the kerMAR image as prior. Shown for reference is the FBP images (left) and the priors (far right). Comparing nMAR and nMAR-k, the blue arrows show artifacts/anatomical deformations that were either introduced or left behind by K-means nMAR due to a flawed prior, but that are absent with nMAR-k.

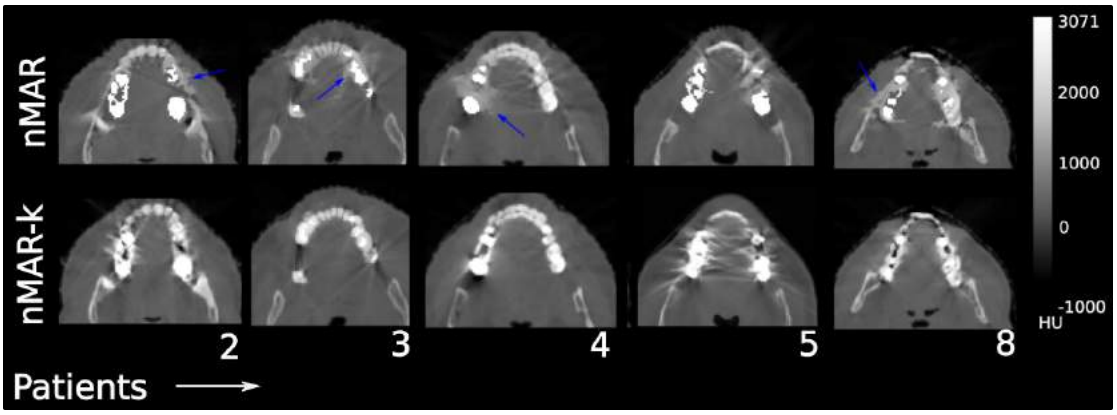


FIGURE 4.12: *Source:* Paper B. Additional axial CT slices for 5 patients. Shown are results for nMAR and nMAR-k.

4.5.3.2 Benefits over image inpainting

As we saw in the previous subsection, kerMAR worked relatively well by itself. There are however clear benefits to including it as a prior for nMAR, as the MRI-based prediction with kerMAR sometimes may become overzealous and interfere with uncorrupted regions. Comparing the MRI-based prior and nMAR-k in fig. 4.11, we for instance see how nMAR-k localized the artifact reduction to the corrupted regions, undoing some filling-in of the windpipe that occured with kerMAR.

4.5.4 MBIR with MRI-based prior

We finally used our posterior predictive model to define an image prior for MBIR using the MLTR algorithm[49]. Note that the experiments on this differ between paper A and B: in A, we

generated a kerMAR image and defined a Gaussian prior given the estimate, while we in paper B used the full posterior predictive model to define the prior distribution. We here report the more succesful results of paper B.

4.5.4.1 The prior step in MLTR

The update equation in eqn. (3.5) requires the first and and second derivatives of the prior. Using eqn. (4.7), the first and second derivatives of $\ln p(y_i|t_i, \mathbf{m}_i, \boldsymbol{\beta})$ may be calculated by the steps in appendix A.6 of paper B. They become:

$$\ln(p(y_i|\mathbf{m}_i, t_i, \boldsymbol{\beta}))' = (\beta_t + \beta_y)[\sum_{n\in\mathcal{A}_i} \tilde{v}_n^i \mu_n^i - y_i] \quad \text{and}$$

$$\ln(p(y_i|t_i, \mathbf{m}_i, \boldsymbol{\beta}))'' = (\beta_t + \beta_y)^2 \left(\sum_{n\in\mathcal{A}_i} \tilde{v}_n^i {\mu_n^i}^2 - [\sum_{n\in\mathcal{A}_i} \tilde{v}_n^i \mu_n^i]^2 \right) - (\beta_t + \beta_y),$$

$$\text{where} \quad \tilde{v}_n^i = \frac{\mathcal{N}\{y_i|\mu_n^i, (\beta_t + \beta_y)^{-1}\} v_n^i}{\sum_{n'\in\mathcal{A}_i} \mathcal{N}\{y_i|\mu_{n'}^i, (\beta_t + \beta_y)^{-1}\} v_{n'}^i}.$$

These expressions are calculated in each iteration during step 4 of algorithm 1 and substituted in eqn. (3.5); this defines the *MLTR-k* algorithm.

The first derivative here points toward a CT value estimate that is similar to kerMAR in eqn. (4.9), except for the weights that are modified by a Gaussian around the current reconstructed attenuation coefficient, y_i. This is similar to how the observation of t_i modified the pCT algorithm to yield kerMAR; to see this, compare $\tilde{v}_n^i$ to the kerMAR weights v_n^i in eqn. (4.9) that may be related to the purely pCT (MRI-based) weights using eqn. (4.10):

$$v_n^i = \frac{\mathcal{N}\{t_i|y_n, \beta_y^{-1} + \beta_t^{-1}\}\mathcal{N}\{\mathbf{m}_i|\mathbf{m}_n, \beta_m^{-1}\mathbf{I}_m\}}{\sum_{n'\in\mathcal{A}_i} \mathcal{N}\{t_i|y_{n'}, \beta_y^{-1} + \beta_t^{-1}\}\mathcal{N}\{\mathbf{m}_i|\mathbf{m}_{n'}, \beta_m^{-1}\mathbf{I}_m\}} = \frac{\mathcal{N}\{t_i|y_n, \beta_y^{-1} + \beta_t^{-1}\} w_n^i}{\sum_{n'\in\mathcal{A}_i} \mathcal{N}\{t_i|y_{n'}, \beta_y^{-1} + \beta_t^{-1}\} w_{n'}^i}.$$

A way to interpret MLTR-k is therefore as follows: In each iteration, it constructs an updated posterior given the current observation y_i and calculates a point estimate from it; the prior step in the algorithm pushes toward this estimate.

4.5.4.2 Improvement over prior-free MLTR

We compared MLTR with the MRI-based prior to prior-free MLTR. We used the same flat initial images and terminated the algorithm when the average change in the reconstructed attenuation coefficients decreased below $10^{-6} mm^{-1}$.

Figs. 4.13 and 4.14 show the results for a few images. While MLTR clearly provided benefits over FBP, a noticeable number of streaks are left behind and, as visible particularly in the poor quality of the teeth, the algorithm is a while from true convergence. Using the prior in MLTR-k

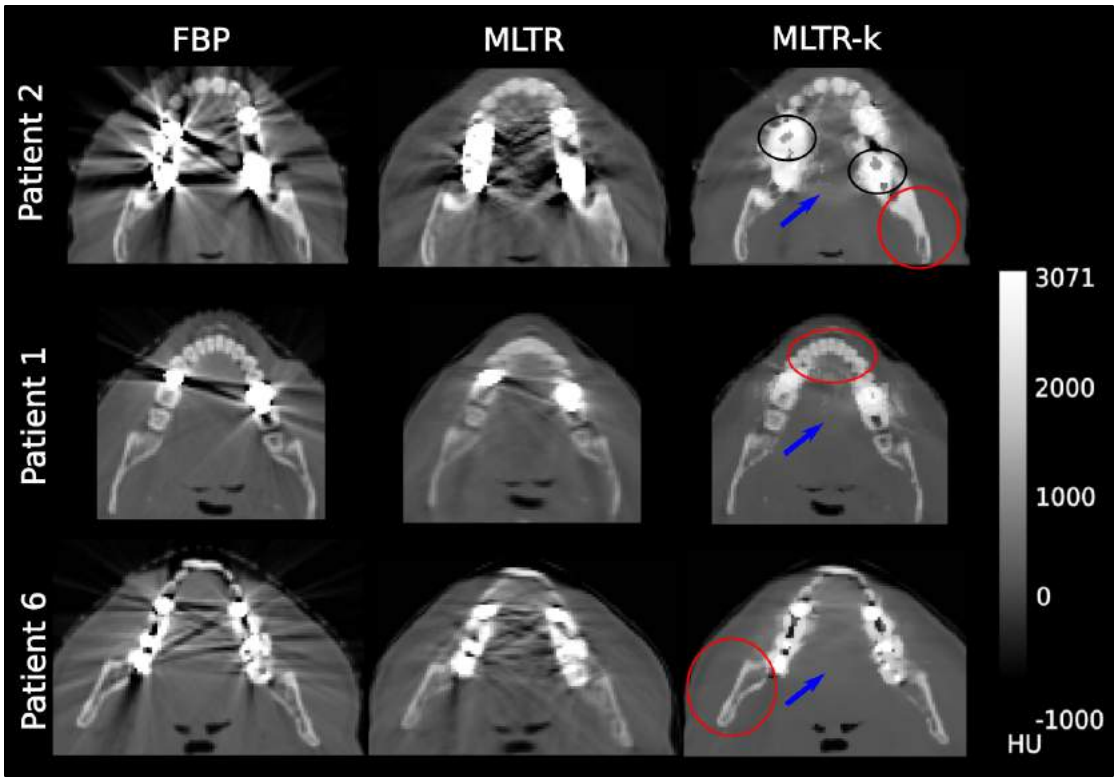


FIGURE 4.13: *Source:* Paper B. a) MLTR and MLTR-k results shown beside the FBP images. The MLTR-k displays fewer artifacts than the MLTR while and a sharper image (red ring and blue arrows). A potential danger with MLTR-k, however, is that the prior may become too strong and lead over-introduction of the MRI anatomy, as we see in the black rings for patient 2

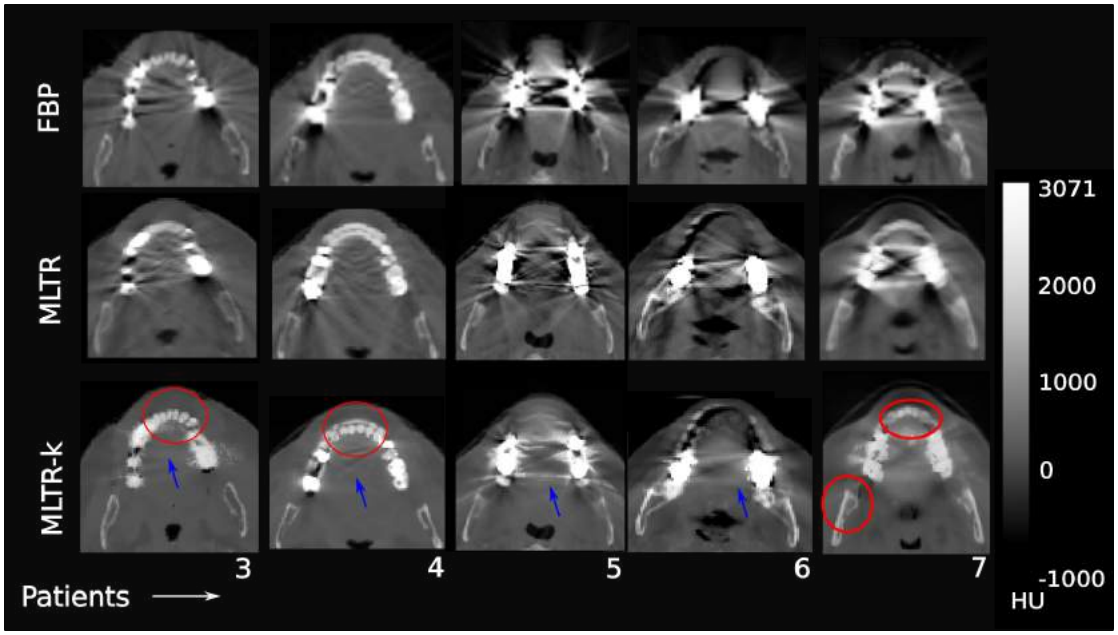


FIGURE 4.14: *Source:* Paper B. Additional axial CT slices for 5 patients, showing results for MLTR and MLTR-k.

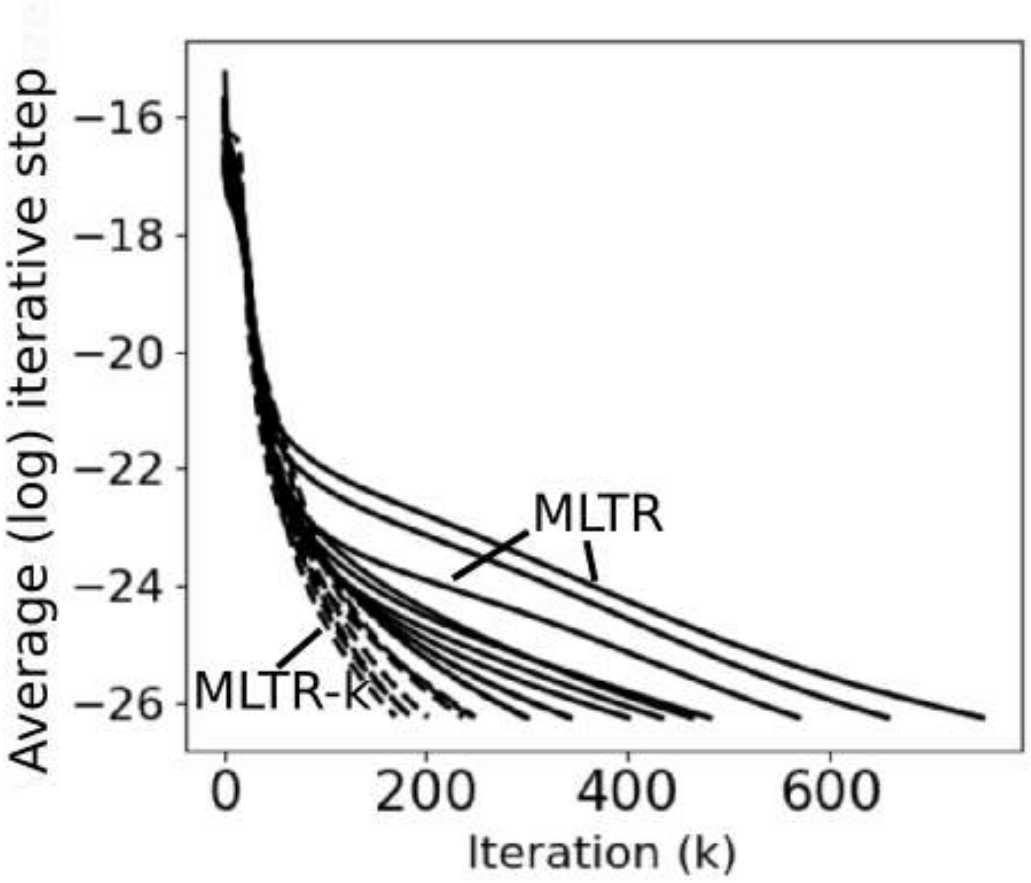


FIGURE 4.15: *Source:* Paper B. Convergence plot of MLTR and MLTR-k for the 9 patients. Solid and dashed curves show the log of the absolute, voxel averaged change between iterations at iteration k for MLTR and MLTR-k respectively. While the curves start out similarly, the MLTR-curves soon break off, leading to slower convergence.

by contrast helped reduce the streaks while improving the image resolution at the stopping point of the iterations. A potential cost associated with this benefit is the possible over-introduction of features in the prior, as may be seen in the top-right corner. This occurred in a few cases.

The improvement in convergence of MLTR-k may be seen in fig. 4.15, which shows the voxel-averaged change in atteunation coefficient in each iteration of the algorithm. For MLTR-k, it took essentially half the number of iterations to reach the stop-criterion. Since MBIR is rather slow, this is an important speed gain for practical applications.

4.5.5 Numerical evaluation and conclusions

We have now presented the image results of three MRI-based MAR algorithms, one in each main category: an image inpainting algorithm, a sinogram inpainting and an MBIR method. To back up our observations about their performance, we now consider how our algorithms performed

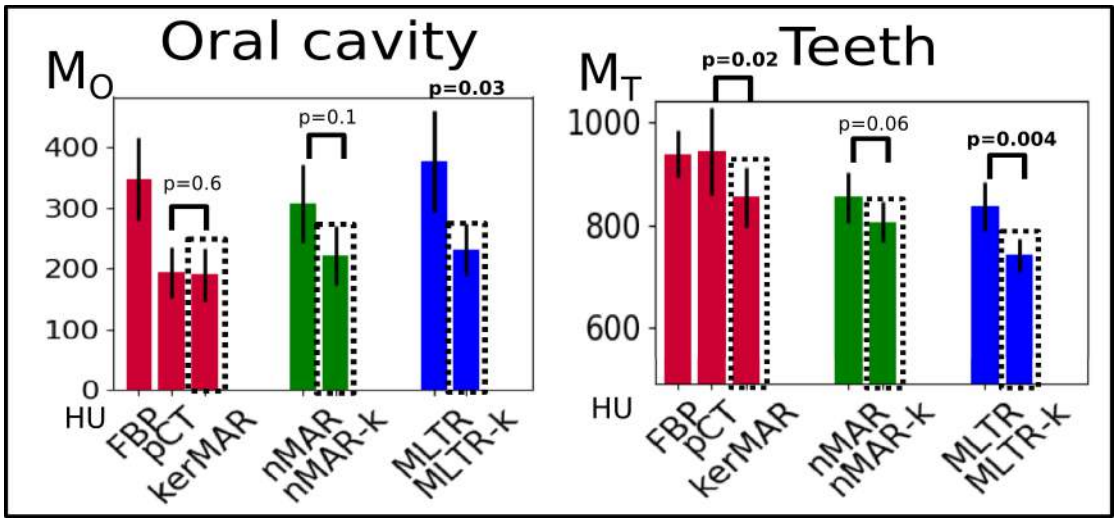


FIGURE 4.16: *Source:* Paper B. Results of the quantitative analysis of the oral cavity and teeth ROIs for the various MARs. The shown metric is the standard deviation of the corrupted CT values around a reference mean estimated from uncorrupted CT values. kerMAR, nMAR-k and MLTR-k are respectively our MRI-based image inpainting, sinogram inpainting and MBIR algorithms. The p-values are the results of two-sided Student's t-tests for paired, repeated observations ($N = 9$), which test the significance of the difference between the MRI-based and conventional algorithms.

within delineations of the oral cavity and teeth. We quantifed this by the CT-value standard deviation (STD) around an estimated true mean CT value that we calculated from uncorrupted parts of the delineations; fig. 4.16 shows the results. The p-values stem from a two-sided repeat observations Student's t-test[64] that compared the results with and without using our model.

We summarize our observations in relation to these results and derive some conclusions:

4.5.5.1 Benefits in image inpainting

Including the CT value to perform Bayesian inference in our image inpainting algorithm kerMAR led to decent prediction of bone and air CT values, leading to good results in the teeth compared to purely MRI-based pCT. This additionally helped resolve coregistration issues. The numerical results in fig. 4.16 show that the main numerical benefits occured in the teeth, while the softer tissue in the oral cavity was similarly well addressed by the purely MRI-based prediction; here, both approaches led to improvements over the FBP of around $150HU$ in the STD, corresponding to $\sim 1.5\%$ of the mean CT value.

Conclusion: By addressing two of the main challenges in MRI-based MAR, i.e. the contrast disparity between bone and air and inaccuracies in the coregistration, kerMAR may lead to anatomically correct images while providing significant numerical benefits through artifact reduction.

4.5.5.2 Benefits in sinogram inpainting

When considering sinogram inpainting using nMAR, the MRI-based kerMAR prior proved superior to what we could achieve by processing the CT, leading to better artifact reduction and fewer anatomical errors in the images. Considering the numerical results, these improvements translated to some numerical benefits in both the oral cavity and the teeth, but at p-values of 0.06 and 0.1 that were just beyond statistical significance.

Conclusion: The main benefit of using our MRI-based prior for nMAR lies in more accurately reconstructing anatomical details in corrupted regions; the over-all numerical benefits are smaller.

4.5.5.3 Benefits in MBIR

Using our MRI-based predictive model as a prior for MLTR led to improved artifact reduction and faster convergence with better image resolution. This led to clear numerical benefits in both the oral cavity and the teeth, similar to the improvements of kerMAR over the FBP.

Conclusion: Our MRI-based predictive model is a promising prior for MBIR that may both speed up convergence and suppress artifacts in the final reconstruction. Since this worked with MLTR that only includes a noise model, our model may achieve this together with even incomplete likelihood models.

Chapter 5

Radiotherapy validation

The particular focus of the MAR algorithms we developed in chapter 4 is the application to radiotherapy when an MRI has been acquired for tumor delination[65]. We in particular focus on head-and-neck RT, since here the implants both occur frequently and are close to both critical organs and the target tumors. It is also a relatively challenging target for MRI-based MAR due to its small, complex structures, the highly attenuating teeth and the frequent occurence of multiple implants.

In this chapter, we thus consider the performance of the MRI-based MAR in the context of head-and-neck RT treatment planning, comparing in particular our image inpainting method kerMAR to conventional clinically applied MAR approaches. Before that, we provide the background knowledge of RT treatment planning that is necessary to interpret our findings.

5.1 RT planning

The principle behind external beam RT is to bombard the tumor site with radiation, thereby killing the cancer cells and stopping their reproduction. There are three forms of radiation in widespread clinical use: Photon and electron radiation with energies in the MeV range; and proton radiation with energies of a few hundreds of MeV in energy[66, 67]. Of the three, photons are the most used, while protons are gaining widespread popularity in recent times[66, 68]. Electrons have more situational applicability to in particular skin cancer treatment.

5.1.1 Dose delivery

In the RT planning phase, a medical physicist or a dosimetrist defines the placement and orientations of the radiation source during therapy. He or she then optimizes parameters such as source

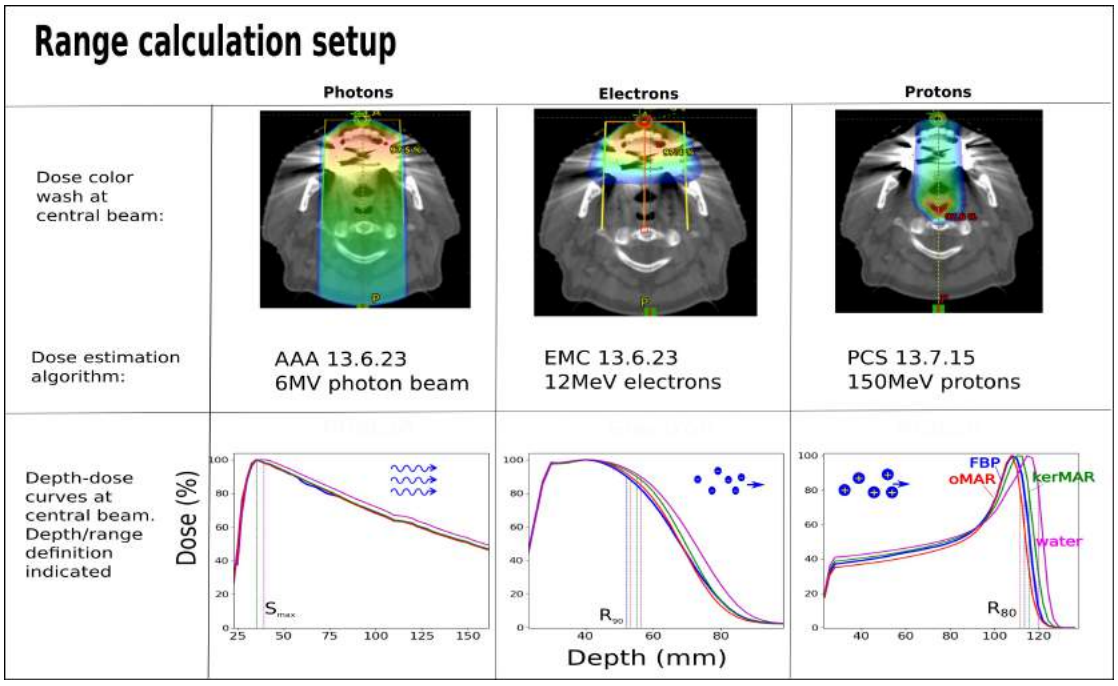


FIGURE 5.1: *Source:* Paper C. Description of the depth/range estimation setup. In an RT planning program (Eclipse v. 13.6, Varian Medical Systems), photon, electron and proton beams with the indicated specifications were angled through the oral cavity (for the patients) and oral cavity-like regions (for the phantom). Dose was calculated using the indicated algorithms and the central profile depth-dose curves were exported, from which the ranges were derived. Note that our Electron Monte Carlo (EMC) algorithm is (currently) invalid in the region of dose build-up near the surface, which led to the unphysical linear segments of the electron depth-dose curve.

output using complex constrained optimization algorithms and a combination of analytical and Monte Carlo dose calculations to realize the *dose target* prescribed by the oncologist, while sparing the normal tissue as much as possible. Dose is mostly deposited via secondary electrons emitted through interaction processes in the tissue that vary between beam types, leading to characteristic variations of the deposited dose along the path of the radiation as characterized by the *depth-dose* profile. For our validation experiments, to which we return later, we performed dose calculations using single beams of photons, electrons and protons for four different CT image sets (FBP, oMAR, kerMAR and manual water override), yielding the depth-dose profiles in fig. 5.1[14, 67].

Photon radiation lead to relatively flat depth-dose profiles, as in fig. 5.1(a)[14, 67]: the dose builds up from the surface of the patient to a characteristic point S_{max} that depends on the traversed tissue as well as the energies in the photon beam spectrum.

The dose from electrons used for electron therapy quickly reach a plateau and then afterward see a steep drop-off in dose (fig. 5.1(b)), and do not penetrate deeply at the practically used energies; they are therefore used in skin cancer treatment. A way to quantify the location of the distal dose drop-off is using the *therapeutic range*, the depth at 90% dose (R_{90}).

Protons are larger, more energetic particles that deposit most of their dose in a relatively narrow region called a *Bragg* peak (fig. 5.1(c)) close to their maximum range. The position of the peak may be quantified by the distal depth at 80% dose R_{80}[6, 66, 69].

5.1.2 Plan calculation and associated errors

The dose calculations require patient-specific maps of physical quantities, which are estimated from the CT. This allows for the calculation of quantities related to the penetration depth of the radiation in the tissue, such as the water equivalent thickness (WET) of the material:

$$WET = d_m \rho, \tag{5.1}$$

where d_m is the thickness of the tissue and ρ is estimated from the CT. In photon RT, ρ is the electron density relative to water; in proton RT, using the stopping power ratio calculates it approximately[70]; in electron RT, the electron densities are used to look up mass densities and other tissue features for Monte Carlo calculation of dose in the tissue.

We now consider the plan errors introduced by errors in the estimates of the mentioned quantities. Since electron RT is a more specialized application, we from here on focus on photon and proton RT.

5.1.2.1 Electron density estimation for photon RT

For photon RT, the electron density relative to water,

$$\rho_e = \frac{\rho_m}{\rho_w},$$

where m and w respectively denote the material and water, is estimated from the CT values in HU via an empirically determined multi-segmented linear curve. The electron density distribution is then used to simulate the dose distribution by a range of calculation methods, such as the *Analytical Anisotropic Algorithm* (AAA) algorithm provided by the Eclipse (Varian Medical Systems) *treatment planning system* (TPS). Under normal circumstances, the errors introduced by inaccuracies in the estimation of ρ_e propagate to errors in the dose estimates of less than 0.8%[1]. In the presence of metal artifacts, the errors may however increase to 5-10%[31, 71], depending on their severity and the treatment plan[40].

The beam positions of some exemplary photon plans for three different head-and-neck tumor types are shown in fig. 5.2(a). The gold markers illustrate the beam positions during treatment. The large number of beams makes the plans relatively insensitive to metal artifacts, as generally only a subset of beams are affected by them[40, 72].

5.1.2.2 Relative stopping power (RSP) and Water Equivalent Thickness (WET) estmation for proton RT

In proton RT, the proton stopping power relative to water is empirically determined from the CT values via a similar, multi-segmented calibration curve to the one used for electron density estimation; it is in theory approximately related to the relative electron density by a linear map [2]:

$$\rho_s = \rho_e \cdot K(v, I_m),$$

where $K(v, I_m)$ is a constant that depends on the proton velocity v (kinetic energy) and the mean ionization energy I_m of the target atoms in the tissue. An algorithm such as *Proton Convolution Superposition* (PCT) in Eclipse is then used to calculated the dose.

Errors in the stopping power estimates affect the WET directly and roughly translate the position of the Bragg peak, which is of great potential consequence for the proton dose distribution. By comparison, fluctuations in ρ_e are less impactful for the photon dose due to its more gently varying depth-dose profile. The effect of tissue inhomogeneities are therefore considered greater for protons than for photons[5, 66, 73], making proton therapy particularly sensitive to e.g. metal artifacts. The influence of this on our calculations may already be seen in fig. 5.1, where the proton depth-dose curves vary noticeably more between the different MAR algorithms than the photon ones.

Various sources attribute $1.1\% - 2.0\%$[2, 5, 6] proton range errors to even the inherent variations in stopping power, which may be owed to variations in the ionization energy that are not well captured by the CT values. To ensure tumor coverage in the presence of such errors, clinical practice in proton RT is to assign an additional distal margin of 3.5% of the range to the *Plan Target Volume* (PTV), i.e. the delineated region to which the dose is planned to be delivered. This is in addition to other margins that account for patient motion and other uncertainties[66]. In the presence of metal artifacts, the errors are naturally amplified, and WET estimate and particle range errors of $5 - 20mm$ have been reported[74, 75]. The recommended clinical practice to deal with such errors is to increase the distal margin from 3.5% to 5.0% after manual cleanup of the image[66]. It is additionally a priority that the beams are 1) multiple (within the limits of the treatment planning system), 2) critical normal tissue regions are not along the beam lines, and 3) the path travelled by the radiation is short and relatively homogeneous[5, 66].

The beam positions of some exemplary proton plans are shown in fig. 5.2(b), where the red arrows denote the entry angles of the beams. Due to limitations in the treatment planning system, more than about 3 beams is currently not feasible within the standard time allocated to treatment planning[5]. This makes the proton plans inherently more sensitive to errors introduced by e.g. metal artifacts. Additionally, some of these beams may have to be targeted towards critical

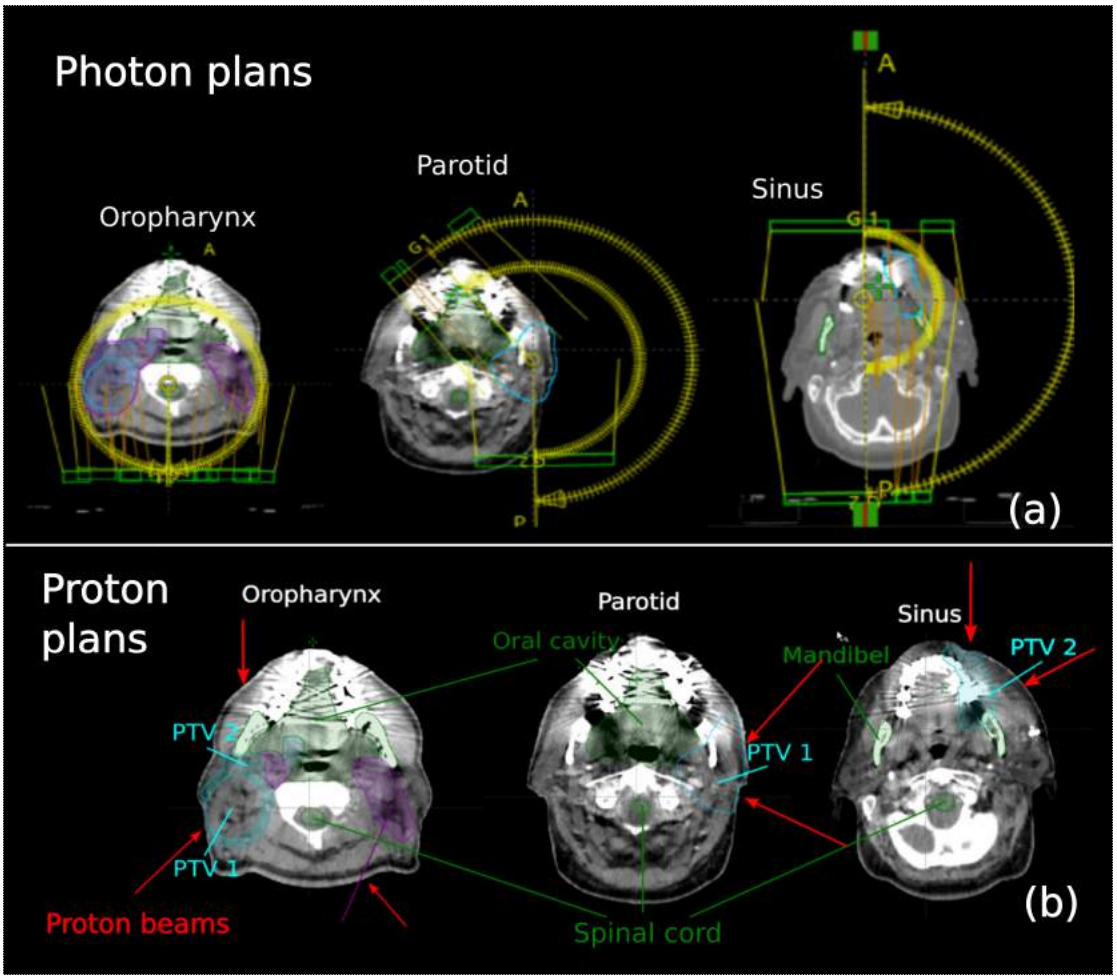


FIGURE 5.2: Illustration of photon (a) and proton (b) dose plans for three types of head-and-neck cancer (oropharynx, parotid, and sinus). Shown are also delineations of regions of some regions of interest: Plan target volumes (PTV 1, 2 and 3), spinal cord, oral cavity and the mandibel (lower jaw). The main difference between the photon and proton setups is the number of beams, which is far larger for photons.

regions, especially if the other concerns are to be satisfied, as is the case for about 2/3 of beams in practice[5].

5.2 Influence of MAR in the RT setting

Artifact reduction is thus important in RT. For this purpose, MAR algorithms may be provided as commercial plugins bundled with the CT scanner software. We had access to the Philips oMAR algorithm (described in chapter 3), whose performance in the context of RT has been considered in the literature. In summary, we have noted the following findings:

1. oMAR visually improves the image, although visible visual streaks persist[72, 76].

2. oMAR increases the accuracy of the average reconstructed soft tissue CT values, and reduces the tissue variance by about 25%; also provides improved HU accuracy in bone[76].

3. oMAR improves dose calculation accuracy for head-and-neck photon RT when the mouth is open, but not when it is closed[72, 77].

4. oMAR provides decreases in WET and range estimate errors when used for proton RT of several mms, though with errors persisting of up to $\sim 4mm$ even with oMAR[74].

While oMAR thus leads to accuracy improvements, it 1) leaves behind residual streaks and 2) still comes with WET estimate errors with proton RT. The two are interrelated, as streaks may lead to large errors in particular when the RT beams run parallel to them.

Manual intervention such as the water override we illustrated in fig. 3.4(c) is considered a prerequisite for angling beams through artifact corrupted areas. This process is however made difficult by 1) delineation problems due to the artifacts, in particular in corrupted high intensity regions such as the teeth and consequently the bordering oral cavity; 2) potential systematic errors in the CT values used for replacement; and 3) variations in practice and conventions between treatment planners.

5.3 Evaluation of MRI-based MAR for RT

Our analysis of RT planning errors and the way such errors are mitigated suggests the following three areas of inquiry:

1. The oMAR appears to not provide dose improvements in all situations, and may leave behind significant errors in the case of protons. To what extent MAR helps in RT over doing nothing thus warrants further investigation.

2. Since our MRI-based MARs use superior anatomical information in severely corrupted regions, they might outperform oMAR by addressing the residual streaks that it leaves behind.

3. Manual intervention methods such as water override are a common practice in RT; does it provide benefits compared to the automatic algorithms, or may it in fact lead to increased error compared to relying on the automatic methods?

We investigated these three questions in paper C, where we performed a retrospective study on our 9 head-and-neck patients as well as a veal shank phantom with and without a set of 6 metal markers.

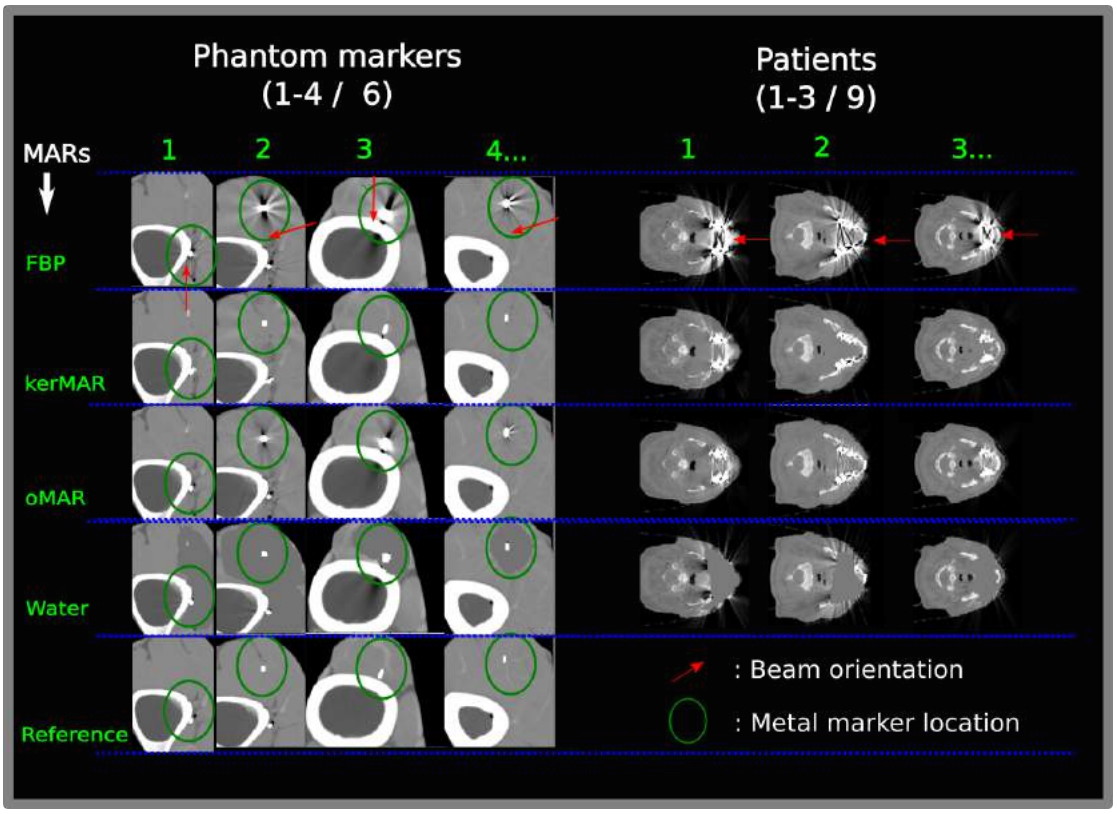


FIGURE 5.3: Axial slices of the veal shank phantom (markers 1-4) and 3 out of 9 patients, for different MAR algorithms. The metal implants were introduced computationally in the artifact-free reference images for the phantom. The red arrows show the orientations of the beam lines in the study.

Rather than consider clinical plans such as those shown in fig. 5.2 in our final experiments, we chose to consider artifical plans with beams angled through the artifact corrupted oral cavity; we made this choice because initial experiments revealed the subjects to be too varied in the amount of artifacts within the treatment regions, leading to little statistical power of such experiments.

We thus created artificial plans for our subjects that angled beams through the oral cavity (patients) and like areas (phantom) using photons and mono-energetic electrons and protons. Fig. 5.3 shows some exemplary images and the orientations of the beamlines (red arrows). We calculated the dose distributions with the exact same beam parameters on our kerMAR, the original FBP, the oMAR and manually water-overridden images. We then performed both dose calculations and image analysis to evaluate the quality of the MAR images.

5.3.1 Dose calculations

We extracted depth-dose curves from the dose distributions along the path corresponding to the center of the beam, leading to curves such as the ones in fig. 5.1, and extracted the particle range measures (R_{90} and R_{80} for respectively electrons and protons) and the photon depth at maximum dose (S_{max}).

