

**Analysis methods for abdominal imaging  
Quantifying Crohn's disease activity**

Naziroglu, Robiel

**DOI**

[10.4233/uuid:e84230f0-19c6-40f2-892d-48a9986f257c](https://doi.org/10.4233/uuid:e84230f0-19c6-40f2-892d-48a9986f257c)

**Publication date**

2016

**Document Version**

Final published version

**Citation (APA)**

Naziroglu, R. (2016). *Analysis methods for abdominal imaging: Quantifying Crohn's disease activity*. [Dissertation (TU Delft), Delft University of Technology]. <https://doi.org/10.4233/uuid:e84230f0-19c6-40f2-892d-48a9986f257c>

**Important note**

To cite this publication, please use the final published version (if applicable).  
Please check the document version above.

**Copyright**

Other than for strictly personal use, it is not permitted to download, forward or distribute the text or part of it, without the consent of the author(s) and/or copyright holder(s), unless the work is under an open content license such as Creative Commons.

**Takedown policy**

Please contact us and provide details if you believe this document breaches copyrights.  
We will remove access to the work immediately and investigate your claim.

# Analysis methods for abdominal imaging: quantifying Crohn's disease activity



# Analysis methods for abdominal imaging: quantifying Crohn's disease activity

## **Proefschrift**

ter verkrijging van de graad van doctor  
aan de Technische Universiteit Delft,  
op gezag van de Rector Magnificus prof. ir. K.C.A.M. Luyben;  
voorzitter van het College voor Promoties,  
in het openbaar te verdedigen op maandag 21 november 2016 om 15:00 uur

door

**Robiel Estifanas NAZIROGLU**

Biomedisch ingenieur  
geboren te Utrecht, Nederland

Dit proefschrift is goedgekeurd door de  
promotoren: Prof. dr. ir. L.J. van Vliet en Prof. dr. J. Stoker  
copromotor: Dr. F.M. Vos

**Samenstelling promotiecommissie:**

Rector Magnificus,	Voorzitter
Prof. dr. ir. L.J. van Vliet,	Technische Universiteit Delft, promotor
Prof. dr. J. Stoker,	Academisch Medisch Centrum, promotor
Dr. F. M. Vos,	Technische Universiteit Delft, copromotor

***Onafhankelijke leden:***

Prof. dr. J. Dankelman	Technische Universiteit Delft
Prof. dr. ir. M.J.T. Reinders	Technische Universiteit Delft
Prof. dr. J.M. Buhmann	Eidgenössische Technische Hochschule Zürich
Dr. K. Horsthuis	Vrije Universiteit Medisch Centrum
Prof. dr. A.M. Vossepoel	Technische Universiteit Delft en Erasmus Universiteit Rotterdam, reservelid



The work in this thesis was conducted at the Quantitative Imaging Group (QI), Faculty of Applied Sciences, Delft University of Technology.

The work was financially supported by VIGOR++ (European Union's Seventh Framework Program, No. 270379).

Cover design: R.E. Naziroglu

Printed by: Proefschrift AIO

Copyright © 2016 by R.E. Naziroglu, All rights reserved.

ISBN

A digital copy can be downloaded from [repository.tudelft.nl](http://repository.tudelft.nl).

# Contents

---

1	Introduction .....	1
1.1.	Assessment of Crohn’s disease severity by MRI.....	2
1.1.1.	Magnetic resonance imaging.....	5
1.1.2.	MRI-based features to assess Crohn’s disease.....	6
1.1.3.	Segmentation of the bowel wall.....	8
1.2.	Patient specific simulation of low-dose CT .....	8
1.3.	Objectives.....	9
1.3.1.	Bowel lumen .....	9
1.3.2.	Bowel wall segmentation and thickness measurements .....	10
1.3.3.	Quantifying Crohn’s disease activity .....	10
1.3.4.	A low dose CT simulator .....	10
1.4.	Outline of the thesis.....	11
2	An active contour algorithm for segmenting the bowel lumen from T1-weighted MR images.....	13
2.1	Introduction.....	14
2.1.1	Objective and approach.....	16
2.2	Preliminaries.....	17
2.3	Methods .....	18
2.3.1	Data driven term.....	19
2.3.2	Initialization .....	22
2.3.3	Updating the model parameters .....	24
2.3.4	Shape driven term and energy function .....	28
2.4	Materials .....	29
2.4.1	Patient preparation and MRI acquisition .....	29
2.4.2	Data annotation .....	30
2.5	Experiments and Results.....	31

2.5.1	Validating the consistency of the ratio between the intensities of fecal residue and water .....	32
2.5.2	Parameter calibration.....	33
2.5.3	Performance testing.....	35
2.5.4	Robustness towards initialization.....	42
2.6	Discussion.....	43
	Appendix .....	44
3	Semi-automatic bowel wall thickness measurement on MR enterography in patients with Crohn’s disease.....	49
3.1	Introduction.....	50
3.2	Methods .....	51
3.2.1	Data .....	51
3.2.2	Annotations .....	52
3.2.3	Semi-automatic thickness measurement .....	52
3.2.4	Evaluation measures.....	55
3.2.5	Statistical analysis .....	56
3.3	Results.....	57
3.3.1	Evaluation of the semi-automatic segmentations .....	57
3.3.2	Evaluation of the semi-automatic thickness measurement .....	59
3.4	Discussion.....	61
	Appendix I: scan protocol .....	64
	Appendix II: technical description of the segmentation algorithm.....	65
4	Semi-automatic assessment of the small bowel and colon in Crohn’s disease patients using MRI (the VIGOR++ project) .....	68
4.1	Introduction.....	69
4.2	Methods .....	70
4.2.1	Retrospective cohort.....	70
4.2.2	Prospective cohort.....	70
4.2.3	MRI protocol .....	71

4.2.4	Image analysis .....	72
4.2.5	Semi-automatic measurements.....	74
4.2.6	Reference standard .....	75
4.2.7	Model development.....	75
4.2.8	Validation .....	76
4.2.9	Statistical analysis .....	76
4.3	Results.....	77
4.3.1	Retrospective cohort.....	77
4.3.2	Prospective cohort.....	77
4.3.3	Model development.....	80
4.3.4	Model validation and comparison.....	81
4.3.5	Analysis on all evaluated segments .....	82
4.4	Discussion.....	82
	Appendix 1 .....	86
	Appendix 2: Prospective cohort: individual MRI features .....	86
5	Simulation of scanner- and patient-specific low-dose CT imaging from existing CT images.....	89
5.1	Introduction.....	91
5.1.1	Related Work.....	91
5.1.2	Objective .....	93
5.2	Lower-dose CT simulator.....	94
5.2.1	The virtual sinogram (Steps 1-2).....	95
5.2.2	Adding noise to the virtual sinogram.....	96
5.2.3	Reconstructing the image from the noisy sinogram (Steps 7-9). 100	
5.3	Parameter estimation .....	101
5.3.1	The reconstruction filter coefficients .....	101
5.3.2	The bowtie filter .....	104
5.3.3	The X-ray tube output parameter and the readout noise level ....	106
5.4	Results.....	108



5.4.1	Measurement data .....	108
5.4.2	Parameter estimation .....	110
5.4.3	Validation of the low-dose CT model .....	118
5.5	Discussion .....	126
6	Conclusion .....	129
6.1	Conclusions .....	130
6.2	Future work on MRI-based assessment of Crohn's disease severity ...	132
6.3	Future work on simulation of low-dose CT .....	134
	References .....	135
	Summary .....	142
	Samenvatting .....	145
	Dankwoord .....	148
	Curriculum Vitae .....	151

# 1 Introduction

---

Cross-sectional medical imaging techniques have become indispensable in assessing abnormalities in the anatomy and physiology of the abdomen[1]. Ultrasonography (US), Computerized Tomography (CT) and Magnetic Resonance Imaging (MRI), Single Photon Emission Computerized Tomography (SPECT) and Positron Emission tomography (PET), are minimal invasive and able to image any part of the abdomen in 3D[2]. Nuclear imaging modalities SPECT and PET are mainly used to examine the pathophysiology, but needs to be combined with CT for high-resolution imaging of the anatomy. Instead, US is often the most cost-effective technique, most portable and has the highest temporal resolution of all the mentioned techniques, but is unable to produce a 3D volume of the full abdomen without complicated reconstruction algorithms. In addition to this, sound waves are unable to pass air (such as air-filled bowel), hence US is unable to image beyond pockets containing air and its performance highly depends on the operator.

Alternatively, MRI and CT are relatively costly with respect to US, but they provide 3D high-resolution image volumes. Furthermore, MRI offers excellent soft-tissue contrast by a suite of scan sequences. Therefore, it is highly appropriated for abdominal imaging, such as of the bowel in patients suffering from Crohn's disease[3]. Whereas MRI offers superior soft-tissue contrast and a much larger versatility, CT has an excellent acquisition speed, a slightly higher spatial resolution. A disadvantage of CT is that it comes at the cost of using ionizing radiation. The latter prohibits its application to clinical conditions that require frequent imaging[4], such as Crohn's disease. Recently, an increasing amount of evidence was given that the exposure can be reduced while maintaining the same diagnostic accuracy[5, 6]. For each diagnostic task, one would be interested to know the minimum amount of radiation exposure before the accuracy deteriorates due to noise and artefacts. Unfortunately, this minimum required radiation dose can only be determined by either scanning a patient multiple times or by simulations of a low-dose scan based on high-dose retrospective data.

This thesis addresses two specific and unrelated challenges concerning cross-sectional abdominal imaging. The first part of the thesis investigates the application of MRI and image processing techniques to assess Crohn's disease activity. The second part presents a realistic low-dose CT-image simulator to facilitate in-silico dose optimization based on available "high-dose" data sets.

## **1.1. Assessment of Crohn's disease severity by MRI**

Crohn's disease is a chronic inflammatory disease of the gastrointestinal tract with a relatively high incidence in the Western world, affecting 700 thousand individuals in Europe alone[7]. Familial and epidemiological studies have stressed the involvement of genetic factors and have also shown the critical role of environmental factors in the development of IBD. It is now believed that intestinal microbes may invoke a disrupted inflammatory response in a genetically susceptible host. This inflammatory response can cause abdominal pain, fever, diarrhea and/or clinical signs of bowel obstruction to the patients. The disease is characterized by a chronic relapsing and remitting course, i.e. periods of exacerbations are alternated by episodes of diminished disease activity. Accordingly, the mere presence of the symptoms that might be related to the disease (other diseases might cause similar symptoms) or sequels of the disease (e.g. strictures) must be distinguished from active disease, which can occur at varying levels of severity and specifying the level of disease activity is necessary for choosing an appropriate treatment strategy[8]. Moreover, determining disease activity is important for monitoring treatment response and following the course of the disease activity over time as well, because the score objectively quantifies any changes over time. Consequentially, the efficacy of treatments can be assessed and compared objectively with each other. Currently, several disease activity scoring systems are available that can be used by gastroenterologists.