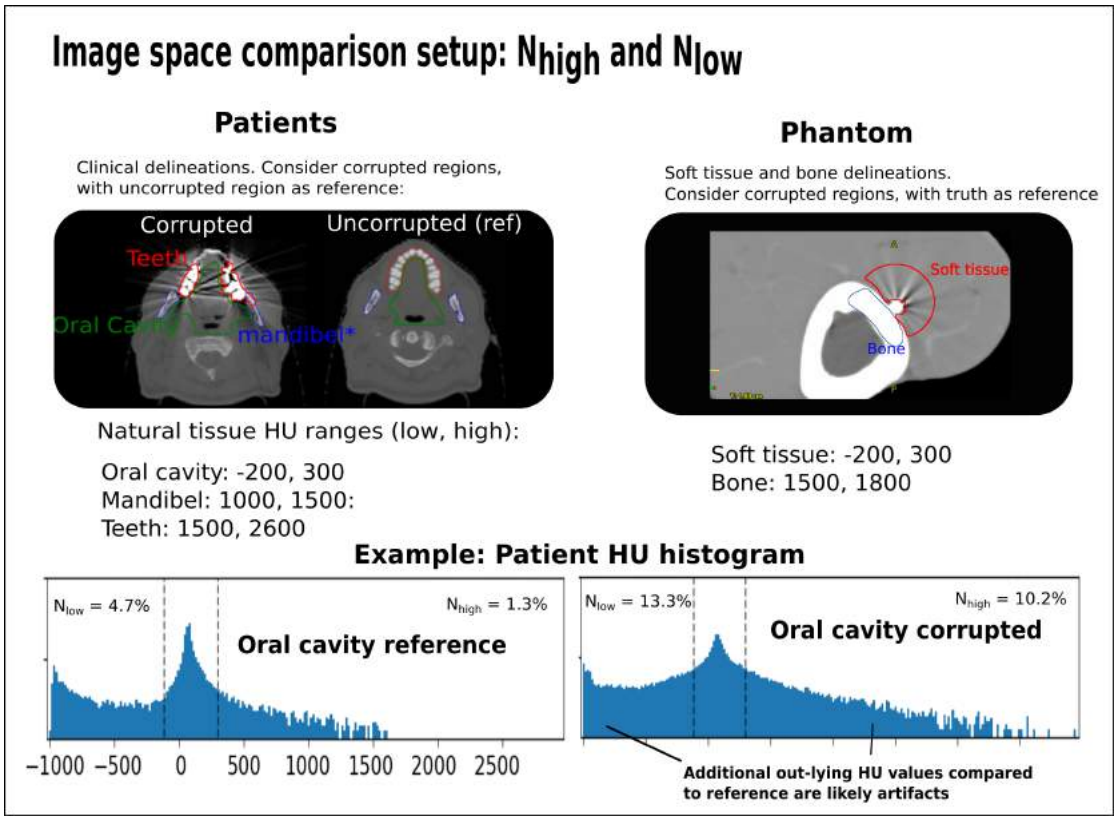


FIGURE 5.4: *Source:* Description of the setup for image analysis of the metal artifacts. Increases in the corrupted part compared to reference in voxel count fractions with HU lower and higher than the expected range in the ROI (respectively N_{low} and N_{high}.

For the phantoms, we calculated the absolute differences from the uncorrupted reference, $|\delta R_{90}|/|\delta R_{80}|/|\delta S_{max}|$. For the patients, having no such ground truth, we considered the raw numbers and simply looked for significant, absolute deviations between MAR methods in the range/depth estimates.

5.3.2 Image analysis

For the image space evaluation, we extracted the CT value distributions from delineated ROIs within each patient, from both corrupted and uncorrupted regions (fig. 5.4). As a metric of the amount of artifacts in the images, we as illustrated considered the tails of the HU-value distributions. For delineations of the oral cavity, mandibel (lower jaw) and teeth, we then defined normal tissue thresholds in reference to measured values of the attenuation coefficients of different tissue types, and denoted by respectively N_{high} and N_{low} the number of unnaturally high and unnaturally low CT values for voxels in the regions of interest. By subtracting these numbers from the same counts acquired from uncorrupted parts of the ROIs (patients) and from the metal-free reference (phantom), we defined the image corruption metrics δN_{low} and δN_{high}, which quantify the amount of respectively low and high intensity artifacts; the lower they are, the better the artifact reduction.

5.3.3 Hypothesis testing

We formulated three hypotheses:

1. Hypothesis I: oMAR and our MRI-based kerMAR is superior to using the uncorrected FBP.
2. Hypothesis II: kerMAR is superior to oMAR.
3. Hypothesis III: Manual override is superior to kerMAR, oMAR and FBP.

We tested these hypotheses using Student's t-test for repeat, dependent observations on both the range calculations and the image quality metrics. This statistical method is commonly used for data points acquired for the same subject at different time points (e.g. before and after intake of a pharmaceutical drug), and serves to desentisize the test to unimportant between-subject variations[64]. In our case, such variations are in particular the amount of artifacts within the path of the beamline and systematic differences in how the ROIs were delineated. For hypothesis I and III where collections of MARs were compared against each other, we aggregated the results by averaging the datapoints over the MARs.

5.4 Results and discussion

Figs. 5.5-5.7 show the mean (denoted by Δ) and standard errors (error bars) used to calculate the t-statistics for the hypothesis tests for the different ROIs and metrics, in both the phantom and the patients. The stars denote the significance level of the results: 1 star: $p < 0.05$; 2 stars: $p < 0.01$; no stars: $p > 0.05$. To interpret the results, note that positive values support the hypothesis while negative values reject the hypothesis, for all but the depths and ranges for the patients (the bottom right figures); for the patient depth and ranges, all values are positive and significant results merely indicate a significant difference between the MARs (which may be positive or negative).

5.4.1 Hypothesis 1: oMAR and kerMAR superior to FBP

The results for hypothesis I in fig. 5.5 may be summarized as follows:

1. The MARs improved upon FBP in image space to a highly significant degree in the soft tissue, but not in the bone; this occured both in the phantom and the patients.
2. Depth/range improvements in the phantom were not significant.

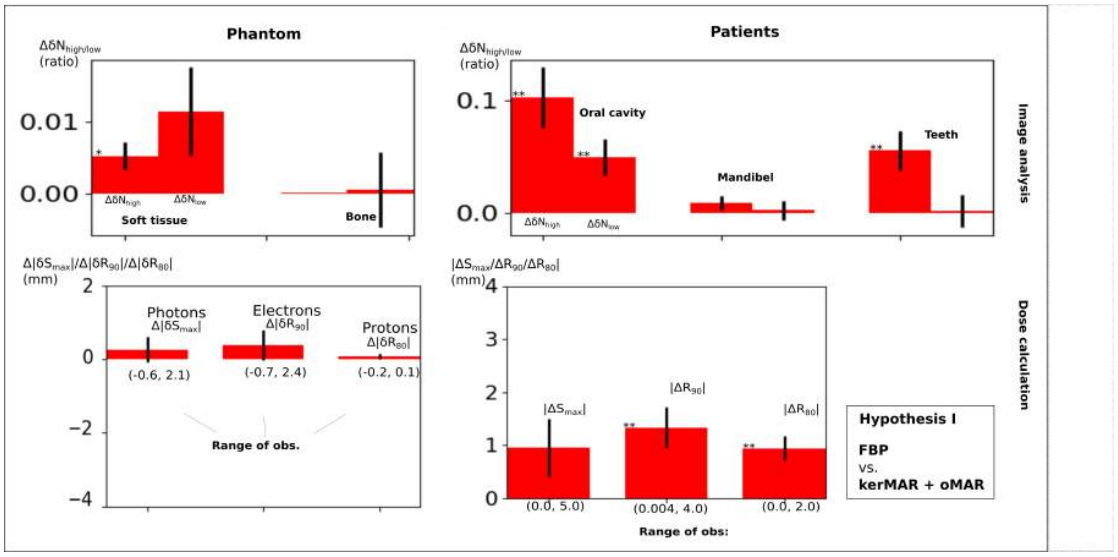


FIGURE 5.5: *Source:* Paper C. Mean (bar heights) and SDE (error bars) of the variations between the datasets contrasted for testing hypothesis I: The uncorrected FBP vs. the aggregate of oMAR and kerMAR. The top row shows $\Delta|\delta N_{high/low}|$ with a column for respectively the phantom and the patients. The second row shows $\Delta|\delta S_{max}|/\Delta|\delta R_{90}|/\Delta|\delta R_{80}|$ and $|\Delta S_{max}|/|\Delta R_{90}|/|\Delta R_{80}|$ for respectively the phantom and the patients. Asterisks denote significance of the paired observations student's t-test (one asterisk significant at $p < 0.05$, two asterisks at $p < 0.01$).

3. The depth/range in the patients were affected to a highly significant degree for electron and proton radiation, leading to absolute range differences of respectively $1.5 \pm 0.4mm$ and $1.0 \pm 0.3mm$.

The most likely reason for the insignificance of the phantom depth/range results is its relative homogeneity combined with the relatively light artifact corruption (see e.g. fig. 5.3). The dental areas in the head-and-neck patients were by contrast more corrupted due to the often multiple implantss, and here the influence of the MAR on the depth/range was correspondingly more significant, at least with the particle modalities. Taken together with the image quality improvements, this suggests that the MAR algorithms led to significant improvements on the order of $\sim 1mm$ for particle therapy. This is in agreement with our expectations from the literature both in in the magnitude of the particle range improvement and in that photons were relatively unaffected; for particles, we conclude that our results support hypothesis I.

5.4.2 Hypothesis 2: kerMAR superior to oMAR

The results for hypothesis II in fig. 5.6 are similar to the results for hypothesis I, and may similarly be summarized as follows:

1. The kerMAR improved upon oMAR in image space to a significant degree in the patient soft tissue as well as mandibel and teeth δN_{high}, while the phantom results were insignificant.

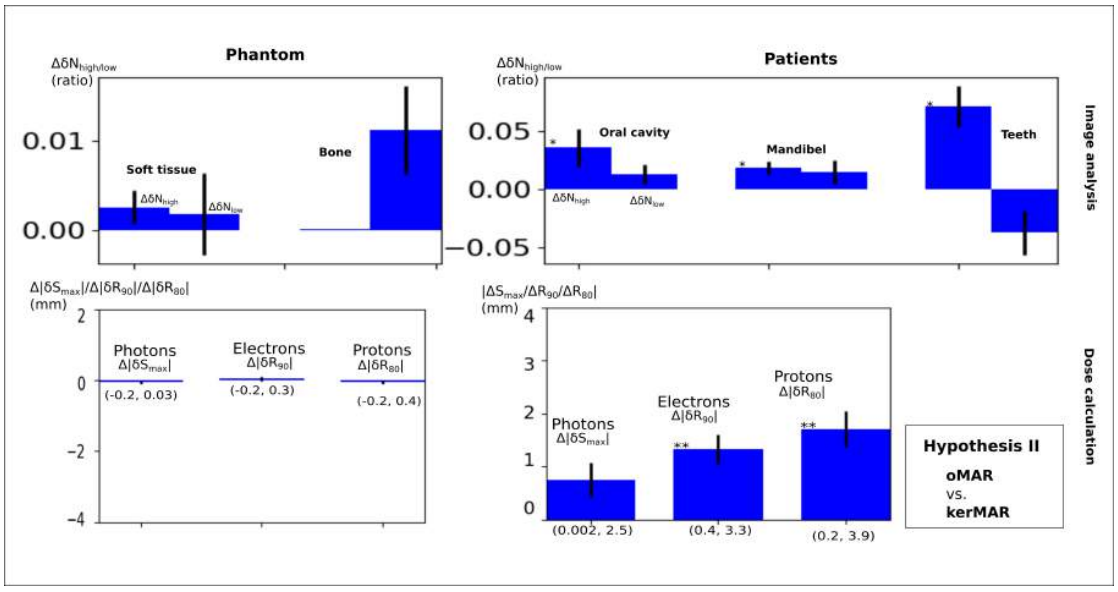


FIGURE 5.6: *Source:* Paper C. Mean (bar heights) and SDE (error bars) of the variations between the datasets contrasted for testing hypothesis II: oMAR vs. kerMAR. The layout details and are identical to fig. 5.5, as is the interpretation of the results.

2. Depth/range improvements in the phantom were again not significant.

3. The depth/range in the patients were affected to a highly significant degree for electron and proton radiation, leading to range differences of respectively $1.3 \pm 0.2mm$ and $1.8 \pm 0.3mm$.

Again, likely due to the relative homogeneity of the phantom, the algorithms performed similarly. In the more complex patients, however, kerMAR appears to have improved upon oMAR in terms of both soft tissue streak reduction and reduction of high intensity streaks in the teeth and mandibel (lower jaw).

The patient ranges were affected by these improvements to an even greater extent than when we compared the aggregate of oMAR and kerMAR to the FBP, suggesting that the improved streak suppression with kerMAR indeed may be beneficial for the RT dose calculations compared to oMAR. The amount of potential improvement ($\sim 1 - 2mm$) is smaller than the maximal $4mm$ WET errors that we noted had been found with oMAR, and are thus consistent with them; we conclude that our results support hypothesis II for the particles.

5.4.3 Hypothesis 3: Water override superior to oMAR, kerMAR and FBP

The results for hypothesis III in fig. 5.7 may be summarized as follows:

1. Water override led to improvements in the soft tissue as well as high intensity streak suppression in the teeth.

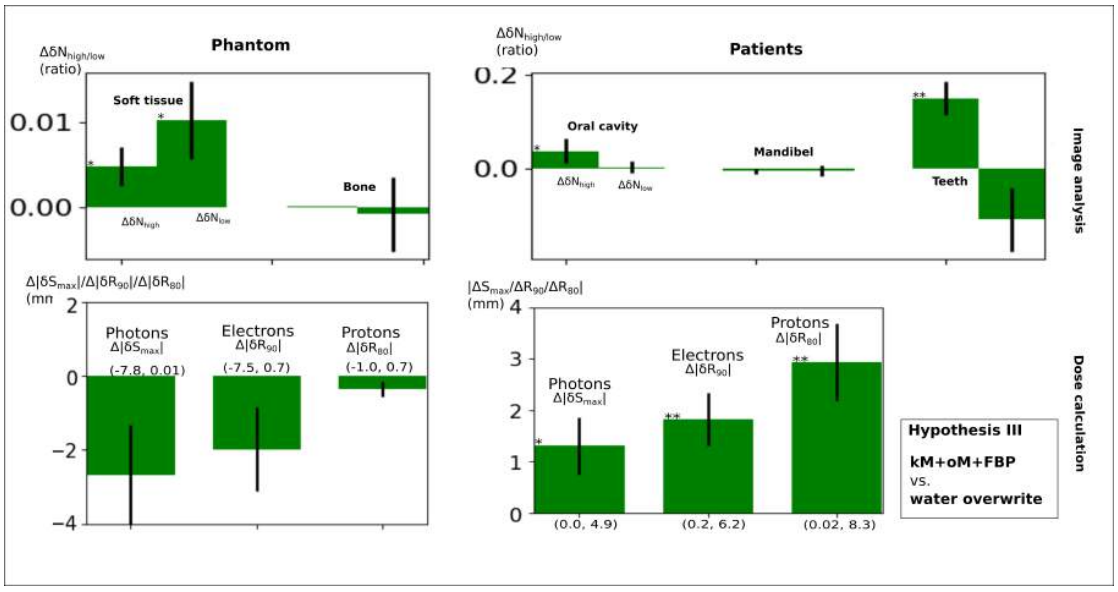


FIGURE 5.7: *Source:* Paper C. Mean (bar heights) and SDE (error bars) of the variations between the datasets contrasted for testing hypothesis III: The aggregate of kerMAR, oMAR and FBP vs. water override. The layout details are identical to fig. 5.5, as is the interpretation of the results.

2. Depth/range improvements in the phantom were once again not significant, but were much closer to significantly *negative* than we saw for hypothesis I and II.

3. The depth/range in the patients were affected to a highly significant degree for electrons and protons radiation, leading to large range differences of respectively $1.9 \pm 0.2mm$ and $3.0 \pm 0.3mm$. The range of the observations further spanned $\sim 7mm$.

The image quality improvements in the soft tissue demonstrate the succesful removal of artifacts that may be achieved using a manual override technique. The improvement in the teeth additionally demonstrates its capacity for removing the high intensity streaks; however, knowing the mechanism by which this occured, namely override with CT values of $0HU$, this masks the introduction of errors due to teeth override that is reflected in the almost significant negative impact on δN_{low}.

In the phantom depth/range results, we see further the potential effects of the systematic errors that may be introduced by this use of a preconceived HU-value for inpainting, as the phantom depth/range results are almost significantly negative: as may indeed be seen in the phantom images (fig. 5.3), the water override is not entirely appropriate for the veal shank, as it contained mostly muscle of $\sim 40 - 100HU$, which may explain the phantom results.

The highly significant patient depth/range results may in turn be due both to improved streak suppresion, which would cast them in a positive light, or due to similar systematic errors, due to override of both the muscle-filled tongue and the teeth. The large range of observations suggests the latter was the case for some of the patients, and we conclude that while hypothesis III is

supported in terms of artifact reduction in the soft tissue, the water override may introduce a large error potential in complex head-and-neck cases.

Chapter 6

Discussion, conclusion and future work

6.1 Summary and main contributions

We have in this thesis considered the problem of using the coacquired anatomical information in an MRI for CT metal artifact reduction, with the purpose of improving the dosimetric estimates in radiotherapy. We put great emphasis throughout the project on pursuing methods that would not interfere unnecessarily in the clinical workflow, and so restricted ourselves in notable ways. Most importantly, we chose to develop methods that only required imaging material from a single patient, and that did not require specially taylored image acquisition parameters, such as UTE sequences[59, 78]. Early investigations revealed that the main challenges of this task were 1) the contrast disparity between the MRI and CT and 2) the generally imperfect coregistration of the MRI and CT images that puts a hard restriction on the potential accuracy of purely MRI-based MAR. We also found that the few existing MRI-based algorithms out there did not adequately account for these issues[54–56]. These two challenges are what inspired the development of our generative model, which is the main methodological contribution of this project.

We generally deem MAR to be a CT reconstruction problem that for optimal performance should be done in reference to the CT raw data. While we used our generative model to perform image inpainting, without interfering in the CT reconstruction process, this motivated us to also consider ways to perform sinogram inpainting and MBIR. As we saw in chapter 4 and paper B, this was not only of benefit to the algorithms that were thus enhanced, we also found that the combination of our MRI-based predictions with the existing CT algorithms may outperform either in isolation.

The evaluation of our image inpainting algorithm for use in head-and-neck radiotherapy finally suggested significant potential improvements compared to a clinical, state-of-the-art algorithm in terms of in particular proton range estimation for treatment beams angled through the corrupted oral cavity.

6.2 Discussion and suggestions for future work

6.2.1 Clinical evaluation

We evaluated our methods on 9 head-and-neck RT patients with beams angled through the corrupted oral cavity. While we believe that this small cohort was sufficient as a proof of concept, a follow-up study on the performance of our MRI-based methods would be required for clinical application. Such a study might consider the following points:

1. Having no ground truth about the dose calculations for the patients, we had to simultaneously perform image analysis and phantom investigations to gauge whether the observed effects on the calculated particle range in the real patients were positive or not. A potential limitation of this procedure was the relatively less extensive corruption of the phantom that led to small dosimetric differences when comparing our MAR algorithm to the competition. Future investigations might include experiments on more sophisticated phantoms, such as the phantom used by Kwon et al. in ref. [76].

2. Our experiments used single beams angled through the oral cavity. We chose this approach after initial experiments with real clinical plans, which suggested that between-patient variations in the influence of the artifacts were too large for this to yield sufficient statistical power; within our cohort of 9 patients alone we encountered 4 different cancer types with according variations in plan designs. A future study might consider more patients (e.g. $N = 30$) and evaluate the effects of the artifact reduction on the dose coverage of the PTV and the dose to normal tissue.

3. We focused on head-and-neck RT patients. Since the CTs of such patients are commonly metal artifact corrupted, they are both clinically relevant and are relatively easy to acquire data for. However, patients with hip implants undergoing e.g. prostate cancer treatment are another important application. Especially with dual implants, such patients exhibit severe artifacts due to in particular photon starvation, in which case prior-based methods such as ours may be of benefit.

   We have not yet tried our methods on such cases, but we believe they would be an interesting challenge for MRI-based MAR, since the MRI metal artifacts are here somewhat more

destructive than in the head-and-neck MR images which were, in our experience, generally only lightly artifact corrupted owing to the metal implants. The hip implant application may therefore potentially require more sophisticated prior modelling than our patch-based kernel density estimation.

6.2.2 Improved image inpainting

Patches have in the fields of remote sensing and computer vision been used to great effect for image inpainting of photographs with missing information[79–82]. Some methods use probabilistic graphical models to optimize the allocation of image patches, taken from the surrounding image, to positions in the missing region. A component in their success is the use of clique potentials that impose spatial relationships between the assigned patches, making the inpainted region appear natural[82].

On one hand, our generative model might be improved by adapting modelling details from such methods; our model currently assumes the CT values and MRI intensities in different voxels to be independent, which given the success of the mentioned algorithms may have been an opportunity cost.

On the other hand, state-of-the-art graph-based algorithms have apparently not been applied to the removal of CT artifacts, in particular not in forms that capitalize on the information in a coacquired MRI. Given their success in other areas, such methods might be interesting to explore.

6.2.3 Bayesian sinogram inpainting

Our approach to using the MRI information for sinogram inpainting was to estimate projections from a prior image calculated by computing a point estimate from our predictive model. The nMAR algorithm then detected the assumedly fully corrupted metal projections by segmentation and replaced them with scaled versions of the prior ones.

Some modern sinogram inpainting algorithms however rather approach the inpainting task as the optimization of an objective function that contains a regularization term on the projection data along with a data fidelity term that measures the difference between the unaffected and the measured sinogram[34, 83]. Such a modelling approach may incorporate detailed forward models of the unaffected projections given the image[34], such as a beam hardening model, and may lead to algorithms without in particular the error-prone metal segmentation step. Our model may complement such methods by contributing a prior term to the objective function. This may

either be defined from our projection estimates calculated on the estimated image or it may be derived from our predictive model of the CT values.

6.2.4 Improved application to MBIR

In this project we had little success with using complex MBIR algorithms such as the IMPACT[38, 50] algorithm that includes a beam hardening model, and settled on the simpler and faster MLTR[49, 50] algorithm that yielded similar results. We attributed this to 1) inaccuracies in our knowledge of the implant material and scanner spectrum and 2) the presence of other artifact sources such as noise and photon starvation.

Given our positive results with the MLTR algorithm, it would be interesting to investigate whether combining our prior with a highly accurate MBIR algorithm[45, 46] in a more controlled setting would lead to greater improvements. The question is in particular whether our prior would help the algorithm converge, or if it would force a compromise between the likelihood and prior terms. While our experiments with MLTR seem to suggest that the first would be the case, given that the prior step is dynamically altered in each iteration in reference to the current estimate, a proper study, simulational or otherwise, might consider this.

Appendix A

Paper A: CT metal artifact reduction using MR Image Patches

CT Metal Artifact Reduction Using MR Image Patches

Jonathan S. Nielsen[a,b], Jens M. Edmund[b,c] and Koen Van Leemput[a,d]

[a]Department of Applied Mathematics and Computer Science, Technical University of Denmark, Lyngby, Denmark.
[b]Radiotherapy Research Unit, Department of Oncology, Gentofte and Herlev Hospital, Herlev, Denmark.
[c] Niels Bohr Institute, University of Copenhagen, Copenhagen, Denmark.
[d] Department of Radiology, Massachusetts General Hospital, Harvard Medical School, Boston MA, USA.

ABSTRACT

Metal implants give rise to metal artifacts in computed tomography (CT) images, which may lead to diagnostic errors and erroneous CT number estimates when the CT is used for radiation therapy planning. Methods for reducing metal artifacts by exploiting the anatomical information provided by coregistered magnetic resonance (MR) images are of great potential value, but remain technically challenging due to the poor contrast between bone and air on the MR image. In this paper, we present a novel MR-based algorithm for automatic CT metal artifact reduction (MAR), referred to as kerMAR. It combines kernel regression on known CT value/MR patch pairs in the uncorrupted patient volume with a forward model of the artifact corrupted values to estimate CT replacement values. In contrast to pseudo-CT generation that builds on multi-patient modelling, the algorithm requires no MR intensity normalisation or atlas registration.
Image results for 7 head-and-neck radiation therapy patients with T1-weighted images acquired in the same fixation as the RT planning CT suggest a potential for more complete MAR close to the metal implants than the oMAR algorithm (Philips) used clinically. Our results further show improved performance in air and bone regions as compared to other MR-based MAR algorithms. In addition, we experimented with using kerMAR to define a prior for iterative reconstruction with the maximum likelihood transmission reconstruction algorithm, however with no apparent improvements

Keywords: Computed Tomography, Metal Artifact Reduction, Bayesian modeling, Radiation Therapy

1. INTRODUCTION AND PURPOSE

CT images of patients with metal implants often suffer from severe streak and cupping artifacts, potentially leading to dosimetric errors in radiation therapy (RT) where the CT is used for patient specific electron density or mass stopping power estimation.[1] Metal implants amplify the effects of beam hardening and noise, which are both major sources of artifacts in filtered back projection (FBP),[2] the most widespread CT reconstruction algorithm. For this reason, a number of MAR algorithms have been proposed in the literature.[3] The most straightforward type of method is image-based, aiming to directly replace CT values in corrupted regions of already reconstructed CT images, typically using segmentation or anatomical prior knowledge.[3] A far more widespread method is sinogram (raw CT data) correction, which treats the x-ray measurements acquired through metal as missing data. The missing data are sometimes simply interpolated, sometimes estimated by forward projecting through a prior image generated by an image-based method.[1,4,5] The clinically used oMAR algorithm (MAR for Orthopedic implants, Philips Healthcare), for instance, combines image space segmentation and projection completion to post-process artifact-corrupted images in an iterative algorithm.[6]
Such methods potentially introduce new artifacts as they impose prior information of limited quality on either the image or sinogram.[7] A newer family of methods, model-based iterative reconstruction techniques (MBIR), attempt to include the underlying causes of the artifacts in the data acquisition model used for CT reconstruction. Although such techniques are slow and may sometimes require unavailable information such as the x-ray source spectrum and implant metal composition, incorporating prior knowledge of the expected CT reconstruction has been shown to help mitigate such issues.[7–11]
A promising source of such prior information is a coregistered MR scan that is acquired for e.g., tumor delineation in head-and-neck patients planned for RT. In such patients, the CT often suffers from metal artifacts due to dental implants.[1] Since metal artifacts are often less pronounced and more localised on MR images, a coregistered MR scan can provide useful tissue information in areas where the CT is corrupted. Using MR for MAR is a relatively new idea that has mainly been investigated with an image-based approach, e.g. by finding replacement CT values in local windows guided by MR voxel intensity differences,[12] or by creating a *pseudo-CT* (pCT) in which replacement CT values are assigned to discrete MR image segmentations.[13] Because these

methods only use local MR intensity similarity to predict CT values, however, they typically produce errors in bone and air regions which both appear dark on MR images acquired with conventional sequences. We therefore propose in this paper a novel MR-based MAR algorithm that combines MR information in larger larger spatial neighborhoods with the local, corrupted CT values to make more accurate CT value predictions. Our approach is based on methods developed for pseudo-CT generation in MR-only radiotherapy[14] and PET/MR attenuation correction[15] applications. Here, CT values are predicted from *image patches* (clusters of neighboring voxels) in coregistered MR scans, using regression models learned from a database of matching MR patches and CT values. Our method combines such a regression model trained on CT value/MR patch pairs from the uncorrupted part of the patient, with a probabilistic forward model of the artifact-corrupted measurements to predict the true CT values using Bayesian inference.

We further provide an expectation maximisation (EM) algorithm for automatically choosing the regression model's hyperparameters using Empirical Bayes estimation. This renders the method fully automatic and optimises the algorithm for each individual patient. Since the resulting method only requires data from the patient targetted for MAR, it avoids any inter-subject issues that arise during pCT generation[14,16]such as the requirement for MR intensity normalisation[17–19] or time-consuming atlas registration of the target patient.

We will refer to our method as "kernel regression MAR" (kerMAR) in the remainder. It can be used both as-is or to generate a prior for subsequent model based iterative reconstruction. In this paper we will use the maximum likelihood transmission reconstruction (MLTR) algorithm for this,[20] and refer to the resulting combination of kerMAR with MLTR as MLTR-k.

The artifact reduced images resulting from applying both algorithms are presented for 7 head-and-neck RT patients, and compared to those of the clinically used oMAR[6] algorithm. To investigate the benefits of including the corrupted CT measurements for MAR, we additionally compare against pCT.

2. METHOD

2.1 Generative model

Consider a patient for which a medical CT and MR volume have been acquired and (assumedly) perfectly coregistered, associating each voxel pair with a unique index i from the set $\mathcal{T}$. In the presence of metal implants, the CT is corrupted by artifacts and so the CT measurements $\{t_i\}_{i\in\mathcal{T}}$ may be incorrect; the problem of metal artifact reduction can then be viewed as estimating the true, unknown CT values $\{y_i\}$, $\forall i \in \mathcal{T}$.

To achieve this given our data, we establish the probabilistic relationship between y_i and t_i as well as the MR measurements. Contrary to CT, MR provides little contrast between bone and air, so the single voxel MR measurement is ambiguously related to y_i in such regions. We therefore extract larger spatial contexts from the MR image, using as our MR measurement for voxel i the *patch* $\mathbf{m}_i$ of size $d = M^{1/3}$, an M-dimensional vector of MR intensity values from a $d \times d \times d$ cuboidal window centered on the voxel. We then model the joint distribution of $\{\mathbf{m}_i, y_i, t_i\}$, $\forall i \in \mathcal{T}$, given hyperparameters $\boldsymbol{\lambda} = \{\beta_m, \beta_y, \beta_t^*\}$:

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

where

$$p(y_i, t_i, \mathbf{m}_i|\boldsymbol{\lambda}) = p(t_i|y_i, \mathbf{m}_i, \boldsymbol{\lambda})p(y_i, \mathbf{m}_i|\boldsymbol{\lambda}).$$

We learn $p(y_i, \mathbf{m}_i|\boldsymbol{\lambda})$ from samples. For this purpose we pick for each voxel $i \in \mathcal{T}$ a subset of indices $\mathcal{A}_i$ far from the metal implants with assumed uncorrupted CT values ($\forall n \in \mathcal{A}_i : t_n = y_n$) and extract the CT value / MR patch pairs $\{y_n, \mathbf{m}_n\}_{n\in\mathcal{A}_i}$. $p(y_i, \mathbf{m}_i|\boldsymbol{\lambda})$ is then estimated using kernel density estimation [21, p. 301-304] on this dataset. Using Gaussian kernels with diagonal covariance matrices with components $\beta_m^{-1}\mathbf{I}_M$ and β_y^{-1}, that thus factor into separate Gaussians, we get the kernel density estimate:

$$p(y_i, \mathbf{m}_i|\boldsymbol{\lambda}) = \frac{1}{|\mathcal{A}_i|} \sum_{n\in\mathcal{A}_i} \mathcal{N}(\mathbf{m}_i|\mathbf{m}_n, \beta_m^{-1}\mathbf{I}_M)\mathcal{N}(y_i|y_n, \beta_y^{-1}). \tag{2}$$

$\mathcal{N}(\cdot|\boldsymbol{\nu}, \boldsymbol{\Sigma})$ here denotes a Gaussian with mean $\boldsymbol{\nu}$ and covariance matrix $\boldsymbol{\Sigma}$, while $\mathbf{I}_M$ is the identity matrix of size M. $|\mathcal{A}_i|$ denotes the number of elements in the set $\mathcal{A}_i$. β_m and β_y are the precisions (reciprocal variances) of the separate kernels and are hyperparameters of the model.

To model $p(t_i|y_i, \mathbf{m}_i, \boldsymbol{\lambda})$, we assume the artifacts add zero mean Gaussian noise to y_i, making t_i independent of $\mathbf{m}_i$ given y_i. We let the variance of the artifact noise be position dependent, letting it vary with position as $\beta_t^{-1}(\mathbf{x}_i) = f(\mathbf{x}_i)[\beta_t^*]^{-1}$, where β_t^* is one of the hyperparameters of the model. $f(\mathbf{x})$ is here a user-specified function with $0 \leq f(\mathbf{x}_i) \leq 1$ that scales the variance of the artifact noise and thus the credibility of the CT measurements, quantifying the observation that artifact corruption decays with distance to the metal implants. We define $f(\mathbf{x})$ in section 3. Suppressing the position dependence of β_t in the notation, the measured CT value is thus modelled as:

$$p(t_i|y_i, \mathbf{m}_i, \boldsymbol{\lambda}) = \mathcal{N}(t_i|y_i, f(\mathbf{x}_i)[\beta_t^*]^{-1}) = \mathcal{N}(t_i|y_i, {\beta_t}^{-1}),$$

where the $\mathbf{m}_i$ dependence disappeared from the right-hans side due the conditional independence of t_i and $\mathbf{m}_i$.

2.2 Inference of the uncorrupted CT values for kerMAR

Given the set of MR patches and CT values $\{\mathbf{m}_i, t_i\}$, $\forall i \in \mathcal{T}$, we wish to infer $\{y_i\}$. Using Bayes' rule, we have that:

$$p(\{y_i\}|\{\mathbf{m}_i, t_i\}, \boldsymbol{\lambda}) = \frac{p(\{\mathbf{m}_i, y_i, t_i\}|\boldsymbol{\lambda})}{p(\{\mathbf{m}_i, t_i\}|\boldsymbol{\lambda})} \tag{3}$$

with

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

and the marginal likelihood

$$\begin{aligned}
p(\mathbf{m}_i, t_i|\boldsymbol{\lambda}) &= \int_{y_i} p(\mathbf{m}_i, y_i, t_i|\boldsymbol{\lambda}) \mathrm{d}y_i \\
&= \int_{y_i} \mathcal{N}(t_i|y_i, \beta_t^{-1}) \left[\frac{1}{|\mathcal{A}_i|} \sum_{n \in \mathcal{A}_i} \mathcal{N}(y_i|y_n, \beta_y^{-1}) \mathcal{N}(\mathbf{m}_i|\mathbf{m}_n, \beta_m^{-1}\mathbf{I}_M) \right] \mathrm{d}y_i \\
&= \frac{1}{|\mathcal{A}_i|} \sum_{n \in \mathcal{A}_i} \left[\int_{y_i} \mathcal{N}(t_i|y_i, \beta_t^{-1}) \mathcal{N}(y_i|y_n, \beta_y^{-1}) \mathrm{d}y_i \right] \mathcal{N}(\mathbf{m}_i|\mathbf{m}_n, \beta_m^{-1}\mathbf{I}_M) \\
&= \frac{1}{|\mathcal{A}_i|} \sum_{n \in \mathcal{A}_i} \mathcal{N}(t_i|y_n, \beta_t^{-1} + \beta_y^{-1}) \mathcal{N}(\mathbf{m}_i|\mathbf{m}_n, \beta_m^{-1}\mathbf{I}_M).
\end{aligned} \tag{4}$$

where we in the last step recognised the integral as a convolution over Gaussians, leading to a new Gaussian with the variances added (see e.g. [21, p. 112]). Inserting this result along with eqn. (1) in eqn. (3) yields the final posterior:

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

with

$$\begin{aligned}
p(y_i|\mathbf{m}_i, t_i, \boldsymbol{\lambda}) &= \frac{\mathcal{N}(t_i|y_i, \beta_t^{-1}) \left[\frac{1}{|\mathcal{A}_i|} \sum_{n \in \mathcal{A}_i} \mathcal{N}(y_i|y_n, \beta_y^{-1}) \mathcal{N}(\mathbf{m}_i|\mathbf{m}_n, \beta_m^{-1}\mathbf{I}_M) \right]}{\frac{1}{|\mathcal{A}_i|} \sum_{n' \in \mathcal{A}_i} \mathcal{N}(t_i|y_{n'}, \beta_t^{-1} + \beta_y^{-1}) \mathcal{N}(\mathbf{m}_i|\mathbf{m}_{n'}, \beta_m^{-1}\mathbf{I}_M)} \\
&= \frac{\sum_{n \in \mathcal{A}_i} \left[\mathcal{N}(t_i|y_i, \beta_t^{-1}) \mathcal{N}(y_i|y_n, \beta_y^{-1}) \right] \mathcal{N}(\mathbf{m}_i|\mathbf{m}_n, \beta_m^{-1}\mathbf{I}_M)}{\sum_{n' \in \mathcal{A}_i} \mathcal{N}(t_i|y_{n'}, \beta_t^{-1} + \beta_y^{-1}) \mathcal{N}(\mathbf{m}_i|\mathbf{m}_{n'}, \beta_m^{-1}\mathbf{I}_M)} \\
&= \frac{\sum_{n \in \mathcal{A}_i} \left[\mathcal{N}(t_i|y_n, \beta_t^{-1} + \beta_y^{-1}) \mathcal{N}(y_i|\mu_n^i, (\beta_y + \beta_t)^{-1}) \right] \mathcal{N}(\mathbf{m}_i|\mathbf{m}_n, \beta_m^{-1}\mathbf{I}_M)}{\sum_{n' \in \mathcal{A}_i} \mathcal{N}(t_i|y_{n'}, \beta_t^{-1} + \beta_y^{-1}) \mathcal{N}(\mathbf{m}_i|\mathbf{m}_{n'}, \beta_m^{-1}\mathbf{I}_M)} \\
&= \sum_{n \in \mathcal{A}_i} v_n^i \mathcal{N}(y_i|\mu_n^i, (\beta_y + \beta_t)^{-1}),
\end{aligned} \tag{5}$$

where we have defined

$$\mu_n^i = \frac{\beta_t}{\beta_t+\beta_y} t_i + \frac{\beta_y}{\beta_t+\beta_y} y_n \quad \text{and} \quad v_n^i = \frac{\mathcal{N}(t_i|y_n, \beta_t^{-1}+\beta_y^{-1})\mathcal{N}(\mathbf{m}_i|\mathbf{m}_n, \beta_m^{-1}\mathbf{I}_M)}{\sum_{n'\in\mathcal{A}_i}\mathcal{N}(t_i|y_{n'}, \beta_t^{-1}+\beta_y^{-1})\mathcal{N}(\mathbf{m}_i|\mathbf{m}_{n'}, \beta_m^{-1}\mathbf{I}_M)}. \tag{6}$$

We now estimate the undistorted CT values $\{y_i\}$ as the mean of the distribution $p(\{y_i\}|\{\mathbf{m}_i, t_i\}, \boldsymbol{\lambda})$, which yields

$$\bar{y}_i = \int_{y_i} y_i p(y_i|\mathbf{m}_i, t_i, \boldsymbol{\lambda}) \mathrm{d}y_i = \sum_{n\in\mathcal{A}_i} v_n^i \mu_n^i, \quad \forall i. \tag{7}$$

We refer to this as the kerMAR estimate ("kernel regression MAR"). It is instructive to consider the following special cases for various β_t:

$\beta_t \to 0$ **(CT measurement t_i fully corrupted):**

$$\bar{y}_i = \sum_{n\in\mathcal{A}_i} w_n^i y_n, \quad w_n^i = \frac{\mathcal{N}(\mathbf{m}_i|\mathbf{m}_n, \beta_m^{-1}\mathbf{I}_M)}{\sum_{n'\in\mathcal{A}_i}\mathcal{N}(\mathbf{m}_i|\mathbf{m}_{n'}, \beta_m^{-1}\mathbf{I}_M)}, \tag{8}$$

which corresponds to conventional kernel regression [21, 301-304]; the CT measurement t_i is completely discarded. This method has previously been used for MR-based pseudo-CT generation,[14,15] and so we will in this paper refer to it as the pCT estimate.

$\beta_t \to \infty$ **(CT measurement t_i not corrupted):**

$$\bar{y}_i = t_i,$$

where the CT measurement is used as-is and the MR measurement $\mathbf{m}_i$ is completely discarded.

$0 < \beta_t < \infty$ **(General case):**

For general hyperparameter settings, y_i is estimated according to eqn. (7) using a combination of the CT measurement t_i and the MR measurement $\mathbf{m}_i$.