Ideally, any activity score should be objective, reproducible, quantifiable, non-invasive and comprehensive. Currently, the disease activity is assessed by imperfect scores such as Crohn's Disease Activity Index (CDAI)[9], Crohn's Disease Endoscopic Index of Severity (CDEIS)[10] and D'Haens index[11]. CDAI incorporates the patient's perception of the disease and therefore is not objective for determining the activity of the disease (patients' well-being might be influenced by many factors). CDEIS requires ileocolonoscopy, which necessitates extensive bowel purgation, is invasive, and only permits inspection of the bowel's lumen-surface. D'Haens index scores histological abnormalities in biopsy specimens, thereby focusing solely on microscopic aspects of Crohn's disease. Table 1.1 summarizes the limitations for the described disease indices. The complex nature of Crohn's disease and the poor correlation between the disease activity and the symptoms hinder the development of an improved index[3, 12]. Magnetic resonance imaging (MRI) has the potential to overcome these limitations[1, 13]. First, the technique is non-invasive. Second, it produces

3D image volumes in which the entire abdomen and particularly the full bowel wall is visualized.

Recently, several MRI disease activity scores incorporating a variety of these features have subsequently been developed and externally validated[14-17], and are slowly disseminating in clinical practice. The Magnetic Resonance Index of Activity (MaRIA) for example has been developed using the CDEIS as reference and incorporates quantitative measurement of relative bowel wall contrast enhancement along with subjective evaluation of the presence of mural ulceration and abnormal T2 signal[14]. Other indices, such as the London score also rely on qualitative grading of various features by reporting radiologists[16, 17]. Before widespread adoption for evaluating disease activity and therapeutic monitoring, MRI activity scores must be demonstrably accurate across the spectrum of disease severity, and reproducible between radiologists. The current literature, however, reports variable interobserver agreement for many features used in MRI activity scores[17, 18]. Moreover, although MRI shows high accuracy for severe disease activity (91% accuracy), diagnostic performance drops for mild disease or disease in remission (62% accuracy)[19]. Accordingly, the Virtual Gastrointestinal Tract (VIGOR++) project aims to overcome these drawbacks by proposing a better score for assessing Crohn's disease activity at MRI.

Index\requirement	Objective	Reproducible	Quantifiable	Noninvasive	Comprehensive
CDAI	-	+/-	-	+	-
CDEIS	+/-	+	-	--	-
D'Haens	+/-	+	-	--	-
VIGOR++	+	+	+	+	+

Table 1.1 Synopsis of Crohn's disease indices. A qualitative comparison between the different scores show that VIGOR++ is the most promising score to assess the severity of Crohn's disease. Notice that CDAI, CDEIS and D'Haens are existing scores. The VIGOR++ score is the aim of the project (Table from [www.vigorpp.eu/facts.php](http://www.vigorpp.eu/facts.php)). - is bad, +/- is okay and + is good.

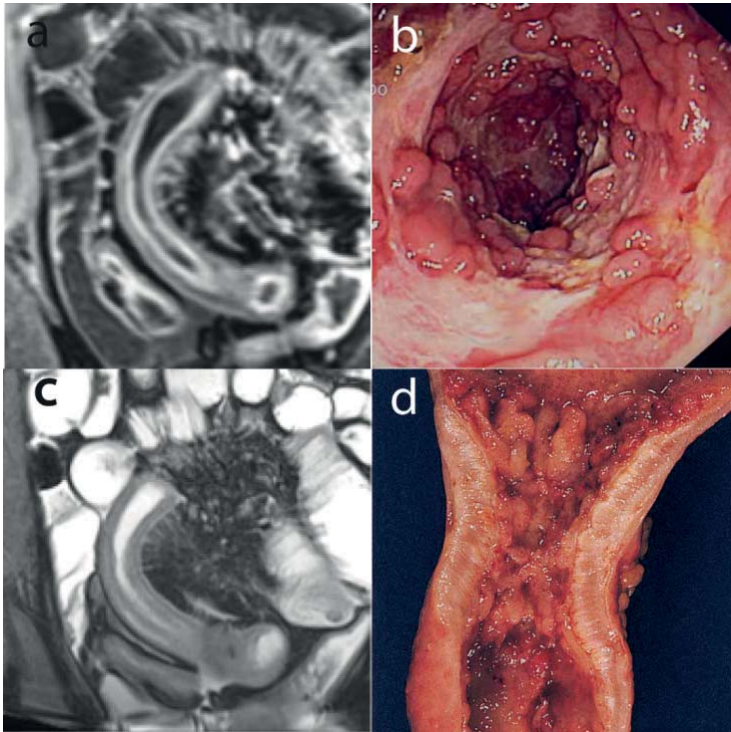


Figure 1.1 [20] A patient with severe Crohn's disease on MRI and colonoscopy. a) A coronal contrast enhanced fat saturated T1-weighted volume interpolated breath hold imaging (VIBE) sequence shows a diseased small bowel segment. The bowel wall is thickened and a layered enhancement pattern is clearly visible. Note that especially the mucosa is enhanced after administration of contrast agent. b) Colonoscopy of the same bowel segment as in a). The bowel surface shows large ulcerations. c) A coronal Balanced Steady State Free Precession sequence (true-FISP, a combination of T1 and T2 contrast) depicting a diseased bowel segment. Note that figure (c) shows the same region as figure (a). d) The diseased bowel segment that was removed by surgery. In contrast to colonoscopy the transmural properties, such as fibrosis, are clearly visible.

### 1.1.1. Magnetic resonance imaging

Magnetic resonance imaging is a medical imaging technique that is widely used in radiology to investigate abnormalities in the anatomy and physiology of patients[21]. MRI images a volume by exciting the Hydrogen nuclei in the body by a carefully timed sequence of radiofrequency pulses while applying magnetic gradients. The contrast highly depends on the selected sequence and the local environment of the hydrogen nuclei. Accordingly, MRI sustains the acquisition

of a number of sequences with different contrasts by adjusting the sequence parameters (see e.g. Figure 1.2).

The patients included in the VIGOR++ study were examined by a multi-sequence protocol. This gives the radiologists the ability to determine a large number of different features that are used to determine Crohn's disease activity. The image processing and image analysis techniques presented in this thesis solely focus on the T1-weighted Gradient echo sequences as it is the only sequence with both a high isotropic resolution and a strong soft-tissue contrast after intravenous administration of Gadolinium.

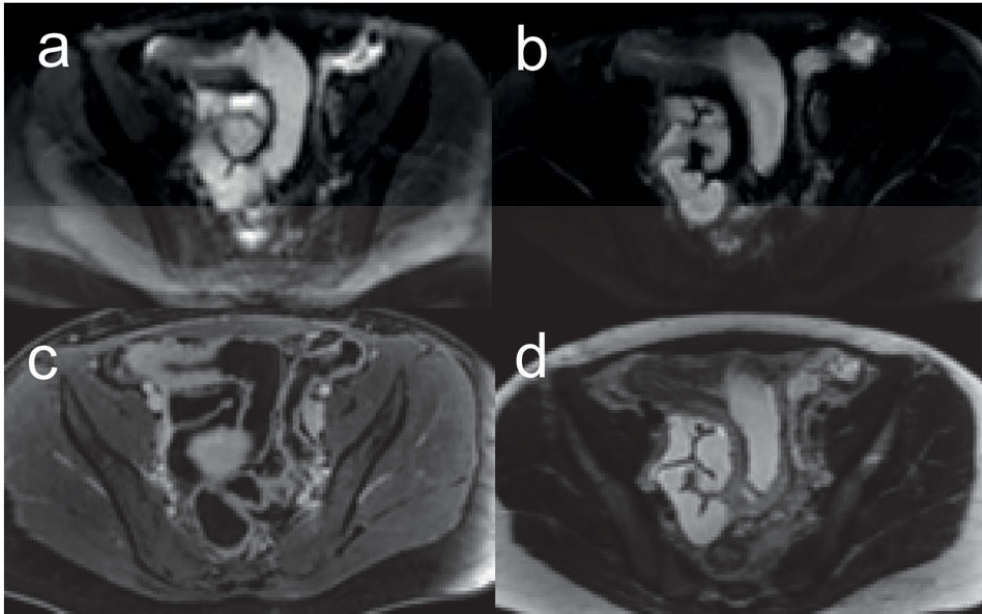


Figure 1.2: Axial slices of the abdomen acquired of the same patient using different MRI sequences. Note that the contrast can be drastically changed by simply modifying the scan sequence. a) A diffusion-weighted image. b) T2-weighted Single-Shot Fast Spin-Echo with fat suppression. c) T1-weighted Spoiled Gradient Echo sequence after intravenous contrast medium. d) T2-weighted Single-Shot Fast Spin-Echo.

### 1.1.2. MRI-based features to assess Crohn's disease

Currently, radiologists use a wide variety of MRI features to grade the disease severity. For instance, Ziech et al reported a total of 18 different features that were used in grading[22]. Some of these features are extra-mural such as lymph-node

enhancement, comb sign and creeping fat, but the largest part of the features are derived directly from the bowel wall itself. A further classification can be made between features describing the geometry of the bowel (e.g. wall thickness) and those quantifying the signal intensity of the wall (e.g. T1-enhancement and T2-signal intensity). Today, these features are more and more frequently weighted into severity scores consisting of two to up to five grades.

Recently, multiple studies used multivariate analyses to correlate manually determined parameters that were extracted from MRI data to other modalities, for instance the Crohn's Disease Endoscopic Index of Severity (CDEIS) [14] and endoscopic biopsy acute inflammatory score (eAIS) [16]. Rimola et al. found that parameters wall thickening, wall signal intensity, and relative contrast enhancement provided a good correlation with the endoscopic CDEIS ( $r = 0.82$ ) [14]. Steward et al. reported a correlation of 0.72 between their MRI index and the eAIS at histopathology [16]. However, MRI showed to be accurate for severe disease cases (91% accuracy), but mediocre for mild disease and remission (62% accuracy) [19]. Furthermore, a recent study by Ziech et al. reported a weak to moderate inter-observer agreement for most of the manually acquired MRI-based features [18]. Another study by Tielbeek et al. confirmed that the inter-observer agreement was weak to moderate for most features [17]. They also found that certain features (e.g. wall thickness) and the actual score itself showed a higher correlation among experienced radiologists than among other groups.

The reproducibility of the features is essential for reliable disease monitoring and treatment planning [3]. Unfortunately, measuring features of the bowel wall is far from trivial. Many features are measured locally at a few handpicked representative points, while the bowel is examined at a segment level (e.g. terminal ileum). Features such as thickness and T1-enhancement vary widely within a diseased bowel segment. Increasing the number of points or averaging over the entire annotated volume is more time-consuming. Nonetheless, techniques that base their measurement on a larger volume are preferred over ones based on a handful of points. An automated method to delineate diseased bowel wall segments could provide a first step for such a feature measurement scheme. Subsequently, an array of features might be computed from the delineated structure. Particularly, features describing the geometry of the bowel such as thickness and length can be derived directly after accurate delineation. The geometrical features such as bowel wall thickness and stratification are commonly



used in grading Crohn's disease [22] and are explicitly used in the methods found by [14, 16, 17].

### **1.1.3. Segmentation of the bowel wall**

The goal of image segmentation is to partition an image into a disjoint set of informative sub domains. In medical imaging analysis, segmentation generally aims to yield the delineation of anatomical structures the user is interested in. Unfortunately, the segmentation problem is often ill-posed, as a typical scene or data volume contains many different potentially relevant ways to partition the image. The task at hand, the properties of the dataset and prior information (shape, intensity, manual interaction) can make the problem tractable.

Generally, segmentation algorithms assign a certain cost/energy to each possible partitioning. The mapping captured by the energy-function is chosen in such a manner that its (preferably global-) minimum corresponds to the desired segmentation (partitioning). Thus, the energy-function must explicitly model prior knowledge of the segmentation problem like the regional statistics of the different image regions and the shape characteristics. Next to defining an energy function, a segmentation algorithm consists of an optimization scheme. Many of these schemes have been developed and some of the most popular are: classification based, graph-based, atlas-based, active shape model ASM and active contours. Each of these methods has its advantages and disadvantages and the application is decisive regarding what optimization scheme should be used.

In this thesis, the active contours approach was chosen as it offers the following advantages: 1. The geometry and topology of the bowel wall is easily preserved (during evolution), which is necessary for robustly measuring geometrical features of Crohn's disease activity such as thickness and length. 2. No extensively labelled training set is required to learn the shape of the bowel, while it can have many degrees of freedom. 3. The background does not need to be defined a-priori.

## **1.2. Patient specific simulation of low-dose CT**

Computed tomography (CT) has established itself as one of the most important medical imaging modalities [4]. In fact, the number of CT examinations is still increasing [23]. A CT-scanner acquires X-ray images under many angles and reconstructs from these images a 3D volume. A particular disadvantage of CT, however, is the exposure to ionizing radiation that is inherent to the technique.

Accordingly, it is a necessity to keep the radiation dose as low as reasonable achievable (ALARA). Unfortunately, lowering the dose yields a lower signal-to-noise ratio and thus a poorer image quality, which may hamper subsequent diagnosis. Optimization of the dose/quality trade-off is a far from trivial problem as one cannot simply expose subjects to a range of radiation doses for ethical reasons. Therefore, a lower-dose CT image is usually simulated by adding noise to the underlying projection data, i.e. the sinogram[24-26]. Subsequently, the lower-dose image is reconstructed from these noisy projections using the scanner's software. However, this approach is often not achievable in practice as the projection data are not routinely saved. Consequently a method to simulate low-dose CT images from the reconstructed high-dose images would be highly valuable since it is more generally applicable. Additionally, retrospective studies and cross-institutional scanner studies would become possible.

### 1.3. Objectives

The objective of this thesis was to develop tools for abdominal radiology and it consists of two separate parts. In the first part, the development of novel image analysis tools is described for grading Crohn's disease activity using MRI. This was one of the goals in the VIGOR++ project. Here, the focus is on obtaining image-based descriptors based on MRI data that are accurate, robust and reproducible. In the second part, a method is presented to realistically simulate low-dose CT-images from reconstructed high-dose images.

#### 1.3.1. Bowel lumen

Segmentation of the bowel's lumen surface is an important step to automatically assess Crohn's disease activity based on magnetic resonance (MR) images. Segmenting the lumen is challenging due to: (1) the large variation in bowel shape and diameter, e.g. the lumen diameter varies from 0 mm in stenosis to more than 75mm for widened segments, (2) partial volume effects, i.e. the bowel wall thickness is on the order of the voxel size, (3) spatially varying signal intensities due to magnetic field heterogeneities, (4) a highly irregular, variable and complex exterior, and (5) a heterogeneous lumen-content. The proposed method is based on an active contour initialized by a manually annotated centerline through the lumen of a diseased segment. The method employs a multi-resolution strategy. This multi-scale approach is necessary to overcome the challenging aspects (1), (2) and (3). Furthermore, we model the intensity inside a neighborhood by a

spatially-variant function instead of a single value to accommodate various material compositions. This offers the flexibility to solve challenges (4) and (5).

### **1.3.2. Bowel wall segmentation and thickness measurements**

An important goal is to evaluate a semi-automatic method to measure the bowel wall thickness in order to enhance the reproducibility of this feature. We have implemented an active contour approach to semi-automatically segment both the bowel wall's inner and outer surface, taking into account the heterogeneous bowel content. Our algorithm uses as input a (small piece of) centerline through the gastrointestinal tract that completely passes by the diseased part of interest. Subsequently, we derive the bowel wall thickness from the acquired segmentation and validated these measurements with respect to manual measurements.

### **1.3.3. Quantifying Crohn's disease activity**

The goal is was to develop and validate a predictive MRI model for enteric Crohn's disease activity incorporating novel software assisted semi-automatic measurement of MRI features using an ileocolonoscopy standard of reference, and to compare performance with existing MRI activity scores. Introducing these semi-automatic measurement to scoring indices raises novel questions: For instance, does a scoring system consisting of automatic features outperforms one consisting of manual measurements? Or does a hybrid model combining automated and manual features provide the best outcome? Hence, the aim is to develop and validate a model based on manual and automated MRI features for grading active Crohn's disease in the small bowel and colon using ileocolonoscopy as reference standard. Here, the automated features are derived from bowel wall segmentations and include geometry-based features such as thickness, length, excess volume and enhancement based features (e.g. A1[27]).

### **1.3.4. A low dose CT simulator**

Simulating low-dose Computed Tomography (CT) facilitates in-silico studies into the required dose for a diagnostic task. Conventionally, low-dose CT images are created by adding noise to the recorded projection data. However, this is not always achievable in practice as the raw data are simply not available. This thesis aims to present a new method for simulating patient-specific, low-dose CT images without the need of the original projection data. The methods assume fan beam imaging and image reconstruction by parallel-beam filtered-backprojection merely to proof the principles of our method and to show that representative noise distributions are obtained.

The low-dose CT simulation method included the following steps: (1) computation of a virtual sinogram from a high-dose CT image by means of the radon transform; (2) simulation of an 'reduced'-dose sinogram with appropriate amounts of noise; (3) subtraction of the high-dose virtual sinogram from the reduced-dose sinogram; (4) reconstruction of a noise volume via filtered back-projection; (5) addition of the noise image to the original high-dose image. The required scanner-specific parameters were retrieved from calibration images of a water cylinder. The apodization window was estimated from the noise power spectrum (NPS) in a small region of interest in the center of those images. Furthermore, the bowtie filter attenuation characteristics were derived from the pixel variance. Finally, the X-ray tube output parameter (reflecting the photon flux) and the detector read-out noise were computed from the pixel variance at various exposure levels.

We believe that unifying the calibration- and simulation-procedure in a single framework, strengthens our approach. The technique enables scientists, clinicians and engineers to simulate low-dose CT-image without the need of the raw projection data. Therefore, it will become easier and less cumbersome to conduct multi-center and cross-scanner retrospective studies on CT radiation dose

## 1.4. Outline of the thesis

The content of this thesis contains two parts, namely quantification of Crohn's disease activity using MR-enterography (chapters 2, 3 and 4) and the simulation of realistic low-dose CT images (Chapter 5).

In **Chapter 2**, a novel region-based active contour model was introduced to segment the bowel lumen in contrast-enhanced MRI given a centerline as an input. The method incorporates prior knowledge of the content of the bowel lumen and the structures of the background.

In **Chapter 3**, the active contour algorithm presented in Chapter 2 was extended to segment the bowel wall. The outer surface was not as well defined as the inner-surface as the background consists of structures with a similar intensity as the bowel wall. Furthermore, the bowel wall thickness was derived from these wall segmentations and compared to the manual measurements of three radiologists.

In **Chapter 4**, image-based descriptors of the bowel wall are derived from the segmentations like wall thickness, disease length, surface and volume. These

features were incorporated together with manual features into a multi-variate model to predict CDEIS. An extensive search was conducted to select the optimal model and a comparison was made whether adding automated features would improve performance. This linear model was one of the end-points of the VIGOR++ project.

In **Chapter 5**, a novel method was introduced and extensively validated to realistically simulate low-dose CT-images based on an existing high-dose image. Additionally, procedures were developed to measure scanner-specific parameters from the set of calibration images of a water-cylinder phantom.

Finally in Chapter 6, the merits and limitations of this thesis are discussed.

## 2 An active contour algorithm for segmenting the bowel lumen from T1-weighted MR images

---

### Abstract

Segmentation of the bowel's lumen surface is an important step to automatically assess Crohn's disease activity based on magnetic resonance (MR) images. This is challenging due to MR signal variations, heterogeneous lumen content, the presence of stenosis, and the diversity of the surrounding tissues. We present a region-based active contour algorithm that can handle these challenges by incorporating prior knowledge about properties of the lumen content (e.g. the stratified appearance) and knowledge of the adjacent anatomy in a space-variant regional mixture model. This information was added to the model by estimating material fractions inside each voxel. A level set representation employing a gradient-descent scheme was used to obtain the actual segmentation. Our method was tested on 61 regions from 59 patients presenting Crohn's disease activity. Two expert radiologists delineated these regions. The average distance between our segmentation and each of the two annotators was  $1.43 \pm 0.55\text{mm}$  and  $1.48 \pm 0.68\text{mm}$ , which was similar to the inter-observer variability  $1.52 \pm 0.70\text{mm}$ . Furthermore, a Wilcoxon-signed rank test showed no significant difference ( $p > 0.7$ ) between our segmentation and the manually obtained ones. Finally, a different initialization yielded a mean distance between the segmentations of  $0.39 \pm 0.31\text{ mm}$ . The segmentations obtained by our method will ease the measurement of features such as the bowel wall thickness.