2.3 Empirical Bayes hyperparameter estimation

The results generated by kerMAR depend directly on the settings of the hyperparameters $\boldsymbol{\lambda} = \{\beta_t^*, \beta_y, \beta_m\}$. We seek to learn these hyperparameters automatically from our data using empirical Bayes estimation, finding the set of hyperparameters that best explain the data $\{t_i, \mathbf{m}_i\}$, $\forall i \in \mathcal{T}$, by maximising their marginal likelihood (eqn. (4)):

$$\boldsymbol{\lambda}^* = \arg\max_{\boldsymbol{\lambda}} \log(p(\{\mathbf{m}_i, t_i\}|\boldsymbol{\lambda})).$$

To simplify this optimisation problem, we make the approximation during hyperparameter estimation that all data belong to either an uncorrupted set $\mathcal{T}_u$ or a fully corrupted set $\mathcal{T}_c$, with respectively $\beta_t \to \infty$ and $\beta_t = \beta_t^*$. These subsets are defined by thresholding $f(\mathbf{x}_i)$ such that $\mathcal{T}_u = \{i \in \mathcal{T} | f(\mathbf{x}_i) \leq 0.5\}$ and $\mathcal{T}_c = \{i \in \mathcal{T} | f(\mathbf{x}_i) > 0.5\}$.

For the optimisation, we use an expectation maximisation algorithm (EM)[22,23] which iteratively performs two steps: The E-step, where for all $i \in \mathcal{T}$ are assigned probabilities to the $n \in \mathcal{A}_i$ regression points based on the current hyperparameter estimates by calculating the weights v_n^i (eqn. (6)); and the M-step where the hyperparameters are updated accordingly:

$$\beta_y^{-1} \leftarrow \frac{1}{|\mathcal{T}_u|} \sum_{i\in\mathcal{T}_u} \sum_{n\in\mathcal{A}_i} v_n^i (t_i - y_n)^2,$$

$$\beta_t^{-1*} \leftarrow \frac{1}{|\mathcal{T}_c|} \sum_{i\in\mathcal{T}_c} \sum_{n\in\mathcal{A}_i} v_n^i (t_i - y_n)^2 - \beta_y^{-1} \quad \text{and}$$

$$\beta_m^{-1} \leftarrow \frac{1}{|\mathcal{T}|} \sum_{i\in\mathcal{T}} \sum_{n\in\mathcal{A}_i} v_n^i (\mathbf{m}_i - \mathbf{m}_n)^T (\mathbf{m}_i - \mathbf{m}_n).$$

The algorithm is initialised by a guess at $\boldsymbol{\lambda}$. We chose $\boldsymbol{\lambda} = \{0, 0, 0\}$, corresponding to flat initial weights.

2.4 Maximum likelihood transmission reconstruction using kerMAR as prior (MLTR-k)

In addition to using the kerMAR algorithm alone, we also consider using it to define an image prior distribution for MBIR using the MLTR algorithm. Specifically, we use the kerMAR estimates of the voxels, $\{\bar{y}_i\}_{i\in\mathcal{T}}$, to define the following Gaussian prior distribution on the CT image $\{z_i\}_{i\in\mathcal{T}}$:

$$p(\{z_i\}_{i\in\mathcal{T}}) = \prod_{i\in\mathcal{T}} \mathcal{N}(z_i|\bar{y}_i, \kappa^{-1}),$$

where z_i is the (unknown) CT value in voxel i, κ the precision of the prior and $\bar{y}_i$ is the kerMAR estimate obtained as explained earlier.

The MLTR[20,24] algorithm is a gradient-based optimisation algorithm that iteratively maximises the Poisson likelihood of the x-ray intensity measurements $\{\mathbf{\Lambda}_j\}_{j\in\mathcal{S}}$, related to the sinogram by an exponential transform.[2] Starting from an initial estimate of $\{z_i\}$, $\forall i \in \mathcal{T}$, MLTR iteratively improves the image estimate via an additive step. Including the image prior alters this step, which, in terms of the system matrix[2] $\mathbf{L}$ with entries $l_{j,i}$,[24] becomes:

$$z_i \leftarrow z_i + \frac{\sum_{j\in\mathcal{S}} l_{j,i}[C\mathsf{e}^{-\sum_{i\in\mathcal{T}} l_{j,i}z_i} - \Lambda_i] + 2\kappa(\bar{y}_i - z_i)}{\sum_{j\in\mathcal{S}} l_{j,i}[\alpha_j C\mathsf{e}^{-\sum_{i\in\mathcal{T}} l_{j,i}z_i}] + 2\kappa}, \quad \forall i \in \mathcal{T} \quad \mathsf{with} \quad \alpha_j = \sum_{i\in\mathcal{T}} l_{j,i}. \tag{9}$$

This update step is performed in parallel for all voxels $\{z_i\}_{i\in\mathcal{T}}$.

Ideally, C should here be the emitted x-ray intensity; in practice, we did not have access to this information, and so we used the Noise Equivalent Count (NEC) scaling coefficient, which scaled the exponentially transformed sinogram[2] such that the measurements became approximately Poisson distributed.

3. EXPERIMENTS

Image material and processing We considered CTs and T1-weighted MRs from 7 head and neck RT patients. The CTs were acquired on a Philips Brilliance Big Bore helical CT scanner at a KVP of 120kEV and tube currents from 272-433mA. They were reconstructed in 512x512 2D slices by the scanner software using FBP at resolutions of $(1.2 \times 1.2 \times 2.0mm)$. We will refer to these images as FBPs in the remainder. The MRs were acquired by a Philips Panorama 1.0T HFO scanner at resolutions of $(0.5 \times 0.5 \times 5.5mm)$ except patient 7 with $(0.5 \times 0.5 \times 6.5mm)$. For patients 2-6 the MR repetition- and echo times were $(TR/TE) = (520.2ms/10ms)$, for patient 1,7 $(TR/TE) = (572.2ms/10ms)$. The MRs were rigidly coregistered to the CTs using mutual information coregistration[25,26] and resampled to the CT resolution. oMAR reconstructions were made by the scanner software and available alongside the FBPs.

The unit of the CT measurements was displaced Hounsfield Units [2, p. 475] (HU + 1024), so the minimum CT measurement was 24. The kerMAR/pCT estimates and their hyperparameters were accordingly calculated in displaced HU and afterwards subtracted by 1024 to yield HU. The unit of the MR measurements had no particular physical meaning other than depending on TR and TE.

kerMAR implementation We used cubic patches with dimensions in units of voxel size of $7 \times 7 \times 7$, or $8.4 \times 8.4 \times 14mm$, chosen in preliminary investigations to provide better results than smaller patches while being computationally favourable. To evaluate the kerMAR and pCT estimates we used the FBPs and T1w MRs described in the previous paragraph.

The FBPs were reconstructed in axial 2D slices, only some of which were reconstructed using any metal projection data, which led us to expect a sharp boundary in the superior-inferior direction between slices containing corrupted CT values and those that did not. To approximately find these potentially corrupted slices, we noted that the intra-slice maximum HU value increased abruptly to the scanner cut-off value of 4095 ($HU = 3071$) in a subset of consecutive slices. Assuming these to be corrupted, 24 slices centered on them were chosen as the voxel set $\mathcal{T}$, excluding air voxels outside the body outline using a watershed segmentation.

Next we segmented the metal voxels by thresholding using Otsu's method on an FBP calculated by the ASTRA toolbox[27] that allowed higher CT values than the clinical FBP and thus for better distinction between high intensity streaks and the implant metal. We excluded these metal voxels from the kerMAR/pCT artifact reduction and used them to define the variance scaling function $f(\mathbf{x})$ as $f(\mathbf{x}) = 1 + \tanh(-s(\mathbf{x})^2/500)$, where $s(\mathbf{x})$ is the Euclidean distance from position $\mathbf{x}$ to the nearest metal voxel in mm. The characteristic squared distance of $500mm^2$ was chosen by experimentation.

We next found $|\mathcal{A}_i| = 500$ kerMAR regression points for each $i \in \mathcal{T}$ among the metal free axial slices. We ideally wanted $\mathcal{A}_i$ to contain the points with most similar MR patches, i.e. smallest $(\mathbf{m}_i - \mathbf{m}_n)^T(\mathbf{m}_i - \mathbf{m}_n)$. Since finding these by direct calculation was too time consuming we used the approximate search algorithm "Fast PatchMatch" presented by Ta et. al. in.[28] To further increase speed, we only used PatchMatch for $i \in \mathcal{T}_c$, selecting instead based on the CT distance $(t_i - y_n)^2$ for $i \in \mathcal{T}_u$ where the CT was less corrupted. We specifically used a K-means clustering with 6 clusters to classify the voxels, picking the 500 closest points from the assigned class.

Having determined $\mathcal{A}_i$, the hyperparameter estimates were then calculated with 3 iterations of the EM-algorithm, at which point the hyperparameters changed by less than 0.1%. Using these hyperparameter estimates, the kerMAR (eqn. (7)) and pCT (kerMAR with $\beta_t = 0$ applied to $\mathcal{T}_c$, eqn. (8)) estimates were finally calculated.

MLTR-k implementation and raw data preprocessing The CT sinograms were interpolated from helical to planar sinograms, exponentially transformed to transmissions and NEC scaled[2] to make the data approximately Poisson distributed for use as intensity measurements for MLTR. The NEC scaling coefficient C was calculated for each patient on an air scan as ~ 3500, varying little between patients.

We chose the MLTR-k precision $\kappa = 5 \cdot 10^6$ by experimentation, finding its exact value within an order of magnitude non-critical for the results. We used the clinical FBP as initialisation, stopping the algorithm after 300 iterations at which point the voxel averaged step was smaller than $\sim 10^{-10}$ compared to the initial $\sim 10^{-7}$. The MLTR iterative step eqn. (9) was calculated using the GPU accelerated primitives of the ASTRA[27] toolbox to evaluate the forward projections $\sum_{i\in\mathcal{T}} l_{j,i}(\cdot)$ and back projections $\sum_{j\in\mathcal{S}} l_{j,i}(\cdot)$.

Since MLTR reconstructs the image in attenuation coefficients [2, p. 475], the kerMAR estimate $\bar{y}_i$ and the FBP used for initialisation were linearly transformed from displaced HU to attenuation coefficients for use in MLTR-k. We found thus transform by linearly fitting the FBP to an MLTR reconstruction with 2000 iterations calculated on a set of axial slices far from the metal implants. We then applied it to the kerMAR and FBP before using them as prior and initialisation respectively, and applied the inverse to the final MLTR-k reconstruction.

4. RESULTS

We calculated three sets of artifact-reduced axial CT images for the 7 head-and-neck RT patients: 1) kerMAR, 2) MLTR-k using kerMAR as prior and 3) pCT (kerMAR with $\beta_t = 0$ applied to $\mathcal{T}_c$), along with the uncorrected FBPs and MLTRs (MLTR-k with $\kappa = 0$). For comparison to the clinical practice we additionally show the oMARs.

Some representative axial images are shown in figure 1 with arrows pointing to regions of interest. The blue arrows point to highly corrupted regions close to metal implants; the yellow arrows to bone/air regions where the MR-based kerMAR and pCT algorithms are error prone; the green arrows to large discrepancies on the pCT for two patients with a poorly registered MR due to its low longitudinal resolution; and the red arrows to a case where MLTR-k potentially show improvement over kerMAR.

The empirical Bayes estimated kerMAR precisions are of similar magnitude over the patients, with means $\pm$ standard deviation of $\langle\beta_t^*\rangle = (0.27 \pm 0.06) \cdot 10^{-5}$, $\langle\beta_y\rangle = (4.3 \pm 0.9) \cdot 10^{-5}$ and $\langle\beta_m\rangle = (0.0097 \pm 0.009) \cdot 10^{-5}$.

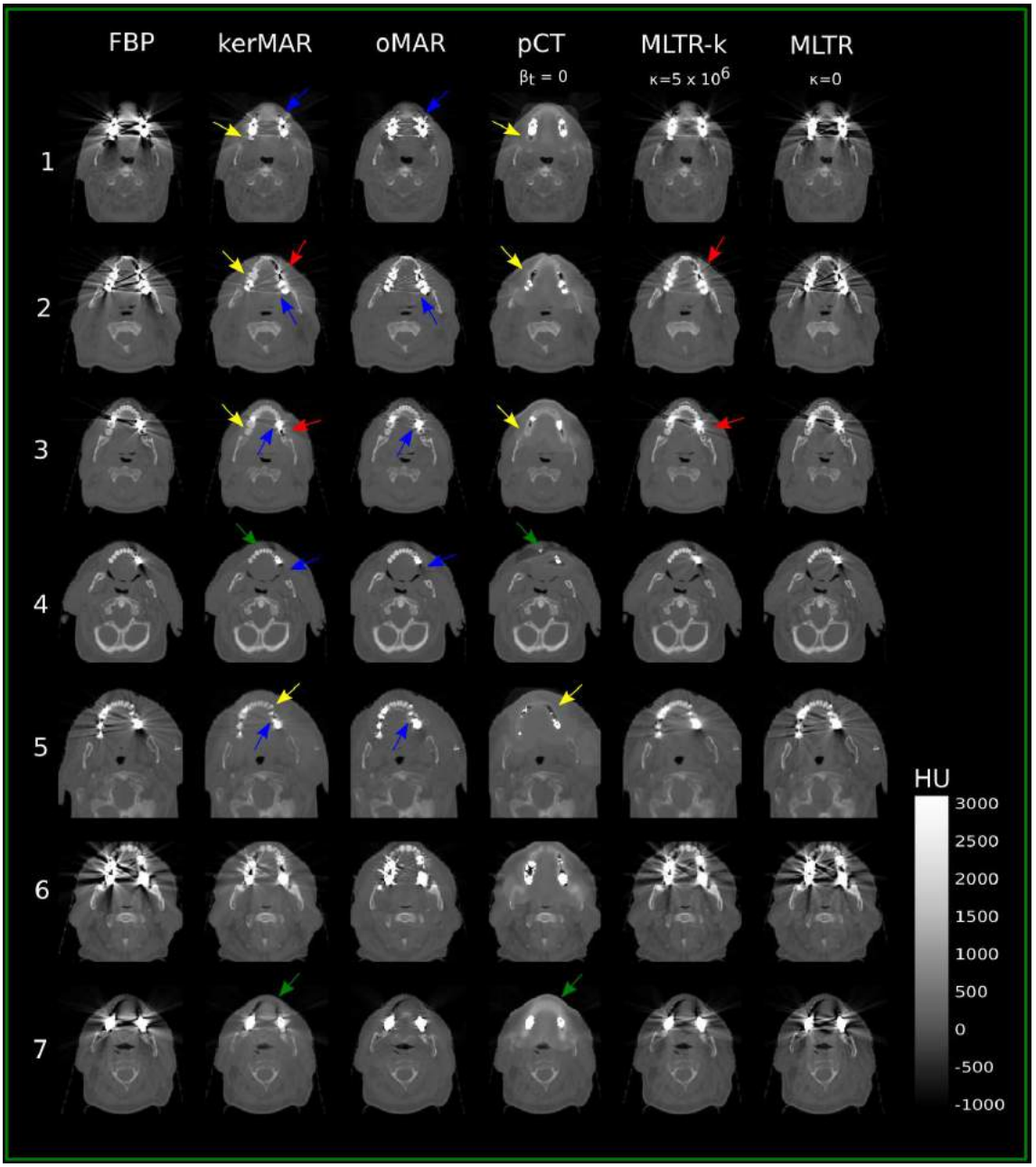


Figure 1: Representative axial images of the 7 patients. Left-Right: The uncorrected FBP; the kerMAR algorithm; the commercial, clinically used oMAR algorithm; kerMAR with $\beta_t = 0$ applied to the corrupted set of voxels $\mathcal{T}_c$ defined during hyperparameter estimation (pCT); the MLTR algorithm initialised with the FBP and with the kerMAR as prior (MLTR-k); the MLTR algorithm without a kerMAR prior. The hyperparameters for patients 1-7 were: $\beta_t^* = (0.13, 0.31, 0.30, 0.28, 0.27, 0.23, 0.35) \cdot 10^{-5}$, $\beta_y = (3.8, 6.5, 5.0, 4.2, 3.5, 4.2, 2.9) \cdot 10^{-5}$, $\beta_m = (0.0048, 0.0084, 0.0061, 0.0080, 0.0034, 0.032, 0.0053) \cdot 10^{-5}$.

5. DISCUSSION

In this paper, we presented a novel MR-based metal artifact reduction algorithm (kerMAR) that used patches drawn from a coregistered MR and the corrupted CT measurements to predict the uncorrupted CT values. It automatically optimised its hyperparameters on each individual patient and only required data from the patient itself, thus requiring no MR intensity normalisation or database registration in contrast to pCT generation algorithms. We additionally experimented with using it as a prior for MBIR using the MLTR algorithm (MLTR-k).

Comparison to existing MR-based MARs The main challenge when using MR images from conventional sequences for CT value prediction is the difficulty of disambiguating air and bone. Our kerMAR algorithm first addresses this issue by employing larger spatial contexts in the form of image patches for this prediction rather than single voxel intensities. As evident on the pCT results (fig. 1 column 4), however, we found the use of MR patches only to be insufficient. kerMAR therefore further included the corrupted measurements in the prediction to help resolve this disambiguation issue. As seen by the yellow arrows, this led to far better bone/air disambiguation.

The few previous MR-based MARs in the literature that we are aware of[12,13] suffer from this disambiguation issue. The algorithm introduced by Anderla et al.[12] works by looking in a 5x5x5 voxel window on the MR around each corrupted voxel (classified using Otsu's thresholding method), finding the voxel with smallest MR intensity difference to the window center and assigning its CT value to the voxel center on the CT. While this effectively constrains the search space to the very local 5x5x5 window and thus improves the bone/air disambiguation, it limits the potential accuracy of the algorithm and leads to failure in heavily corrupted regions.[12] The algorithm presented by Delso et al.[13] is more promising in this regard. It relies on threshold based implant segmentation and water/fat classification on a 3D MR dataset to create a pCT, replacing only soft tissue regions with database values empirically corrected on the patient data. This approach allows for improved handling of heavily corrupted soft tissue regions, but cannot address corrupted high or low intensity regions, in contrast to our proposed kerMAR algorithm. Additionally, coregistration errors between the CT and MR can lead to serious misestimations using the method, as reported by the authors. The green arrows on fig. 1 point to a case where we encountered such coregistration errors, thus leading to a poor pCT, but where the inclusion of the corrupted measurement in kerMAR led to good results.

Comparison of kerMAR to oMAR The commercial oMAR algorithm (Philips Healthcare) has been found in a few other studies to improve dosimetric accuracy when used for photon RT dose planning[1,29,30] and generally improving the image quality.[30,31] In this study, we found comparable visual improvement to oMAR with our kerMAR algorithm in terms of artifact reduction, with the following potential improvement as pointed to by the blue arrows on fig. 1: Our MR based kerMAR appears to better suppress high intensity streaks close to the implants than oMAR.

Kidoh et al. found in[31] a statistically significant tendency with oMAR to introduce additional artifacts in the form of image blur and unnatural features. In kerMAR, the inclusion of the corrupted measurement in the prediction seems to act as a barrier against such tendencies, albeit an imperfect one. In particular, some erosion of thin low and high intensity areas is visible on the kerMAR when the MR and CT were imperfectly coregistered due to the low axial resolution of the MR. A higher resolution MR that provides better coregistration could potentially increase the accuracy of kerMAR.

Benefit of kerMAR as a prior in MLTR-k In addition to using kerMAR as an image based MAR algorithm, we used it to define an image prior for MBIR with the MLTR algorithm (MLTR-k). Since such prior modelling may be used to integrate prior known features of the image in the reconstruction,[2,7,8,24] our hypothesis was that using kerMAR as a prior would alter the likelihood function such that the reconstructed image contained fewer artifacts.
We observe this on fig. 1 near the red arrows where MLTR-k provided streak suppression over MTLR and FBP

while disagreeing with its prior, kerMAR, on the dental CT values. In terms of image quality, however, MLTR-k did not lead to improvement over kerMAR as the MLTR-k results in general are of lower visual quality and only improved slightly upon the MLTR and FBP.

Clinical feasibility The oMAR algorithm runs on the order of minutes per patient,[31] which is suitable for clinical use. On our system (Dell Precision M3100 Laptop, CPU: Intel Core i7-4712HQ @ 2.3GHz, RAM: 16Gb) and our largely unoptimised Python implementation, kerMAR takes around 10-30min. per patient, the most time consuming part by far being the search for the regression point sets in the corrupted volume using Fast PatchMatch. Considering the results of the algorithm by Ta et al.,[28] optimised code on a dedicated system should take on the order of tens of seconds per patient.

KerMAR only uses data from a single patient and in principle works independently of the specific MR sequence employed, since all parameters are picked on a per-patient basis, as mentioned earlier. kerMAR thus excludes issues of MR intensity normalisation[17–19] and inter-patient registration issues, adding to its clinical attractiveness; the main requirement for its clinical implementation is an infrastructure for acquisition in similar patient orientations, coregistration and joint storage of the MR and CT volumes, which is clinical routine in RT clinics that uses both MR and CT for RT planning.

6. CONCLUSION

We presented a novel, clinically feasible MR-based algorithm for automatic CT metal artifact reduction (MAR), referred to as kerMAR. It requires no MR intensity normalisation or atlas registration. Image results for 7 head-and-neck RT patients suggest a potential for better suppression of high intensity streaks near the metal implants in the oral cavity as compared to the oMAR algorithm (Philips) used clinically. Our results further suggest improved performance in air and bone regions as compared to existing MR-based MAR algorithms. Using kerMAR as a prior for MLTR (MLTR-k) did not provide apparent improvement.

REFERENCES

[1] Giantsoudi, D., De Man, B., Verburg, J., Trofimov, A., Jin, Y., Wang, G., Gjesteby, L., and Paganetti, H., "Metal artifacts in computed tomography for radiation therapy planning: dosimetric effects and impact of metal artifact reduction," *Physics in Medicine and Biology* **62**(8), R49–80, R49–R80 (2017).

[2] Buzug, T. M., [*Computed Tomography - From Photon Statistics to Modern Cone-Beam CT*], Springer, Berlin (2008).

[3] Abdoli, M., Dierckx, R. a. J. O., and Zaidi, H., "Metal artifact reduction strategies for improved attenuation correction in hybrid PET/CT imaging," *Medical Physics* **39**(6), 3343 (2012).

[4] Karimi, S. and Cosman, P., "Using segmentation in CT metal artifact reduction," *Proceedings of the IEEE Southwest Symposium on Image Analysis and Interpretation* (Di), 9–12 (2012).

[5] Li, M., Zheng, J., Zhang, T., Guan, Y., Xu, P., and Sun, M., "A prior-based metal artifact reduction algorithm for x-ray CT," *Journal of X-Ray Science and Technology* **23**(2), 229–241 (2015).

[6] Philips Healthcare, "Metal artifact reduction for orthopedic implants (omar)." `http://clinical.netforum.healthcare.philips.com/us_en/Explore/White-Papers/CT/Metal-Artifact-Reduction-for-Orthopedic-Implants-(O-MAR)` (2012).

[7] Zhang, X., Wang, J., and Xing, L., "Metal artifact reduction in x-ray computed tomography (CT) by constrained optimization.," *Medical physics* **38**(2), 701–711 (2011).

[8] Lemmens, C., Faul, D., and Nuyts, J., "Suppression of metal artifacts in CT using a reconstruction procedure that combines MAP and projection completion.," *IEEE transactions on medical imaging* **28**(2), 250–60 (2009).

[9] Sadiq, M. U., Simmons, J. P., and Bouman, C. A., "Model based image reconstruction with physics based priors," *ICIP* **2**(1), 2–5 (2016).

[10] Webster Stayman, J. and Fessler, J. A., "Regularization for uniform spatial resolution properties in penalized-likelihood image reconstruction," *IEEE Transactions on Medical Imaging* **19**(6), 601–615 (2000).

[11] Xu, Q., Yu, H., Mou, X., Zhang, L., Hsieh, J., and Wang, G., "Low-dose X-ray CT reconstruction via dictionary learning.," *IEEE transactions on medical imaging* **31**(9), 1682–97 (2012).

[12] Anderla, A. A., Culibrk, D. R., and Delso, G., "Metal artifact reduction from ct images using complementary mr images," *2013 11th International Conference on Telecommunication in Modern Satellite, Cable and Broadcasting Services (telsiks) Vols 1 and 2* , 337–340 (2013).

[13] Delso, G., Wollenweber, S., Lonn, a., Wiesinger, F., and Veit-Haibach, P., "MR-driven metal artifact reduction in PET/CT.," *Physics in medicine and biology* **58**(7), 2267–80 (2013).

[14] Andreasen, D., Van Leemput, K., Hansen, R. H., Andersen, J. A. L., and Edmund, J. M., "Patch-based generation of a pseudo CT from conventional MRI sequences for MRI-only radiotherapy of the brain," *Med Phys* **42**(4), 1596–1605 (2015).

[15] Hofmann, M., Steinke, F., Scheel, V., Charpiat, G., Farquhar, J., Aschoff, P., Brady, M., Scholkopf, B., and Pichler, B. J., "MRI-based attenuation correction for PET/MRI: A novel approach combining pattern recognition and atlas registration," *Journal of Nuclear Medicine* **49**(11), 1875–1883 (2008).

[16] Sjolund, J., Forsberg, D., Andersson, M., and Knutsson, H., "Generating patient specific pseudo-CT of the head from MR using atlas-based regression," *Physics in medicine and biology* **60**(2), 825–839 (2015).

[17] Sun, X., Shi, L., Luo, Y., Yang, W., Li, H., Liang, P., Li, K., Mok, V. C. T., Chu, W. C. W., and Wang, D., "Histogram-based normalization technique on human brain magnetic resonance images from different acquisitions," *Biomedical Engineering Online* **14**(1), 73 (2015).

[18] Jog, A., Roy, S., Carass, A., and Prince, J. L., "Pulse sequence based multi-acquisition MR intensity normalization.," *Proc. SPIE Medical Imaging* **8669**(March 2013), 1–8 (2013).

[19] Roy, S., Carass, A., and L, J., "Patch based intensity normalization of brain mr images," *IEEE* (2013).

[20] Slambrouck, K. V., Nuyts, J., Van Slambrouck, K., and Nuyts, J., "Metal artifact reduction in computed tomography using local models in an image block-iterative scheme," *Medical Physics* **39**(11), 7080–7093 (2012).

[21] Bishop, C. M., [*Pattern recognition and machine learning*], Springer (2006).

[22] P. Minka, T., "Expectation-maximization as lower bound maximization." `https://tminka.github.io/papers/minka-em-tut.pdf` (2009).

[23] Borman, S., "The expectation maximization algorithm: A short tutorial. unpublished paper available at http://www.seanborman.com/publications." `http://www.seanborman.com/publications/EM_algorithm.pdf` (2009).

[24] De Man, B., [*PhD Thesis: Iterative Reconstruction for Reduction of Metal Artifacts in Computed Tomography*], Katholieke Universitiet Leuven, Leuven (2001).

[25] Wells, W. M. I., Viola, P., Atsumi, H., Nakajima, S., and Kikinis, R., "Multi-Modal image registration by maximization of mutual information," *Medical Image Analysis* **1**(1), 35–51 (1996).

[26] Maes, F., Collignon, A., Vandermeulen, D., Marchal, G., and Suetens, P., "Multimodality image registration by maximization of mutual information," *IEEE Transactions on Medical Imaging* **16**(2), 187–198 (1997).

[27] van Aarle, W., Palenstijn, W. J., De Beenhouwer, J., Altantzis, T., Bals, S., Batenburg, K. J., and Sijbers, J., "The astra toolbox: A platform for advanced algorithm development in electron tomography," *Ultramicroscopy* **157**, 35–47 (2015).

[28] Ta, V.-t., Collins, D. L., and Coup, P., "Optimized PatchMatch for near real time and accurate label fusion," *MICCAI* **2**, 105–112 (2014).

[29] Schoenfeld, A., Crilly, R., Poppe, B., and Laub, W., "Su-e-i-39: Experimental study on the performance of the omar ct artifact correction algorithm near titanium and stainless steel," *Medical Physics* **40**(6Part5), 133–134 (2013).

[30] Kwon, H., Kim, K. S., Chun, Y. M., Wu, H.-G., Carlson, J. N. K., Park, J. M., and Kim, J.-I., "Evaluation of a commercial orthopaedic metal artefact reduction tool in radiation therapy of patients with head and neck cancer," *The British Journal of Radiology* **88**(1052), 20140536 (2015). PMID: 25993487.

[31] Kidoh, M., Nakaura, T., Nakamura, S., and Tokuyasu, S., "Reduction of dental metallic artefacts in CT : Value of a newly developed algorithm for metal artefact reduction (O-MAR)," *Clinical Radiology* **69**(1), e11–e16 (2014).

Appendix B

Paper B: MRI-based CT metal artifact reduction using Bayesian modelling

MRI-based CT 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 Applied Mathematics and Computer Science, 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] *Department of Radiology, Massachusetts General Hospital, Harvard Medical School, Boston MA 02114, USA.*

May 5, 2019

Metal artifact reduction (MAR) algorithms reduce the errors in Computed Tomography (CT) images and are an important part of error management in radiotherapy (RT). A promising approach to MAR in RT is to leverage the information in a Magnetic Resonance Image (MRI) acquired for tumor delineation. By focusing on image inpainting, existing algorithms however do not leverage the benefits of different kinds of MAR algorithms, i.e. sinogram inpainting and model-based iterative reconstruction (MBIR) algorithms, which in turn may use the MRI prior information to respectively lessen the potential for introduction of new artifacts and to speed up convergence. Additionally, by using conventional-sequence MR intensities to pick replacement values for inpainting, they face difficulties in 1) bone/air disambiguation and 2) handling co-registration errors between the MRI and CT. This paper therefore develops a novel approach to MRI-based MAR, defining a patient-specific Bayesian generative model that provides a posterior distribution of the unknown, uncorrupted CT image given the simultaneous observation of the MRI and corrupted CT. We use this distribution to perform image inpainting, and integrate it in existing sinogram inpainting and MBIR methods. Considering a set of nine head-and-neck RT patients, we show how our image inpainting method leverages the corrupted CT information to resolve the mentioned bone/air disambiguation and co-registration issues, leading in the teeth to a reduction of the CT value standard deviation (STD) around a reference mean of ~ 100 Hounsfield Units (HU), at a significance level of $p = 0.02$. We also show how using it to generate a prior for sinogram inpainting leads to less introduction of artifacts, accompanied by STD decreases in the oral cavity and teeth by $\sim (100, 30)HU$, at $p = (0.1, 0.06)$. We finally show how using our predictive distribution for prior modelling improves the speed and quality of MBIR, the first by $\sim 50\%$ and the second by decreasing the STD in the oral cavity and teeth by $\sim (150, 110)HU$, at $p = (0.03, 0.004)$. We conclude that our MRI-based predictive model offers a versatile way to use the anatomical information of a co-acquired MRI to boost the performance of MAR algorithms.

1 Introduction and Purpose

Background

Medical x-ray Computed Tomography (CT) images of patients with metal implants often display major corruption from streak and cupping artifacts[1, 2]. These artifacts affect both the visual quality of the images, compromising their diagnostic use, and the quantitative CT value accuracy. The latter is of potential consequence for radiotherapy (RT), where the CT values are used in treatment planning to provide electron density

and relative stopping power estimates[3, 4]. 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. Metal artifact reduction (MAR) therefore plays an important role in the application of CT to image-guided medicine, in particular to RT[5–8].

MAR is in general a difficult problem, as demonstrated by its approximately 40-year long history that has spawned numerous algorithmic approaches[3, 9, 10]. A reason for this is that the artifacts do not come from a single source, but rather arise from multiple contributions that are amplified in the presence of metal[10, 11]. Some of these artifact sources are incorrect assumptions in the CT reconstruction model that relates the image coefficients to the x-ray projection data (sinogram) [12, 13], such as the mono-energetic approximation of the x-ray source spectrum; in e.g. the wide-spread *Filtered Back Projection* (FBP) algorithm, this leads to incorrect modelling of the *metal projections* that are acquired through the metal, and thus beam hardening artifacts[14, 15]. Other sources are more model-independent, such as the photon starvation of the metal projections that leads to noise artifacts.

MAR algorithms may be categorized in three categories that are more or less suited to handling the various artifact sources. *Image inpainting* algorithms replace corrupted CT values with better estimates by post-processing the images[16]; *sinogram inpainting* algorithms replace the metal projections by estimates that better fit the reconstruction model, which may be particularly effective in dealing with photon starvation[17–20]; and *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[21–29].

An important part of some MAR algorithms is the inclusion of prior information about the image. While this is especially true for the image inpainting algorithms that directly impose such information on the image, prior information is also used in some of the most succesful sinogram inpainting algorithms. An example is "normalized MAR" (nMAR)[17, 18], which replaces the metal projections by (scaled) estimates simulated on a prior image. Further, MBIR models also include prior information as they consist of two parts: A sinogram data likelihood that may address 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, potentially compensating for the information loss due to e.g. photon starvation.[21–29]

In a general CT setting, the quality of the available prior information is however limited. In sinogram inpainting, the prior therefore tends to be generated by post-processing the corrupted CT image[17–19]. Being generated from corrupted data, such a prior may, depending on the care that went into its construction, be compromised in the complex, highly corrupted regions where its quality is most important, such as the head and neck near the teeth and oral cavity. In MBIR, the limited availability of accurate prior information in turn motivates the use of relatively simple functional priors that e.g. impose regularities in the reconstructed image[30–33].

In the specific context of RT, on the other hand, a promising source of prior information is available in the Magnetic Resonance Image (MRI) that is commonly acquired for tumor delineation; since the metal artifacts are potentially more localized in the MRI, this may provide superior anatomical prior information in corrupted regions[34, 35]. In addition, since co-registration and acquisition of the MRI and CT in the same patient fixation is already part of the tumor-delineation process, this can be done with little interruption in the existing clinical workflow.

Previous work

Using the MRI to reduce metal artifacts has been approached in three existing image inpainting approaches[36–38]. The first method by Anderla et al. [36] uses the MRI to filter the CT image in 2D by looking for replacement CT values near to the corrupted pixels, choosing the replacements based on MR intensity differences. The second method by Delso et al. [37] uses bulk CT value replacement of corrupted regions in a segmentation calculated using MRIs from three modalities (a 3D-MRI set). The third method by Park et al. [38] replaces the corrupted target CT image by applying a non-rigid deformation map to a less corrupted axial source image. This source image is taken from the nearby CT volume and the deformation is calculated by non-rigid co-registration of the corresponding MRIs.

By focusing on image inpainting, these algorithms do not leverage the potential benefits of using the MRI for prior modelling in sinogram inpainting and MBIR. Additionally, they are purely MRI-based in the sense that they choose CT values for inpainting by referring only to the MRI. They thus do not incorporate the information in the corrupted CT image, which, though corrupted, may be useful for two purposes: First, the MRI and CT may not be perfectly co-registered, which motivates leveraging the anatomical information in the corrupted CT; and second, while the contrast between bone and air is large in the CT, it is generally small in the MRI unless a specifically taylored ultra-short-echo time (UTE)[39] sequence is used. The ensuing contrast disparity makes the MRI an ambiguous CT-predictor in such regions, which we propose to improve by leveraging the information in the corrupted CT.

Novelties

Building on work that we presented in an earlier conference proceedings [40], this paper therefore develops a novel approach to MRI-based MAR. In the methods section, it defines a Bayesian generative model[41–43] that provides a predictive probability distribution of the unknown, uncorrupted CT image given the simultaneous observation of a conventional-sequence (T1-weighted) MRI and the corrupted CT; it achieves this by combining artifact noise modelling with kernel density estimation[41]. It uses data only from the patient in consideration and requires no user-intervention as it automatically tunes its hyperparameters on the image data using empirical Bayes estimation, and is therefore easy to use in varying and changing clinical environments.

For the experiments, we use this predictive distribution to calculate CT value replacements and thus perform image inpainting, and also integrate it in existing MAR methods to make up for their potential inadequacies. In particular, considering a set of nine head-and-neck RT patients, we show how our proposed image inpainting algorithm uses the anatomical information in the corrupted CT to 1) resolve the mentioned contrast disparity issues between the CT and MRI, and 2) correct for co-registration errors. We then show how our proposed sinogram inpainting algorithm leverages the anatomical information in the MRI for improved performance by using our image inpainting algorithm as a prior. We finally show how using our prior for MBIR leads to both speed and quality improvements.

To summarize the structure of the paper, we first present the generative model, and derive the mentioned posterior distribution and our hyperparameter estimation algorithm. We then use the posterior with its tuned hyperparameters to perform the MAR experiments and show the results.

2 Methods

2.1 The generative model

Consider a CT volume reconstructed using Filtered Back Projection (FBP), with an assumed perfectly coregistered MRI. The observed CT values $\{t_i\}_{i\in\mathcal{T}}$, where $\mathcal{T}$ denotes the set of voxel indices, are potentially corrupted by artifact noise but remain statistically related to the underlying true CT values and thus contain information about them. Our task is to estimate these underlying true CT values, $\{y_i\}_{i\in\mathcal{T}}$, given the imaging data. To accomplish this, we model the probabilistic relationship between y_i, t_i and anatomical features extracted from the MRI.

For the model to be useful for CT value estimation, we aim for MRI features that provide unambiguous information about the CT values. The MRI and CT contrasts being vastly different in especially the bone and air regions, this is not achieved by using e.g. single voxel MR intensities; therefore, we rather use anatomical features in the form of *patches* $\{\mathbf{m}_i\}_{i\in\mathcal{T}}$, i.e. vectors of voxel intensities from local spatial contexts extracted from a cuboidal window centered on voxel i.

We thus model the joint distribution $p(\{t_i, y_i, \mathbf{m}_i\})$, letting in the following $\{\cdot\} \equiv \{\cdot\}_{i\in\mathcal{T}}$ for ease of notation, unless otherwise is stated. To factorize it, we specify the statistical dependency relations between the parameters. First, the artifact corruption is purely a product of the CT reconstruction, and so the knowledge of $\{y_i\}$ should entirely define the distribution of $\{t_i\}$, which are thus conditionally independent of $\{\mathbf{m}_i\}$ given $\{y_i\}$. Second, we assume that the observations for different voxels are statistically independent. These relations allow for the following factorization:

$$p(\{t_i, y_i, \mathbf{m}_i\}|\boldsymbol{\beta}) = \prod_{i\in\mathcal{T}} p(t_i, y_i, \mathbf{m}_i|\boldsymbol{\beta})$$
$$\text{with} \quad p(t_i, y_i, \mathbf{m}_i|\boldsymbol{\beta}) = p(t_i|y_i, \boldsymbol{\beta})p(y_i, \mathbf{m}_i|\boldsymbol{\beta}), \quad (1)$$

and where the precisions (reciprocal variances) $\boldsymbol{\beta} = \{\beta_t^*, \beta_y, \beta_m\}$ are the hyperparameters of the model, to be explained shortly.

The artifact noise model The first factor, $p(t_i|y_i, \boldsymbol{\beta})$, models the distribution of the corrupted CT value given the underlying true CT value, and constitutes an artifact noise model. We model this noise as additive Gaussian, with variance β_t^{-1} depending on the voxel position $\mathbf{x}_i$ through a sigmoidal function $f(\mathbf{x}_i)$ that decreases with the distance to the metal, leading to:

$$p(t_i|y_i, \boldsymbol{\beta}) = \mathcal{N}(t_i|y_i, \beta_t^{-1}) \quad \text{with} \quad \beta_t^{-1} = \beta_t^{*-1} f(\mathbf{x}_i) \quad (2)$$
$$\text{and} \quad f(\mathbf{x}_i) = 1 + \tanh(-\frac{D_\perp^2(\mathbf{x}_i)}{\kappa}),$$

where $D_\perp(\mathbf{x}_i)$ is the perpendicular spatial distance to the set of metal voxels, found by image thresholding using Otsu's method[44], and $\mathcal{N}(\cdot|\psi, \nu^2)$ denotes a Gaussian with mean ψ and variance ν^2. κ determines the rate of the sigmoidal decrease of the scaling function $f(\mathbf{x}_i)$ with distance to the metal; we chose $\kappa = (10mm)^2$ in all our experiments.

The hyperparameter β_t^* quantifies the artifact noise level near the implants, and is automatically tuned on the observed data $\{t_i, \mathbf{m}_i\}$ using Empirical Bayes estimation[42, 43], as explained in section 2.3.