## **2.1 Introduction**

Crohn's disease is a chronic inflammatory disease of the gastrointestinal tract with a relatively high incidence in the Western world[7]. Grading Crohn's disease activity is important for monitoring and treatment planning[8]. Traditionally, the disease activity is assessed by ileocolonoscopy, which requires extensive bowel purgation, is invasive, and only permits inspection of the bowel's lumen-surface. Magnetic resonance imaging (MRI) has the potential to overcome these limitations[1, 13]. Rimola et al. used multivariate analyses to correlate manually determined parameters that were extracted from MRI data to the Crohn's Disease Endoscopic Index of Severity (CDEIS)[14]. They found that parameters such as wall thickening, wall signal intensity, and relative contrast enhancement provided a good correlation with CDEIS ( $r = 0.82$ ). However, two recent studies by Ziech et al. and Tielbeek et al. reported a weak to moderate inter-observer agreement for most of the manually acquired MRI-based features[18],[28]. Furthermore, MRI showed to be accurate for severe disease cases (91% accuracy), but mediocre for mild disease or remission (62% accuracy)[19].

To overcome the aforementioned limitations, a (preferably automatic) computer-aided system is needed that sustains accurate and reproducible grading of Crohn's disease severity over the entire range of disease activity[29, 30]. To do so, automatic segmentation of abnormal bowel wall is an important first step. Segmentation is a challenging task, because it is hampered by inhomogeneous lumen intensities, space-varying wall thickness, local signal enhancement, and a wide variety of surrounding tissues.

Previously, bowel lumen segmentation algorithms have been integrated in computer-aided detection systems for finding polyps in computed tomography colonography (CTC). The solutions can be differentiated in methods for patients with extensive [31] [32] and limited [33-36] bowel preparation. To the best of our knowledge, such algorithms have not been developed to segment the lumen in magnetic resonance colonography (MRC). Presumably, this is because MRC has a poorer spatial resolution, a weaker contrast-to-noise ratio, suffers from signal fluctuations (e.g. the bias field) and other imaging artifacts[37]. For similar reasons, the CTC algorithms are not applicable to segment the bowel lumen in MR images of patients with Crohn's disease.

Recently, a method was described to automatically segment the small bowel from CT-images with the purpose of detecting carcinoma's[38]. The patient preparation consisted of administration of an oral and an intravenous contrast agent to facilitate segmentation of the small intestine based on the mesenteric vasculature's anatomy. A full distention of the bowel lumen was not required, which is necessary though for assessment of Crohn's disease severity. Alternatively, a method was developed to automatically detect and segment diseased pieces of bowel wall from MRI volumes through a machine learning approach[39, 40]. Although this approach aids in the detection of the disease, clinicians still need to manually measure relevant features, such as the wall thickness and the length of the diseased segment, for severity assessment. Essentially, a lumen segmentation algorithm that accurately captures the full geometry of the bowel is required to obtain such geometric information.

Active contours enable to segment objects while implicitly adapting the topology. They were initially introduced by Kass et al.[41], after which Osher et al. proposed an embedding in level sets[42]. Active contours have evolved into a powerful segmentation technique in medical image analysis[43]. Applications of level sets can be partitioned into approaches relying on edge detection [44-47] and methods employing differences in regional statistics[48-51]. Statistical approaches to the application of level sets were described in[52-56]. Initially, [49] proposed to segment an object from the background based on its mean intensity. Such an approach is insufficient, however, when the image intensity is distorted by slowly varying intensity fluctuations (e.g. B0-field inhomogeneities in MRI).

Mumford and Shah proposed a framework to partition an image into piecewise, smoothly varying regions[57]. Although such a model can handle smooth intensity fluctuations, finding an efficient implementation proved to be cumbersome[50]. Alternatively, several recent methods measured the mean intensity locally, inside a neighborhood[50, 58-60]. In such a way slowly varying intensity fluctuations can be handled as long as the intensity is *locally* constant. Particularly, Brox et al [50] showed that such an approach is a first order, smooth approximation to the Mumford-Shah model. This local means method was extended by adding the variance as a descriptor[50, 61]. Particularly, the mean and the variance were empirically determined in[50], while this method was shown to be equivalent to locally fitting a Gaussian distribution[61].



Obviously, a single Gaussian distribution cannot model large intensity discontinuities associated with multiple anatomical structures, either in the foreground or background. This problem may be dealt with using multiple smaller regions to yield a piecewise constant model. For instance, Vese and Chan [48] presented a method to segment  $N$  regions using multiple level sets and indicator functions, i.e. one for each region. A far from trivial challenge that remains, however, is to merge the different regions to obtain the object of interest. In our case, we are interested in segmenting the bowel's inner surface. The lumen consists of a mixture of air, water and fecal remains. Unfortunately, many anatomical structures *surrounding* the bowel have a similar intensity distribution as the materials found *inside* the lumen, which complicates a merging procedure. Additionally, the number of regions that is necessary to model the background is a-priori unknown.

Strong discontinuities can also be modeled by directly estimating the posterior probability density function of the underlying data. Both parametric (e.g. mixture of Gaussians model) [62, 63] and non-parametric methods (e.g. histogram) [64-66] have been suggested. Unfortunately, each material that is modeled by this probability density function influences the discriminatory power of the other materials, as the integral of the density function needs to be one. This may become problematic whenever two materials with slightly overlapping intensity need to be segmented, while one of the two has a low prevalence in a mixture of materials.

The most intuitive technique for bowel lumen segmentation perhaps is to take a multi-region segmentation approach ([67] contains an elaborate overview of such methods). However, it is not possible to directly use these techniques due to the large inhomogeneities in signal intensity. Essentially, we propose a solution to this problem.

### **2.1.1 Objective and approach**

This chapter presents a novel method to segment the bowel lumen from T1-weighted images of patients with Crohn's disease. The method is based on an active contour initialized by manually annotated centerlines through the lumen of a diseased segment. Segmenting the lumen is challenging due to: (1) the large variation in bowel shape and diameter, e.g. the lumen diameter varies from 0 mm in stenosis to 75mm for widened pieces, (2) partial volume effects, i.e. the bowel

wall thickness is on the order of the voxel size, (3) spatially varying signal intensities due to magnetic field inhomogeneities, (4) a highly irregular, variable and complex exterior, and (5) a heterogeneous lumen-content. The proposed method employs a multi-resolution strategy. This multi-scale approach is necessary to overcome the challenging aspects (1), (2) and (3) (see Figure 2.1a). Furthermore, we model the intensity inside a neighborhood by a spatially-variant function instead of a single value to accommodate various material compositions. This offers the flexibility to solve challenges (4) and (5) (see Figure 2.1b).

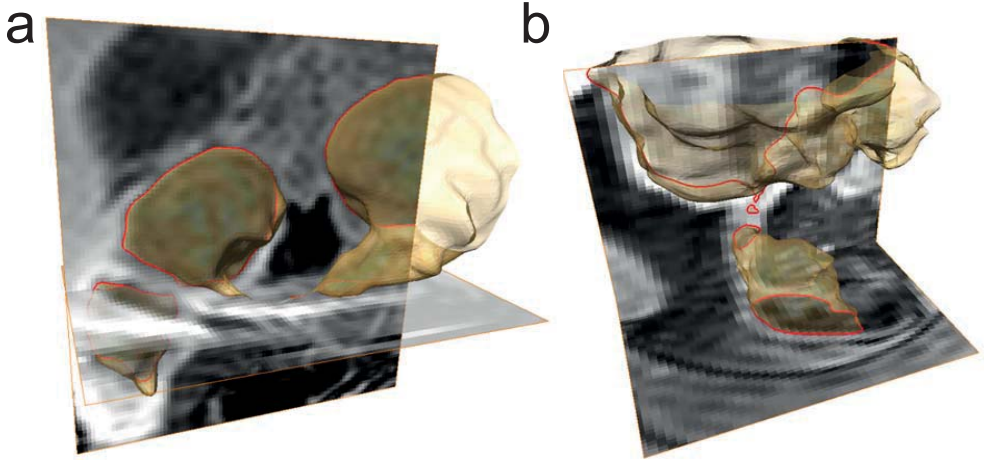


Figure 2.1. Two regions of interest focusing on diseased pieces of bowel and showing the segmentation of the bowel lumen by our method. Notice that the shape of the lumen differs dramatically between these regions.

## 2.2 Preliminaries

Previously, several active contour approaches were presented to cope with smooth signal inhomogeneities[59, 60]. Let  $\Omega \subset \mathbb{R}^D$  denote the  $D$ -dimensional image domain,  $I(x): \Omega \rightarrow \mathbb{R}$  the image itself and  $\Omega_i$  a partitioning of  $\Omega$  into  $N$  regions with  $\Omega = \bigcup_N \Omega_i$  and  $\Omega_k \cap \Omega_l = \emptyset$ . Most techniques essentially model the intensity in each region  $\Omega_i$  by a smooth function[50, 68]. The Local Binary Fitting (LBF) model[58] searches for the partitioning that minimizes the following energy function

$$E_{LBF} = \int_{\Omega} \sum_{i=1}^N \lambda_i \int_{\Omega} K_{\rho}(\mathbf{x} - \mathbf{y}) (I(\mathbf{y}) - \mu_i(\mathbf{x}))^2 M_i(\mathbf{y}) d\mathbf{y} d\mathbf{x}, \quad (2.1)$$

where  $\mathbf{x}$  is an image coordinate,  $\mathbf{x} - \mathbf{y}$  a neighborhood coordinate and  $K_\rho(\mathbf{x} - \mathbf{y})$  a weight function (e.g. a Gaussian) imposing locality around voxel  $\mathbf{x}$ . The summation is over a presumed partitioning of ( $N$ ) image regions;  $\lambda_i$  is a weight related to the importance of the region denoted by label  $i$ , and  $M_i$  an indicator function which is one if  $\mathbf{y} \in \Omega_i$  and zero otherwise. The weighted average intensity  $\mu_i(\mathbf{x})$  in a neighborhood of size  $\rho$  is defined by

$$\mu_i(\mathbf{x}) = \frac{\int_{\Omega} K_\rho(\mathbf{x} - \mathbf{y}) I(\mathbf{y}) M_i(\mathbf{y}) d\mathbf{y}}{\int_{\Omega} K_\rho(\mathbf{x} - \mathbf{y}) M_i(\mathbf{y}) d\mathbf{y}}. \quad (2.2)$$

Notice that the intensity  $I(\mathbf{y})$  in a neighborhood  $\rho \cap \Omega_i$  around  $\mathbf{x}$  is modeled by a constant  $\mu_i(\mathbf{x})$  in (2.1), hence implicitly assuming that the intensity inside this neighborhood is constant. The initial LBF model ignores the variance, which was later added by Wang et al[61]. Assuming  $N = 2$  and  $\lambda_i = 1$  we get

$$E_{Wang} = \int_{\Omega} \sum_{i=1}^2 \int_{\Omega} K_\rho(\mathbf{x} - \mathbf{y}) \times \left( \frac{(I(\mathbf{y}) - \mu_i(\mathbf{x}))^2}{2\sigma_i^2(\mathbf{x})} + \log \sigma_i(\mathbf{x}) \right) M_i(\phi(\mathbf{y})) d\mathbf{y} d\mathbf{x}, \quad (2.3)$$

where  $\phi(\mathbf{y})$  represents a level set function that is defined to be negative for all  $\mathbf{y}$  inside the lumen and positive elsewhere, and  $\sigma_i^2(\mathbf{x})$  is the weighted variance defined by

$$\sigma_i^2(\mathbf{x}) = \frac{\int_{\Omega} K_\rho(\mathbf{x} - \mathbf{y}) (I(\mathbf{y}) - \mu(\mathbf{x}))^2 M_i(\mathbf{y}) d\mathbf{y}}{\int_{\Omega} K_\rho(\mathbf{x} - \mathbf{y}) M_i(\mathbf{y}) d\mathbf{y}}. \quad (2.4)$$

## 2.3 Methods

Figure 2.2 presents a flow diagram of the proposed method, which is initialized by a manually annotated centerline through a diseased section of the gastrointestinal tract.

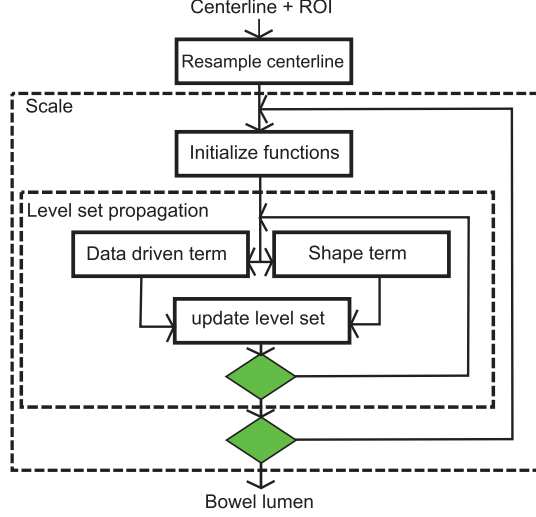


Figure 2.2. The flow diagram of the proposed method comprises two loops. The inner-loop implements the level set evolution, the outer-loop iterates over scales from coarse to fine

The method consists of a coarse to fine implementation of an active contour model. The energy is first optimized at the coarsest scale  $\rho$ , before progressing to the finer scales. The coarse to fine approach allows the algorithm to sense the lumen exterior over a large distance at a large scale whenever the lumen diameter is very large, while still being able to segment structures such as Haustral folds at the finest scale.

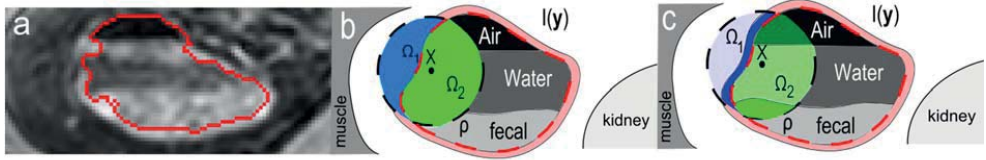


Figure 2.3. (a) A bowel segment containing air, water, and fecal residue. (b) A neighborhood  $\rho$  around  $x$  is conventionally partitioned into two regions  $\Omega_i$  (green/blue) in which the intensity is assumed constant. (c) The proposed method models each region by neighborhood varying functions (reflected in the color tone), instead of a constant.

### 2.3.1 Data driven term

The data driven part of the energy function is a modified version of (2.3). Importantly, the intensity  $I(\mathbf{y})$  in a neighborhood  $\rho$  around  $\mathbf{x}$  is modelled by a locally varying function  $\xi_i(\mathbf{x}, \mathbf{y})$ :

$$E_{data} = \int_{\Omega} \sum_{i=1}^2 \int_{\Omega} K_{\rho}(\mathbf{x} - \mathbf{y}) \times \left( \frac{(I(\mathbf{y}) - \xi_i(\mathbf{x}, \mathbf{y}))^2}{2\eta_i^2(\mathbf{x})} + \log \eta_i(\mathbf{x}) \right) M_i(\phi(\mathbf{y})) d\mathbf{y} d\mathbf{x} \quad (2.5)$$

where  $\xi_i(\mathbf{x}, \mathbf{y})$  denotes the modelled value at image coordinate  $\mathbf{y}$  in the neighborhood around  $\mathbf{x}$  and  $\eta_i^2(\mathbf{x})$  the squared residual w.r.t. the model (i.e. the local variance). Notice that  $\xi_i(\mathbf{x}, \mathbf{y})$  is now a function of image coordinates  $\mathbf{x}$  and  $\mathbf{y}$ . This is different compared to previous work[59-61], in which a slowly varying constant is used, i.e.  $\xi_i(\mathbf{x}, \mathbf{y}) = \mu_i(\mathbf{x})$  and  $\eta_i(\mathbf{x}) = \sigma_i(\mathbf{x})$ .

Essentially,  $\xi_i(\mathbf{x}, \mathbf{y})$  splits the inhomogeneity in signal intensity into two parts, which are modeled independently. The slowly varying inhomogeneities are modeled (as before) by a locally varying constant  $\mu_i(\mathbf{x})$ , while the abrupt transitions are modelled by a space-variant material fraction  $C_{i,m}(\mathbf{y})$ , predicting the fraction of a specific material  $\{m\}$  in that particular voxel. The signal intensity  $\xi_i(\mathbf{x}, \mathbf{y})$  of a voxel  $\mathbf{y}$  in the neighborhood around  $\mathbf{x}$  is represented by a volume-weighted sum of material specific contributions:

$$\xi_i(\mathbf{x}, \mathbf{y}) = \sum_{m=1}^L C_{i,m}(\mathbf{y}) \mu_{i,m}(\mathbf{x}), \quad (2.6)$$

where  $\mu_{i,m}(\mathbf{x})$  is the mean intensity of material  $\{m\}$  at location  $\mathbf{x}$ , and  $L$  the number of materials in compartment  $\Omega_i$ . Note that  $\mu_{i,m}(\mathbf{x})$  is similar to the mean intensity defined in (2.2). Additionally, the material fractions per volume add up to one. Figure 2.3 (middle) illustrates the difference between the mixture model for  $L=3$  and the single material per compartment approach as used before[59]. The neighborhood  $\rho$  around a voxel  $\mathbf{x}$ , depicted by the dashed circle, is split into two regions  $\Omega_1 \cap \rho$  and  $\Omega_2 \cap \rho$ , corresponding to the region inside, respectively outside the level set boundary. Previously, both regions would have been modeled by a single mean intensity  $\mu_i(\mathbf{x})$  and variance  $\sigma_i^2(\mathbf{x})$ , which are indicated in the middle drawing by the uniform colors. In the new approach, the signal is modeled by a mixture model  $\xi_i(\mathbf{x}, \mathbf{y})$  and a space-variant variance  $\eta_i^2(\mathbf{x})$ . This is indicated by the different tones of green and blue.

The next paragraphs explicitly describe how the intensity distribution of the bowel exterior and the bowel lumen are modeled using (2.6). Section 2.3.3 describes how the material fractions  $C_{i,m}(\mathbf{y})$ , the intensity distribution  $\mu_{i,m}(\mathbf{x})$  of a material  $\{m\}$  and the residual  $\eta_i(\mathbf{x})$  are determined in an iterative way.

### 2.3.1.1 Modeling the intensity distribution of the bowel exterior

The bowel exterior consists of many different tissues of which the intensities show large variations. We assume that the exterior can be locally partitioned into three materials (tissues) corresponding to dark (e.g. fat, air), medium (muscles, healthy bowel wall, thin arteries) and bright signal intensities (part of the liver, kidney, inflamed wall and large arteries). Accordingly, the intensity  $\xi_{ext}(\mathbf{x}, \mathbf{y})$  of a neighborhood voxel  $\mathbf{y}$  around voxel  $\mathbf{x}$  is modeled by

$$\xi_{ext}(\mathbf{x}, \mathbf{y}) = \sum_{m=1}^3 C_{ext,m}(\mathbf{y}) \mu_{ext,m}(\mathbf{x}), \quad (2.7)$$

where  $C_{ext,m}(\mathbf{y})$  is the fraction of material  $m \in \{\text{dark}, \text{medium}, \text{bright}\}$  in a voxel and  $\mu_{ext,m}(\mathbf{x})$  the mean intensity of material  $m$  outside the lumen.