MRI-based kernel density estimation $p(y_i, \mathbf{m}_i|\boldsymbol{\beta})$ is learned from uncorrupted samples taken from the patient volume, using *kernel density estimation*[41] (KDE) as illustrated in fig. 1: We first locate a set of uncorrupted voxels $\mathcal{T}_u \subseteq \mathcal{T}$ by thresholding $f(\mathbf{x}_i)$ such that $\mathcal{T}_u \equiv \{i \in \mathcal{T} | f(\mathbf{x}_i) < 0.5\}$. For each voxel $i \in \mathcal{T}$, we then extract a sample set of CT values and MRI patches $\{y_i, \mathbf{m}_i\}_{i\in\mathcal{A}_i}$, with indices in $\mathcal{A}_i \subseteq \mathcal{T}_u$; we use

the MRI patch matching algorithm in ref. [45] for this purpose, referred to as *Fast PatchMatch*, selecting a set of patches that are similar to $\mathbf{m}_i$ in the sense that the quantity $(\mathbf{m}_i - \mathbf{m}_n)^T(\mathbf{m}_i - \mathbf{m}_n)$ is small. Since we will use Gaussian kernels for the KDE, this ensures that the most important patches are included in the model.

Additionally using diagonal kernel covariances, KDE on $\{y_i, \mathbf{m}_i\}_{i\in\mathcal{A}_i}$ yields the distribution:

$$p(y_i, \mathbf{m}_i|\boldsymbol{\beta}) = |\mathcal{A}_i|^{-1} \sum_{n\in\mathcal{A}_i} \mathcal{N}(y_i|y_n, \beta_y^{-1})\mathcal{N}(\mathbf{m}_i|\mathbf{m}_n, \beta_m^{-1}\mathbf{I}_M) \quad (3)$$

$\mathcal{N}(\cdot|\boldsymbol{\psi}, \boldsymbol{\Sigma})$ here denotes a multivariate Gaussian with mean $\boldsymbol{\psi}$ 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. β_y and β_m are the kernel precisions that, like β_t^*, are tuned on the data $\{t_i, \mathbf{m}_i\}$ using Empirical Bayes (section 2.3).

In general, $p(y_i, \mathbf{m}_i|\boldsymbol{\beta})$ defines a hyperplane, but it may be visualized as a 2D surface in the special case of 1x1x1 patches, which reduce to MR intensities (fig. 1). The surface displays several peaks along the CT-axis due to the mentioned contrast disparity between the MRI and CT, and highlights the ambiguous relationship between the MRI and CT values; the use of larger patches partially remedies this issue, but it persists especially in cases where the structures in the image are larger than the patch size. As we will see in the next section, where we derive the posterior distribution of y_i given the observations, the purpose of the artifact noise model is to resolve this ambiguity.

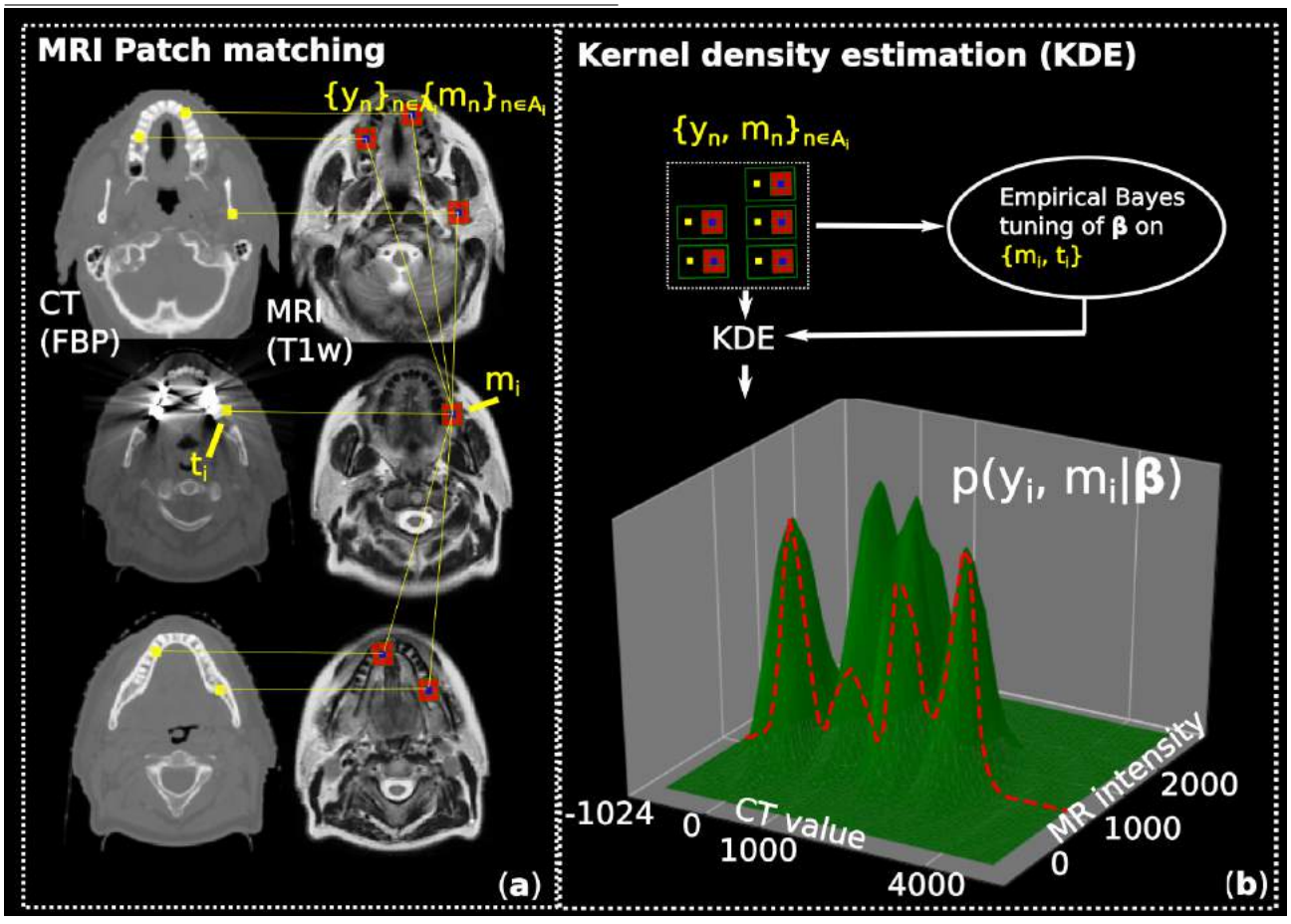


Figure 1: *(a): For each corrupted voxel, a regression point set of CT value / MRI patch pairs, $\{y_n, \mathbf{m}_n\}$, is found by matching the observed MRI patch $\mathbf{m}_i$ to patches at uncorrupted locations (far from the implants). (b): On the regression point set, kernel density estimation (KDE) is used to estimate the joint distribution $p(y_i, \mathbf{m}_i|\boldsymbol{\beta})$ (shown on the surfaces for $1 \times 1 \times 1$ patches). The KDE results depend directly on the hyperparameters $\boldsymbol{\beta}$, which are tuned on the data using empirical Bayes estimation. The red curve is a trace on the surface at a specific $\mathbf{m}_i$, whose relevance is explained in fig. 2.*

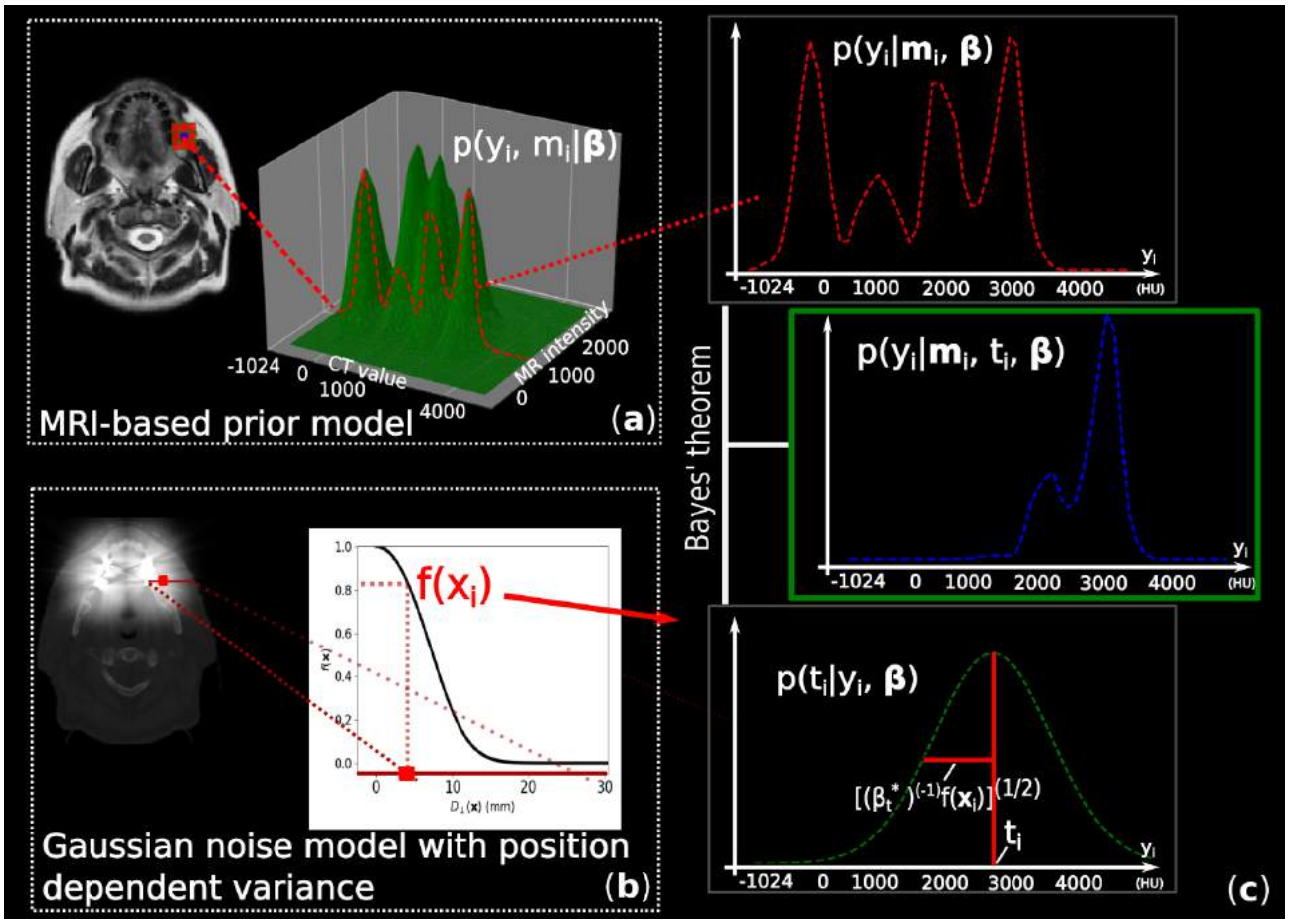


Figure 2: *(a): Given an observed MRI patch, the prior distribution $p(y_i|\mathbf{m}_i, \boldsymbol{\beta})$ corresponds to tracing a curve on the KDE hypersurface (fig. 1 (b)) at the observation. (b): The function $f(\mathbf{x}_i)$ decreases sigmoidally from the metal implants. Its value at the position of the voxel i, $\mathbf{x}_i$, is used to scale the variance of artifact noise model. (c): The artifact noise model $p(t_i|y_i, \boldsymbol{\beta})$ is Gaussian with a variance that decreases sigmoidally with the distance to the metal implants as $\beta_t^{-1} = \beta_t^{*-1} f(\mathbf{x}_i)$; the hyperparameter β_t^* is tuned on the data. The noise model acts as a likelihood function in y_i centered at t_i together with $p(y_i|\mathbf{m}_i, \boldsymbol{\beta})$ to define an improved posterior predictive model, $p(y_i|\mathbf{m}_i, t_i, \boldsymbol{\beta})$.*

2.2 The posterior predictive distribution

The principle behind our model may be seen in the following expression for the posterior distribution of y_i:

$$p(y_i|\mathbf{m}_i, t_i, \boldsymbol{\beta}) \propto p(t_i|y_i, \boldsymbol{\beta}) p(y_i|\mathbf{m}_i, \boldsymbol{\beta}),$$

where we directly applied Bayes theorem and used the independence relations defined in section 2.1.

As illustrated in fig. 2, $p(y_i|\mathbf{m}_i, \boldsymbol{\beta})$ corresponds to an MRI-based prior distribution found by tracing a curve on the KDE hypersurface at the observation $\mathbf{m}_i$. The artifact noise model $p(t_i|y_i, \boldsymbol{\beta})$ in turn acts as a likelihood function in y_i, and the posterior $p(y_i|t_i, \mathbf{m}_i, \boldsymbol{\beta})$ is arrived at by multiplying the two and renormalizing. In the given example, this multiplication effectively picks out a mode from $p(y_i|\mathbf{m}_i, \boldsymbol{\beta})$ near the observed t_i, and thus the observations of the MRI and corrupted CT collaborate to construct an improved predictive distribution. This distribution may then be used for MAR, as we consider in the experiments section.

To calculate the posterior distribution, we write it in terms of eqns. (1)-(3):

$$p(\{y_i\}|\{t_i, \mathbf{m}_i\}, \boldsymbol{\beta}) = \prod_{i \in \mathcal{T}} p(y_i|t_i, \mathbf{m}_i, \boldsymbol{\beta})$$
$$\text{with} \quad p(y_i|t_i, \mathbf{m}_i, \boldsymbol{\beta}) = \frac{p(t_i, y_i, \mathbf{m}_i|\boldsymbol{\beta})}{p(t_i, \mathbf{m}_i|\boldsymbol{\beta})}. \tag{4}$$

We start by treating the numerator. Inserting eqns. (2) and (3) in eqn. (1), and rearranging:

$$p(t_i, y_i, \mathbf{m}_i|\boldsymbol{\beta}) = |\mathcal{A}_i|^{-1} \sum_{n \in \mathcal{A}_i} \left[\mathcal{N}(t_i|y_i, \beta_t^{-1}) \mathcal{N}(y_i|y_n, \beta_y^{-1})\right] \mathcal{N}(\mathbf{m}_i|\mathbf{m}_n, \beta_m^{-1}\mathbf{I}_M) \tag{5}$$

The product of Gaussians in the square brackets may be simplified as follows, dropping y_i-independent constants:

$$\mathcal{N}(t_i|y_i, \beta_t^{-1}) \mathcal{N}(y_i|y_n, \beta_y^{-1}) \tag{6}$$
$$\propto \mathcal{N}(y_i|\mu_n^i, (\beta_t + \beta_y)^{-1}) \mathcal{N}(t_i|y_n, \beta_t^{-1} + \beta_y^{-1}), \tag{7}$$

where we defined

$$\mu_n^i = \frac{\beta_t}{\beta_t + \beta_y} t_i + \frac{\beta_y}{\beta_t + \beta_y} y_n. \quad (8)$$

Consider now the marginal likelihood in the denominator of eqn. (4), which may be calculated by marginalizing eqn. (5). This corresponds to a convolution over the Gaussians in the square brackets, which adds up their variances[46], and we get:

$$p(t_i, \mathbf{m}_i|\boldsymbol{\beta}) = \int_{-\infty}^{\infty} p(t_i, \mathbf{m}_i, y_i|\boldsymbol{\beta}) \mathrm{d}y_i$$
$$= |\mathcal{A}_i|^{-1} \sum_{n \in \mathcal{A}_i} \mathcal{N}(t_i|y_n, \beta_t^{-1} + \beta_y^{-1}) \mathcal{N}(\mathbf{m}_i|\mathbf{m}_n, \beta_m^{-1}\mathbf{I}_M). \quad (9)$$

We now substitute the appropriate part in eqn. (5) with the factorization in eqn. (7), and insert it alongside eqn. (9) in eqn. (4). This leads to the following, normalized distribution:

$$p(y_i|t_i, \mathbf{m}_i, \boldsymbol{\beta}) = \sum_{n \in \mathcal{A}_i} v_n^i \mathcal{N}(y_i|\mu_n^i, (\beta_y + \beta_t)^{-1}), \quad (10)$$

where we defined:

$$v_n^i \equiv \frac{\mathcal{N}(t_i|y_n, \beta_t^{-1} + \beta_y^{-1}) \mathcal{N}(\mathbf{m}_i|\mathbf{m}_n, \beta_m^{-1}\mathbf{I}_M)}{\sum_{n' \in \mathcal{A}_i} \mathcal{N}(t_i|y_{n'}, \beta_t^{-1} + \beta_y^{-1}) \mathcal{N}(\mathbf{m}_i|\mathbf{m}_{n'}, \beta_m^{-1}\mathbf{I}_M)}. \quad (11)$$

This distribution may now be calculated except that we do not know the hyperparameters $\boldsymbol{\beta}$, which directly impact its shape. Of particular importance is the relative weight of t_i and $\{y_n\}_{n \in \mathcal{A}_i}$ in μ_n^i, as directly determined by β_t and β_y. To understand how, it is instructive to consider the following special cases for β_t:

$\beta_t \to 0$ **(CT measurement t_i fully corrupted)**
For this setting, $\mu_n^i = y_n^i$ (see eqn. (8)) and:

$$p(y_i|\mathbf{m}_i, t_i, \boldsymbol{\beta}) = p(y_i|\mathbf{m}_i, \boldsymbol{\beta}) = \sum_{n \in \mathcal{A}_i} w_n^i \mathcal{N}(y_i|y_n, \beta_y^{-1})$$
$$\text{with} \quad w_n^i = \frac{\mathcal{N}(\mathbf{m}_i|\mathbf{m}_n, \beta_m^{-1}\mathbf{I}_M)}{\sum_{n' \in \mathcal{A}_i} \mathcal{N}(\mathbf{m}_i|\mathbf{m}_{n'}, \beta_m^{-1}\mathbf{I}_M)}.$$

The observed CT measurement t_i is here completely discarded, and the posterior becomes a Gaussian mixture with MRI-determined weights.

$\beta_t \to \infty$ **(CT measurement t_i not corrupted)**
For this setting, $\mu_n^i \to t_i$ and:

$$p(y_i|\mathbf{m}_i, t_i, \boldsymbol{\beta}) = \delta(y_i - t_i).$$

Here, the MR measurement $\mathbf{m}_i$ is discarded and the only probable configuration is $y_i = t_i$.

$0 < \beta_t < \infty$ **(General case)**
For general hyperparameter settings, the posterior predictive distribution is given by eqn. (10) and is constructed in reference to both t_i and $\mathbf{m}_i$.

The settings of β_t thus determine where the model lies between pure MRI-based prediction and "naive acceptance" of the observed CT values. We made use of this when we scaled β_t with position as $\beta_t^{-1} = \beta_t^{*-1} f(\mathbf{x}_i)$, as this then ensured a transition to naive acceptance far away from the metal implants.

Near the implants, β_t in turn attains a minimum value of β_t^*, the value of which impacts the model in the corrupted regions that we particularly want to address: For a highly corrupted image, the observed CT values are untrustworthy, and so we want to rely more on the MRI and thus be close to case 1; for a less corrupted image, we want to be closer to case 2. To automatically make this decision given our observed data, while simultaneously picking the kernel precisions β_y and β_m to get a reasonable KDE, we use Empirical Bayes hyperparameter estimation.

2.3 Empirical Bayes hyperparameter estimation

Empirical Bayes maximizes the marginal log likelihood of the data (the CT and MRI) given the hyperparameters ($\log p(\{t_i, \mathbf{m}_i\}|\boldsymbol{\beta})$, eqn. (9))[42, 43]. Thus, after having found the regression point sets $\{y_n, \mathbf{m}_n\}_{n \in \mathcal{A}_i}$, $\forall i \in \mathcal{T}$, and evaluated the function $f(\mathbf{x}_i)$, we solve the following maximization problem for $\boldsymbol{\beta} = \{\beta_t^*, \beta_y, \beta_m\}$:

$$\arg\max_{\boldsymbol{\beta}} \Phi(\boldsymbol{\beta}) \quad \text{with} \quad \Phi(\boldsymbol{\beta}) = \sum_{i \in \mathcal{T}} \sum_{n \in \mathcal{A}_i} \log \phi_n^i(\boldsymbol{\beta}) \quad \text{and}$$
$$\phi_n^i(\boldsymbol{\beta}) = \mathcal{N}(t_i|y_n, f(\mathbf{x}_i)\beta_t^{*-1} + \beta_y^{-1}) \mathcal{N}(\mathbf{m}_i|\mathbf{m}_n, \beta_m^{-1}\mathbf{I}_M)]. \quad (12)$$

In defining $\Phi(\boldsymbol{\beta})$, we here assumed a constant $|\mathcal{A}_i|$ that was accordingly dropped as a constant term.

This objective function cannot be maximized in closed form, and so we do it by sequentially improving an estimate in an iterative algorithm, starting from an initial guess. To simplify this task and avoid potentially slower gradient-based optimization, we derive an Expectation Maximization (EM)-algorithm[47] that approximately solves this problem in appendix A.4.

The approximation consists of a truncation of $f(\mathbf{x}_i)$ such that $f(\mathbf{x}_i) = 0$ for $i \in \mathcal{T}_u$, and 1 elsewhere, leading to a new, approximate objective. Given a hyperparameter estimate, an iteration of the algorithm then maximizes a lower bound to this objective that is constructed to be equal to the objective at the current estimate; the maximization is thus guaranteed to increase the objective[47]. The truncation of $f(\mathbf{x}_i)$ allows for this optimization to be done in closed form, and the algorithm becomes fast, simple and gradient-free. The truncation also does not noticeably change the results compared to using exact gradient-based optimization, as we found in initial experiments, owing to the sharp sigmoidal nature of $f(\mathbf{x}_i)$.

Using the average change in the hyperparameter estimates as a stopping criterion, we get algorithm 1. The

Algorithm 1 Empirical Bayes Hyperparameter estimation

1: Choose an initial estimate of the hyperparameters (e.g. $\boldsymbol{\beta} \leftarrow \{0, 0, 0\}$, and set $\delta \leftarrow 0$
2: **while** $\delta > 10^{-3}$ **do**
3: $\boldsymbol{\beta}_0 \leftarrow \boldsymbol{\beta}$
4: **E-step**: Calculate v_n^i using eqn. (11).
5: **M-step**: Update the hyperparameter estimates:

$$[\beta_y]^{-1} \leftarrow \frac{1}{|\mathcal{T}_u|} \sum_{i \in \mathcal{T}_u} \sum_{n \in \mathcal{A}_i} v_n^i (t_i - y_n)^2$$

$$[\beta_t{}^*]^{-1} \leftarrow \frac{1}{|\mathcal{T}_c|} \sum_{i \in \mathcal{T}_c} \sum_{n \in \mathcal{A}_i} v_n^i (t_i - y_n)^2 - [\beta_y]^{-1}$$

$$[\beta_m]^{-1} \leftarrow \frac{1}{|\mathcal{T}|} \sum_{i \in \mathcal{T}} \sum_{n \in \mathcal{A}_i} v_n^i \frac{(\mathbf{m}_i - \mathbf{m}_n)^T (\mathbf{m}_i - \mathbf{m}_n)}{M}.$$

$\delta \leftarrow \sqrt{(\boldsymbol{\beta} - \boldsymbol{\beta}_0)^T (\boldsymbol{\beta} - \boldsymbol{\beta}_0)}/3$
6: **end while**

right-hand side (RHS) of each update equation resembles an empirical variance, showing how this algorithm picks an appropriate β_t^* for the observed image material: The RHS for β_y^{-1} quantifies the variance of the regression point set $\{y_n\}_{n \in \mathcal{A}_i}$ around t_i that occurs naturally even in uncorrupted regions. The RHS for ${\beta_t^*}^{-1}$ in turn quantifies the added variance in the corrupted region above the uncorrupted level. A highly artifact corrupted image accordingly pushes toward large ${\beta_t^*}^{-1}$ and thus small β_t^*, bringing the model closer to purely MRI-based prediction.

Our algorithm in this way provides hyperparameter estimates that are automatically tuned to the patient in consideration, which, in addition to fully defining our predictive model, has several benefits in terms of both accuracy and practical application. We return to these benefits in section 3.1, where we see how they affect our proposed image inpainting algorithm.

3 Experiments

Having covered all the parts, we may now summarize how to calculate the posterior predictive distribution in eqn. (10) (algorithm 2). This algorithm provides an analytical expression that may be used for further development.

We will now use this distribution to define an algorithm in each of the three MAR categories: An image inpainting method that uses Bayesian estimation; a sinogram inpainting method that uses the image inpainting result as a prior; and an MBIR algorithm that uses the model as a prior distribution along with a Poisson likelihood model. While presenting these methods, we show in a retrospective review the results of applying them to an anonymized set of nine head-and-neck RT patients containing dental implants and/or fillings.

Algorithm 2 Calculation of the posterior predictive distribution

1: Calculate the FBP.
2: Threshold the FBP using Otsu's heuristic to define a metal segmentation.
3: Calculate $f(\mathbf{x}_i) = 1 + \tanh(-\frac{D_\perp(\mathbf{x}_i)^2}{\kappa})$ ($\kappa = (10mm)^2$) using the metal segmentation.
4: Threshold $f(\mathbf{x}_i)$ (≤ 0.5) to define the set of uncorrupted voxel indices, $\mathcal{T}_u$.
5: Find $\mathcal{A}_i \subseteq \mathcal{T}_u$, $\forall i \in \mathcal{T}$, using Fast PatchMatch.
6: Store $\mathcal{A}_i$ along with $(\mathbf{m}_i - \mathbf{m}_n)^T (\mathbf{m}_i - \mathbf{m}_n)$ ($\forall n \in \mathcal{A}_i$).
7: Estimate $\boldsymbol{\beta}$ using algorithm 1.
8: The posterior may now be evaluated using the analytical expression in eqn. (10).

For each MAR category, we consider the improvements provided by our model over the conventional alternative: For image inpainting, we compare to purely MRI-guided image inpainting; for sinogram inpainting, to using a prior generated only from the CT; and for MBIR, to the same algorithm with a flat prior.

Technical details For our experiments, we used regression sets of size $|\mathcal{A}_i| = 200$, $\forall i \in \mathcal{T}$, and used cuboidal patches with 5 voxels on each side on image sets with a resolution of $1.2 \times 1.2 \times 2.0mm$ (CT) and $0.5 \times 0.5 \times 5.5mm$ (T1w MRI). The MRIs were resampled to the CT resolution after co-registration using mutual information[48, 49], and the spatial dimensions of the patches were thus $6 \times 6 \times 10mm$. As mentioned earlier, the parameter κ in $f(\mathbf{x}_i)$ was chosen as $(10mm)^2$ for all patients.

For additional details on the image set (scanner models, sequence parameters, etc.), refer to appendix A.1.

Quantitative evaluation Metal artifacts increase the CT value variance within expectedly homogeneous tissue regions, which may be used to quantify their severity[11, 28, 50]. We in particular investigate the artifact-induced variance in the oral cavity and teeth, which were delineated in our patient cohort; for example delineations, see appendix A.2. During contouring of the teeth, we sought to avoid the metal implants and also used the metal mask calculated by Otsu's method to further avoid including them in our analysis.

We indicate these ROIs by the labels $X \in \{\mathsf{O}, \mathsf{T}\}$ and split them into an uncorrupted and corrupted region. We denote these regions as respectively $\mathcal{T}_u^X$ and $\mathcal{T}_c^X$ and define them as the intersection between delineation X and respectively the uncorrupted set $\mathcal{T}_u$, which we defined earlier by truncating $f(\mathbf{x}_i)$, and its complement, the corrupted set $\mathcal{T}_c$.

As an image quality metric, we then calculate the standard deviation of the CT values in the corrupted parts around the mean value in the uncorrupted part, which we denote as $M_{X \in \{O,T\}}$; it is the square root of

the following variance:

$$M_X^2 = \frac{1}{|\mathcal{T}_c^X|} \sum_{i \in \mathcal{T}_c^X} (t_i - \frac{\sum_{j \in \mathcal{T}_u^X} t_j}{|\mathcal{T}_u^X|})^2,$$

Smaller values of M_X suggest less severe artifacts. We show the results averaged over the nine patients for all MARs in fig. 3, where the error bars are the standard errors on the means and the dashed boxes show the results of the MRI-based algorithms. We further use two-sided Student's t-tests for paired, repeated observations ($N = 9$) to test the significance of the improvement with the MRI-based algorithms over the conventional alternatives; the p-values are indicated in the figure. We will in the following refer back to figure 3 several times.

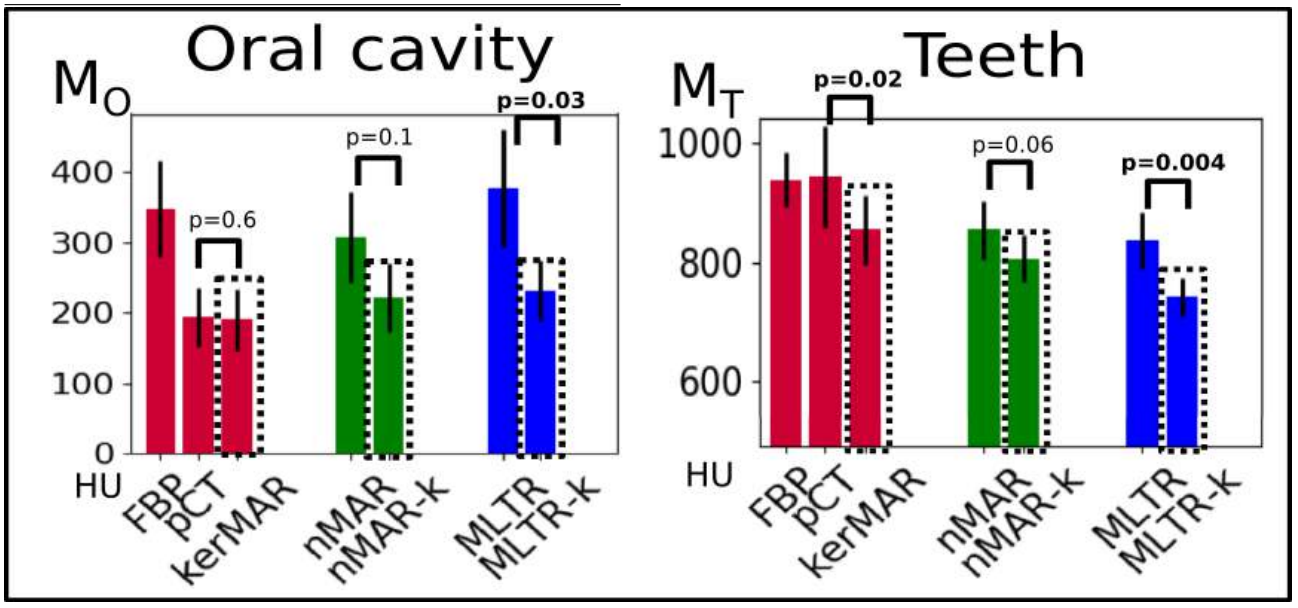


Figure 3: *Results of the quantitative analysis of the oral cavity and teeth ROIs for the various MARs. The shown metric is the standard deviation of the corrupted CT values around a reference mean estimated from uncorrupted CT values. kerMAR, nMAR-k and MLTR-k are respectively our MRI-based image inpainting, sinogram inpainting and MBIR algorithms. The p-values are the results of two-sided Student's t-tests for paired, repeated observations ($N = 9$), which test the significance of the difference between the MRI-based and conventional algorithms.*

3.1 Image inpainting: Kernel regression MAR (kerMAR)

Our image inpainting algorithm "kernel regression MAR" (kerMAR) calculates a Bayesian estimate of the uncorrupted CT value y_i from the posterior distribution. In particular, given a tuned hyperparameter set $\boldsymbol{\beta}$, we calculate the conditional expectation over the posterior in all voxels except those containing metal, as determined by the automatic thresholding. We avoid the metal voxels since the MRI-based prior model may contain no modes for metal, and so they may otherwise be replaced by teeth. kerMAR is defined in algorithm 3 and the graphical interpretation of the estimate in step 2 as the mean of the posterior is illustrated in fig. 4 (left).

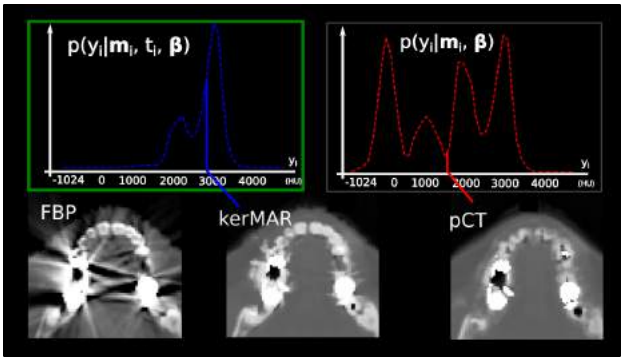


Figure 4: *Illustration of our inpainting algorithm kerMAR (kernel regression MAR) and purely MRI-guided image inpainting (pseudo-CT, pCT). kerMAR calculates the mean of $p(y_i|t_i, \mathbf{m}_i, \boldsymbol{\beta})$, while pCT does the same on $p(y_i|\mathbf{m}_i, \boldsymbol{\beta})$.*

kerMAR vs. purely MRI-based inpainting (pCT) Eqn. (13) references both the MRI and the corrupted CT, providing a potential improvement over purely MRI-

Algorithm 3 Kernel regression MAR (kerMAR)

1: Perform steps 1-7 in algorithm 2.
2: Calculate the posterior expectation:

$$\bar{y}_i = \int_{-\infty}^{\infty} y_i p(y_i|t_i, \mathbf{m}_i, \boldsymbol{\beta}) dy_i = \sum_{n \in \mathcal{A}_i} v_n^i \mu_n^i, \quad \forall i \in \mathcal{T}. \tag{13}$$

3: Use the metal mask in step 2, algorithm 2, to inpaint the metal implants on the kerMAR image with values from the FBP.

based inpainting, which is similar in principle to existing pseudo-CT (pCT) generation algorithms for MRI-only RT[51, 52]. We compare kerMAR to such purely MRI-based inpainting by downscaling β_t^* with a large factor (1000), such that $p(y_i|t_i, \mathbf{m}_i, \boldsymbol{\beta}) \approx p(y_i|\mathbf{m}_i, \boldsymbol{\beta})$ in the corrupted region. The resulting "pCT" estimate is illustrated in fig. 4, which shows how it may become a mean over a multimodal distribution.

Fig. 5 shows the results of applying kerMAR and pCT. The MRI-based pCT estimate provided a superficially decent correction; however, it misestimated tooth regions of sizes similar to the patch size (circles) and introduced errors where the CT and MRI were imperfectly co-registered, as may be seen by comparing between the MRI, pCT and FBP at the arrows. In the kerMAR image, however, both these issues were better handled due to the incorporation of the CT information.

As seen in the quantitative results in fig. 3, this difference in performance between kerMAR and pCT is reflected by the similarity in the oral cavity ($p = 0.6$) combined with the significant difference in the teeth ($\sim 100HU$ at $p = 0.02$): Compared to FBP, we particularly note that both performed well in the oral cavity, while kerMAR did relatively better in the teeth. More examples of pCT vs. kerMAR are additionally shown in fig. 11 in appendix A.3.

Impact of the $\boldsymbol{\beta}$ settings The performance of kerMAR depends on the appropriate settings of $\boldsymbol{\beta}$, which in particular controls the degree to which the corrupted CT was included in the prediction. By using empirical Bayes hyperparameter estimation, these settings were tuned to each individual patient.

We considered the benefit of this patient-specific tuning by calculating the kerMAR images with the hyperparameters swapped between patients with different levels of artifact corruption, as shown in fig. 6. The kerMARs on the top were calculated with the tuned hyperparameters, the bottom using hyperparameters swapped (7 with 2, 6 with 9). 2 and 9 were cases with comparatively minor corruption, 7 and 6 with more severe corruption, such that the hyperparameters upon swapping led to respectively an effective under and over-estimation of the artifact noise by the model. Accordingly, we see a lessening in artifact removal for patients 7 and 6; and vice versa, the less corrupted cases 2 and 9 show less succesful bone/air disambiguation, similar but less extensive compared to what we saw with pCT. This shows how the automatic tuning of the hyperparameters optimized the model to the individual patients.

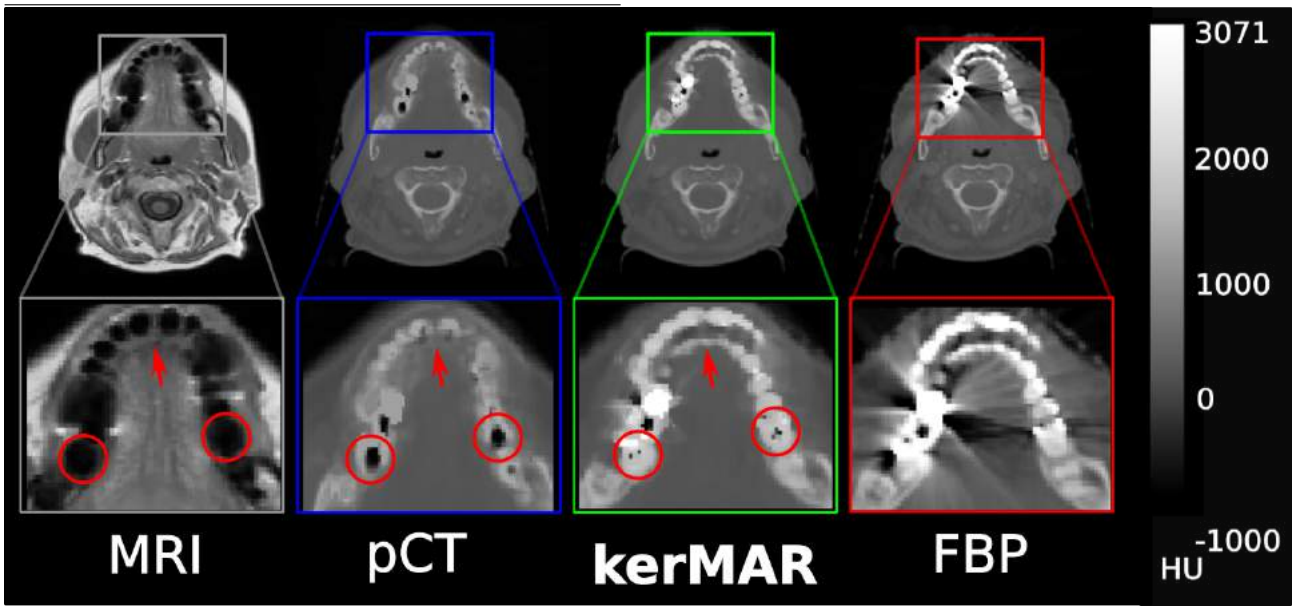


Figure 5: *Axial slices of kerMAR and pCT (kerMAR with infinite artifact noise variance) shown together with the FBP for a head-and-neck patient. The tendency to miss-classify bone with pCT is much decreased through the use of the CT information using kerMAR (blue arrows), as are anatomical errors due to an imperfect co-registration (red arrow). The latter may be seen by comparing the MRI to the FBP, noting how the pCT imposes the MRI anatomy to a larger degree than kerMAR.*

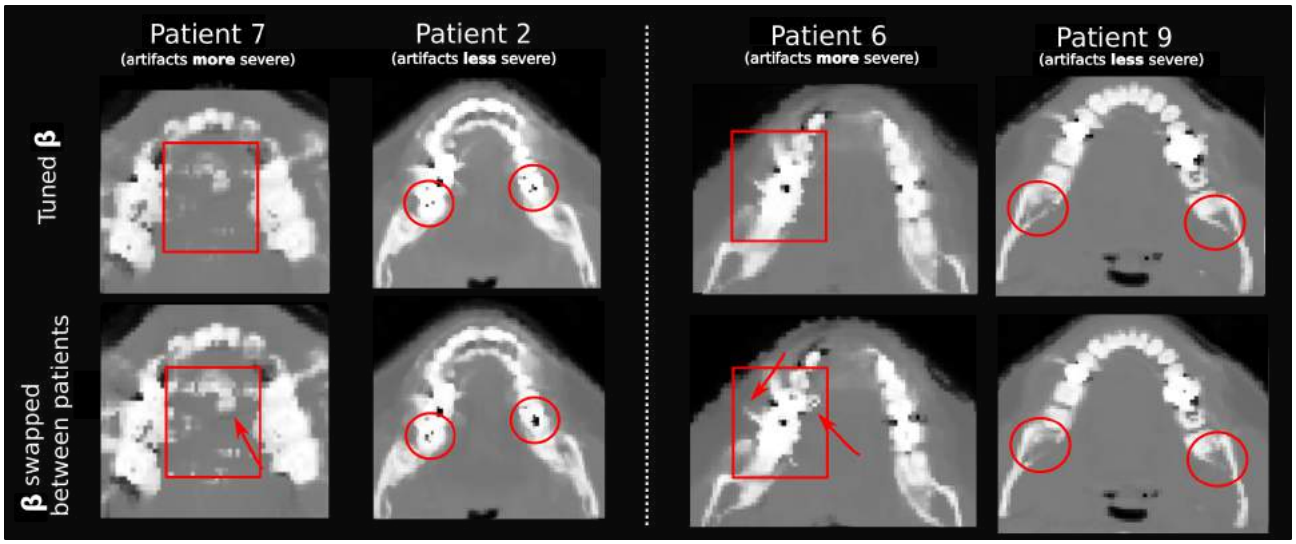


Figure 6: *kerMAR calculated for 4 head-and-neck patients using (top) the tuned hyper-parameters β and (bottom) hyper-parameters swapped (7 with 2, 6 with 9). Patients 7 and 6 showed more severe artifacts than respectively 2 and 9. The arrows point to increases in anatomical errors and artifacts, which occurred for patients 7 and 6 due to a relative under-estimation of the artifact noise level. The circles show bone/air/soft tissue disambiguation issues, which occured for patients 2 and 9 due to a relative* over-*estimation of the artifact noise level.*

3.2 Sinogram inpainting: nMAR with a kerMAR prior (nMAR-k)

As mentioned in the introduction, CT images are reconstructed from x-ray measurements acquired at a set of view angles around the patient that are stored as projections in a sinogram[13]. Sinogram inpainting algorithms replace the corrupted metal projections[13] by estimates that better fit the reconstruction model. Normalized MAR (nMAR) in particular calculates these estimates by simulating a CT acquisition on a prior image, generated by processing the corrupted CT, and then uses them for interpolation-based sinogram inpainting. We describe in appendix A.5 the nMAR implementation that we use in this paper and provide details on the way that we calculated such a CT-based prior: We settled on our method by experimenting with the recommendations in refs. [17] and [19]. Briefly, we did as follows: An initial artifact reduced image was calculated using linear interpolation MAR[20], which was then segmented using K-means clustering. The segmented regions were finally assigned their cluster mean values.