### 2.3.1.2 Modeling the intensity distribution of the bowel lumen

Much in the same way,  $\xi_{lumen}(\mathbf{x}, \mathbf{y})$  is modeled by a volume-weighted sum of the constituting material intensities. We assume that the lumen consists of air, water, and fecal remains and that the intensity ratio between the different materials is locally constant. This implies that the lumen intensity can be written as the product of a weight and the reference intensity. By taking water as the reference material, the lumen intensity becomes

$$\xi_{lumen}(\mathbf{x}, \mathbf{y}) = w(\mathbf{y}) \mu_{water}(\mathbf{x}), \quad (2.8)$$

with

$$\begin{aligned} w(\mathbf{y}) &= \frac{\mu_{air}}{\mu_{water}} C_{air}(\mathbf{y}) + C_{water}(\mathbf{y}) + \frac{\mu_{fecal}}{\mu_{water}} C_{fecal}(\mathbf{y}) \\ &= \alpha_{air} C_{air}(\mathbf{y}) + C_{water}(\mathbf{y}) + \alpha_{fecal} C_{fecal}(\mathbf{y}) \end{aligned} \quad (2.9)$$

We dropped the subscript *lumen* from these equations just for the compactness of the notation. Again,  $C_k(\mathbf{y})$  denotes the fraction of material  $k$  and  $\alpha_k$  represents the

ratio between the local intensity of material  $k \in \{\text{air}, \text{water}, \text{fecal}\}$  and  $\mu_{\text{water}}(\mathbf{x})$ . Notice that  $w(\mathbf{y})$  enables us to vary  $\xi_{\text{lumen}}(\mathbf{x}, \mathbf{y})$ . Simultaneously,  $\mu_{\text{water}}(\mathbf{x})$  is allowed to fluctuate slowly, so that in turn the mean material intensities may vary due to field inhomogeneities and space-variant pick-up coil sensitivities.

We assert that  $\alpha_{\text{air}}$  is 0 as air voxels contain a negligible amount of protons. Furthermore,  $\alpha_{\text{water}}=1$  by definition. Essentially, the material fractions  $C_k(\mathbf{y})$ , the ratio's  $\alpha_k$  and the local intensity of water  $\mu_{\text{water}}(\mathbf{x})$  determine how well  $\xi_{\text{lumen}}(\mathbf{x}, \mathbf{y})$  approximates  $I(\mathbf{y})$ . The values for these parameters are also iteratively found (see 2.3.3). Particularly, the ratio between intensities of water and fecal residue is locally estimated from the data. As such, the assumption that this ratio is locally constant is not compromised by inherent differences in signal intensity (e.g. from patient to patient).

### 2.3.2 Initialization

To initialize the level set function, only a small track is drawn manually along a region supposedly affected by Crohn's disease (see Figure 2.4d). The user needs to label those points that pass through stenotic parts. This was efficiently implemented by using different mouse buttons. The algorithm removes the stenotic part from the initial centerline and reinserts it at the finest scale. The remaining centerline segments are dilated by a small structuring element of  $3 \times 3 \times 3$  voxels, after which the boundary serves to initialize the level set's signed distance transform: a negative distance to the contour denotes the inside and a positive distance the outside of the lumen.

Several parameters need to be initialized upon entering the scheme depicted in Figure 2.2. The parameters not set at this stage are defined during the update step (see below).

Initialization of the model parameters is visualized in Figure 2.4. A mixture of five Gaussians was fitted to the histogram of the entire image. The first peak of the histogram ( $\mu_1$ ) did not differ significantly from zero and indeed represented the signal intensity of air. Furthermore, the second ( $\mu_2$ ) and third ( $\mu_3$ ) peak roughly corresponded to fat and water, while the fourth ( $\mu_4$ ) one corresponded to fecal matter, respectively. The fifth peak usually resulted from the kidneys, arteries and the bladder.

The extra-luminal material fractions  $C_m(\mathbf{y})$  are fixated at

$$C_{dark} = \begin{cases} 1 & I(\mathbf{y}) < \mu_2 \\ 0 & \text{otherwise} \end{cases}, \quad (2.10)$$

$$C_{bright} = \begin{cases} 1 & I(\mathbf{y}) > \mu_5 \\ 0 & \text{otherwise} \end{cases}, \quad (2.11)$$

and

$$C_{medium}(\mathbf{y}) = 1 - C_{dark}(\mathbf{y}) - C_{bright}(\mathbf{y}), \quad (2.12)$$

which is simply a nearest mean classifier. We do not update these material fraction, but stick to the initializations. We have found that these initial estimates are very robust since they are derived from the entire volume. Subsequently, we initialize the luminal material intensities:  $\mu_{air} = 0$ ,  $\mu_{water} = \mu_3$  and  $\mu_{fecal} = \mu_4$ , after which we set  $\alpha_{fecal} = \mu_{fecal}/\mu_{water}$  and we initialize the luminal material fractions by:

$$\hat{C}_k(\mathbf{y}) = \begin{cases} 1 & \begin{aligned} & (I(\mathbf{y}) - \alpha_k \mu_{water}(\mathbf{y}))^2 \\ & < \min_{l \neq k} (I(\mathbf{y}) - \alpha_l \mu_{water}(\mathbf{y}))^2 \end{aligned} \\ 0 & \text{otherwise} \end{cases}. \quad (2.13)$$

Please notice the “hat” (^) above the variable  $C_k(\mathbf{y})$  indicating the initial value for the material fractions. The width of the first and largest peak of the histogram ( $\mu_1$ ) was also used to estimate the overall noise level  $\eta_{noise}$ . This parameter will be used in a later stage of the algorithm (see Table 2.3, (2.27)). Finally, a low-pass filter resembling the Point Spread Function (PSF) of the MRI system blurs the material fractions. We approximate the MRI system by a Gaussian kernel instead of the sinc-function, in order to avoid negative material fractions and to still be able to approximate the first lobe of the true PSF. Figure 2.4 illustrates the initialization procedure.



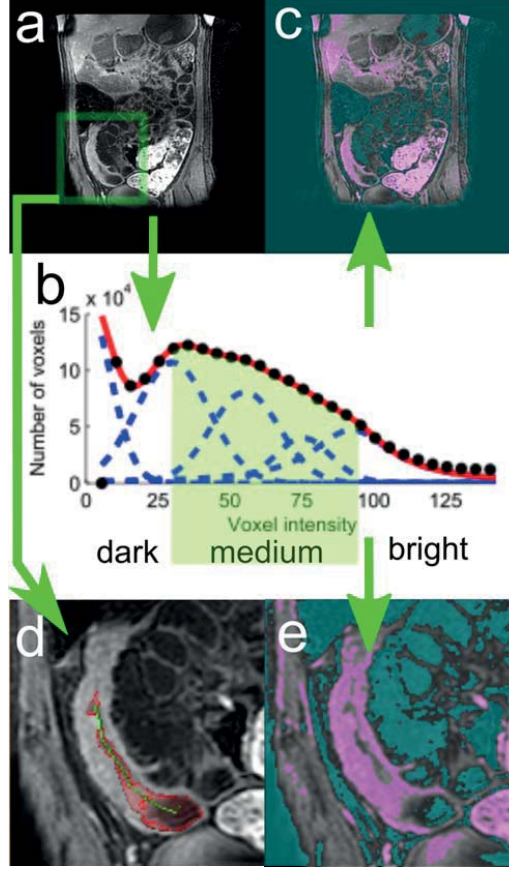


Figure 2.4. Illustration of the procedure to initialize the material fractions of the bowel's exterior. (a) A coronal T1-weighted MR enterography showing Crohn's disease in a manually annotated region-of-interest (ROI). (b) The intensity histogram of the full MR volume; black points denote the raw data, the red line the sum of five Gaussians fitted to the data, and the dashed blue lines each of the five Gaussians. The green shading indicates the segmentation thresholds. (c) Result of segmenting the slice into three areas based on the thresholds indicated in subfigure (b); bright voxels are masked with magenta, dark voxels with cyan and medium voxels are not masked. (d) The ROI from subfigure (a) containing the projected centerline in green and the final segmentation in (transparent) red. (e) Result of segmenting the ROI into three areas based on the fit of five Gaussians as in c.

### 2.3.3 Updating the model parameters

#### 2.3.3.1 Updating the model parameters for the bowel exterior

The mean material intensities  $\mu_{ext,m}(\mathbf{x})$  and the variance  $\eta_{ext}^2(\mathbf{x})$  are the only model parameters of the exterior that are updated. First,  $\mu_{ext,m}(\mathbf{x})$  is updated by

computing the derivative of  $E_{data}$  with respect to  $\mu_{ext,m}(\mathbf{x})$  and setting it to zero gives

$$\int_{\Omega} K_{\sigma}(\mathbf{x}-\mathbf{y}) C_m(\mathbf{y}) I(\mathbf{y}) H_{\varepsilon}(\varphi(\mathbf{y})) d\mathbf{y} = \int_{\Omega} K_{\sigma}(\mathbf{x}-\mathbf{y}) C_m(\mathbf{y}) \sum_{l=1}^3 C_l(\mathbf{y}) \mu_{ext,l}(\mathbf{x}) H_{\varepsilon}(\varphi(\mathbf{y})) d\mathbf{y}, \quad (2.14)$$

in which  $H_{\varepsilon}()$  represents the regularized Heaviside function, defined by

$$H_{\varepsilon}(x) = \begin{cases} 1, & z > \varepsilon \\ 0, & z < -\varepsilon \\ \frac{1}{2} \left( 1 + \frac{z}{\varepsilon} + \frac{1}{\pi} \sin\left(\frac{\pi z}{\varepsilon}\right) \right), & |z| \leq \varepsilon \end{cases} \quad (2.15)$$

and where (2.14) can be simplified by using the following vector notation:

$$I_m(\mathbf{x}) = \mathbf{w}_m(\mathbf{x}) \boldsymbol{\mu}_{ext}(\mathbf{x}), \quad (2.16)$$

where  $I_m(\mathbf{x})$  is defined by the left hand side of (2.14),  $\mathbf{w}_m(\mathbf{x})$  is a 1x3 vector with element  $l$  equal to

$$w_{ml}(\mathbf{x}) = \int_{\Omega} K_{\sigma}(\mathbf{x}-\mathbf{y}) C_m(\mathbf{y}) C_l(\mathbf{y}) H(\varphi(\mathbf{y})) d\mathbf{y}, \quad (2.17)$$

and  $\boldsymbol{\mu}_{ext}(\mathbf{x})$  is a 3x1 vector containing the material intensities (dark, medium and bright). Now, stacking the equations of the materials results in a linear system of three equations with three variables defined by

$$\mathbf{I}_{ext}(\mathbf{x}) = \mathbf{W}(\mathbf{x}) \boldsymbol{\mu}_{ext}(\mathbf{x}), \quad (2.18)$$

in which  $\mathbf{I}_{ext}(\mathbf{x})$  and  $\mathbf{W}(\mathbf{x})$  are 3x1 and 3x3 matrices. Solving this system yields the updated  $\mu_{ext,m}(\mathbf{x})$ . Subsequently, the exterior's variance is updated via

$$\eta_{ext}^2(\mathbf{x}) = \frac{1}{N_{ext}(\mathbf{x})} \int_{\Omega} K_{\rho}(\mathbf{x} - \mathbf{y}) \times \left( I(\mathbf{y}) - \sum_{m=1}^L C_m(\mathbf{y}) \mu_{ext,m}(\mathbf{x}) \right)^2 H(\phi(\mathbf{y})) d\mathbf{y}, \quad (2.19)$$

in which  $N_{ext}(\mathbf{x})$  normalizes over the integration volume of  $\mathbf{y}$

$$N_{ext}(\mathbf{x}) = \int_{\Omega} K_{\rho}(\mathbf{x} - \mathbf{y}) H(\phi(\mathbf{y})) d\mathbf{y} \quad (2.20)$$

Note that (2.19) is similar to (2.4) with the addition of the material fraction  $C_m(\mathbf{y})$ . Figure 2.5 illustrates the estimation of the mean material intensities of the bowel exterior.

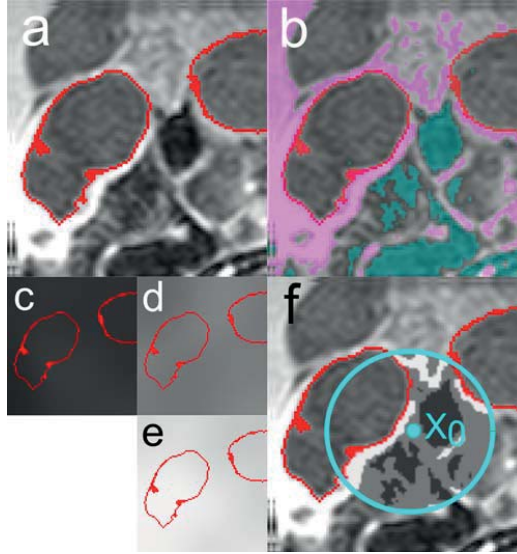


Figure 2.5. Modelling the materials of the bowel exterior; the segmented boundary is superimposed in red. (a) An ROI depicting a piece of diseased bowel. (b) The results of a voxel-wise classification into three materials, computed using (2.10)-(2.12); the three colours indicate dark (cyan), medium (none), bright (magenta). (c-e) The mean intensity  $\mu_{ext,m}(\mathbf{x})$  of the dark structures (c), the medium intensity structures (d), and the bright structures (e) computed by solving (2.18). (f) The modelled intensity  $\xi_{ext}(\mathbf{x}_0, \mathbf{y})$  around  $\mathbf{x}_0$  (cyan circle), i.e. outside the lumen, but inside the neighbourhood, indicated by three shades of grey.

### 2.3.3.2 Updating the model parameters for the bowel lumen

One may observe that if  $w(\mathbf{y})$  is known,  $\mu_{water}(\mathbf{x})$  can be updated by substituting (2.8) into (2.5) and solving for  $\mu_{water}(\mathbf{x})$  by setting the derivative with respect to  $\mu_{water}(\mathbf{x})$  to zero. This yields

$$\mu_{water}(\mathbf{x}) = \frac{\int_{\Omega} K_{\rho}(\mathbf{x} - \mathbf{y}) w(\mathbf{y}) I(\mathbf{y}) (1 - H_{\varepsilon}(\phi(\mathbf{y}))) d\mathbf{y}}{\int_{\Omega} K_{\rho}(\mathbf{x} - \mathbf{y}) w(\mathbf{y})^2 (1 - H_{\varepsilon}(\phi(\mathbf{y}))) d\mathbf{y}}, \quad (2.21)$$

Note the similarity between (2.2) and (2.21); they are identical if  $w(\mathbf{y}) = 1$ . Accordingly, the variance is modelled by a function  $\eta_{lumen}^2(\mathbf{x})$ , which is similar to (2.4)

$$\eta_{lumen}^2(\mathbf{x}) = \frac{1}{N_{lumen}(\mathbf{x})} \int_{\Omega} K_{\rho}(\mathbf{x} - \mathbf{y}) \times (I(\mathbf{y}) - \xi_{lumen}(\mathbf{x}, \mathbf{y}))^2 (1 - H_{\varepsilon}(\phi(\mathbf{y}))) d\mathbf{y}, \quad (2.22)$$

in which  $N_{lumen}(\mathbf{x})$  normalizes over the integration volume of  $\mathbf{y}$

$$N_{lumen}(\mathbf{x}) = \int_{\Omega} K_{\rho}(\mathbf{x} - \mathbf{y}) (1 - H_{\varepsilon}(\phi(\mathbf{y}))) d\mathbf{y}. \quad (2.23)$$

Next,  $\alpha_{fecal}$  is updated by

$$\alpha_{fecal} = \frac{\int_{\Omega} \frac{I(\mathbf{x})}{\mu_{water}(\mathbf{x})} C_{fecal}(\mathbf{x}) H_{\varepsilon}(-\phi(\mathbf{x})) d\mathbf{x}}{\int_{\Omega} C_{fecal}(\mathbf{x}) H_{\varepsilon}(-\phi(\mathbf{x})) d\mathbf{x}}. \quad (2.24)$$

Basically, (2.24) computes the average  $\alpha_{fecal}$  inside the volume defined by  $\Omega_{lumen}$ .

The update of fractions  $C_k(\mathbf{y})$  consists of two steps: First, a threshold initially classifies the voxels into air, water or fecal residue (just as before) by (2.13). Second, we assume that the ordering of materials is fixed: air always resides on top of water or fecal residue and water generally resides on top of fecal residue or bowel wall. Importantly, we define criteria for transitions through which the level set should not propagate. This is prevented by updating the material fractions

as described in the Appendix. Figure 2.6 illustrates the estimation of the bowel lumen's material intensities.

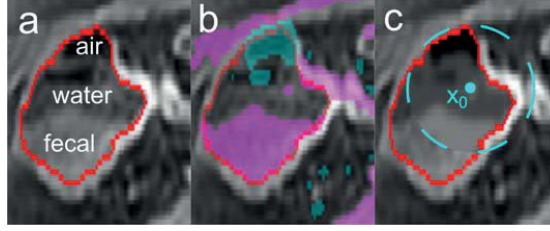


Figure 2.6. Modelling the materials inside the bowel lumen; the segmented boundary is superimposed in red. (a) An ROI depicting a piece of bowel with a heterogeneous lumen content containing fecal residue, water and air. (b) The results of a voxel-wise classification into three materials, computed using (2.13), (2.39), (2.43) and (2.44); the three colours indicate air (cyan), water (none), fecal (magenta). (c) The modelled lumen intensities  $\xi_{lumen}(\mathbf{x}_0, \mathbf{y})$  inside a neighbourhood (cyan circle) around  $\mathbf{x}_0$  computed using (2.8).

### 2.3.4 Shape driven term and energy function

In this chapter, the shape driven term consists of the commonly used weighted minimal length term [44] defined by

$$E_{shape} = \int_{\Omega} g(\mathbf{x}) |\nabla \phi(\mathbf{x})| \delta_{\epsilon}(\phi(\mathbf{x})) d\mathbf{x}, \quad (2.25)$$

where  $\delta_{\epsilon}()$  represents a regularized Dirac function[49]:

$$\delta_{\epsilon}(x) = \begin{cases} \frac{1}{2\epsilon} \left( 1 + \cos\left(-\frac{\pi x}{\epsilon}\right) \right) & |x| \leq \epsilon \\ 0 & else \end{cases}, \quad (2.26)$$

in which  $\epsilon$  specifies the width of the narrowband. The weight function, which was based on[44, 46, 47], is defined as

$$g(\mathbf{x}) = \exp\left(-\frac{\langle \nabla I(\mathbf{x}), \nabla \phi(\mathbf{x}) \rangle}{\gamma |\nabla \phi(\mathbf{x})|}\right). \quad (2.27)$$

Essentially, this is an edge detector that compares the gradient of the image to the normal of the level set function. Here, we assume that the transition from lumen to bowel wall is from dark to bright.

The total energy function consists of combining (2.5) and (2.25) into

$$E_{total} = E_{data} + \lambda E_{shape}, \quad (2.28)$$

where  $\lambda$  is a parameter that weights the contribution of both terms. Minimizing (2.28) involves iteratively updating all the model parameters, alternated by adjusting  $\phi$ .  $E_{total}$  is minimized with respect to  $\phi(\mathbf{y})$  after restructuring (2.5) by means of an Euler-Lagrange equation. The minimization is implemented using a gradient descent approach[50]. The resulting level set equation becomes

$$\begin{aligned} \frac{d\phi}{dt} = & \delta(\phi(\mathbf{x})) \left( e_{lumen}(\mathbf{x}) - e_{ext}(\mathbf{x}) + \right. \\ & \left. \left\langle \nabla g(\mathbf{x}), \frac{\nabla \phi}{|\nabla \phi|} \right\rangle + g(\mathbf{x}) \kappa(\mathbf{x}) \right), \end{aligned} \quad (2.29)$$

where  $\kappa$  represents the mean isophote curvature of  $\phi(\mathbf{x})$ . The terms  $e_{lumen}(\mathbf{x})$  and  $e_{ext}(\mathbf{x})$  denote the data-driven parts of the differential equation defined by

$$\begin{aligned} e_{lumen}(\mathbf{x}) = & \int_{\Omega} K_{\sigma}(\mathbf{x} - \mathbf{y}) \times \\ & \left( \frac{(I(\mathbf{x}) - \xi_{lumen}(\mathbf{x}, \mathbf{y}))^2}{2\eta_{lumen}^2(\mathbf{y})} + \log(\eta_{lumen}(\mathbf{y})) \right) d\mathbf{y}, \end{aligned} \quad (2.30)$$

for the lumen and

$$\begin{aligned} e_{ext}(\mathbf{x}) = & \int_{\Omega} K_{\sigma}(\mathbf{x} - \mathbf{y}) \times \\ & \left( \frac{(I(\mathbf{x}) - \xi_{ext}(\mathbf{x}, \mathbf{y}))^2}{2\eta_{ext}^2(\mathbf{y})} + \log(\eta_{ext}(\mathbf{y})) \right) d\mathbf{y}, \end{aligned} \quad (2.31)$$

for the exterior.

## 2.4 Materials

### 2.4.1 Patient preparation and MRI acquisition

The data employed in this chapter were taken from two separate studies on Crohn's disease. The data from the first study will be called the retrospective data

[69] and the data from the second study will be called the prospective data. The local Medical Ethics Committee approved both studies. All patients had given informed consent to usage of their data for future investigations.