We will refer to nMAR using the CT-based prior in appendix A.5 as conventional nMAR, or simply nMAR. By relying on a potentially flawed prior image based on the corrupted CT, nMAR risks the introduction of new artifacts to the image. This forms the motivation to use our potentially higher quality MRI-based kerMAR prior in step 2.

Comparing MRI-based nMAR to conventional nMAR We denote nMAR with our MRI-based kerMAR prior as nMAR-k and show the results of applying it in fig. 7 alongside nMAR. The figure shows how nMAR introduced artifacts in the oral cavity that were absent with nMAR-k. The origin of these artifacts may be seen by comparing the priors (far right in fig. 7): The K-means prior appears directly responsible for introducing the artifacts. This highlights the utility of incorporating information from the relatively artifact-free MRI. The additional images in fig. 7, appendix A.3 shows additional cases where nMAR-k provides improved artifact reduction as well as less introduction of artifacts.

Considering the quantitative results in fig. 3, nMAR-k yielded decreases in M_O and M_T of $\sim (100, 30) HU$, albeit at strictly insignificant levels ($p > 0.05$), with p-values of respectively 0.1 and 0.06. We note that the results are closer to significance in the teeth, as expected from in particular the images in the appendix: Most of the introduction of artifacts occurs close to the implants where the prior is most difficult to create from the corrupted CT.

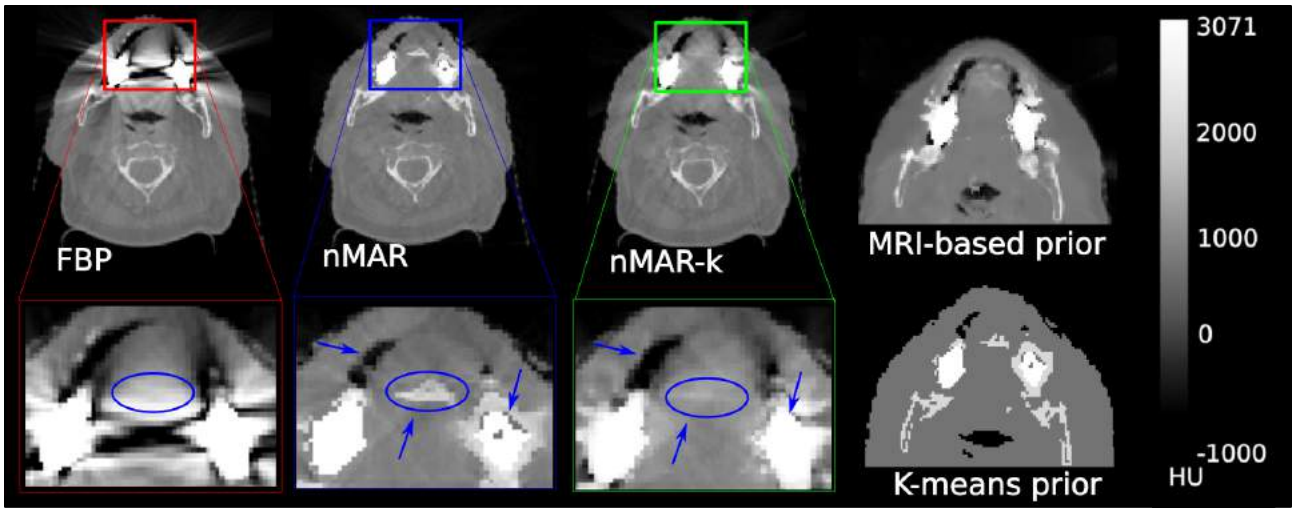


Figure 7: *Results using the nMAR sinogram inpainting algorithm with (middle) a K-means clustering based prior and (center-right) with the kerMAR image as prior. Shown for reference is the FBP images (left) and the priors (far right). Comparing nMAR and nMAR-k, the blue arrows show artifacts/anatomical deformations that were either introduced or left behind by K-means nMAR due to a flawed prior, but that are absent with nMAR-k.*

3.3 MBIR: MLTR with prior (MLTR-k)

Model-based iterative reconstruction (MBIR) iteratively maximizes a probabilistic model of the image coefficients that consists of a sinogram data likelihood and an image prior distribution. In the Maximum Likelihood Transmission Reconstruction (MLTR) algorithm, the prior is assumed flat and the log likelihood is maximized with respect to the image in an EM-like algorithm[53, 54] (see appendix A.6). The likelihood is a Poisson distribution centered on the average x-ray intensities, calculated given the image using the Lambert-Beers law of x-ray attenuation[53, 54]. By itself, the Poisson likelihood provides artifact reduction by down-weighing the contribution of the low intensity, noisy measurements to the final image, e.g. those through the highly attenuating metal[13]. This however comes at a cost of convergence speed for voxels near the metal implants that rely on such measurements, and does not necessarily lead to removal of all artifacts. The prior distribution thus becomes important, which motivates using our MRI-based predictive model.

The MLTR algorithm may be modified to include a general image prior, which changes its update equation to include a prior term that depends on the first and second derivative of the log prior[33, 54]. Using eqn. (10), we therefore calculate the first and second derivatives of our MRI-based model, $\ln(p(y_i|t_i, \mathbf{m}_i, \boldsymbol{\beta}))$, which become:

$$\ln(p(y_i|t_i, \mathbf{m}_i, \boldsymbol{\beta}))' = (\beta_t + \beta_y)[\sum_{n \in \mathcal{A}_i} \tilde{v}_n^i \mu_n^i - y_i] \quad \text{and}$$

$$\ln(p(y_i|t_i, \mathbf{m}_i, \boldsymbol{\beta}))'' = (\beta_t + \beta_y)^2 \Big(\sum_{n \in \mathcal{A}_i} \tilde{v}_n^i {\mu_n^i}^2 - [\sum_{n \in \mathcal{A}_i} \tilde{v}_n^i \mu_n^i]^2 \Big) - (\beta_t + \beta_y)$$

$$\text{where} \quad \tilde{v}_n^i = \frac{v_n^i \mathcal{N}(y_i|\mu_n^i, (\beta_t + \beta_y)^{-1})}{\sum_{n' \in \mathcal{A}_i} \mathcal{N}(y_i|\mu_{n'}^i, (\beta_t + \beta_y)^{-1}) v_{n'}^i}.$$

We show the steps to calculate these derivatives along with a definition of the MLTR algorithm in appendix A.6.

The derivatives are evaluated during each iteration and push the image estimate in the direction of the first order derivative. At the root of the derivative, y_i is matched to a weighted average that is similar to the kerMAR estimate in eqn. (13), but where the weights have been updated in reference to the estimated y_i. This is directly analogous to how our posterior predictive model arose from modifying an MRI-based prior, and serves in a similar way to pick out modes from the distribution (see figs. 1 and 4). If, for example, the slightly multi-modal example posterior in figs. 1 and 4 was used as our MLTR prior, the estimate of y_i would pick which of the two modes the prior term pushes toward.

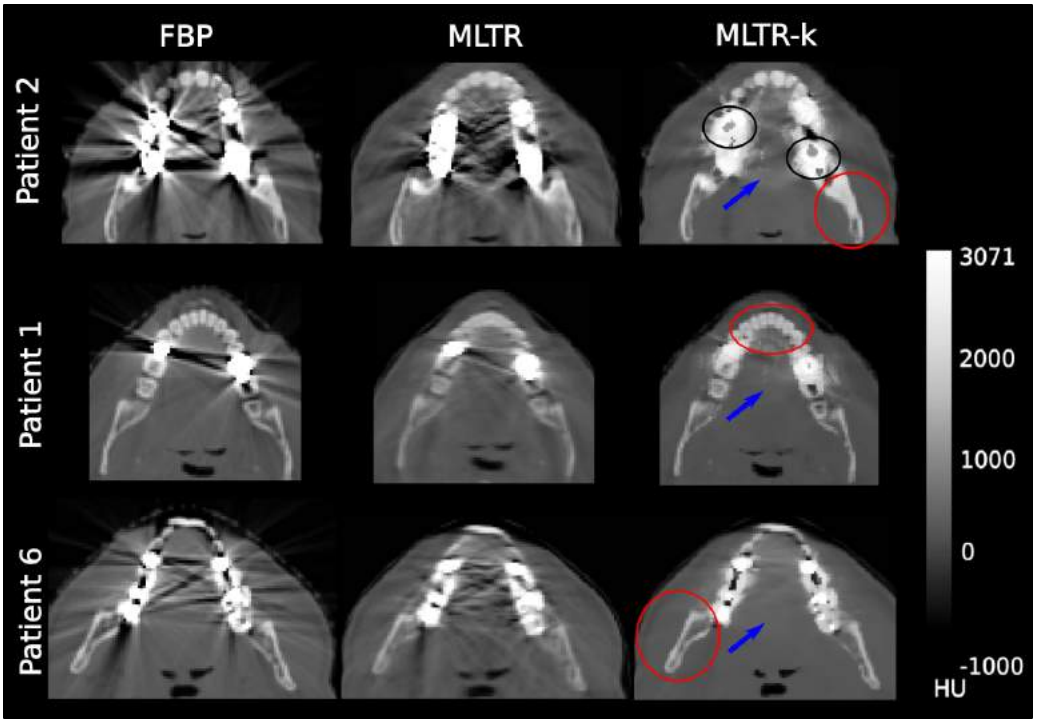


Figure 8: *a) MLTR and MLTR-k results shown beside the FBP images. The MLTR-k displays fewer artifacts than the MLTR while and a sharper image (red ring and blue arrows). A potential danger with MLTR-k, however, is that the prior may become too strong and lead over-introduction of the MRI anatomy, as we see in the black rings for patient 2*

MLTR with MRI-based prior vs. prior-free MLTR We will refer to MLTR using our MRI-based prior model as MLTR-k. We ran MLTR and MLTR-k for the nine head and neck patients at the same image resolution as the FBP, using a flat initial image estimate with attenuation coefficients of 10^{-4} and terminated the algorithms once the voxel-averaged change in the volume between iterations went below 10^{-6}.

Fig. 8 a) shows an example of a reconstruction using MLTR-k alongside one using prior-free MLTR. The prior led to positive results in terms of image sharpness (red circles), in particular for patients 5 and 6, as well as improved artifact reduction (blue arrows). As may be seen in fig. 9, which shows the voxel-averaged image change between iterations, it also led to markedly faster convergence; the number of required iterations was effectively cut in half from $\sim 400 - 600$ to $\sim 200 - 300$.

On the negative side, use of the prior also led to introduction of artifacts in some cases: The black circles for patient 2 show some bloating and miseestimation in the teeth, while the oral cavity for patient 5 shows some added noise. In general, however, we found improvements for our patients, as also reflected in the quantitative results in fig. 3 that show significant improvements both in the teeth ($\sim 150HU$, $p = 0.004$) and oral cavity ($\sim 110HU$, $p = 0.03$), as well as in the additional images in appendix A.3, fig. 13.

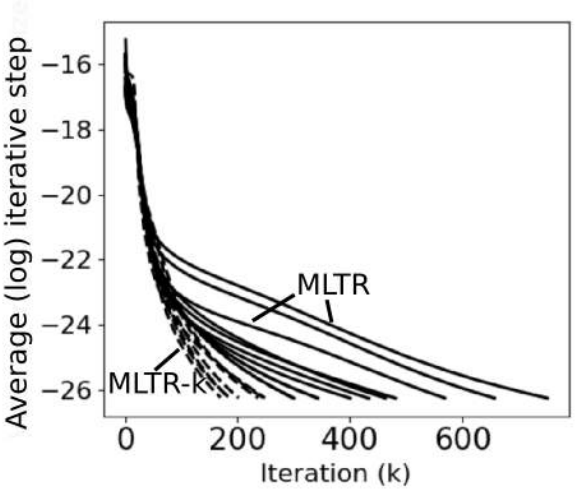


Figure 9: *Convergence plot of MLTR and MLTR-k for the 9 patients. Solid and dashed curves show the log of the absolute, voxel averaged change between iterations at iteration k for MLTR and MLTR-k respectively. While the curves start out similarly, the MLTR-curves soon break off, leading to slower convergence.*

4 Discussion

We have presented a Bayesian, MRI-based predictive model for uncorrupted CT value prediction from an observed, conventional-sequence MRI and artifact corrupted CT. We then presented three methods that applied it to metal artifact reduction: An image inpainting algorithm, a sinogram inpainting algorithm and an MBIR-algorithm. Considering nine head-and-neck RT patients, we compared our image inpainting algorithm (kerMAR) to purely MRI-based image inpainting (pCT), and saw a benefit of including the corrupted CT to disambiguate bone and air and correct co-registration errors. This manifested chiefly as quantitative improvements in the teeth, while the soft tissue in the oral cavity was handled well by both pCT and kerMAR. Then, comparing our sinogram inpainting algorithm nMAR-k to the same algorithm in the same implementation, but with a conventionally generated, CT-based prior image, we saw quality improvements as the MRI information improved the prior in highly corrupted regions. Finally, comparing our MBIR algorithm with the MRI-based prior (MLTR-k) to MBIR with the same Poisson likelihood model, but with no prior, we saw both an increase in convergence speed and increased artifact reduction.

These benefits are relevant improvements to the state of the art in MRI-based MAR, which consists of the three image inpainting algorithms that we briefly covered in the introduction, for the following reasons: First, our method proved able to address the bone/air disambiguation issue despite using a T1w MRI, which is challenging for existing algorithms. The authors of the MRI segmentation-based algorithm (Delso et al.) proposed in [37] to resolve the issue by using ultra short echo time (UTE) MRI sequences for better bone/air contrast, but our ability to forego such specific alterations to the MRI sequence leaves it free to be chosen in reference to other treatment concerns. Such potential, clinically motivated changes to the MRI sequence are in addition accounted for automatically by our algorithm due to the hyperparameter tuning, which makes our approach adaptable to a changing clinical environment.

Second, the method by Park et al.[38] relies on the presence of an anatomically similar, uncorrupted slice in the volume that may succesfully be deformed to the corrupted slice. Our method is relatively more adaptable, as it uses a regression model on anatomical features that may be found anywhere in the patient volume. Together with the automatic tuning of its hyperparameters that accounts for the extent of the artifact corruption in each individual patient, this makes it applicable to a wider range of patients.

Third, the incorporation of our model in sinogram inpainting and MBIR methods allowed us to supplement its performance by that of existing MAR methods, and, as we have shown, in turn improve their performance.

Our experiments with using our model as a prior for MBIR focused on the relatively simple MLTR-algorithm, which chiefly handles the noise artifacts that stem from photon starvation of the metal projections: As mentioned in paragraph 3.3 and briefly explained in appendix A.6, the Poisson model down-weighs the noisy, photon-starved measurements that are acquired through the metal, such that the prior distribution becomes of greater relative importance[13, 21, 25, 33, 53]. Using our MRI-based model as a prior led to general improvements, but since the Poisson likelihood is not a perfect solution to the photon starvation problem, and does not account for all artifacts, in particular beam hardening, it also led to potential introduction of artifacts: The noise for patient 5 and the bloating for patient 2 (fig. 8) are the result of the likelihood pushing toward erroneous CT values and thus picking out wrong modes from the prior distribution.

Our results may however generalize to MBIR algorithms with more complex likelihoods, and if the photon starvation does not entirely dominate as an artifact source, this may help to avoid the artifact introduction issues that our prior faced with MLTR. On the other hand, when photon starvation does dominate, the benefit of using a complex forward model may reduce to the effects of the noise model: For instance, a beam hardening model would lose relevance when the error in the metal projections due to noise was much larger than that due to beam hardening; this may e.g. be the case with multiple dental implants[10]. In such cases, the likelihood model cannot solve the artifact reduction problem by itself, and using our proposed MRI-based distribution for prior modelling becomes particularly relevant.

Clinical implementation As mentioned in the introduction, we presented an earlier conception of our image inpainting method, kerMAR, in ref. [40], and compared to the clinically oMAR algorithm (MAR for orthopedic implants, Philips Healthcare)[55]. We here saw a benefit in terms of artifact reduction in particular near the implants, showing a benefit of our method compared to a state-of-the-art method in clinical use.

A potential barrier to clinical implementation is time constraints, which in the future are bound to become increasingly more relevant as other parts of the RT workflow, e.g. organ/tumor delineation and treatment planning, become faster and more automatized. The time-consuming part of our model is finding the sets $\mathcal{A}_i$, $\forall i \in \mathcal{T}$, using Fast PatchMatch[45] (step 5 in algorithm 2). In our Python implementation on a single CPU (Intel Core i7-4712HQ @ 2.30GHz), this took between 10-30min. The algorithm is however parallelizable, and on a similar-sized dataset, the authors of Fast PatchMatch reported results on the order of $\sim$ 1min on a multi-CPU cluster[45]. In the future, we intend to look into speeding up our algorithm in a similar fashion.

Despite the time required for the patch matching in

its current unoptimized state, we still found a general speed-up of the iterative MLTR-k algorithm compared to MLTR: MLTR-k saved between 200-300 iterations, which on our system with $\sim$ 20 sec per MLTR iteration translated to approximately 66-100min. In the worst case sceneario (66 min gain, 30min loss), we may thus expect a speed increase of about 50%.

Limitations of the study A limitation in our study is the difficulty of defining the state of the art of the nMAR sinogram inpainting algorithm, since the method of creating the prior image is not directly a part of its definition: For this paper, we experimented with the recommendations by Meyer et al. in [17, 18] and Lell et al. in [19] to create the prior, and settled on the method in appendix A.5. Our results with this prior are however less impressive than those reported in references [17] and [19]. While this may be partially explained by the relatively large and numerous implants that we faced in our study, which led to high demands of the prior quality, this suggests a potential for over-estimating the relative benefit of our MRI-based prior.

One reason for the inferior nMAR results in this study is the use of liMAR in the prior generation, which is prone to introducing artifacts[19, 20]; we however found that this was necessary for the most corrupted patients, and that not using it led to over-all worse results for our patients. Better results would be possible if the prior generation method was altered between patients, but this would cause the method to be non-automatic, and so we decided against this.

Another possible limitation is the low resolution of our MRI dataset along the axial (z-)direction (5.5mm vs. 2.0mm on the CT). While this allowed us to show the ability of our methods to resolve co-registration issues, it led to results that may actually lead to an underestimation of their performance. An example is patient 5 for kerMAR (fig. 11 in appendix A.3) where the regions around the windpipe were poorly co-registered between MRI and CT to an extent that 1) would be unlikely with a higher resolution MRI acquired using e.g. a 3D reconstruction sequence[35], and 2) in clinical practice might have prompted a more accurate and manually guided deformable co-registration.

5 Conclusion

We have presented a novel MRI-based Bayesian generative model of uncorrupted CT images, and derived the associated posterior predictive distribution of the uncorrupted image given the observed corrupted CT and a co-registered, conventional-sequence MRI. This model is versatile as it requires no user input and automatically tunes its hyperparameters on the available image data. We used this distribution to define three automatic MAR approaches: 1) The image inpainting algorithm, kernel regression MAR (kerMAR), which, as we showed, leveraged the information in corrupted CT to help with bone/air disambiguation and fix co-registration errors and decreased the standard deviation in the teeth around a reference mean by $\sim 100HU$ at a significance level of $p = 0.02$; 2) The sinogram inpainting algorithm nMAR-k that used the kerMAR image as a prior for normalized MAR (nMAR), which led to improvements over using a conventional, CT-based prior by introducing fewer artifacts and decreasing the aforementioned standard deviation by $(100, 30)HU$ at $p = (0.1, 0.06)$ in respectively the oral cavity and teeth; and 3) MLTR-k that used our predictive distribution as a prior for the MBIR algorithm MLTR, leading to improvements in terms of both speed (of $\sim 50\%$) and artifact reduction with standard deviation decreases in the oral cavity and teeth of $(150, 110)HU$ at $p = (0.03, 0.004)$. We conclude in general that our method provides a versatile way to use the anatomical information in the MRI to boost the performance of MAR.

6 Acknowledgments

The authors would like to express their appreciation to Johan Nuyts (UZ Leuven) and Tao Sun (MGH) for valuable discussions.

7 References

Bibliography

[1] M. Bal and L. Spies, “Metal artifact reduction in CT using tissue-class modeling and adaptive prefiltering”, *Medical Physics*, vol. 33, no. 8, pp. 2852–2859, 2006, ISSN: 00942405. DOI: 10.1118/1.2218062.

[2] O. Jäkel and P. Reiss, “The influence of metal artefacts on the range of ion beams”, *Physics in Medicine and Biology*, vol. 52, no. 3, pp. 635–644, 2007, ISSN: 00319155. DOI: 10.1088/0031-9155/52/3/007.

[3] K. M. Andersson, A. Ahnesjö, and C. V. Dahlgren, “Evaluation of a metal artifact reduction algorithm in CT studies used for proton radiotherapy treatment planning”, vol. 15, no. 5, pp. 112–119, 2014.

[4] D. Giantsoudi, B. De Man, J. Verburg, A. Trofimov, Y. Jin, G. Wang, L. Gjesteby, and H. Paganetti, “Metal artifacts in computed tomography for radiation therapy planning: dosimetric effects and impact of metal artifact reduction”, eng, *Physics in Medicine and Biology*, vol. 62, no. 8, R49–80, R49–R80, 2017, ISSN: 13616560, 00319155. DOI: 10.1088/1361-6560/aa5293.

[5] Y. Kim and W. A. Tomé, "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)", *Medical and Biological Engineering and Computing*, vol. 45, no. 11, pp. 1045–1051, 2007, ISSN: 01400118. DOI: 10.1007/s11517-007-0196-8.

[6] I. Dietlicher, M. Casiraghi, C. Ares, A. Bolsi, D. C. Weber, A. J. Lomax, and F. Albertini, "The effect of surgical titanium rods on proton therapy delivered for cervical bone tumors: Experimental validation using an anthropomorphic phantom", *Physics in Medicine and Biology*, vol. 59, no. 23, pp. 7181–7194, 2014, ISSN: 13616560. DOI: 10.1088/0031-9155/59/23/7181.

[7] H. Kwon, K. S. Kim, Y. M. Chun, H. G. Wu, J. N. Carlson, J. M. Park, and J. I. Kim, "Evaluation of a commercial orthopaedic metal artefact reduction tool in radiation therapy of patients with head and neck cancer", *British Journal of Radiology*, vol. 88, no. 1052, pp. 13–22, 2015, ISSN: 00071285. DOI: 10.1259/bjr.20140536.

[8] M. Maerz, O. Koelbl, and B. Dobler, "Influence of metallic dental implants and metal artefacts on dose calculation accuracy", *Strahlentherapie und Onkologie*, vol. 191, no. 3, pp. 234–241, 2015, ISSN: 0179-7158. DOI: 10.1007/s00066-014-0774-2. [Online]. Available: http://link.springer.com/10.1007/s00066-014-0774-2.

[9] T Koehler, B Brendel, and K. Brown, "A New Method for Metal Artifact Reduction in CT", *The International Conference on Image Formation in X-ray Computed Tomography*, 2011. [Online]. Available: http://repository.tudelft.nl/assets/uuid:22d5815d-dcfe-48df-93a4-9d1c4e8e85fc/MS-33.229.pdf.

[10] L. Gjesteby, B. D. E. Man, Y. Jin, H. Paganetti, J. Verburg, D. Giantsoudi, and G. E. W. Fellow, "Metal Artifact Reduction in CT : Where Are We After Four Decades ?", vol. 4, 2016.

[11] M. Beister, D. Kolditz, and W. A. Kalender, "Iterative reconstruction methods in X-ray CT", *Physica Medica*, vol. 28, no. 2, pp. 94–108, 2012, ISSN: 11201797. DOI: 10.1016/j.ejmp.2012.01.003. [Online]. Available: http://dx.doi.org/10.1016/j.ejmp.2012.01.003.

[12] L. W. Goldman, "Principles of CT and CT technology.", *Journal of nuclear medicine technology*, vol. 35, no. 3, pp. 115–130, 2007, ISSN: 0091-4916. DOI: 10.2967/jnmt.107.042978. [Online]. Available: http://www.ncbi.nlm.nih.gov/pubmed/17823453.

[13] T. M. Buzug, "Computed Tomography", p. 526, 2008. DOI: 10.1007/978-3-540-39408-2.

[14] J. P. Stonestrom, R. E. Alvarez, and A. Macovski, "A Framework for Spectral Artifact Corrections in X-Ray CT", *IEEE Transactions on Biomedical Engineering*, vol. BME-28, no. 2, pp. 128–141, 1981, ISSN: 15582531. DOI: 10.1109/TBME.1981.324786.

[15] J Hsieh, R. C. Molthen, C. a. Dawson, and R. H. Johnson, "An iterative approach to the beam hardening correction in cone beam CT.", *Medical physics*, vol. 27, no. 1, pp. 23–29, 2000, ISSN: 00942405. DOI: 10.1118/1.598853.

[16] M. M. Lell, E. Meyer, M. Schmid, R. Raupach, M. S. May, M. Uder, and M. Kachelriess, "Frequency split metal artefact reduction in pelvic computed tomography", *European Radiology*, vol. 23, no. 8, pp. 2137–2145, 2013, ISSN: 09387994. DOI: 10.1007/s00330-013-2809-y.

[17] E. Meyer, R. Raupach, M. Lell, B. Schmidt, and M. Kachelrieß, "Normalized metal artifact reduction „ NMAR ... in computed tomography", pp. 5482–5493, 2010. DOI: 10.1118/1.3484090.

[18] E. Meyer, R. Raupach, M. Lell, B. Schmidt, and M. Kachelriess, "Edge-preserving metal artifact reduction", vol. 8313, 83133A, 2012. DOI: 10.1117/12.906392. [Online]. Available: http://proceedings.spiedigitallibrary.org/proceeding.aspx?doi=10.1117/12.906392.

[19] M. M. Lell, E. Meyer, M. A. Kuefner, M. S. May, R. Raupach, M. Uder, and M. Kachelriess, "Normalized metal artifact reduction in head and neck computed tomography", *Investigative Radiology*, vol. 47, no. 7, pp. 415–421, 2012, ISSN: 00209996. DOI: 10.1097/RLI.0b013e3182532f17.

[20] W. A. Kalender, R Hebel, and J Ebersberger, "Reduction of ct artifacts caused by metallic implants", eng, *Radiology*, vol. 164, no. 2, pp. 576–577, 1987, ISSN: 15271315, 00338419. DOI: 10.1148/radiology.164.2.3602406.

[21] B. De Man, *Iterative Reconstruction for Reduction of Metal Artifacts in Computed Tomography*. 2001, ISBN: 9056823000. [Online]. Available: http://scholar.google.com/scholar?hl=en{\&}btnG=Search{\&}q=intitle:Iterative+Reconstruction+for+Reduction+of+Metal+Artifacts+in+Computed+Tomography{\#}3.

[22] I. A. Elbakri and J. a. Fessler, "Statistical image reconstruction for polyenergetic X-ray computed tomography", *IEEE Trans Med Imaging*, vol. 21, no. 2, pp. 89–99, 2002, ISSN: 0278-0062. DOI: 10.1109/42.993128. [Online]. Available: http://www.ncbi.nlm.nih.gov/pubmed/12953909{\%}5Cnhttp://www.ncbi.nlm.nih.gov/pubmed/11929108.

[23] J. A. O'Sullivan and J. Benac, "Alternating minimization multigrid algorithms for transmission tomography.", *IEEE transactions on medical imaging*, vol. 26, no. 3, pp. 283–97, 2007, ISSN: 0278-0062. DOI: 10.1109/TMI.2006.886806. [Online]. Available: http://www.ncbi.nlm.nih.gov/pubmed/17354635.

[24] B. Hamelin, Y. Goussard, D. Gendron, J.-P. Dussault, G. Cloutier, G. Beaudoin, and G. Soulez, "Iterative CT reconstruction of real data with metal artifact reduction", *IEEE International Symposium on Biomedical Imaging*, no. 1, pp. 1453–1456, 2008. DOI: 10.1109/ISBI.2008.4541281. [Online]. Available: http://ieeexplore.ieee.org/lpdocs/epic03/wrapper.htm?arnumber=4541281.

[25] K. Van Slambrouck and J. Nuyts, "Local models for artefact reduction in iterative CT reconstruction", *IEEE Nuclear Science Symposium Conference Record*, no. 2, pp. 4346–4350, 2012, ISSN: 10957863. DOI: 10.1109/NSSMIC.2011.6153837.

[26] J. Nuyts, B. De Man, J. A. Fessler, W. Zbijewski, and F. J. Beekman, "Modelling the physics in the iterative reconstruction for transmission computed tomography.", *Physics in medicine and biology*, vol. 58, no. 12, R63–96, 2013, ISSN: 1361-6560. DOI: 10.1088/0031-9155/58/12/R63. [Online]. Available: http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=3725149{\&}tool=pmcentrez{\&}rendertype=abstract.

[27] K. Li, J. Tang, and G.-H. Chen, "Statistical model based iterative reconstruction (MBIR) in clinical CT systems: Experimental assessment of noise performance", *Medical Physics*, vol. 41, no. 4, p. 041 906, 2014, ISSN: 00942405. DOI: 10.1118/1.4867863. [Online]. Available: http://doi.wiley.com/10.1118/1.4867863.

[28] S. T. II, J. H. Siewerdsen, and J. W. Stayman, "Model-based iterative reconstruction for flat-panel cone-beam CT with focal spot blur, detector blur, and correlated noise", *Physics in Medicine and Biology*, vol. 61, no. 1, p. 296, 2016, ISSN: 0031-9155. DOI: 10.1088/0031-9155/61/1/296. [Online]. Available: http://stacks.iop.org/0031-9155/61/i=1/a=296.

[29] L. Fu, T.-c. Lee, S. M. Kim, A. M. Alessio, P. E. Kinahan, Z. Chang, K. Sauer, M. K. Kalra, and B. D. Man, "Comparison Between Pre-Log and Post-Log Statistical Models in Ultra-Low-Dose CT Reconstruction", vol. 36, no. 3, pp. 707–720, 2017.

[30] A. Makeev and S. J. Glick, "Investigation of statistical iterative reconstruction for dedicated breast CT", *Medical Physics*, vol. 40, no. 8, p. 081 904, 2013, ISSN: 00942405. DOI: 10.1118/1.4811328. [Online]. Available: http://doi.wiley.com/10.1118/1.4811328.

[31] J. Webster Stayman and J. Fessler, "Nonnegative definite quadratic penalty design for penalized-likelihood reconstruction", *2001 IEEE Nuclear Science Symposium Conference Record (Cat. No.01CH37310)*, vol. 2, pp. 1060–1063, 2001, ISSN: 1082-3654. DOI: 10.1109/NSSMIC.2001.1009735. [Online]. Available: http://ieeexplore.ieee.org/lpdocs/epic03/wrapper.htm?arnumber=1009735.

[32] Y. Pan, R. Whitaker, A. Cheryauka, and D. Ferguson, "TV-regularized iterative image reconstruction on a mobile C-ARM CT", *Imaging*, no. March 2010, pp. 76222L–76222L–12, 2010, ISSN: 16057422. DOI: 10.1117/12.844398. [Online]. Available: http://link.aip.org/link/PSISDG/v7622/i1/p76222L/s1{\&}Agg=doi.

[33] B. De Man, J. Nuyts, P. Dupont, G. Marchal, and P. Suetens, "Reduction of metal streak artifacts in X-ray computed tomography using a transmission maximum a posteriori algorithm", *IEEE Transactions on Nuclear Science*, vol. 47, no. 3, pp. 977–981, 2000, ISSN: 00189499. DOI: 10.1109/23.856534. [Online]. Available: http://ieeexplore.ieee.org/document/856534/.

[34] J. H. Jonsson, M. G. Karlsson, M. Karlsson, and T. Nyholm, "Treatment planning using MRI data: an analysis of the dose calculation accuracy for different treatment regions.", *Radiation oncology (London, England)*, vol. 5, p. 62, 2010, ISSN: 1748-717X. DOI: 10.1186/1748-717X-5-62. [Online]. Available: http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=2909248{\&}tool=pmcentrez{\&}rendertype=abstract.

[35] U. B. e. a. Daniel J Bell, *Magnetic susceptibility artifact*, https://radiopaedia.org/articles/magnetic-susceptibility-artifact, Accessed 2018. [Online]. Available: https://radiopaedia.org/articles/magnetic-susceptibility-artifact.

[36] A. A. Anderla, D. R. Culibrk, and G. Delso, "Metal Artifact Reduction from CT Images Using Complementary MR Images", 2013.

[37] G Delso, S Wollenweber, a Lonn, F Wiesinger, and P Veit-Haibach, "MR-driven metal artifact reduction in PET/CT.", *Physics in medicine and biology*, vol. 58, no. 7, pp. 2267–80, 2013, ISSN: 1361-6560. DOI: 10.1088/0031-9155/58/7/2267. [Online]. Available: http://www.ncbi.nlm.nih.gov/pubmed/23478566.

[38] P. C. Park, E. Schreibmann, J. Roper, E. Elder, I. Crocker, T. Fox, X. R. Zhu, L. Dong, and A. Dhabaan, "MRI-based computed tomography metal artifact correction method for improving proton range calculation accuracy", *International Journal of Radiation Oncology Biology Physics*, vol. 91, no. 4, pp. 849–856, 2015, ISSN: 1879355X. DOI: 10.1016/j.ijrobp.2014.12.027. [Online]. Available: http://dx.doi.org/10.1016/j.ijrobp.2014.12.027.

[39] D. J. Tyler, M. D. Robson, R. M. Henkelman, I. R. Young, and G. M. Bydder, "Magnetic resonance imaging with ultrashort te (ute) pulse sequences: technical considerations", *Journal of Magnetic Resonance Imaging*, vol. 25, no. 2, pp. 279–289, DOI: 10.1002/jmri.20851. eprint: https://onlinelibrary.wiley.com/doi/pdf/10.1002/jmri.20851. [Online]. Available: https://onlinelibrary.wiley.com/doi/abs/10.1002/jmri.20851.

[40] J. Nielsen, J. Edmund, and K. Van Leemput, "Ct metal artifact reduction using mr image patches", in *Proceedings of SPIE*. SPIE - International Society for Optical Engineering, 2018, vol. 10573. DOI: 10.1117/12.2293815.

[41] C. M. Bishop, *Pattern recognition and machine learning*, eng. Springer, 2006, 738 s. ISBN: 0387310738, 9780387310732.

[42] D. J. C. MacKay, *Information Theory, Inference, and Learning Algorithms David J.C. MacKay*. 2005, vol. 100, pp. 1–640, ISBN: 9780521642989. DOI: 10.1198/jasa.2005.s54. arXiv: arXiv:1011.1669v3. [Online]. Available: http://pubs.amstat.org/doi/abs/10.1198/jasa.2005.s54{\%}5Cnhttp://www.cambridge.org/0521642981.

[43] C. Robert, *The Bayesian choice. A decision-theoretic motivation*, eng. Springer-Verlag, 1994, 436 s. ISBN: 0387942963, 3540942963.

[44] N. Otsu, "A threshold selection method from gray-level histograms", *IEEE Transactions on Systems, Man, and Cybernetics*, vol. 9, no. 1, pp. 62–66, 1979, ISSN: 0018-9472. DOI: 10.1109/TSMC.1979.4310076.

[45] V.-t. Ta, D. L. Collins, and P. Coup, "Optimized PatchMatch for Near Real Time and Accurate Label Fusion", vol. 2, pp. 105–112, 2014.

[46] P. A. Bromiley, "Products and Convolutions of Gaussian Distributions", *Tina Vision Memo*, vol. 2003, no. 003, pp. 1–13, 2014. [Online]. Available: http://tina.wiau.man.ac.uk/docs/memos/2003-003.pdf.

[47] T. P. Minka, "Expectation-Maximization as lower bound maximization", 1998.

[48] W. M. I. Wells, P. Viola, H. Atsumi, S. Nakajima, and R. Kikinis, "Multi-Modal image registration by maximization of mutual information", *Medical Image Analysis*, vol. 1, no. 1, pp. 35–51, 1996, ISSN: 13618415. DOI: 10.1016/S1361-8415(01)80026-8. [Online]. Available: http://linkinghub.elsevier.com/retrieve/pii/S1361841501800268{\%}5Cnhttp://linkinghub.elsevier.com/retrieve/pii/S1361841501800049.

[49] F Maes, A Collignon, D Vandermeulen, G Marchal, and P Suetens, "Multimodality image registration by maximization of mutual information", eng, *IEEE Transactions on Medical Imaging*, vol. 16, no. 2, pp. 187–198, 1997, ISSN: 1558254x, 02780062. DOI: 10.1109/42.563664.

[50] Y. Lin and E. Samei, "A fast poly-energetic iterative FBP algorithm", *Physics in Medicine and Biology*, vol. 59, no. 7, pp. 1655–1678, 2014, ISSN: 13616560. DOI: 10.1088/0031-9155/59/7/1655.

[51] D Andreasen, K Van Leemput, R. H. Hansen, J. A. L. Andersen, and J. M. Edmund, "Patch-based generation of a pseudo CT from conventional MRI sequences for MRI-only radiotherapy of the brain", *Med Phys*, vol. 42, no. 4, pp. 1596–1605, 2015, ISSN: 0094-2405. DOI: 10.1118/1.4914158.

[52] A. Torrado-Carvajal, J. L. Herraiz, E. Alcain, A. S. Montemayor, L. Garcia-Canamaque, J. A. Hernandez-Tamames, Y. Rozenholc, and N. Malpica, "Fast Patch-Based Pseudo-CT Synthesis from T1-Weighted MR Images for PET/MR Attenuation Correction in Brain Studies", *Journal of Nuclear Medicine*, vol. 57, no. 1, pp. 136–144, 2015, ISSN: 0161-5505. DOI: 10.2967/jnumed.115.156299. [Online]. Available: http://jnm.snmjournals.org/cgi/doi/10.2967/jnumed.115.156299.

[53] J. Nuyts, B. D. Man, P. Dupont, M. Defrise, P. Suetens, and L. Mortelmans, "Iterative reconstruction for helical CT: a simulation study", *Physics in Medicine and Biology*, vol. 43, no. 4, p. 729, 1998, ISSN: 1361-6560. DOI: 10.1088/0031-9155/43/4/003. [Online]. Available: http://stacks.iop.org/0031-9155/43/i=4/a=003.

[54] B. De Man, *Iterative Reconstruction for Reduction of Metal Artifacts in Computed Tomography*. 2001, ISBN: 9056823000. [Online]. Available: http://scholar.google.com/scholar?hl=en{\&}btnG=Search{\&}q=intitle:Iterative+Reconstruction+for+Reduction+of+Metal+Artifacts+in+Computed+Tomography{\#}3.

[55] Philips Healthcare, *Metal artifact reduction for orthopedic implants (omar)*, http://clinical.netforum.healthcare.philips.com/us_en/Explore/White-Papers/CT/Metal-Artifact-Reduction-for-Orthopedic-Implants-(O-MAR), 2012.

[56] W. van Aarle, W. J. Palenstijn, J. De Beenhouwer, T. Altantzis, S. Bals, K. J. Batenburg, and J. Sijbers, "The astra toolbox: a platform for advanced algorithm development in electron tomography", eng, *Ultramicroscopy*, vol. 157, pp. 35–47, 2015, ISSN: 18792723, 03043991. DOI: 10.1016/j.ultramic.2015.05.002.

[57] W. van Aarle, W. J. Palenstijn, J. Cant, E. Janssens, F. Bleichrodt, A. Dabravolski, J. De Beenhouwer, K. J. Batenburg, and J. Sijbers, "Fast and flexible x-ray tomography using the astra toolbox", eng, *Optics Express*, vol. 24, no. 22, pp. 25 129–25 147, 2016, ISSN: 10944087.

A Appendix

A.1 Image materials

All images were acquired using a Philips Brilliance Big Bore CT scanner at a kVp of 120kV and a resolution of (0.5 x 0.5 x 2.0)mm. The images were reconstructed by the scanner software using a version of FBP. The MRIs were acquired on a Philips Panorama 1.0T HFO scanner using a 2D T1w sequence at a resolution of (0.5 x 0.5 x 5.5mm), with TE = 10ms and TR = 520.2 - 572.2ms.

The sinograms were exported from the CT scanner in a semi-raw format for use in sinogram inpainting and MBIR: The vendor-provided beam hardening and scatter corrections had been performed, while bow-tie filter, detector inhomogeneity and denoising had not; we thus performed these ourselves using empty bore scans, upon which we interpolated and rebinned the spiral sinograms to 2D sinogram with a pencil beam geometry. We then implemented the nMAR and MLTR algorithm in Python using the GPU accelerated forward- and back projection primitives in the ASTRA[56, 57] toolbox.

A.2 Example delineations for the quantitative evaluation

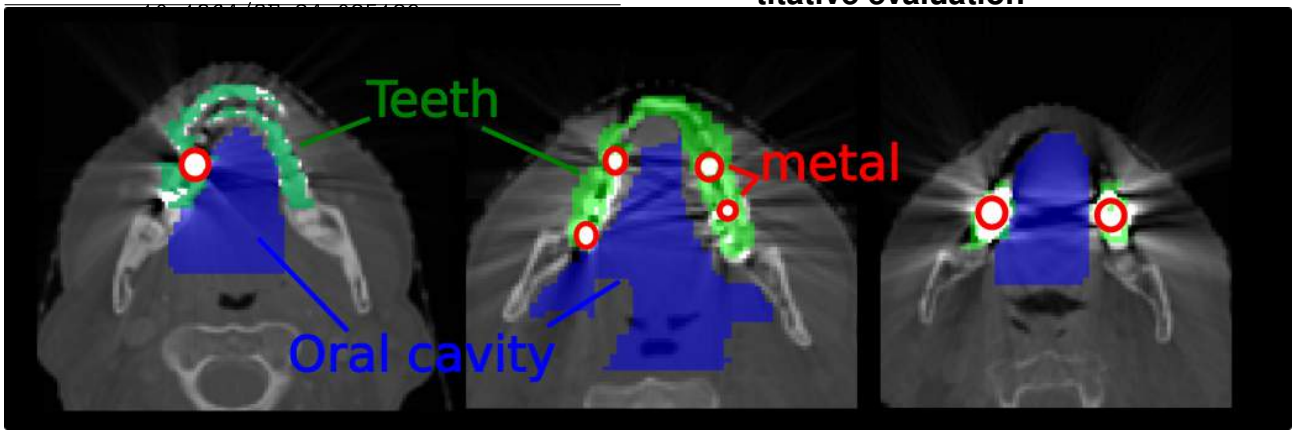


Figure 10: *Example delineations for three patients. The oral cavity delineations were performed by the clinical personnel for RT planning, while the teeth delineations were contoured manually by the authors of this study. We sought to avoid the metal implants during contouring and also used the mask of the metal implants provided by Otsu's method to better exclude them from the delineations.*

A.3 Additional images

Figures 11 - 13 show images from 5 additional head-and-neck patients out of the 9 used in our study.

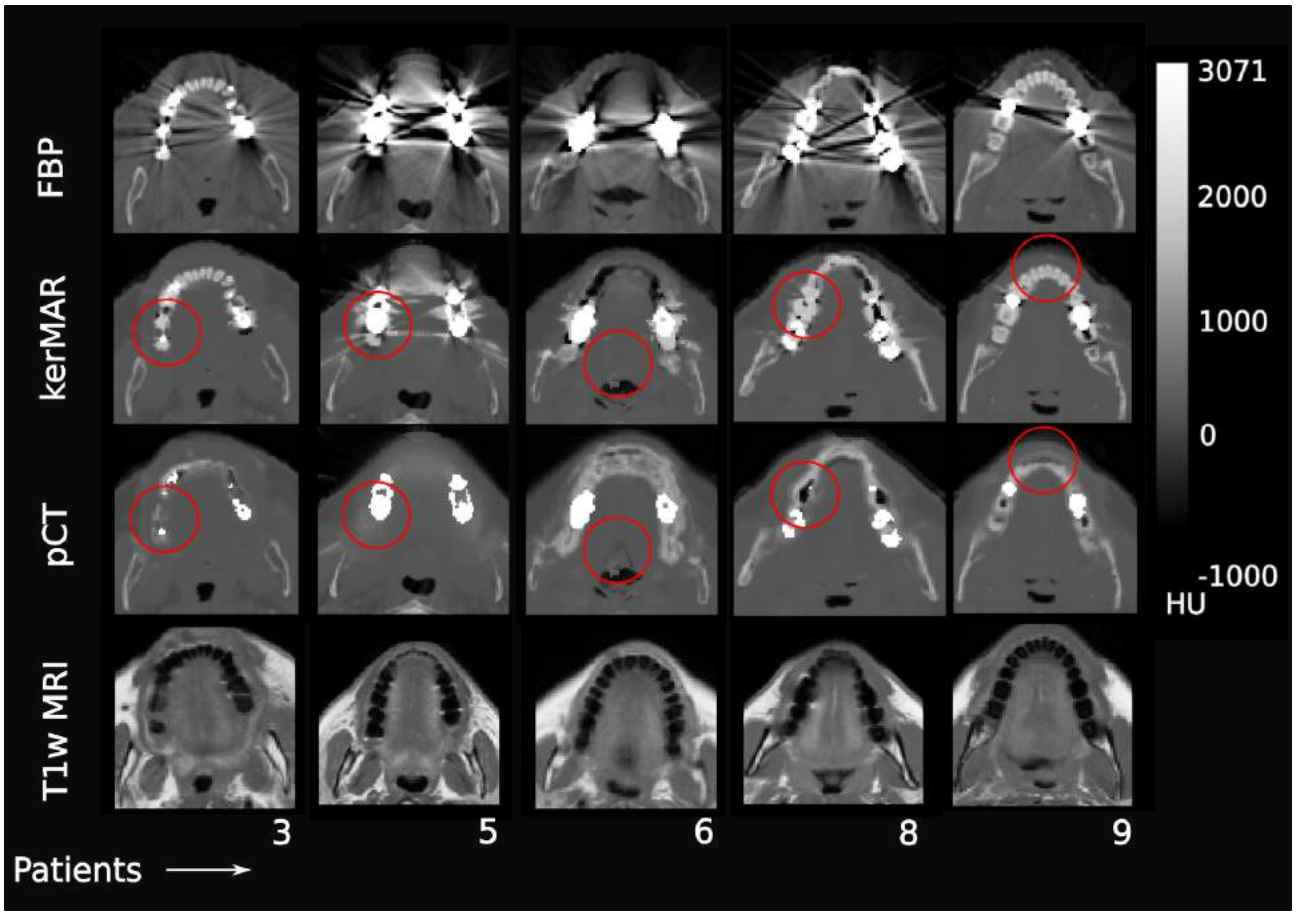


Figure 11: *Additional axial slices for 5 patients, showing results for FBP, kerMAR and oMAR.*

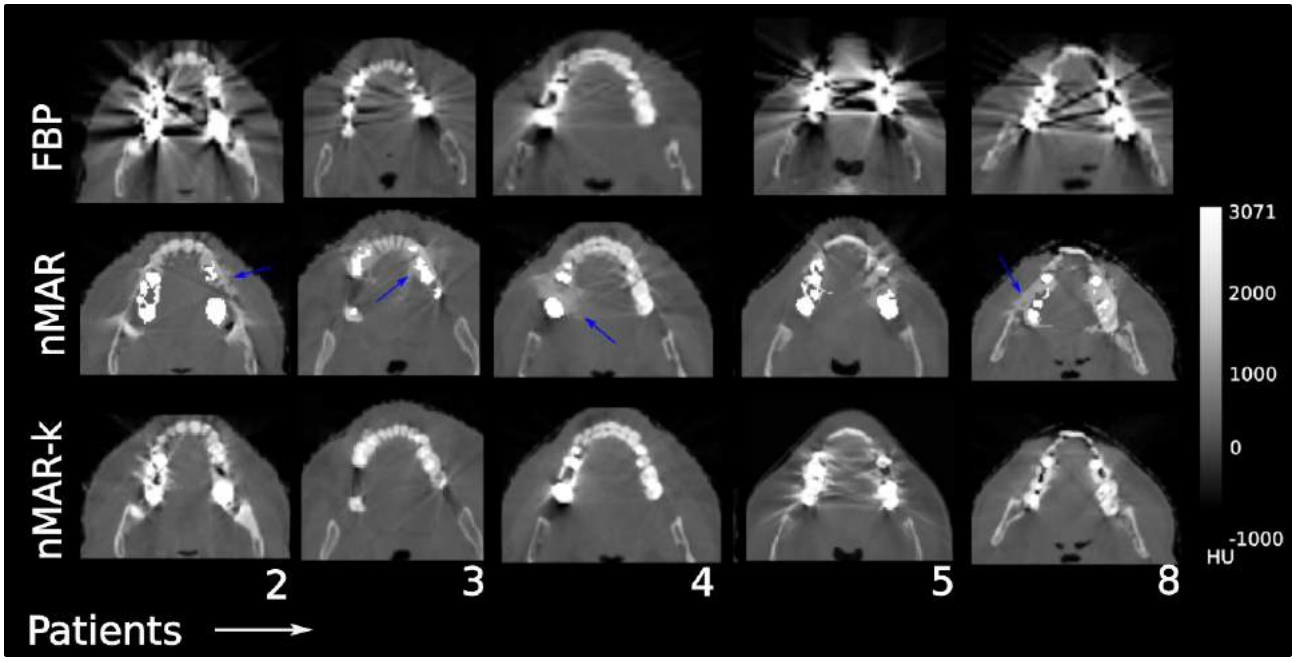


Figure 12: *Additional axial CT slices for 5 patients. Shown are results for nMAR and nMAR-k.*

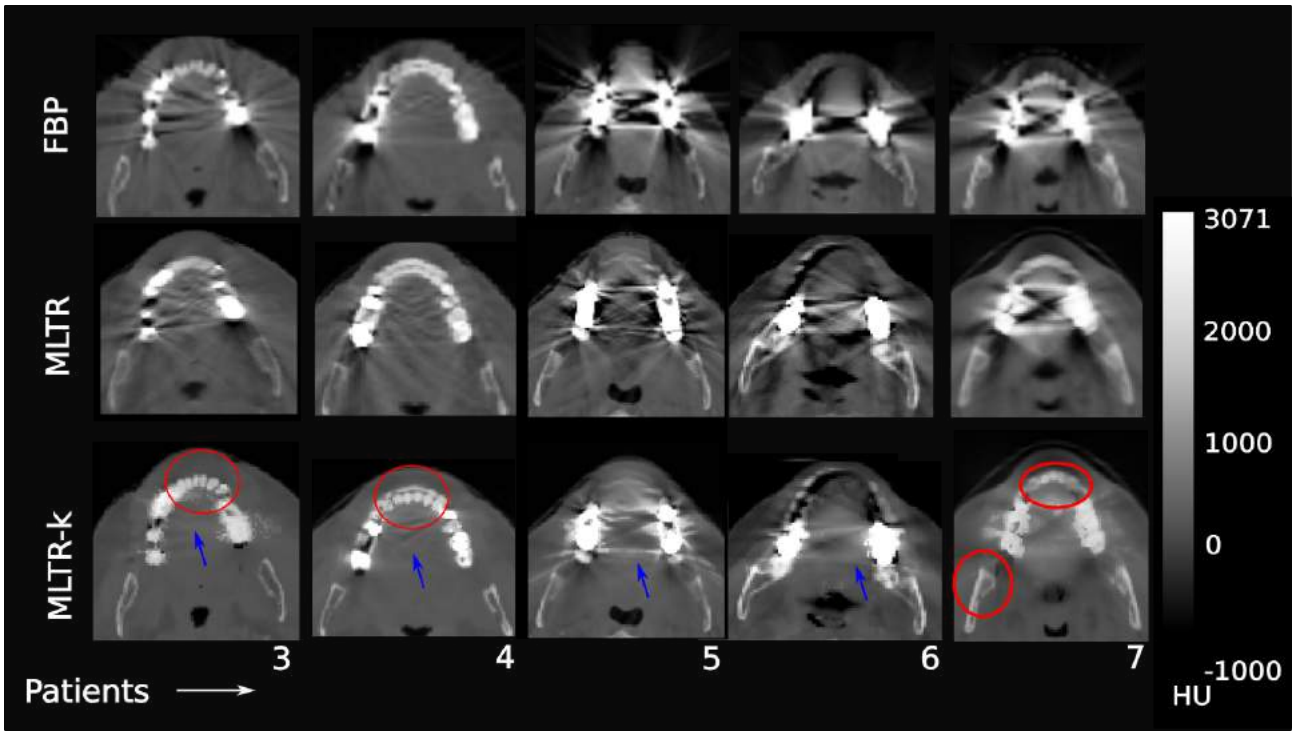


Figure 13: *Additional axial CT slices for 5 patients, showing results for MLTR and MLTR-k.*

A.4 Hyperparameter estimation

We here derive the Expectation Maximization (EM) algorithm for Empirical Bayes hyperparameter estimation. EM is an iterative algorithm that improves a parameter estimate through a sequence of steps, starting from an initial estimate $\boldsymbol{\beta}^{(0)}$. It calculates a set of estimates, $\{\boldsymbol{\beta}_k\}_{k<N}$, for which (ideally) $\Phi(\boldsymbol{\beta}_{k+1}) > \Phi(\boldsymbol{\beta}_k)$, $\forall k < N$, where $\Phi(\boldsymbol{\beta})$ is the objective function. EM in particular does this by maximizing a lower bound $l(\boldsymbol{\beta}|\boldsymbol{\beta}_k)$ that depends on the current parameter estimate, and for which $l(\boldsymbol{\beta}_k|\boldsymbol{\beta}_k) = \Phi(\boldsymbol{\beta}_k)$; due to this stipulation, its maximization is guaranteed to increase the value of the objective function in successive estimates[47], and thus $\Phi(\boldsymbol{\beta}_{k+1}) > \Phi(\boldsymbol{\beta}_k)$.

The benefit of EM arises when the lower-bound maximization is fast and simple, or maybe even has a closed-form expression. We derive a lower bound that, as we will see, may be used to approximately achieve this by applying Jensen's inequality[47] to the objective function. Introducing ${v_n^i}^{(k)}$ to the objective by simultaneously multiplying and dividing it, we get the following inequality after a few manipulations:

$$\begin{aligned}\Phi(\boldsymbol{\beta}) &= \sum_{i\in\mathcal{T}} \log \sum_{n\in\mathcal{A}_i} {v_n^i}^{(k)} \frac{\phi_n^i(\boldsymbol{\beta})}{{v_n^i}^{(k)}} \\ &\geq \sum_{i\in\mathcal{T}} \sum_{n\in\mathcal{A}_i} {v_n^i}^{(k)} \log \phi_n^i(\boldsymbol{\beta}).\end{aligned}$$

where ${v_n^i}^{(k)}$ are the weights in eqn. (11) evaluated at the k^{th} hyperparameter estimate, $\boldsymbol{\beta}_k$. We here used that $\sum_{n\in\mathcal{A}_i} v_n^i = 1$. Equality is achieved at the current estimate $\boldsymbol{\beta} = \boldsymbol{\beta}_k$, and so maximizing the right-hand-side is guaranteed to improve the objective. We therefore define the following lower bound:

$$l(\boldsymbol{\beta}|\boldsymbol{\beta}_k) = \sum_{i\in\mathcal{T}} \sum_{n\in\mathcal{A}_i} {v_n^i}^{(k)} \log \phi_n^i(\boldsymbol{\beta})$$

$$\text{with} \quad \phi_n^i(\boldsymbol{\beta}) = \mathcal{N}(t_i|y_n, f(\mathbf{x}_i)\beta_t^{*-1} + \beta_y^{-1})\mathcal{N}(\mathbf{m}_i|\mathbf{m}_n, \beta_m^{-1}\mathbf{I}_M),$$

The expectation (E-step) of our algorithm calculates the weights v_n^i while the M-step maximizes this function. For notational convenience, we define the variances (reciprocal precisions) $(\sigma_t^{*2}, \sigma_y^2, \sigma_m^2) = (\beta_t^{*-1}, \beta_y^{-1}, \beta_m^{-1})$, as well as the combination $\sigma^2 \equiv f(\mathbf{x}_i)\sigma_t^{*2} + \sigma_y^2$.

Using the chain rule, the derivative of the lower bound with respect to e.g. σ_y^2 becomes:

$$\frac{\partial l(\boldsymbol{\beta}|\boldsymbol{\beta}_k)}{\partial \sigma_y^2} = \sum_{i\in\mathcal{T}} \sum_{n\in\mathcal{A}_i} {v_n^i}^{(k)} \frac{1}{\phi_n^i(\boldsymbol{\beta})} \frac{\partial \phi_n^i(\boldsymbol{\beta})}{\partial \sigma^2} \frac{\partial \sigma^2}{\partial \sigma_y^2}.$$

We therefore calculate the derivatives of $\phi_n^i(\boldsymbol{\beta})$ by eval-

uating:

$$\frac{\partial \mathcal{N}(t_i|y_n,\sigma^2)}{\partial \sigma^2} = \frac{\mathcal{N}(t_i|y_n,\sigma^2)}{2\sigma^2}[\frac{(t_i-y_n)^2}{\sigma^2}-1],$$
$$\frac{\partial \mathcal{N}(\mathbf{m}_i|\mathbf{m}_n,\sigma_m^2)}{\partial \sigma_m^2}$$
$$= \frac{\mathcal{N}(\mathbf{m}_i|\mathbf{m}_n,\sigma_m^2)}{2\sigma_m^2}[\frac{(\mathbf{m}_i-\mathbf{m}_n)^T(\mathbf{m}_i-\mathbf{m}_n)}{M\sigma_m^2}-1],$$
$$\frac{\partial \sigma^2}{\partial {\sigma_t^*}^2} = f(\mathbf{x}_i) \quad \text{and} \quad \frac{\partial \sigma^2}{\partial \sigma_y^2} = 1.$$

Here, M is the number of voxels in a patch, i.e. the dimension of $\mathbf{m}_i$. Using these expressions, we get the gradient:

$$\frac{\partial l(\boldsymbol{\beta}|\boldsymbol{\beta}_k)}{\partial {\sigma_t^*}^2} = \sum_{i\in\mathcal{T}}\sum_{n\in\mathcal{A}_i} {v_n^i}^{(k)} \frac{1}{2\sigma^2} f(\mathbf{x}_i)[\frac{(t_i-y_n)^2}{\sigma^2}-1],$$
$$\frac{\partial l(\boldsymbol{\beta}|\boldsymbol{\beta}_k)}{\partial {\sigma_y}^2} = \sum_{i\in\mathcal{T}}\sum_{n\in\mathcal{A}_i} {v_n^i}^{(k)} \frac{1}{2\sigma^2}[\frac{(t_i-y_n)^2}{\sigma^2}-1],$$
$$\frac{\partial l(\boldsymbol{\beta}|\boldsymbol{\beta}_k)}{\partial {\sigma_m}^2}$$
$$= \frac{1}{2\sigma_m^2}\sum_{i\in\mathcal{T}}\sum_{n\in\mathcal{A}_i} {v_n^i}^{(k)}[\frac{(\mathbf{m}_i-\mathbf{m}_n^T)(\mathbf{m}_i-\mathbf{m}_n)}{M\sigma_m^2}-1].$$

To maximize the lower bound, we seek the root of this gradient. However, this cannot be done in closed form for σ_y^2 and σ_t^2 due to the dependence of σ^2 on $f(\mathbf{x}_i)$. As mentioned in section 2.3, we get around this issue by using the following approximation: Defining the corrupted set as $\mathcal{T}_c \equiv \{i \in \mathcal{T} : f(\mathbf{x}_i) > 0.5\}$ and the uncorrupted set as $\mathcal{T}_u \equiv \{i \in \mathcal{T} : f(\mathbf{x}_i) \leq 0.5\}$, we truncate $f(\mathbf{x}_i)$ such that $\forall i \in \mathcal{T}_c$, $f(\mathbf{x}_i) = 1$, and $\forall i \in \mathcal{T}_u$, $f(\mathbf{x}_i) = 0$.

With this definition, the first equation for ${\sigma_t^*}^2$ becomes a sum over $\mathcal{T}_c$ with $\sigma^2 \leftarrow {\sigma_t^*}^2 + \sigma_y^2$ while the second equation for σ_y^2 becomes a two-part sum, with one part equal to the mentioned sum over $\mathcal{T}_c$, the other a sum over $\mathcal{T}_u$ with $\sigma^2 \leftarrow \sigma_y^2$:

$$\frac{\partial l(\boldsymbol{\beta}|\boldsymbol{\beta}_k)}{\partial {\sigma_t^*}^2} = \frac{1}{2[{\sigma_t^*}^2+\sigma_y^2]}\sum_{i\in\mathcal{T}_c}\sum_{n\in\mathcal{A}_i} {v_n^i}^{(k)}[\frac{(t_i-y_n)^2}{[{\sigma_t^*}^2+\sigma_y^2]}-1],$$
$$\frac{\partial l(\boldsymbol{\beta}|\boldsymbol{\beta}_k)}{\partial {\sigma_y}^2} = \frac{1}{2[{\sigma_t^*}^2+\sigma_y^2]}\sum_{i\in\mathcal{T}_c}\sum_{n\in\mathcal{A}_i} {v_n^i}^{(k)}[\frac{(t_i-y_n)^2}{[{\sigma_t^*}^2+\sigma_y^2]}-1]+$$
$$\frac{1}{2[\sigma_y^2]}\sum_{i\in\mathcal{T}_u}\sum_{n\in\mathcal{A}_i} {v_n^i}^{(k)}[\frac{(t_i-y_n)^2}{\sigma_y^2}-1].$$

To simultaneously make all three equations equal to 0, we do as follows: Pick σ_y^2 such that the second sum in the second equation is 0; then, pick ${\sigma_t^*}^2$ such that the remaining sum that is shared between the two equations is 0; and finally minimize the third equation for σ_m^2. This may easily be seen to be achieved by the scheme in alg. (4):

Algorithm 4 M-step of the hyperparameter estimation

1: $\beta_y^{-1} \leftarrow \frac{1}{|\mathcal{T}_u|}\sum_{i\in\mathcal{T}_u}\sum_{n\in\mathcal{A}_i} {v_n^i}^{(k)}(t_i-y_n)^2$.
2: $\beta_t^{-1} \leftarrow \frac{1}{|\mathcal{T}_c|}\sum_{i\in\mathcal{T}_c}\sum_{n\in\mathcal{A}_i} {v_n^i}^{(k)}(t_i-y_n)^2 - \beta_y^{-1}$.
3: $\beta_m^{-1} \leftarrow \frac{1}{|\mathcal{T}|}\sum_{i\in\mathcal{T}}\sum_{n\in\mathcal{A}_i} {v_n^i}^{(k)}\frac{(\mathbf{m}_i-\mathbf{m}_n)^T(\mathbf{m}_i-\mathbf{m}_n)}{M}$.

A.5 Normalized MAR and CT-based prior

We here give an overview of our normalized MAR (nMAR) implementation in algorithm 5. The reader may further consult the original publication of the algorithm by Meyer et al. in [17], as well as refs. [18, 19]. We calculate the prior in point 2 by the procedure in algorithm 6, following some of the recommendations in refs. [17–19].

Algorithm 5 Normalized MAR (nMAR)[17, 18]

1: Stack the sinogram in projection vector $\mathbf{p}$ and define the scanner system matrix as $\mathbf{L}$.
2: Generate a prior image, stacked in vector $\bar{\mathbf{y}}$, from the FBP as in algorithm 6.
3: Simulate a CT acquisition by *forward projecting* through the prior: $\bar{\mathbf{p}} = \mathbf{L}\bar{\mathbf{y}}$.
4: Calculate the ratios $\mathbf{r} = \mathbf{p} \odot \bar{\mathbf{p}}^{-1}$.
5: Threshold the FBP using Otsu's method to find a metal mask. Forward project through it and label the metal projections by thresholding (> 0).
6: Interpolate $\mathbf{r}$ over the metal-labelled projections $(\mathbf{r}_{\text{int}} \leftarrow \mathbf{r})$.
7: Calculate the inpainted sinogram $\mathbf{p}_{\text{inp}} = \mathbf{r}_{\text{int}}\bar{\mathbf{p}}$.
8: Reconstruct the nMAR image from $\mathbf{p}_{\text{inp}}$ using FBP.

Algorithm 6 Prior generation for nMAR

1: Reconstruct the images from the sinogram using FBP.
2: On the FBP, segment the metal implants using Otsu's[44] thresholding method.
3: Forward project through a binary mask of the metal implants, yielding a metal-only sinogram. Apply a threshold (> 0) to yield a binary mask of the metal projections.
4: Calculate a liMAR (linear interpolation MAR)[20] image as follows:
   1. Linearly interpolate the sinogram over the metal projections, as determined by the binary mask.
   2. Reconstruct the resulting interpolated sinogram using FBP.
5: On the liMAR images, perform K-means clustering[42] with 4 clusters, 20 iterations, yielding a cluster label for each voxel.
6: To the voxels in each class, assign the cluster mean, yielding the final prior image.

A.6 Maximum Likelihood Transmission Reconstruction and MRI prior term

The maximum likelihood transmission reconstruction (MLTR) algorithm uses a Poisson distribution as data likelihood that relates the x-ray intensity measurements $\{n_j\}_{j\in\mathcal{S}}$, where $\mathcal{S}$ contains indices to the x-ray detectors, to the image coefficients $\{y_i\}_{i\in\mathcal{T}}$ through a mean count calculated as $\lambda_j = C\mathsf{e}^{-\sum_{i\in\mathcal{T}} l_{j,i}y_i}$; $l_{j,i}$ is and entry in the *system matrix* $\mathbf{L}$ and C is the (assumed constant) x-ray source intensity. Maximizing the log of this likelihood is performed in an EM-like algorithm: The E-step calculates a quadratic lower bound to the log-likelihood that is equal to it at the current estimate; and the M-step maximizes this lower bound in closed form using an expression that depends on the first and second derivatives of of the log-likelihood. This leads in effect to an additive update equation. When a general prior is used, this update equation is altered to include it, as may be seen in refs. [21, 33]. Stacking the image coefficients and x-ray intensities/intensity estimates in column vectors $\mathbf{y}$ and $\mathbf{n}/\boldsymbol{\lambda}$, and starting from an initial image, this leads to algorithm 7. $\mathsf{e}^{\{\cdot\}}$ is here applied element-wise, $\mathbf{1}$ is an image of ones, $l_{*,i}$ denotes a column of the system matrix, and we use the voxel-averaged magnitude of the iterative step as a stop-criterion.

The likelihood step matches the *back projection* of the simulated intensities $\boldsymbol{\lambda}$ to that of the measured $\mathbf{n}$; this is important for MAR since the contribution of a measurement to this back projection is weighted linearly according to its magnitude, such that e.g. those through the highly attenuating metal get a low weight. In turn, the prior becomes relatively more important for voxels that depend on such measurements.

Algorithm 7 Maximum Likelihood Transmission Reconstruction (MLTR) with an image prior

1: Transform the sinogram to intensities: $\mathbf{n} = C\mathsf{e}^{-\mathbf{p}}$, estimating C by NEC scaling[13, 53].
2: Initialize $\mathbf{y}$ as a small-valued, homogeneous image, and $\delta \leftarrow 1$.
3: Calculate $\boldsymbol{\alpha} = \mathbf{L}\mathbf{1}$.
4: **while** $\delta > 10^{-6}$ **do**

$$\mathbf{y}_0 \leftarrow \mathbf{y} \tag{14}$$

$$y_i \leftarrow y_i + \frac{l^T_{*,i}[\boldsymbol{\lambda} - \boldsymbol{n}] + \frac{\partial \ln p(y_i')}{\partial y_i'}|_{y_i}}{l^T_{*,i}[\boldsymbol{\alpha}\odot\boldsymbol{\lambda}] + \frac{\partial^2 \ln p(y_i')}{\partial {y_i'}^2}|_{y_i}}, \tag{15}$$

$$\forall i \in \mathcal{T} \quad \text{with} \quad \lambda_j = C\mathsf{e}^{-\sum_{i\in\mathcal{T}} l_{j,i}y_i}.$$

5:

$$\delta \leftarrow \frac{\sqrt{(\mathbf{y}-\mathbf{y}_0)^T(\mathbf{y}-\mathbf{y}_0)}}{|\mathcal{T}|}$$

6: **end while**

The update equation requires the first and and second derivatives of the prior. Using eqn. (10), the first derivative of $\ln p(y_i|t_i, \mathbf{m}_i, \boldsymbol{\beta})$ may be calculated:

$$\ln(p(y_i|\mathbf{m}_i, t_i, \boldsymbol{\beta})' = \frac{1}{p(y_i|\mathbf{m}_i, t_i, \boldsymbol{\beta})}\frac{\partial p(y_i|\mathbf{m}_i, t_i, \boldsymbol{\beta})}{\partial y_i},$$

with

$$\frac{\partial p(y_i|\mathbf{m}_i, t_i, \boldsymbol{\beta})}{\partial y_i} = \sum_{n\in\mathcal{A}_i} v_n^i \frac{\partial \mathcal{N}(y_i|\mu_n^i, (\beta_t+\beta_y)^{-1})}{\partial y_i}$$

and

$$\frac{\partial \mathcal{N}(y_i|\mu_n^i, (\beta_t+\beta_y)^{-1})}{\partial y_i}$$
$$= (\beta_t+\beta_y)(\mu_n^i - y_i)\mathcal{N}(y_i|\mu_n^i, (\beta_t+\beta_y)^{-1}).$$

Inserting:

$$\ln(p(y_i|\mathbf{m}_i, t_i, \boldsymbol{\beta}))' = (\beta_t+\beta_y)[\sum_{n\in\mathcal{A}_i}\tilde{v}_n^i\mu_n^i - y_i],$$

$$\text{where} \quad \tilde{v}_n^i = \frac{v_n^i\mathcal{N}(y_i|\mu_n^i, (\beta_t+\beta_y)^{-1})}{\sum_{n'\in\mathcal{A}_i}\mathcal{N}(y_i|\mu_{n'}^i, (\beta_t+\beta_y)^{-1})v_{n'}^i}.$$

We here used the fact that $\sum_{n\in\mathcal{A}_i}\tilde{v}_n^i = 1$. For the second derivative:

$$\ln(p(y_i|\mathbf{m}_i, t_i, \boldsymbol{\beta}))'' = \frac{\partial}{\partial y_i}\sum_{n\in\mathcal{A}_i}\tilde{v}_n^i\mu_n^i - (\beta_t+\beta_y)$$

$$\text{with} \quad \frac{\partial}{\partial y_i}\sum_{n\in\mathcal{A}_i}\tilde{v}_n^i\mu_n^i = \sum_{n\in\mathcal{A}_i}\frac{\partial\tilde{v}_n^i}{\partial y_i}\mu_n^i$$

$$\text{and} \quad \frac{\partial\tilde{v}_n^i}{\partial y_i} = (\beta_t+\beta_y)\tilde{v}_n^i(\mu_n^i - y_i) -$$
$$(\beta_t+\beta_y)\tilde{v}_n^i[\sum_{n\in\mathcal{A}_i}\tilde{v}_n^i\mu_n^i - y_i].$$

Inserting:

$$\ln(p(y_i|t_i, \mathbf{m}_i, \boldsymbol{\beta}))''$$
$$= (\beta_t + \beta_y)^2 \Bigg(\sum_{n \in \mathcal{A}_i} \tilde{v}_n^i {\mu_n^i}^2 - [\sum_{n \in \mathcal{A}_i} \tilde{v}_n^i \mu_n^i]^2 \Bigg) - (\beta_t + \beta_y).$$

These expressions for the first and second derivatives are calculated in each iteration during step 4 of algorithm 7 and substituted in eqn. (15); this defines the MLTR-k algorithm.

Appendix C

Paper C: MR based CT metal artifact reduction for head-and-neck photon, electron and proton radiotherapy

MR based CT metal artifact reduction for head-and-neck photon, electron and proton radiotherapy

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

(1) Department of Applied Mathematics and Computer Science,

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) Department of Radiology, Massachusetts General Hospital,

Harvard Medical School, Boston MA 02114, USA. and

(4) Niels Bohr Institute, University of Copenhagen, 2100 Copenhagen, Denmark.

(Dated: May 5, 2019)

Abstract

Purpose: We investigated the impact on Computed Tomography (CT) image quality and photon, electron and proton radiotherapy (RT) dose calculations of three CT metal artifact reduction (MAR) approaches: The oMAR algorithm (Philips Healthcare), manual water override and our recently presented, MR-based kerMAR algorithm. We considered three hypotheses: I: The automatic algorithms oMAR and kerMAR provide significant benefits over using the uncorrected CT; II: kerMAR improves upon oMAR; III: Manual water override improves upon the alternatives. **Methods:** We considered a veal shank phantom with/without 6 metal inserts, and a retrospective set of 9 head and neck RT patients with dental implants. We quantified the metal artifact reduction capabilities by the increase in the tails of the CT value distribution in regions of interest and the change in the estimates of the particle range and photon depth at maximum dose. **Results:** We found significant improvements in image quality with oMAR and kerMAR for the patients, accompanied by significant impacts on the particle ranges. We saw superior improvements in high intensity streak suppression with kerMAR as compared to oMAR, in particular in the oral cavity, and associated impacts on the particle range estimates. In contrast, while water override provided apparent image improvements in the phantom soft tissue and the oral cavity, the dose calculation impacts were either insignificant or negative. **Conclusion:** We found support for hypotheses I and II in terms of image quality as well as RT particle range estimates, while hypothesis III was tentatively rejected.

Keywords: Computed tomography, CT metal artifact reduction, radiotherapy, proton therapy, Bayesian modelling

I. INTRODUCTION AND PURPOSE

Metal implants in patients subject to x-ray computed tomography (CT) lead to potentially severe cupping and streak artifacts due to breakdown of the model relating the reconstructed image coefficients to the x-ray measurements. Apart from potentially leading to diagnostic errors, such image corruption may be of especial consequence when quantitative CT information is needed for treatment purposes. External beam radiotherapy (RT) in particular relies on CT attenuation information to aid dosimetric treatment planning. The CT may provide both electron density information for photon therapy and relative stopping power (RSP) estimates for particle therapy that directly influence calculated quantities such as water equivalent thickness (WET) and particle range; it also provides image material for organ and lesion delineation.[1–3]

With photon radiation, the dose plan errors caused by the artifacts may be small, since photon RT plans typically use tens of beam positions- and angles and may accordingly be relatively robust to errors in the dose deposited by small subsets of the beams[1,3,4]. The photon absorption is in addition relatively insensitive to tissue variations. For particles, on the other hand, i.e. electrons and protons, the RSP is highly sensitive to tissue variations and the plans typically contain only 1-3 beams.[1,3–5] The ensuing dose plan uncertainties in electron and the increasingly widespread proton therapy can be a major concern for head-and-neck RT patients with tumors simultaneously close to critical organs at risk (OARs) and the frequently corrupted oral region; it also decreases the degrees of freedom during dose planning by e.g. removing the option of angling beams through the oral cavity to potentially spare more critical OARs.[5,6].

To potentially reduce plan uncertainties, metal artifacts may be addressed manually by an experienced dosimetrist, radiographer or similar replacing corrupted regions with bulk CT values. This is a common clinical practice for the, as mentioned, relatively insensitive photon therapy modality. It however requires time and is subject to unpredictable human error.[7,8] Alternatively, automatic metal artifact reduction (MAR) algorithms may be employed, typically supplied by the vendor of the CT scanner as a commercial add-on. Such vendor solutions must be scrutinized before clinical use to gauge their efficacy and possible limitations.

An example is the clinically used oMAR algorithm[9] (MAR for orthopedic implants,

Philips Healtchare) that provides visual improvements, but has been found to leave behind residual streaks close to the implants[6,10,11]. These lead to associated RSP estimation errors and thus imperfect water equivalent thickness (WET) estimates for proton RT[6], as well as findings of minimal photon dose improvement in the oral cavity (with a closed mouth)[1,11]. Consequently, the residual metal artifacts cannot necessarily be safely disregarded during RT dose planning, causing a potential loss of automation and accuracy.

Using more accurate MARs would help alleviate this issue, examples of which may be found among the numerous, well-documented MAR alternatives in the literature. The multiple MAR approaches span from fast and simple raw data interpolation/replacement schemes and image space methods[12–15] to complex and slow full iterative reconstruction algorithms[16,17], and offer alternatives suitable for a wide range of diverse situations and levels of implementational complexity. While some methods, such as the iterative algorithms, are hard to implement by the user as they require low-level access to proprietary information about the CT system, and must intervene in the CT reconstruction process, some of the simpler algorithms are compatible with the vendor-provided reconstruction software (which tends to use a variation of Filtered Back Projection (FBP)[18]).

We recently presented a novel example of one such alternative, which in addition to the corrupted CT information incorporates complementary image information from magnetic resonance imaging (MRI) using kernel regression (kerMAR)[19]. Taking advantage of the coacquired, less artifact-corrupted MRI that is commonly acquired for head and neck RT to aid in tumor delineation, kerMAR uses kernel regression on CT value / MR image patch pairs along with a forward model of the CT artifacts to estimate the true CT values underlying the artifacts. This leads to potential improvements, in particular close to the metal implants where purely CT-based algorithms are the least effective.

In some situations, this algorithm may thus in principle provide a benefit over available alternatives such as oMAR. To decide whether to include such an alternative algorithm in the radiotherapy practice, however, a few questions must be considered: First, there is the question of whether metal artifact reduction is sufficiently important to spend time and effort worrying about, in particular considering findings such as the minimal photon improvements with oMAR in the oral cavity; and if so, how much improvement does the MAR provide? Second, there is the question of whether a change of algorithm provides sufficient benefits, in our case pitting kerMAR against the default oMAR. And third and finally, the conceptually

simple but time-consuming practice of manual override, that also introduces possible human bias, should be questioned and evaluated in comparison to the automatic alternatives.

These questions lead to the following corresponding hypotheses, which we consider in this study: I: The automatic MAR algorithms kerMAR and oMAR provide significant benefits over using the uncorrected FBP; II: The novel, MR-based kerMAR algorithm is superior to the clinically used oMAR algorithm; and III: Manual intervention using a simple water correction technique provides significant benefits over all other alternatives. We investigate these three hypotheses using phantom as well as retrospective head-and-neck patient data, evaluating the level of artifact corruption via image metrics that quantify the amount of low and high intensity artifacts. We also investigate the impact of the MARs on photon, electron and proton maximum depth / particle range estimates in the dose calculations used for radiotherapy.

II. MATERIALS AND METHODS

A. The MAR algorithms

The three algorithms we consider in this study are explained schematically on fig. 1.

a. *kerMAR*[19] is an image space, Bayesian inference algorithm that uses kernel regression[20] on matching cuboidal MRI *patches* (vectors of voxel intensities from local spatial contexts) and uncorrupted CT values in the patient volume; it thus estimates the prior distribution of the true CT value y given the corresponding MRI patch $\mathbf{m}$, $p(y|\mathbf{m})$. Assuming additive Gaussian artifact noise and given an observed $\mathbf{m}$ centered on a corrupted location as well as the corresponding corrupted CT value t, the posterior distribution $p(y|t, \mathbf{m})$ is then constructed. Calculating the expectation of y over this distribution finally yields the kerMAR CT value estimate.[19] Not relying fully on CT information, there arises a potential for artifact reduction in even very corrupted regions.

b. *oMAR*[9] is an iterative algorithm that combines image processing and projection replacement. It functions by iteratively improving a tissue classified image with consequently suppressed artifacts. This classification smooths away artifacts, such that projecting through the classified image (simulating a CT measurement)[18] approximates a CT data acquisition without artifacts, which is expected to improve over the iterations[9].

c. Manual override[7,8] techniques address the metal artifacts by replacing visibly artifact corrupted regions with a bulk CT value. Since the oral cavity is largely filled with water, a plausible CT value is 0 Hounsfield Units (HU)[18], leading to water override. How this override is performed depends on the practitioner; in our approach, we replaced obviously corrupted soft tissue regions as well as severely corrupted high intensity regions.

B. Study overview and materials

The study may be split into two parts, which in similar ways considered two separate data sets and evaluated the MAR algorithms as well as the baseline strategy of using the uncorrected FBP: 1) A phantom study where we evaluated the MAR algorithms on a veal shank with and without a set of inserted metal markers; and 2) a retrospective patient study where we evaluated the algorithms on a set of patients. We undertook both parts at the same clinic with the same equipment and used the same software to create the MAR images.

Our CT images were all acquired with a Philips Brilliance Big Bore third generation CT scanner, with a kVp of 120kV at a resolution of (0.5 x 0.5 x 2.0)mm. While the exact algorithm used for the CT reconstruction is unknown to us, it was based on filtered back projection (FBP); we will thus refer to it as FBP. The MRs were acquired on a Philips Panorama 1.0T HFO scanner using a 2D T1w sequence at a resolution of (0.5 x 0.5 x 5.5mm), with $TE = 10ms$ and $TR = 520.2 - 572.2ms$. For the MR-based MAR algorithm kerMAR, the T1w MRs were rigidly coregistered to the CTs using mutual information coregistration[21,22] and resampled to the CT resolution.

We scanned the veal shank with and without 6 metal pins inserted in separate regions, varying in phantom depth and metal vicinity, acquiring the FBP, the oMAR and the T1w MR. We performed manual water override on the FBP and used the coregistered T1w MR and FBP to calculate the kerMAR images. Axial phantom slices are shown on fig. 2, with the metal pins computationally inserted on the artifact free reference images by delineating the implants on the FBP using threshold-guided manual segmentation. For the retrospective patient study, we accessed CTs, oMARs and T1w MRs from 9 head and neck patients selected for photon radiotherapy, again acquiring FBPs, oMARs and T1w MRs, performing water override on the FBPs and calculating kerMAR using the FBPs and T1w MRs. Axial slices are shown on fig. 3.

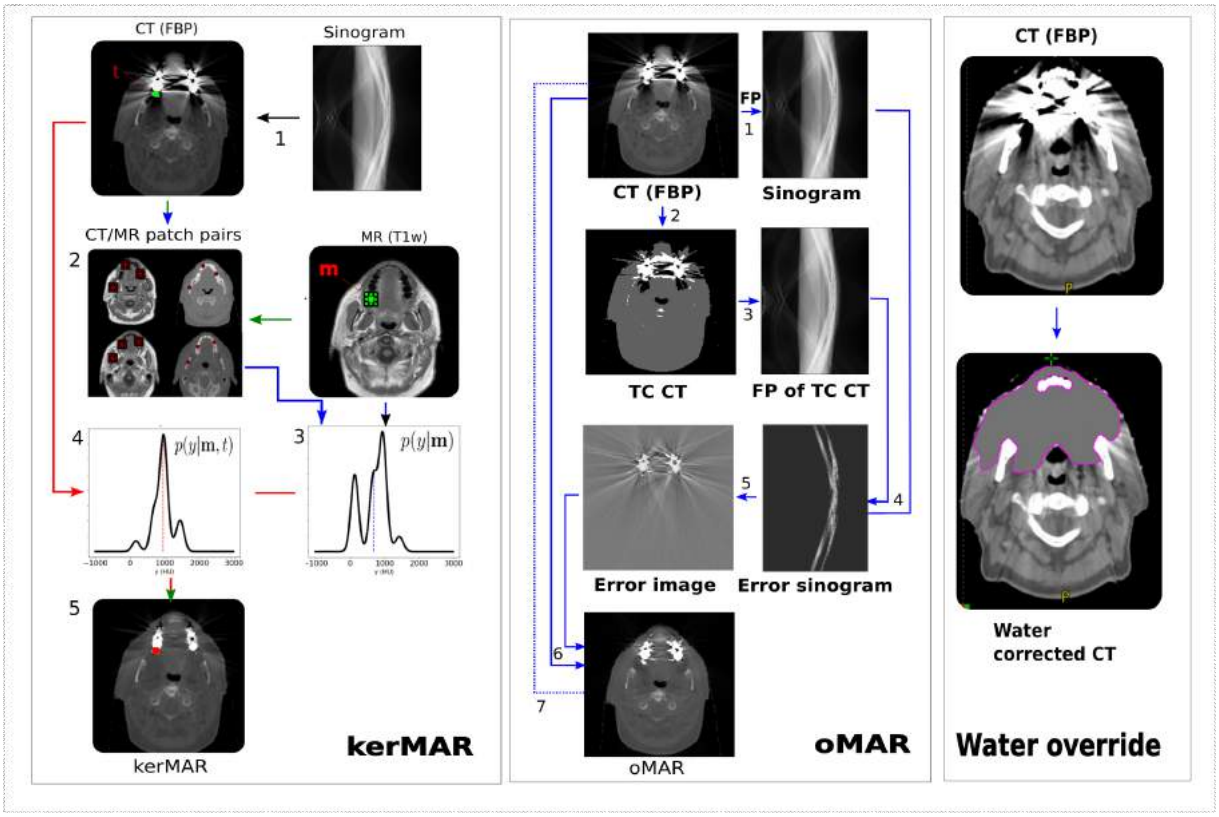


FIG. 1. Schematic illustrations of (left-right) our kerMAR algorithm, the Philips oMAR algorithm and manual override. kerMAR proceeds as follows: 1) Reconstruct the FBP from the raw CT data (sinogram); 2) For each corrupted voxel (example marked by t), find a set of CT value/MR patch regression point pairs in the uncorrupted volume, and estimate the joint distribution $p(y, \mathbf{m})$; 3) Given this distribution and an observed MR patch $\mathbf{m}$ at the corrupted voxel location, construct the prior model of the clean CT values y, $p(y|\mathbf{m})$ (kernel regression); 4) Given the observed corrupted CT measurement t and a Gaussian noise model $p(t|y)$, construct the posterior model $p(y|\mathbf{m}, t)$ using Bayes formula; 5) evaluate the kerMAR estimate as the expectation over the posterior model (red line). oMAR proceeds as follows: 1) Forward project (FP) the input CT image (reconstructed by FBP) to simulate the sinogram; 2) segment the input CT to get a tissue classified prior image (TC CT); 3) forward project through the prior; 4) subtract the prior projection from the original sinogram; 5) reconstruct the error sinogram for an error image; 6) subtract the error image from the FBP; 7) replace the FBP with the updated image for iteration, or stop here. Water override proceeds as follows: Replace any visibly corrupted soft tissue regions, as well as any intensely corrupted regions, with 0HU.

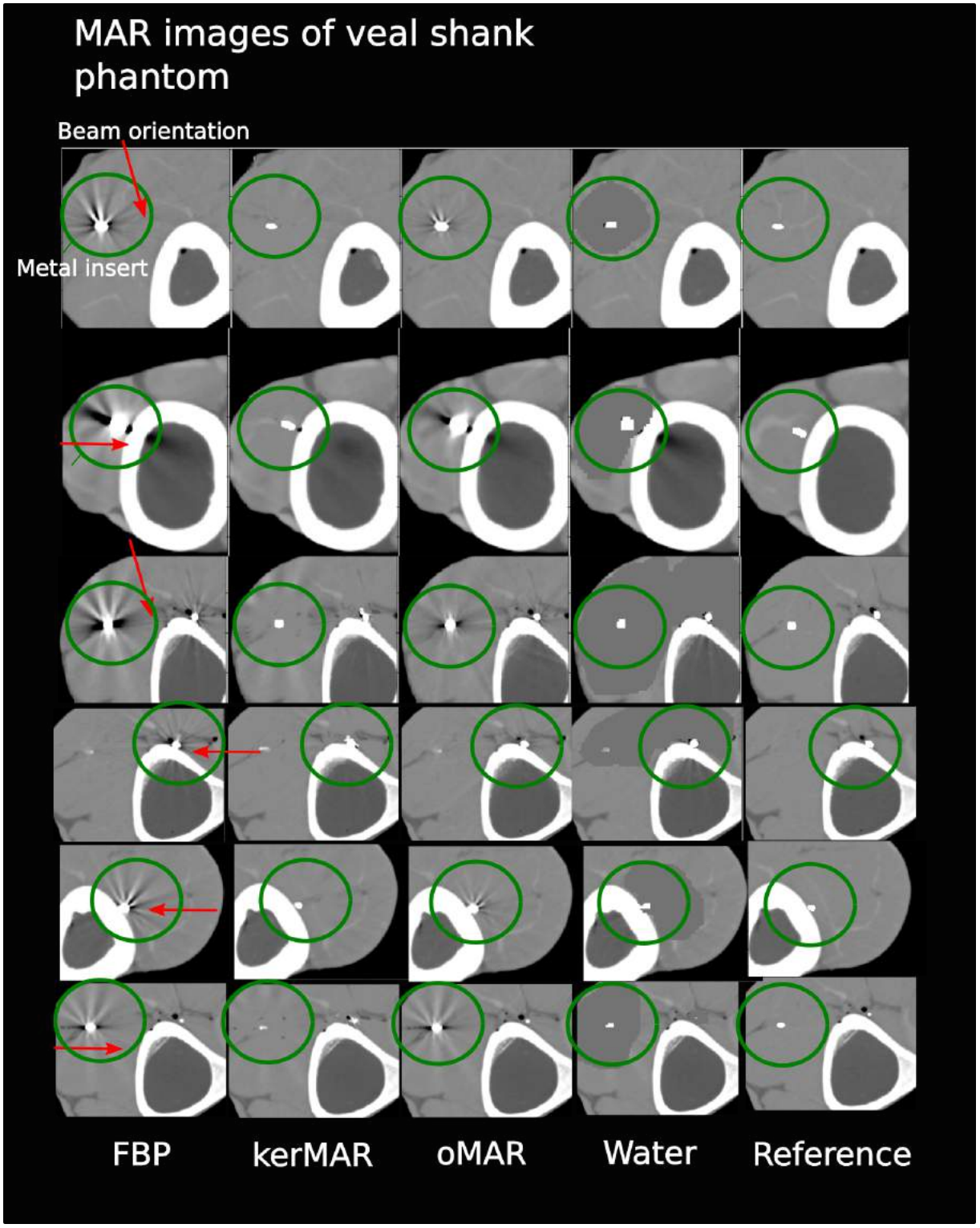


FIG. 2. Axial FBP, MAR image and uncorrupted reference slices of the veal shank phantom in the central plane of the therapeutic beams, whose orientations are shown by red arrows.

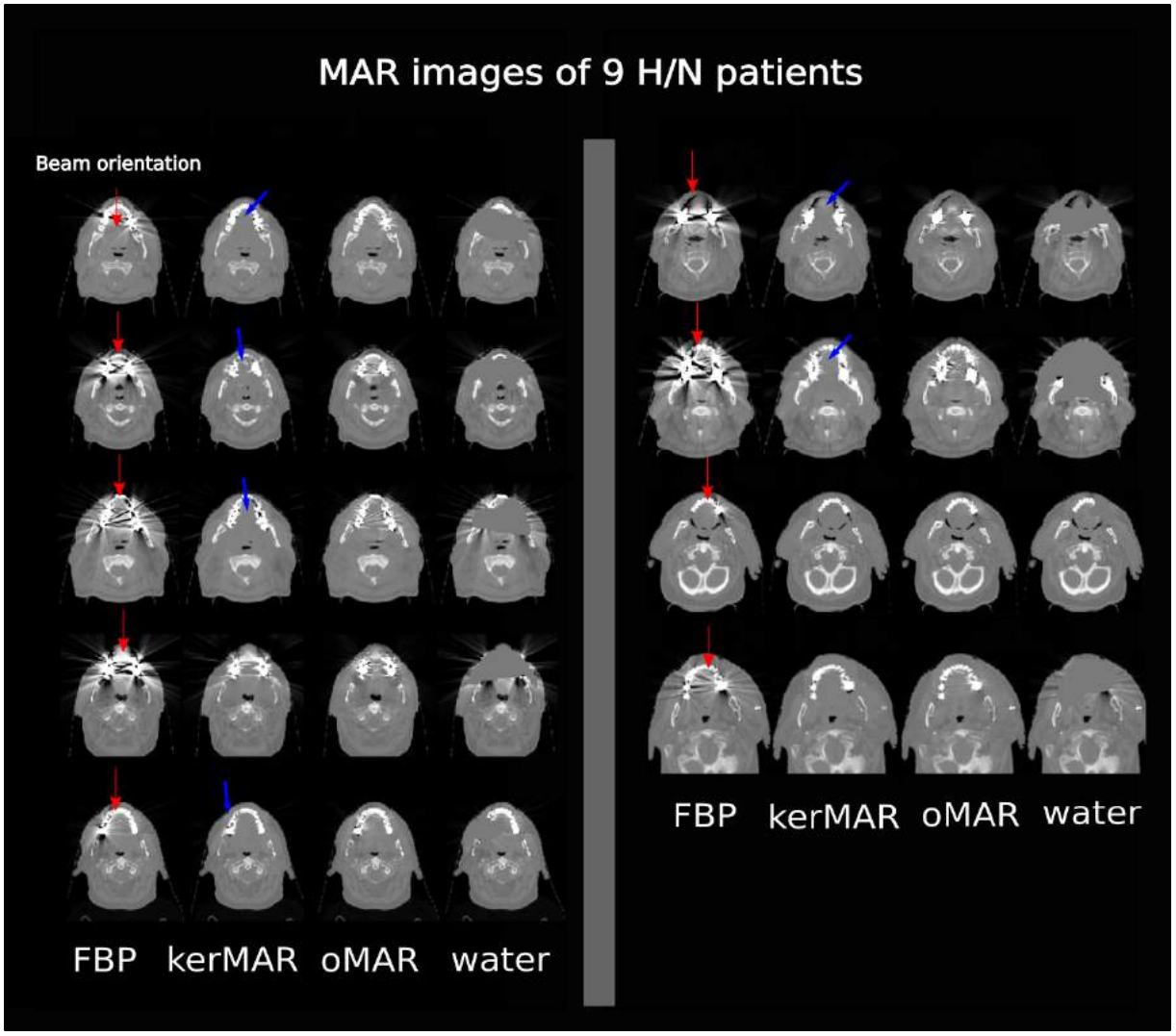


FIG. 3. Axial FBP and MAR image slices of the 9 patients in the central plane of the therapeutic beams (orientations shown by red arrows). The blue arrows show cases with potential benefits of the MR-based kerMAR algorithm over the clinically used oMAR algorithm.

III. EXPERIMENTS

We evaluated the MAR algorithms on two metrics: 1) The amount of artifacts in the images relative to a defined reference; this was quantified by counting the number of voxels with unexpectedly low and high CT values; and 2) the impact in simulated dose plans with beams angled through corrupted regions on the depth at maximum dose (photons) and the effective particle range (particles). More details on these metrics are provided in the next subsections.

A. Image analysis of artifacts

Our method for the image analysis experiments is illustrated on fig. 4. It relied on the finding that metal artifacts tend to move CT values to higher and lower values compared to an uncorrupted reference[5], which can be most easily observed in homogeneous, well-delineated regions by comparison to a reference; the artifacts here lead to noticeable increases in the tails of the CT value distribution. In the head and neck, such homogenous regions are the oral cavity, the mandibel and the teeth, the first two of which had been delineated for dose planning in our patient cohort.

For the patients, we thus acquired the clinical delineations of the oral cavity (including part of the trachea) and mandibel, manually delineating the teeth ourselves, and, by visual inspection, for each patient split the ROIs into a corrupted region and an uncorrupted reference region. This way of defining a separate reference region for each patient ensured that constant contributions to the metrics that were conserved between image sets, such as anatomical inequivalences between the reference and corrupted regions (e.g. a larger proportion of trachea in the reference region) would largely vanish when comparing between MARs.

In order to separately evaluate both the influence on the low and high intensity artifacts, we then defined HU thresholds that spanned the expected range of HU values in uncorrupted tissue. The difference between the corrupted and reference regions of the number of voxels respectively below the lower threshold and above the higher threshold, denoted δN_{low} and δN_{high}, were then used to quantify the amount of low and high intensity artifacts (as illustrated on the histogram). Values closer to 0 of these image corruption metrics imply respectively fewer low and high intensity artifacts.

To define the thresholds, we picked the lowest and highest attenuating common tissue types in each of the ROIs. For all but the tooth enamel, we then used the corresponding tissue composition and mass density data from ICRU 46[23], along with the scanner effective energy ($75.2keV$, estimated by considering the water attenuation coefficient) and the NIST[24] element-specific mass attenuation coefficients to estimate the CT value of each tissue type in HU. For the enamel we found the composition from the chemical formula for hydroxyapatite but otherwise did the same. The HU-values were then rounded to the nearest 100 so as to expand the interval. This process provided the following thresholds: Oral cavity: adipose

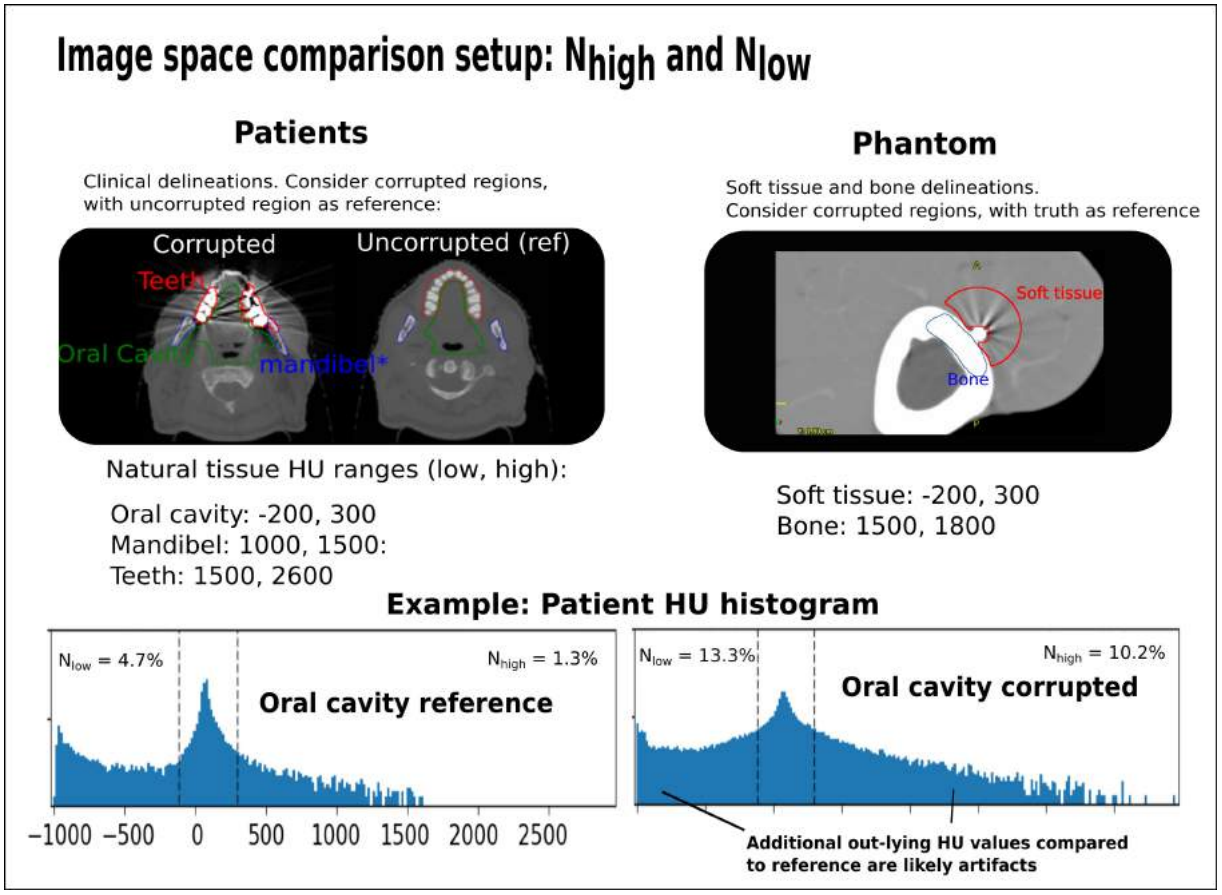


FIG. 4. Description of the setup for image analysis of the metal artifacts. Increases in the corrupted part compared to reference in voxel count fractions with HU lower and higher than the expected range in the ROI (respectively N_{low} and N_{high}) indicate artifacts.

tissue (-200 HU) to average soft tissue (300 HU); mandibel: Mandibular bone (1000 HU) to cortical bone (1500 HU); teeth: Cortical bone (1500 HU) to enamel (2600 HU)[23].

For the phantom, we delineated artifact corrupted soft tissue and bone regions as illustrated. For the soft tissue, we used the oral cavity thresholds. For the shank bone (tibia), having no reliable information on its composition, we considered the HU distribution in uncorrupted regions well separated from the ones used in our experiment and found the thresholds of 1500 and 1800HU to encompass the range of natural CT values. $\delta N_{high/low}$ were, for the phantom, defined as the difference to the uncorrupted ground truth reference.

B. Depth/range experiments

We sought to investigate the influence of the MARs on RT dose plans by considering the impact on the computed depth at maximum dose (photons) and particle range (electrons and protons). RT dose plans are devised by defining a set of beam lines in different orientations. While plan designs vary between cancer types, dose planners and treatment modality, every plan uses such beam lines as the basic building block.

Accordingly, we sought to investigate the influence of artifact reduction on a single beam, particularly in the extreme, but potentially realistic, case where it is angled through the corrupted oral cavity; our setup is explained on fig. 5. For the patients, we created three dose plans in Eclipse 14.6 (Varian Medical Systems) with a single beam through the oral cavity, using 6MV photons, 12MeV electrons and 150MeV protons. The beam orientations are shown on fig. 3 (red arrows).

For the phantom, we created similar plans and chose beam locations and orientations that were similar to the patient experiments, in that they were near to both artifacts and bone. The orientations of these beams are illustrated on fig. 2. To quantify the influence of the MARs on the dose deposited by the beams, we considered a measure of the effective range of the beams derived from the calculated central depth-dose curves, extracted from the dose distributions for all image sets.

For a photon beam, we used the depth at maximum dose[25,26]. For the electron beams, we chose the distal depth at 90% of maximum dose, denoted R_{90} (the *therapeutic range*[27]), which is a typical metric of the distal tumor coverage. For protons, we chose to use the distal depth at 80% dose since this has been found to be relatively independent of the machine-dependent energy spread of the proton beam, making it more desirable for reasons of reproducibility than e.g. R_{90}[28].

C. Statistical analysis

To quantify the significance of the between-MAR variations in our image corruption metrics and depth/range parameters, we used a simple statistical analysis. We considered the phantom data from the separate pins to be independent observations, but the data from the various MARs to be dependent, leading effectively to 4 repeated measurements

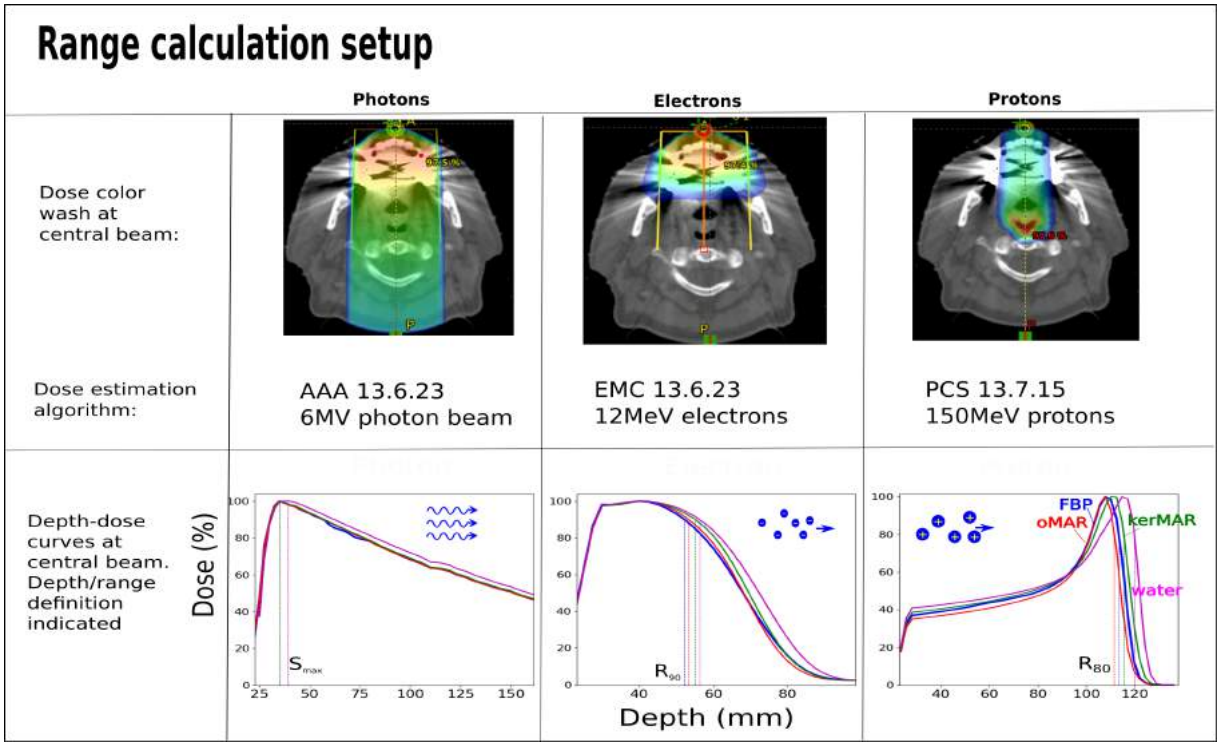


FIG. 5. Description of the depth/range estimation setup. In an RT planning program (Eclipse v. 14.6, Varian Medical Systems), photon, electron and proton beams with the indicated specifications were angled through the oral cavity (for the patients) and oral cavity-like regions (for the phantom). Dose was calculated using the indicated algorithms and the central profile depth-dose curves were exported, from which the ranges were derived.

(FBP, oMAR, kerMAR, and water override) on $N = 6$ subjects (pins). Data points for the different patients were similarly considered independent observations, leading to 4 repeated measurements on $N = 9$ subjects. The statistical analysis thus considered variations between the 4 repetitions over the respectively 6 and 9 subjects. This kind of repeat measurements analysis has the virtue of desentisising the contrast to between-subject variations, such as changes in the relative localisations of the corrupted volumes and variations in delineational practice.

As mentioned in the introduction, this paper focuses on three specific hypotheses, and we thus restricted our attention to the following three orthogonal contrasts: I: FBP versus the aggregate of kerMAR and oMAR; II: oMAR versus kerMAR; III: The aggregate of kerMAR, oMAR and FBP versus water override. The aggregates were calculated by averaging the N

observations. All contrasts being orthogonal, the p-values of the statistical tests described in the next paragraphs were not corrected for multiple comparisons.[29]

For the image analysis, we used a two-tailed Student's t-test for paired (dependent) observations on the absolute values of the image corruption metrics $\delta N_{high/low}$. We looked for significant differences in these quantities between the contrasted MAR approaches and thus calculated the average absolute difference $\Delta|\delta N_{high/low}|$ between the contrasted terms (e.g. oMAR and kerMAR), as well as the standard deviation to calculate the t-statistic; a positive mean value here means a smaller $\delta N_{high/low}$ for the second term (e.g. kerMAR) relative to the first term (e.g. oMAR). Accordingly, a positive test result with a positive mean implies fewer artifacts for the second term in the contrast, thus supporting the tested hypothesis; a negative mean on the other hand implies a smaller value for the first term in the contrast, thus supporting the rejection of the hypothesis.

For the phantom depth/range results, we used this same test and calculated $\Delta|\delta S_{max}/\delta R_{90}/\delta R_{80}|$, leading to the same interpretation of the results. For the patients, however, since no reference for the dose calculations was available, we sought to simply evaluate the significance of the absolute differences between the contrasted terms. We thus considered the mean absolute difference between the depths/ranges, denoted $|\Delta S_{max}/\Delta R_{90}/\Delta R_{80}|$. This quantity being stricly positive, we performed a one-tailed Student's t-test of the hypothesis that it was equal to 0.

IV. RESULTS

A. Hypothesis I: FBP vs. oMAR and kerMAR

Fig. 6 shows the contrast between FBP and the aggregate of oMAR and kerMAR. The first row shows the averages (bar heights) and SDE (error bars) of the quantities $\Delta|\delta N_{high/low}|$. The results are universally positive or near zero, consistently implying smaller values of the image corruption metrics for the MAR image in comparison to the uncorrected FBP. In the phantom, only the soft tissue $\Delta|\delta N_{high}|$ is statistically significant ($p < 0.05$) while the patient $\Delta|\delta N_{high}|$ results, apart from the insignificant mandibel result, are highly significant ($p < 0.01$). $\Delta|\delta N_{low}|$ is significant in the oral cavity, but not elsewhere.

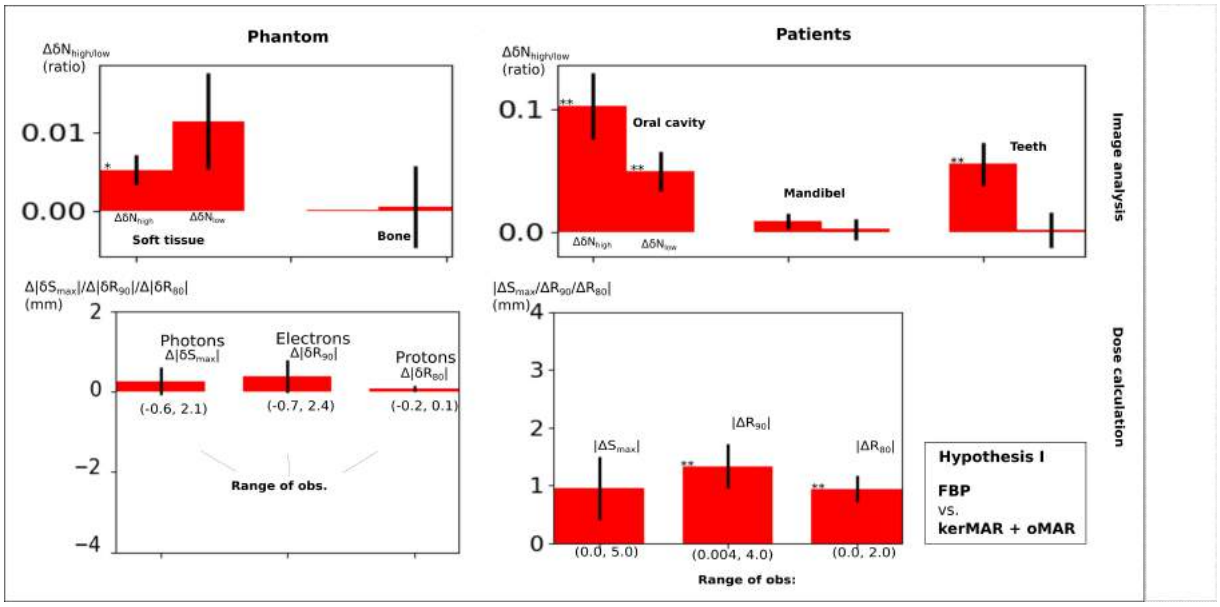


FIG. 6. Mean (bar heights) and SDE (error bars) of the variations between the datasets contrasted for testing hypothesis I: The uncorrected FBP vs. the aggregate of oMAR and kerMAR. The top row shows $\Delta|\delta N_{high/low}|$ with a column for respectively the phantom and the patients. The second row shows $\Delta|\delta S_{max}/\delta R_{90}/\delta R_{80}|$ and $\Delta|S_{max}/R_{90}/R_{80}|$ for respectively the phantom and the patients. Asterisks denote significance of the paired observations student's t-test (one asterisk significant at $p < 0.05$, two asterisks at $p < 0.01$). To interpret these results, note that for all but the bottom right figure, positive variations support the hypothesis while negative variations support its rejection; for the bottom right figure, results are always positive and simply imply different results for the contrasted MAR approaches.

The second row on fig. 6 shows the depth/range variations, $\Delta|\delta S_{max}/\delta R_{90}/\delta R_{80}|$ for the phantom and $|\Delta S_{max}/\Delta R_{90}/\Delta R_{80}|$ for the patients. There are only insignificant results for the phantom, while the range of observations is wide (for the photons and electrons, ~ -0.7 to $2.4mm$). With the patients, on the other hand, we do see a significant difference with electrons and protons (at $p < 0.01$) of respectively $\Delta R_{90} = 1.5 \pm 0.4mm$ and $R_{80} = 1.0 \pm 0.3mm$, but not for photons with $\Delta S_{max} = 1.0 \pm 0.5mm$.

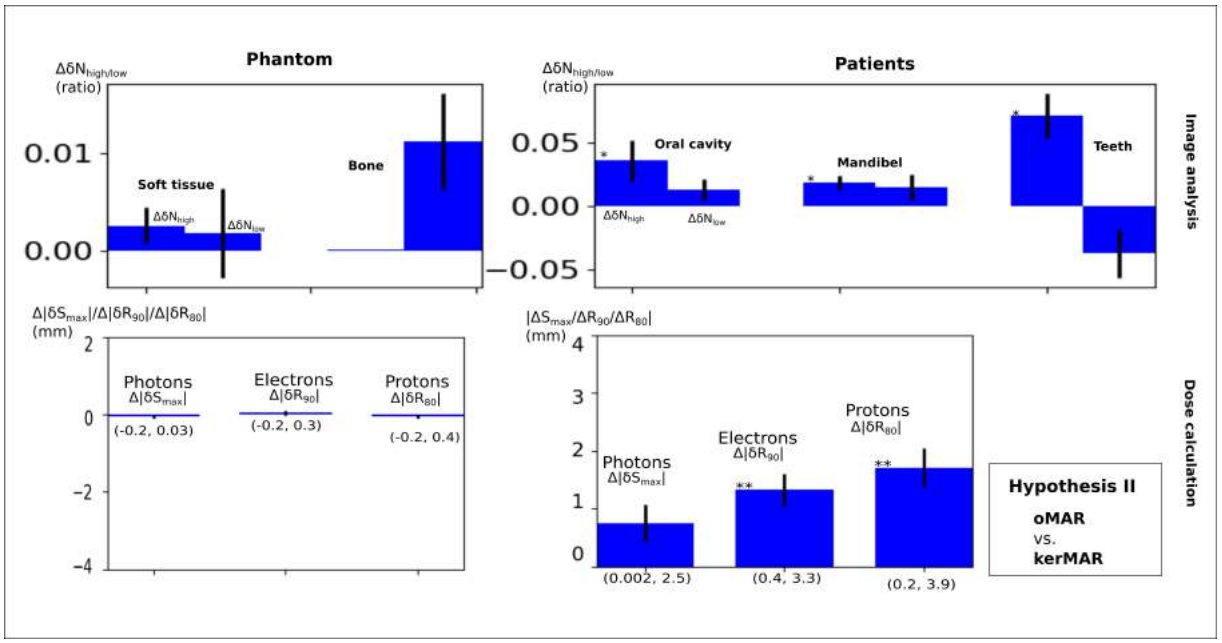


FIG. 7. Mean (bar heights) and SDE (error bars) of the variations between the datasets contrasted for testing hypothesis II: oMAR vs. kerMAR. The layout details and are identical to fig. 6, as is the interpretation of the results.

B. Hypothesis II: oMAR vs. kerMAR

Fig. 7 contrasts oMAR and kerMAR, arranged as for hypothesis I. Both $\Delta|\delta N_{high}|$ and $\Delta|\delta N_{low}|$ are here universally positive except for the teeth $\Delta|\delta N_{low}|$, implying smaller values of the image corruption metrics with kerMAR than oMAR. In the phantom, these results are not statistically significant though the positive variation for bone δN_{low} is close to, while the patient δN_{high} results are universally significant. We additionally see an almost significant negative $\Delta|\delta N_{low}|$ in the teeth.

Considering the depth/range results (second row), we see no significant results in the phantom, while we for the patients see significant differences with electrons and protons of respectively $\Delta R_{90} = 1.3 \pm 0.3mm$ and $\Delta R_{80} = 1.8 \pm 0.4mm$.

C. Hypothesis III: FBP, oMAR and kerMAR vs. water override

Fig. 8 shows the aggregate of kerMAR, oMAR and FBP contrasted with the manual water override. The top row image analysis results show significantly positive phantom

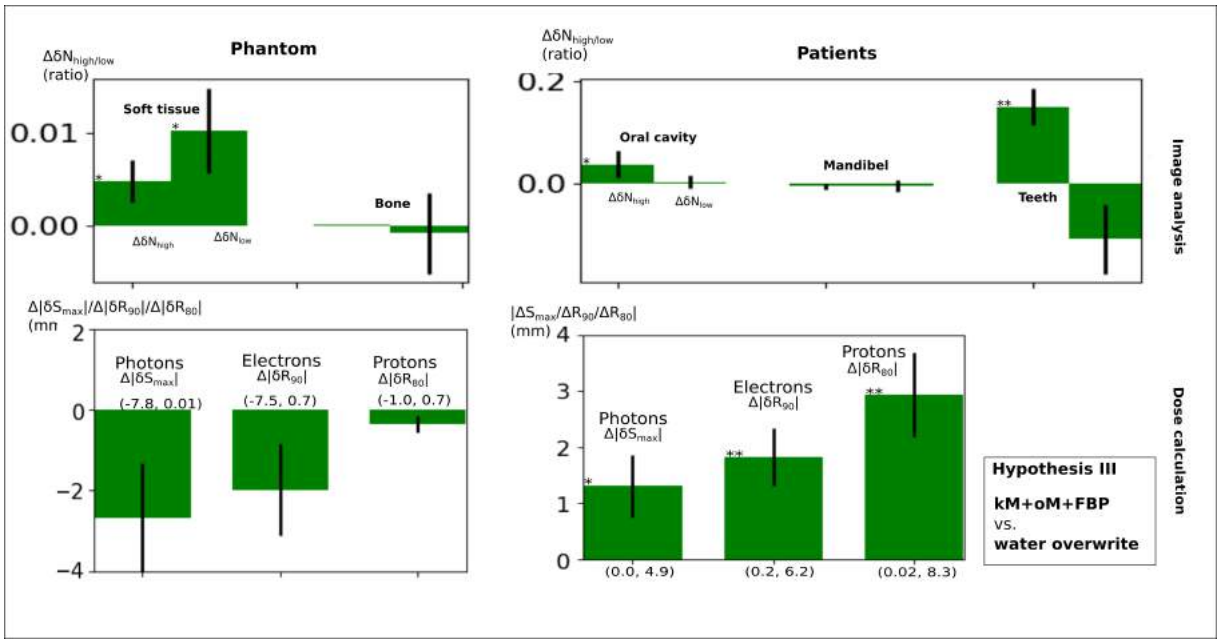


FIG. 8. Mean (bar heights) and SDE (error bars) of the variations between the datasets contrasted for testing hypothesis III: The aggregate of kerMAR, oMAR and FBP vs. water override. The layout details are identical to fig. 6, as is the interpretation of the results.

differences in the soft tissue on both image corruption metrics. In the patients, we see a significant difference only in the oral cavity and teeth δN_{high}, though with an almost significant, negative teeth $\Delta\delta N_{low}$.

The depth/range results on the bottom row are negative for the phantom, though not significantly so, in order of decreasing negativity of photon-electron-proton. The patient results reverse the ordering by being significant for electrons and highly significant for protons at respectively $\Delta R_{90} = 1.9 \pm 0.4$ and $\Delta R_{80} = 3.0 \pm 0.5mm$; photons show a smaller, less significant variation at $\Delta S_{max} = 1.2 \pm 0.4mm$. The ranges of the observations are generally large for this contrast, on the order of $\sim 7mm$ for the particles.

V. DISCUSSION

We have considered four approaches to metal artifact reduction in the context of radiotherapy, using phantom and retrospective patient data to evaluate both the artifact reduction capabilities via image metrics and the impact on photon, proton and electron maximum depth / range estimates. We focused on three hypotheses of potential clinical relevance.

A. Hypothesis I (FBP vs. oMAR and kerMAR)

Our first hypothesis postulated that using the automatic metal artifact reduction algorithms would provide a significant benefit over an uncorrected image, considering the question of how beneficial metal artifact reduction is in the context of RT. In terms of image quality, oMAR has been found in literature phantom studies to provide improvements in average HU on the order of $\sim 25\%$, along with a decrease in variance of a similar magnitude[9,11]. Improved average HU in bone areas has also been shown with oMAR[10]. Such image improvements have been found to be accompanied by proton range estimate improvements of $2-5\%$, or several mms[3,4], as well as WET estimate improvements by a similar amount in a phantom study on hip implants[6]. The photon dose accuracy improvements have in the literature been comparatively modest, with only negligible improvements for beams passing through the oral cavity for patients with closed mouths.[1,11].

In our study, we indeed saw such apparent image space improvements that were accompanied by depth/range impacts, but only at a significant level for the head and neck patients, in contrast to the more homogeneous and less artifact corrupted phantom. The particle range impacts of $\Delta R_{80} \sim 1 \pm 0.3mm$ for protons and $\Delta R_{90} \sim 1.5 \pm 0.4mm$ for electrons, which in light of the image space improvements may signify potential improvements, were of a comparable magnitude to the previous findings. They were additionally accompanied by insignificant improvements with photons, in agreement with the literature finding that photon therapy is relatively less affected by the artifacts.

B. Hypothesis II (oMAR vs. kerMAR)

Our second hypothesis was that the novel kerMAR algorithm would provide benefits over the clinically used oMAR algorithm. In the literature, oMAR has been found to leave behind residual streaks, in particular close to the metal implants, leading some authors to advise caution when using oMAR, especially if beamlines run parallel with the artifacts[5,6,11]. This tendency to leave behind residual streaks may be part of the explanation for why one study only found insignificant photon dosimetric improvements with oMAR for patients with closed mouths[1,11], and why proton WET errors of up to $\sim 4mm$ between oMAR and ground truth persisted in the phantom study on hip implants[6]. Since the MR-based kerMAR algorithm

uses superior anatomical information for the artifact reduction near the implants, it may be expected to better handle the residual streaks and thus lead to both image space and dosimetric improvements.

In our study, while kerMAR performed only similarly to oMAR on the phantom both in image and dose space, our head and neck patient image analysis indeed showed significant improvements with kerMAR in terms of the image corruption metric δN_{high}, consistent over the ROIs. This apparently improved high intensity streak suppression when taken together with the apparent visual improvements (see in particular the blue arrows on fig. 3) implies better handling of the residual oMAR streaks[6,11], consistent with the intention behind the MR based algorithm to improve streak suppression in severely corrupted regions near the metal. Considering the depth/range results for the patients, while the photon S_{max} was not significantly impacted by these improvements, kerMAR did lead to highly significant absolute particle range differences ($\Delta R_{90} = 1.3 \pm 0.2mm$ and $\Delta R_{80} = 1.8 \pm 0.3mm$ for respectively electrons and protons). Proton range differences being roughly equivalent to WET differences, these findings are smaller than and thus consistent with the maximal found error of $\sim 4mm$ in the cited hip implant phantom study[6].

C. Hypothesis III (oMAR, kerMAR and FBP vs. water override)

Our third hypothesis was that manual water override would provide a benefit over the alternatives. This method being manual, time consuming and apparently singularly effective in soft tissue regions, whilst possibly introducing human error, we may expect water override to provide good results for the phantom, but be more inconsistent with the real patients. This is reflected in the literature, where water override has been found to provide benefits that were however sometimes outdone by automatic algorithms in more complex cases[7,8].

In our study, water override showed significant apparent image improvements in the soft tissue and teeth δN_{high}, paired with a statistically insignificant error in the teeth δN_{low}. The apparent improvements however did not lead to dosimetric accuracy increases in the phantom, but rather the opposite, showing almost significant *decreases*. The high variance and thus insignificant results may be explained by the large range of observations over the individual beams, which varied from 0 to more than $-7mm$ for photons and electrons; curiously, the protons were relatively less effected, likely owing to their higher energy and

thus penetration which put the Bragg peaks outside the overridden area. Upon further investigation, the magnitude of the results depended particularly on the depth of overridden soft tissue over which the beam travelled, suggesting a large potential for introduction of systematic errors with water override. The large ranges of observation and mean depth/range impacts (1.3 ± 0.4mm, $2.0 \pm 0.3mm$ and $3.0 \pm 0.4mm$ for photons, electrons and protons) with the patients suggest that this potential carries over to real head and neck cases, at least for the particle modalities.

A possible source of the systematic errors may be that the veal shank tissue as well as the human tongue consist mainly of muscle, and thus display HU values of $\sim$ 40 to $\sim 100HU$ rather than the $0HU$ used in the override. In the phantom, this may obviously have led to errors, and may also, considering that the corrupted part of the oral cavity is at least partially occupied by the tongue, be part of the explanation for the patient results. Consequently, our study advises caution when using water override for particle RT in the head and neck case, due to the possibility of introducing errors by overriding the muscular tongue as well as possible dental areas.

A possible limitation to this part of our study is our choice to also override high intensity regions with a much different average CT value than water, such as the teeth. This means in particular that our results do not reflect the maximum potential of the water override, but may rather be viewed as a worst case impact of the method. There are in addition several advantages to using our approach for the experiments: 1) It is relatively reproducible since it overrides artifacts with little judgment of the anatomy; and 2) it may be the chosen approach of some planners, since when the artifacts are sufficiently severe and the potential error thus large and hard to estimate, replacing all artifacts with bulk CT values may lead to greater ease of planning and error management.

VI. CONCLUSIONS

Using phantom and retrospective head-and-neck patient data, we investigated the impact of MAR when using three MAR strategies, as well as their impact on the calculated head-and-neck RT depth/range estimates on photon, proton and electron beams. We investigated the following three hypotheses of potential clinical relevance: I: The automatic MARs, oMAR and kerMAR, improve upon the FBP; II: kerMAR is superior to oMAR; III: Water

override superior to the alternatives.

Hypothesis I was supported in terms of artifact reduction in soft tissue and the patient teeth. The image improvements were accompanied by a significant impact on particle ranges in the patients of $\sim 1.5 \pm 0.4mm$ and $\sim 1.0 \pm 0.3mm$.

Hypothesis II was supported in terms of high intensity artifact reduction in the same regions as hypothesis I, as well as in the mandibel, potentially owing to better suppression with kerMAR of the residual streaks left behind by oMAR. The image improvements were accompanied by impacts on the electron and proton range estimates of respectively $1.3 \pm 0.3mm$ and $1.8 \pm 0.4mm$.

Hypothesis III was supported in terms of artifact reduction in the soft tissue. However, the water override introduced systematic CT value errors leading to phantom depth/range errors coupled with significant and relatively large patient impacts of $\sim 2 \pm 0.3mm$ and $\sim 3 \pm 0.4mm$ for respectively electrons and protons. This along with a large range of observations suggested a relatively large capacity for error of the manual water override technique.

VII. CONFLICTS OF INTEREST

The authors have no relevant conflicts of interest to disclose.

VIII. REFERENCES

* Email: josni@dtu.dk; Jonathan Scharff Nielsen

[1] Giantsoudi D, De Man B, Verburg J, Trofimov A, Jin Y, Wang G, Gjesteby L, Paganetti H, Metal artifacts in computed tomography for radiation therapy planning: dosimetric effects and impact of metal artifact reduction. *Physics in Medicine and Biology* 2017;62(8):R49–80, R49–R80, doi:10.1088/1361-6560/aa5293.

[2] Kim Y, Tomé WA, 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).

Medical and Biological Engineering and Computing 2007;45(11):1045–1051, doi:10.1007/s11517-007-0196-8.

[3] Maerz M, Koelbl O, Dobler B, Influence of metallic dental implants and metal artefacts on dose calculation accuracy. *Strahlentherapie und Onkologie* 2015;191(3):234–241, doi:10.1007/s00066-014-0774-2, URL `http://link.springer.com/10.1007/s00066-014-0774-2`.

[4] Newhauser WD, Giebeler A, Langen KM, Mirkovic D, Mohan R, Can megavoltage computed tomography reduce proton range uncertainties in treatment plans for patients with large metal implants? *Physics in Medicine and Biology* 2008;53(9):2327–2344, doi:10.1088/0031-9155/53/9/009.

[5] Jäkel O, Reiss P, The influence of metal artefacts on the range of ion beams. *Physics in Medicine and Biology* 2007;52(3):635–644, doi:10.1088/0031-9155/52/3/007.

[6] Andersson KM, Ahnesjo A, Vallhagen Dahlgren C, Evaluation of a metal artifact reduction algorithm in CT studies used for proton radiotherapy treatment planning. *Journal of applied clinical medical physics / American College of Medical Physics* 2014;15(5):4857, doi:10.1120/jacmp.v15i5.4857.

[7] Ziemann C, Stille M, Cremers F, Buzug TM, Rades D, Improvement of dose calculation in radiation therapy due to metal artifact correction using the augmented likelihood image reconstruction 2018;(November 2017):227–233, doi:10.1002/acm2.12325.

[8] Zheng Y, Johnson R, Risalvato D, Is Density Override for Fiducial-Induced CT Artifacts Necessary in Proton Treatment Planning of Prostate Cancer? *Radiation Oncology Biology* 2013;87(2):S741, doi:10.1016/j.ijrobp.2013.06.1963, URL `http://dx.doi.org/10.1016/j.ijrobp.2013.06.1963`.

[9] Philips Healthcare, Metal artifact reduction for orthopedic implants (omar). `http://clinical.netforum.healthcare.philips.com/us_en/Explore/White-Papers/CT/Metal-Artifact-Reduction-for-Orthopedic-Implants-(O-MAR)`, 2012.

[10] Kidoh M, Nakaura T, Nakamura S, Tokuyasu S, Reduction of dental metallic artefacts in CT : Value of a newly developed algorithm for metal artefact reduction (O-MAR). *Clinical Radiology* 2014;69(1):e11–e16, doi:10.1016/j.crad.2013.08.008, URL `http://dx.doi.org/10.1016/j.crad.2013.08.008`.

[11] Kwon H, Kim KS, Chun YM, Wu HG, Carlson JN, Park JM, Kim JI, Evaluation of a commercial orthopaedic metal artefact reduction tool in radiation therapy of patients with head and neck

cancer. *British Journal of Radiology* 2015;88(1052):13–22, doi:10.1259/bjr.20140536.

[12] Abdoli M, Dierckx RaJO, Zaidi H, Metal artifact reduction strategies for improved attenuation correction in hybrid PET/CT imaging. *Medical Physics* 2012;39(6):3343, doi:10.1118/1.4709599.

[13] Lemmens C, Faul D, Nuyts J, Suppression of metal artifacts in CT using a reconstruction procedure that combines MAP and projection completion. *IEEE transactions on medical imaging* 2009;28(2):250–60, doi:10.1109/TMI.2008.929103, URL `http://www.ncbi.nlm.nih.gov/pubmed/19188112`.

[14] Boas FE, Fleischmann D, CT artifacts: causes and reduction techniques. *Imaging in Medicine* 2012;4(2):229–240, doi:10.2217/iim.12.13.

[15] Meyer E, Raupach R, Lell M, Schmidt B, Kachelrie M, Normalized metal artifact reduction (nmar) in computed tomography. *Medical Physics* 2010;37(10):5482–5493, doi:10.1118/1.3484090, URL `https://aapm.onlinelibrary.wiley.com/doi/abs/10.1118/1.3484090`.

[16] Slambrouck KV, Nuyts J, Van Slambrouck K, Nuyts J, Metal artifact reduction in computed tomography using local models in an image block-iterative scheme. *Medical Physics* 2012;39(11):7080–7093, doi:10.1118/1.4762567, URL `http://www.ncbi.nlm.nih.gov/pubmed/23127099`.

[17] Nuyts J, De Man B, Fessler JA, Zbijewski W, Beekman FJ, Modelling the physics in the iterative reconstruction for transmission computed tomography. *Physics in medicine and biology* 2013;58(12):R63–96, doi:10.1088/0031-9155/58/12/R63, URL `http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=3725149{&}tool=pmcentrez{&}rendertype=abstract`.

[18] Buzug TM, *Computed Tomography - From Photon Statistics to Modern Cone-Beam CT*. Springer, Berlin, 2008, doi:10.1007/978-3-540-39408-2.

[19] Nielsen J, Edmund J, Van Leemput K, *CT metal artifact reduction using MR image patches*, vol. 10573. SPIE - International Society for Optical Engineering, 2018;doi:10.1117/12.2293815.

[20] Bishop CM, *Pattern recognition and machine learning*. Springer, 2006.

[21] Maes F, Collignon A, Vandermeulen D, Marchal G, Suetens P, Multimodality image registration by maximization of mutual information. *IEEE Transactions on Medical Imaging* 1997;16(2):187–198, doi:10.1109/42.563664.

[22] Wells WMI, Viola P, Atsumi H, Nakajima S, Kikinis R, Multi-Modal image registration by maximization of mutual information. *Medical Image Analysis* 1996;1(1):35–51,

doi:10.1016/S1361-8415(01)80026-8, URL http://linkinghub.elsevier.com/retrieve/pii/S1361841501800268{%}5Cnhttp://linkinghub.elsevier.com/retrieve/pii/S1361841501800049.

[23] International Commission on Radiation Units and Measurements, ICRU report 46: Photon, Electron, Proton and Neutron Interaction Data for Body Tissues. 1968.

[24] National Institute for Science and Technology, X-ray mass attenuation coefficients. https://physics.nist.gov/PhysRefData/XrayMassCoef/tab3.html, Accessed 2018, specific references at: https://physics.nist.gov/PhysRefData/XrayMassCoef/ref.html.

[25] DAHANCA, Retningslinjer for strålebehandling i DAHANCA 2013;:0–31.

[26] International Commission on Radiation Units and Measurements, ICRU 83. *Journal of the ICRU* 2002;10(1):41, doi:10.1093/jicru/ndn032.

[27] Chang DS, Lasley FD, Das IJ, Mendonca MS, Dynlacht JR, *Dosimetry of Electron Beams*. Springer International Publishing, Cham, 2014;111–123, doi:10.1007/978-3-319-06841-1"˙10, URL https://doi.org/10.1007/978-3-319-06841-1_10.

[28] Hsi WC, Schreuder AN, Moyers MF, Allgower CE, Farr JB, Mascia AE, Range and modulation dependencies for proton beam dose. *Medical physics* 2009;36.2:634–641, doi:10.1118/1.3056466.

[29] Ruxton GD, Beauchamp G, Time for some a priori thinking about post hoc testing. *Behavioral Ecology* 2008;19(3):690–693, doi:10.1093/beheco/arn020, URL http://dx.doi.org/10.1093/beheco/arn020.

Bibliography

[1] S J Thomas. Relative electron density calibration of CT scanners for radiotherapy Relative electron density calibration of CT scanners for radiotherapy treatment planning. (April), 2016. doi: 10.1259/bjr.72.860.10624344.

[2] U Schneider, E Pedroni, and A Lomax. The calibration of ct hounsfield units for radiotherapy treatment planning. *Physics in Medicine and Biology*, 41(1):111–124, 1996. ISSN 13616560, 00319155. doi: 10.1088/0031-9155/41/1/009.

[3] F. Herrera-Martinez, M. Rodriguez-Villafuerte, A. Martinez-Davalos, C. Ruiz-Trejo, M. A. Celis-Lopez, J. M. Larraga-Gutierrez, and A. Garcia-Garduno. Electron density calibration for radiotherapy treatment planning. *AIP Conference Proceedings*, 854(1), 9 2006. doi: 10.1063/1.2356416.

[4] Wen C Hsi, Andries N Schreuder, Michael F Moyers, Chris E Allgower, Jonathan B Farr, and Anthony E Mascia. Range and modulation dependencies for proton beam dose. *Medical physics*, 36.2:634–641, 2009. doi: 10.1118/1.3056466.

[5] Jatinder R Palta PhD. Understanding the Uncertainties in Proton Therapy. http://ptcog.ch/archive/conference_p&t&v/PTCOG52/PresentationsEW/E-10-Palta.pdf, 2013.

[6] Harald Paganetti. Range Uncertainties in Proton Therapy. https://amos3.aapm.org/abstracts/pdf/77-22619-310436-91773.pdf, 2013.

[7] Thorsten M Buzug. Computed Tomography. page 526, 2008. doi: 10.1007/978-3-540-39408-2.

[8] Lars Gjesteby, Bruno D E Man, Yannan Jin, Harald Paganetti, Joost Verburg, Drosoula Giantsoudi, and G E Wang Fellow. Metal Artifact Reduction in CT : Where Are We After Four Decades ? 4, 2016.

[9] Philips Healthcare. Metal artifact reduction for orthopedic implants (omar). http://clinical.netforum.healthcare.philips.com/us_en/Explore/White-Papers/CT/Metal-Artifact-Reduction-for-Orthopedic-Implants-(O-MAR), 2012.

[10] Usman Bashir et al. Daniel J Bell. Magnetic susceptibility artifact. https://radiopaedia.org/articles/magnetic-susceptibility-artifact, Accessed 2018. URL https://radiopaedia.org/articles/magnetic-susceptibility-artifact.

[11] Dwight G. Nishimura. *Principles of Magnetic Resonance Imaging*. Department of Electrical Engineering, Stanford University, 1996.

[12] M Hofmann, F Steinke, V Scheel, G Charpiat, J Farquhar, P Aschoff, M Brady, B Scholkopf, and B J Pichler. MRI-based attenuation correction for PET/MRI: A novel approach combining pattern recognition and atlas registration. *Journal of Nuclear Medicine*, 49 (11):1875–1883, 2008. ISSN 0161-5505. doi: 10.2967/jnumed.107.049353. URL http://jnm.snmjournals.org/content/49/11/1875.full.pdf.

[13] Mika Kapanen and Mikko Tenhunen. T1/T2*-weighted MRI provides clinically relevant pseudo-CT density data for the pelvic bones in MRI-only based radiotherapy treatment planning. *Acta Oncologica*, 52(August):1–7, 2012. ISSN 0284-186X. doi: 10.3109/0284186X.2012.692883. URL http://www.ncbi.nlm.nih.gov/pubmed/22712634.

[14] Herman Cember and Thomas E. Johnson. *Health*. McGraw-Hill, fourth edi edition, 2009. ISBN 9780071643238.

[15] K M Hanson and G W Wecksung. Local basis-function approach to computed tomography. *Applied optics*, 24(23):4028, 1985. ISSN 0003-6935. doi: 10.1364/AO.24.004028.

[16] Michael T. Heath. *Scientific Computing - An introductory survey, Second Edition*. McGraw-Hill, January 2002.

[17] Moore-penrose matrix inverse. From MathWorld - A Wolfram Web Resource, 2019. URL http://mathworld.wolfram.com/Moore-PenroseMatrixInverse.html.

[18] Wim van Aarle, Willem Jan Palenstijn, Jan De Beenhouwer, Thomas Altantzis, Sara Bals, K. Joost Batenburg, and Jan Sijbers. The astra toolbox: A platform for advanced algorithm development in electron tomography. *Ultramicroscopy*, 157:35–47, 2015. ISSN 18792723, 03043991. doi: 10.1016/j.ultramic.2015.05.002.

[19] Wim van Aarle, Willem Jan Palenstijn, Jeroen Cant, Eline Janssens, Folkert Bleichrodt, Andrei Dabravolski, Jan De Beenhouwer, K. Joost Batenburg, and Jan Sijbers. Fast and flexible x-ray tomography using the astra toolbox. *Optics Express*, 24(22):25129–25147, 2016. ISSN 10944087. doi: 10.1364/OE.24.025129.

[20] Yong Long, Jeffrey A Fessler, and James M Balter. 3D Forward and Back-Projection for X-Ray CT Using Separable Footprints. 29(11):1839–1850, 2010.

[21] National Health Service. Report 06014: Wide bore CT scanner comparison report version 14. *Nhs*, (February), 2006.

[22] Centre for Evidence-based Purchasing (CEP). Comparative specifications: Wide bore CT scanners (CEP08029). (March):76, 2009.

[23] Marc Kachelrie??, Michael Knaup, Christian Pen??el, and Willi A. Kalender. Flying focal spot (FFS) in cone-beam CT. *IEEE Transactions on Nuclear Science*, 53(3):1238–1247, 2006. ISSN 00189499. doi: 10.1109/TNS.2006.874076.

[24] Cylindrical coordinates. From MathWorld - A Wolfram Web Resource, 2019. URL http://mathworld.wolfram.com/CylindricalCoordinates.html.

[25] Jiang Hsieh. *Computed Tomography: Principles, Design, Artifacts and Recent Advances*. 2002. ISBN 9780819475336. doi: 10.1117/3.817303.ch9.

[26] Henrik Turbell. *Cone-Beam Reconstruction Using Filtered Backprojectionn*. Number 672. 2001. ISBN 9172199199. URL http://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.134.5224{&}rep=rep1{&}type=pdf.

[27] G Poludniowski, G Landry, F DeBlois, P M Evans, and F Verhaegen. Spekcalc : a program to calculate photon spectra from tungsten anode x-ray tubes. *Physics in Medicine and Biology*, 54(19):N433, 2009. URL http://stacks.iop.org/0031-9155/54/i=19/a=N01.

[28] National Institute for Science and Technology. X-ray mass attenuation coefficients. https://physics.nist.gov/PhysRefData/XrayMassCoef/tab3.html, Accessed 2018. Specific references at: https://physics.nist.gov/PhysRefData/XrayMassCoef/ref.html.

[29] Johan Nuyts, Bruno De Man, Jeffrey A Fessler, Wojciech Zbijewski, and Freek J Beekman. Modelling the physics in the iterative reconstruction for transmission computed tomography. *Physics in medicine and biology*, 58(12):R63–96, 2013. ISSN 1361-6560. doi: 10.1088/0031-9155/58/12/R63. URL http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=3725149{&}tool=pmcentrez{&}rendertype=abstract.

[30] Yuan Lin and Ehsan Samei. A fast poly-energetic iterative FBP algorithm. *Physics in Medicine and Biology*, 59(7):1655–1678, 2014. ISSN 13616560. doi: 10.1088/0031-9155/59/7/1655.

[31] Andrea Schwahofer, Esther Bär, Stefan Kuchenbecker, J. Günter Grossmann, Marc Kachelrieß, and Florian Sterzing. The application of metal artifact reduction (MAR) in CT scans for radiation oncology by monoenergetic extrapolation with a DECT scanner. *Zeitschrift fur Medizinische Physik*, 25(4):314–325, 2015. ISSN 18764436. doi: 10.1016/j.zemedi.2015.05.004. URL http://dx.doi.org/10.1016/j.zemedi.2015.05.004.

[32] F Edward Boas and Dominik Fleischmann. CT artifacts: causes and reduction techniques. *Imaging in Medicine*, 4(2):229–240, 2012. ISSN 1755-5191. doi: 10.2217/iim.12.13.

[33] J Hsieh, R C Molthen, C a Dawson, and R H Johnson. An iterative approach to the beam hardening correction in cone beam CT. *Medical physics*, 27(1):23–29, 2000. ISSN 00942405. doi: 10.1118/1.598853.

[34] G. Van Gompel, K. Van Slambrouck, M. Defrise, K. J. Batenburg, J. de Mey, J. Sijbers, and J. Nuyts. Iterative correction of beam hardening artifacts in CT. *Medical Physics*, 38 (S1):S36, 2011. ISSN 00942405. doi: 10.1118/1.3577758.

[35] J Nuyts, C Michel, and P Dupont. Maximum-likelihood expectation-maximization reconstruction of sinograms with arbitrary noise distribution using nec-transformations. *Ieee Transactions on Medical Imaging*, 20(5):365–375, 2001. ISSN 1558254x, 02780062. doi: 10.1109/42.925290.

[36] Zhiqian Chang, Ruoqiao Zhang, Jean Baptiste Thibault, Debashish Pal, Lin Fu, Ken Sauer, and Charles Bouman. Modeling and pre-treatment of photon-starved CT data for iterative reconstruction. *IEEE Transactions on Medical Imaging*, PP(99):277–287, 2016. ISSN 1558254X. doi: 10.1109/TMI.2016.2606338.

[37] Issei Mori, Yoshio Machida, Makoto Osanai, and Kazuhiro Iinuma. Photon starvation artifacts of x-ray ct: Their true cause and a solution. *Radiological Physics and Technology*, 6(1):130–141, 2013. ISSN 18650341, 18650333. doi: 10.1007/s12194-012-0179-9.

[38] Bruno De Man. *PhD Thesis: Iterative Reconstruction for Reduction of Metal Artifacts in Computed Tomography*. Katholieke Universitiet Leuven, Leuven, 2001. ISBN 9056823000. URL https://lirias.kuleuven.be/handle/123456789/174050.

[39] Y Zheng, R Johnson, and D Risalvato. Is Density Override for Fiducial-Induced CT Artifacts Necessary in Proton Treatment Planning of Prostate Cancer? *Radiation Oncology Biology*, 87(2):S741, 2013. ISSN 0360-3016. doi: 10.1016/j.ijrobp.2013.06.1963. URL http://dx.doi.org/10.1016/j.ijrobp.2013.06.1963.

[40] Manuel Maerz, Oliver Koelbl, and Barbara Dobler. Influence of metallic dental implants and metal artefacts on dose calculation accuracy. *Strahlentherapie und Onkologie*, 191(3): 234–241, 2015. ISSN 0179-7158. doi: 10.1007/s00066-014-0774-2. URL http://link.springer.com/10.1007/s00066-014-0774-2.

[41] Esther Meyer, Rainer Raupach, Michael Lell, Bernhard Schmidt, and Marc Kachelrieß. Normalized metal artifact reduction (nmar) in computed tomography. *Medical Physics*, 37(10):5482–5493, 2010. doi: 10.1118/1.3484090. URL https://aapm.onlinelibrary.wiley.com/doi/abs/10.1118/1.3484090.

[42] M. M. Lell, E. Meyer, M. Schmid, R. Raupach, M. S. May, M. Uder, and M. Kachelriess. Frequency split metal artefact reduction in pelvic computed tomography. *European Radiology*, 23(8):2137–2145, 2013. ISSN 09387994. doi: 10.1007/s00330-013-2809-y.

[43] Esther Meyer, Rainer Raupach, Michael Lell, Bernhard Schmidt, and Marc Kachelriess. Edge-preserving metal artifact reduction. 8313:83133A, 2012. doi: 10.1117/12.906392. URL http://proceedings.spiedigitallibrary.org/proceeding.aspx?doi=10.1117/12.906392.

[44] Stuart Lloyd. Least squares quantization in pcm. *Ieee Transactions on Information Theory*, IT-28(2):129–137, 1982. ISSN 15579654, 00189448.

[45] Steven Tilley II, Jeffrey H Siewerdsen, and J Webster Stayman. Model-based iterative reconstruction for flat-panel cone-beam CT with focal spot blur, detector blur, and correlated noise. *Physics in Medicine and Biology*, 61(1):296, 2016. ISSN 0031-9155. doi: 10.1088/0031-9155/61/1/296. URL http://stacks.iop.org/0031-9155/61/i=1/a=296.

[46] Lin Fu, Tzu-cheng Lee, Soo Mee Kim, Adam M Alessio, Paul E Kinahan, Zhiqian Chang, Ken Sauer, Mannudeep K Kalra, and Bruno De Man. Comparison Between Pre-Log and Post-Log Statistical Models in Ultra-Low-Dose CT Reconstruction. 36(3):707–720, 2017.

[47] David J C MacKay. *Information Theory, Inference, and Learning Algorithms David J.C. MacKay*, volume 100. 2005. ISBN 9780521642989. doi: 10.1198/jasa.2005.s54. URL http://pubs.amstat.org/doi/abs/10.1198/jasa.2005.s54{%}5Cnhttp://www.cambridge.org/0521642981.

[48] C.P. Robert. *The Bayesian choice. A decision-theoretic motivation*. Springer-Verlag,, 1994. ISBN 0387942963, 3540942963.

[49] John Nuyts, Bruno De Man, Patrick Dupont, Michel Defrise, Paul Suetens, and Luc Mortelmans. Iterative reconstruction for helical CT: a simulation study. *Physics in Medicine and Biology*, 43(4):729, 1998. ISSN 1361-6560. doi: 10.1088/0031-9155/43/4/003. URL http://stacks.iop.org/0031-9155/43/i=4/a=003.

[50] Katrien Van Slambrouck and Johan Nuyts. Local models for artefact reduction in iterative CT reconstruction. *IEEE Nuclear Science Symposium Conference Record*, (2):4346–4350, 2012. ISSN 10957863. doi: 10.1109/NSSMIC.2011.6153837.

[51] J Fessler. TU-A-211-01: Iterative Image Reconstruction for CT. *Medical Physics*, 38(6): 3744, 2011. ISSN 00942405. doi: 10.1118/1.3613090.

[52] Thomas P Minka. Expectation-Maximization as lower bound maximization. 1998.

[53] Sean Borman. The Expectation Maximization Algorithm A short tutorial. *Submitted for publication*, 25(x):1–9, 2009. ISSN 15360229. doi: 10.1097/RLU.0b013e3181b06c41\r00003072-200909000-00002. URL http://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.150.8193{&}rep=rep1{&}type=pdf.

[54] Andras A Anderla, Dubravko R Culibrk, and Gaspar Delso. Metal Artifact Reduction from CT Images Using Complementary MR Images. 2013.

[55] G Delso, S Wollenweber, a Lonn, F Wiesinger, and P Veit-Haibach. MR-driven metal artifact reduction in PET/CT. *Physics in medicine and biology*, 58(7):2267–80, 2013. ISSN 1361-6560. doi: 10.1088/0031-9155/58/7/2267. URL http://www.ncbi.nlm.nih.gov/pubmed/23478566.

[56] Peter C. Park, Eduard Schreibmann, Justin Roper, Eric Elder, Ian Crocker, Tim Fox, X. Ronald Zhu, Lei Dong, and Anees Dhabaan. MRI-based computed tomography metal artifact correction method for improving proton range calculation accuracy. *International Journal of Radiation Oncology Biology Physics*, 91(4):849–856, 2015. ISSN 1879355X. doi: 10.1016/j.ijrobp.2014.12.027. URL http://dx.doi.org/10.1016/j.ijrobp.2014.12.027.

[57] D Andreasen, K Van Leemput, R H Hansen, J A L Andersen, and J M Edmund. Patch-based generation of a pseudo CT from conventional MRI sequences for MRI-only radiotherapy of the brain. *Med Phys*, 42(4):1596–1605, 2015. ISSN 0094-2405. doi: 10.1118/1.4914158.

[58] A. Torrado-Carvajal, J. L. Herraiz, E. Alcain, A. S. Montemayor, L. Garcia-Canamaque, J. A. Hernandez-Tamames, Y. Rozenholc, and N. Malpica. Fast Patch-Based Pseudo-CT Synthesis from T1-Weighted MR Images for PET/MR Attenuation Correction in Brain Studies. *Journal of Nuclear Medicine*, 57(1):136–144, 2015. ISSN 0161-5505. doi: 10.2967/jnumed.115.156299. URL http://jnm.snmjournals.org/cgi/doi/10.2967/jnumed.115.156299.

[59] Kuan-Hao Su, Lingzhi Hu, Christian Stehning, Michael Helle, Pengjiang Qian, Cheryl L Thompson, Gisele C Pereira, David W Jordan, Karin A Herrmann, Melanie Traughber, Raymond F Jr Muzic, and Bryan J Traughber. Generation of brain pseudo-CTs using an undersampled, single-acquisition UTE-mDixon pulse sequence and unsupervised clustering. *Medical physics*, 42(8):4974–4986, 2015. ISSN 0094-2405 (Print). doi: 10.1118/1.4926756. URL http://dx.doi.org/10.1118/1.4926756.

[60] Camps S., Van Der Meer S., and Verhaegen F. Various approaches for pseudo-CTscan creation based on ultrasound to ultrasound deformable image registration between different treatment time points for radiotherapy treatment plan adaptation in prostate cancer patients. *Biomedical Physics and Engineering Express*, 2(3):no pagination, 2016. URL http://iopscience.iop.org/article/10.1088/2057-1976/2/3/035018/pdf{%}5Cnhttp://ovidsp.ovid.com/ovidweb.cgi?T=JS{&}PAGE=reference{&}D=emed18b{&}NEWS=N{&}AN=612115192.

[61] Christopher M. Bishop. *Pattern recognition and machine learning*. Springer, 2006. ISBN 0387310738, 9780387310732.

[62] N. Otsu. A threshold selection method from gray-level histograms. *IEEE Transactions on Systems, Man, and Cybernetics*, 9(1):62–66, Jan 1979. ISSN 0018-9472. doi: 10.1109/TSMC.1979.4310076.

[63] Vinh-thong Ta, D Louis Collins, and Pierrick Coup. Optimized PatchMatch for Near Real Time and Accurate Label Fusion. 2:105–112, 2014.

[64] Timothy J Robinson. The r book. *American Statistician*, 62(3):272–273, 2008. ISSN 15372731, 00031305. doi: 10.1198/tas.2008.s263.

[65] Joakim H Jonsson, Magnus G Karlsson, Mikael Karlsson, and Tufve Nyholm. Treatment planning using MRI data: an analysis of the dose calculation accuracy for different treatment regions. *Radiation oncology (London, England)*, 5:62, 2010. ISSN 1748-717X. doi: 10.1186/1748-717X-5-62. URL http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=2909248{&}tool=pmcentrez{&}rendertype=abstract.

[66] Stefan Both. Proton Treatment Planning. http://chapter.aapm.org/GLC/media/2012/Both.pdf, 2012.

[67] I Rosenberg. Radiation Oncology Physics: A Handbook for Teachers and Students. *British journal of cancer*, 98:1020, 2008. ISSN 00942405. doi: 10.1038/sj.bjc.6604224.

[68] Damien C. Weber, André Abrunhosa-Branquinho, Alessandra Bolsi, Andrzej Kacperek, Rémi Dendale, Dirk Geismar, Barbara Bachtiary, Annika Hall, Jens Heufelder, Klaus Herfarth, Jürgen Debus, Maurizio Amichetti, Mechthild Krause, Roberto Orecchia, Vladimir Vondracek, Juliette Thariat, Tomasz Kajdrowicz, Kristina Nilsson, and Cai Grau. Profile of european proton and carbon ion therapy centers assessed by the eortc facility questionnaire. *Radiotherapy and Oncology*, 124(2):185 – 189, 2017. ISSN 0167-8140. doi: https://doi.org/10.1016/j.radonc.2017.07.012. URL http://www.sciencedirect.com/science/article/pii/S0167814017324714.

[69] Harald Paganetti. Range uncertainties in proton therapy and the role of Monte Carlo simulations. *Physics in Medicine and Biology*, 57(11), 2012. ISSN 00319155. doi: 10.1088/0031-9155/57/11/R99.

[70] Wayne D Newhauser and Rui Zhang. The physics of proton therapy. 2015. doi: 10.1088/0031-9155/60/8/R155.

[71] Christian Ziemann, Maik Stille, Florian Cremers, Thorsten M Buzug, and Dirk Rades. Improvement of dose calculation in radiation therapy due to metal artifact correction using the augmented likelihood image reconstruction. (November 2017):227–233, 2018. doi: 10.1002/acm2.12325.

[72] Drosoula Giantsoudi, Bruno De Man, Joost Verburg, Alexei Trofimov, Yannan Jin, Ge Wang, Lars Gjesteby, and Harald Paganetti. Metal artifacts in computed tomography for radiation therapy planning: dosimetric effects and impact of metal artifact reduction. *Physics in Medicine and Biology*, 62(8):R49–80, R49–R80, 2017. ISSN 13616560, 00319155. doi: 10.1088/1361-6560/aa5293.

[73] A. Wambersie. Journal of the ICRU. 10(1):41, 2002. ISSN 17423422. doi: 10.1093/jicru/ndn032.

[74] Karin M Andersson, Anders Ahnesjo, and Christina Vallhagen Dahlgren. Evaluation of a metal artifact reduction algorithm in CT studies used for proton radiotherapy treatment planning. *Journal of applied clinical medical physics / American College of Medical Physics*, 15(5):4857, 2014. ISSN 1526-9914 (Electronic). doi: 10.1120/jacmp.v15i5.4857.

[75] Wayne D. Newhauser, Annelise Giebeler, Katja M. Langen, Dragan Mirkovic, and Radhe Mohan. Can megavoltage computed tomography reduce proton range uncertainties in treatment plans for patients with large metal implants? *Physics in Medicine and Biology*, 53(9): 2327–2344, 2008. ISSN 00319155. doi: 10.1088/0031-9155/53/9/009.

[76] H. Kwon, K. S. Kim, Y. M. Chun, H. G. Wu, J. N.K. Carlson, J. M. Park, and J. I. Kim. Evaluation of a commercial orthopaedic metal artefact reduction tool in radiation therapy of patients with head and neck cancer. *British Journal of Radiology*, 88(1052):13–22, 2015. ISSN 00071285. doi: 10.1259/bjr.20140536.

[77] Zhilei Liu Shen, Ping Xia, Paul Klahr, and Toufik Djemil. Dosimetric impact of orthopedic metal artifact reduction (O-MAR) on spine SBRT patients. 16(5):106–116, 2015.

[78] Damian J. Tyler, Matthew D. Robson, R. Mark Henkelman, Ian R. Young, and Graeme M. Bydder. Magnetic resonance imaging with ultrashort te (ute) pulse sequences: Technical considerations. *Journal of Magnetic Resonance Imaging*, 25(2):279–289. doi: 10.1002/jmri.20851. URL https://onlinelibrary.wiley.com/doi/abs/10.1002/jmri.20851.

[79] Weiliang Fan, Jing M. Chen, and Weimin Ju. A pixel missing patch inpainting method for remote sensing image. *Proceedings - 2011 19th International Conference on Geoinformatics, Geoinformatics 2011*, (40871240):3–6, 2011. ISSN 2161-024X. doi: 10.1109/GeoInformatics.2011.5980782.

[80] Paul A. Ardis and Christopher M. Brown. Self-similarity inpainting. *Proceedings - International Conference on Image Processing, ICIP*, pages 2789–2792, 2009. ISSN 15224880. doi: 10.1109/ICIP.2009.5414184.

[81] Maxime Daisy, David Tschumperle, and Olivier Lezoray. Spatial patch blending for artefact reduction in pattern-based inpainting techniques. *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in*

Bioinformatics), 8048 LNCS(PART 2):523–530, 2013. ISSN 03029743. doi: 10.1007/978-3-642-40246-3_65.

[82] Probabilstic graphical models and their applications.

[83] Hanming Zhang, Lei Li, Linyuan Wang, Yanmin Sun, Bin Yan, Ailong Cai, and Guoen Hu. Computed Tomography Sinogram Inpainting With Compound Prior Modelling Both Sinogram and Image Sparsity. *IEEE Transactions on Nuclear Science*, 63(1):2567–2576, 2016. doi: 10.1109/TNS.2016.2577045.

Danmarks Tekniske Universitet

DTU Health Tech
Department of Health Technology
Ørsteds Plads, Building 345C
DK-2800 Kgs. Lyngby
Denmark

Email: healthtech-info@dtu.dk
www.healthtech.dtu.dk