The retrospective data was from 30 consecutively included patients with luminal Crohn's disease[69]. Patients drank 1600ml of a hyperosmolar fluid (Mannitol, 2.5%, Baxter, Utrecht, The Netherlands) 1 hour before acquiring the MRI scans for optimal distention of the terminal ileum. MR imaging was performed on a 3.0T MRI scanner (Intera, Philips Healthcare, Best, The Netherlands). A contrast agent (Gadovist 1.0 mmol/ml, Bayer Schering Pharma, Berlin, Germany) was injected (0.1 ml/kg bodyweight) prior to acquisition of an image volume by means of a 3D T1-weighted spoiled gradient echo sequence with fat saturation. The size of these images was  $400 \times 400 \times 100$  voxels with a resolution of  $1 \times 1 \times 2$  mm<sup>3</sup>. It was acquired in a breath-hold.

The prospective data was from 29 patients included in an ongoing, prospective study into luminal Crohn's disease. The data was acquired with almost the same imaging protocol as the retrospective data. The only relevant difference with the retrospective data concerned patient preparation, which involved an additional ingestion of 800 ml Mannitol (2.5%) three to six hours prior to the examination to optimize the distension of the colon.

### **2.4.2 Data annotation**

Two abdominal radiologists independently annotated all regions suspected to present Crohn's disease activity. This was done by successively drawing 2D polygons in all slices intersecting a diseased area. In total 103 such 3D regions were extracted from the patient data. A technician randomly selected 12 regions that did not overlap between the experts for parameter tuning: six from the prospective and another six from the retrospective data (6+6). The 91 remaining regions were exclusively used for testing. 30 regions had a large overlap between the experts, so that there were 61 'unique' regions. The non-overlapping annotations generally concerned only mildly diseased areas. The technician, supervised by the radiologists, drew a centerline through the lumen for each annotated ROI. For 15 cases the same technician drew a second centerline approximately three months later (to avoid memory bias), which was used to assess the robustness to initialization. Hence  $61+12+15=88$  segmentations were

made in total. The technician also coarsely drew large 3D bounding boxes around the concerned piece of bowel, which were used as the input for our segmentation algorithm. The latter step was only done to limit the computation time.

## 2.5 Experiments and Results

This section describes the experiments, which were done to tune the parameters of the algorithm and to validate the method. First, the consistency of the intensity ratio of water and fecal residue was empirically evaluated. Second, the model parameters were tuned using the training set. Third, the algorithm was validated quantitatively using the test set. Fourth, the dependence of the method on the initialization was tested by varying both the size of the ROI and by starting from a different manually annotated centerline.

The computation time of the method dataset was less than 10 minutes on a personal computer equipped with an Intel® Core™2 Quad Processor Q8400 clocked at 2.66 GHz and 4GB RAM memory. Figure 2.7 shows segmentations obtained using the LBF model[58], the model proposed by Whang et al. [61] and our approach. Notice the similarity of the outcome in the presence of homogeneous bowel content and how the previous methods grossly fail with inhomogeneous content.

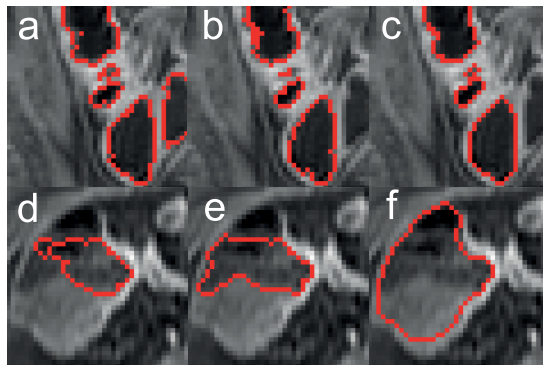


Figure 2.7. Segmentations obtained using the LBF model [58] (a,d), the model proposed by Whang et al.[61] (b,e) and our approach (c,f). The top images focus on a region with homogeneous bowel content (a,b,c) and the bottom images on areas with heterogeneous content (d,e,f).



### 2.5.1 Validating the consistency of the ratio between the intensities of fecal residue and water

The method must be applied to a region where a radiologist has spotted disease activity. One of the model assumptions is that inside such a region (ROI), the ratio between the intensities of water and fecal residue is constant. To validate this we selected ten regions containing fecal residue from the whole patient study that did not overlap with any ROI from the test set. Inside each ROI, 10 small spheres were drawn with a diameter of 10 mm: five such spheres only contained water, and the other five only contained fecal residue (see Figure 2.8). Every water filled sphere was assigned to the closest unassigned fecal residue sphere, so that smooth signal fluctuations, e.g. B0-field inhomogeneities, hardly influenced the measurement. Finally,  $\alpha_{fecal}$  was computed five times by dividing the mean intensity of the sphere containing fecal residue by the average intensity of the corresponding sphere containing water. Figure 2.8 illustrates how the spheres were selected. Table 2.1 shows the mean ratios and corresponding standard deviation per ROI. The table demonstrates that  $\alpha_{fecal}$  was approximately constant within a ROI as the standard deviation was approximately 10% of the average. Notice that the variation across ROIs is larger. This is not problematic, because our algorithm works on a per lesion basis.

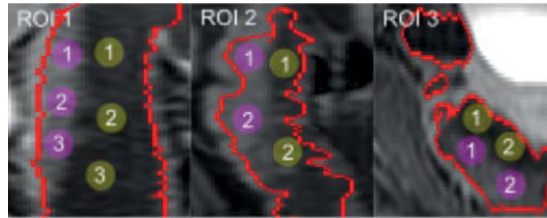


Figure 2.8. Sagittal slices from regions of interest with fecal residue. The magenta circles depict spheres containing fecal residue, the yellow ones spheres with water. The labels depict the correspondence between spheres for computing the intensity ratio between fecal residue and water. Per ROI five pairs of spheres were drawn; pairs not shown are in different image planes.

ROI	$\alpha_{fecal}$	ROI	$\alpha_{fecal}$
I	$2.31 \pm 0.26$	VI	$1.67 \pm 0.05$

II	$2.06 \pm 0.22$	VII	$1.74 \pm 0.08$
III	$2.02 \pm 0.22$	VIII	$2.09 \pm 0.28$
IV	$2.08 \pm 0.18$	IX	$1.36 \pm 0.03$
V	$2.88 \pm 0.37$	X	$1.62 \pm 0.02$

Table 2.1: Mean Ratio of the signal intensity of water and fecal residue and the corresponding standard deviation

## 2.5.2 Parameter calibration

Table 2.2 and Table 2.3 list the model parameters and their values as used in the validation experiments. Table 2.2 contains the parameters that are considered not critical and whose values were manually chosen based on reasoning. The parameters in Table 2.3 were tuned using the training set.

Parameter	Value	Description
$\rho$	64,32,8, 4,2 mm	Size of kernel $K_\rho$ , specifying the neighborhood size around voxel $\mathbf{x}$ (see (2.5)).
$\alpha_{air}$	0	The ratio of the intensities of air and water.
$\theta_{max}$	$30^\circ$	Maximum angle of an image gradient compared to the z-direction (see (2.38) and (2.40)).
$\rho_t$	2mm	Kernel size for the detection of thin structures (see (2.41). It is chosen slightly larger than the thickness of healthy bowel wall (2-3mm).
$\epsilon$	0.25	Minimum required gradient magnitude of edges in the material fraction images (see (2.38) and (2.40)).
$\tau_{thin}$	0.5	Threshold for selection of thin pieces of bowel wall (2.41).

Table 2.2: List of a priori set parameters

Parameter	Value	Description
-----------	-------	-------------

$\lambda$	2	Weight factor for balancing the data and shape terms (see (2.28)).
$\gamma$	$\eta_{noise}$	Scale parameter for comparing the gradient of the image to the normal of the level set function (see (2.27))
$\Delta t$	2	The step size of the discretized gradient descent (see (2.29))

---

Table 2.3 List of tuned parameters

Figure 2.9 shows a segmentation obtained after visually tuning the algorithm's parameters. It depicts a diseased piece of bowel consisting of several stenotic parts. In some parts the bowel lumen is fully occluded in which the centerline serves as the final segmentation. However, the bowel lumen *is* visible in other parts, in which case the centerline is the starting point for the segmentation. Here, the intensity of the stenotic lumen is increased compared to voxels in well-distended parts due to the adjacency of the bowel wall and partial volume effects. This underlines the usefulness of a coarse to fine approach, as the lumen intensity can still be assumed to be locally constant. Consequently, (2.21) still returns an accurate estimate of the local intensity at fine scales.

Manually tuning some of the parameters based on visual interpretation, enabled to produce these visually accurate segmentations. The parameter values thus determined can be found in Table III.

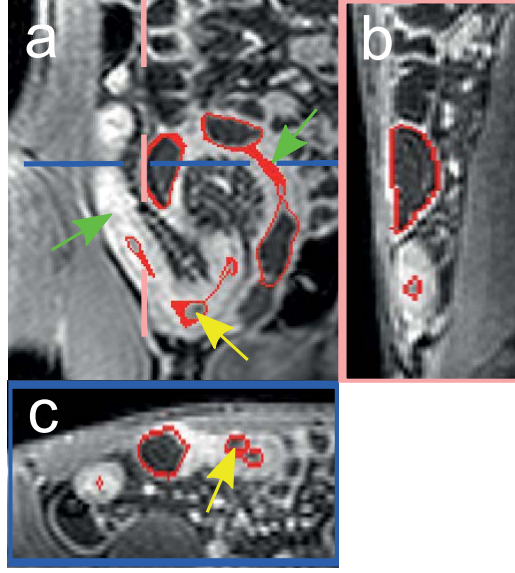


Figure 2.9. Cross-sections of a volume from the training set showing the diseased terminal ileum: (a) coronal (no outline), (b) sagittal (pink outline) and axial (c) (blue outline). The green arrows indicate a stenotic part and the yellow arrow points at a region in which the lumen is still partially visible. The segmentation outcome that was obtained after visual tuning is overlaid in red.

### 2.5.3 Performance testing

The performance of our algorithm was evaluated in two steps. First, the 30 corresponding annotations were used to determine the inter-observer variability of the annotations. Second, our segmentations were compared to the annotations of the experts. Clearly, the highest accuracy of our level set segmentation is needed in regions harboring Crohn's disease. Observe that our level set segments the bowel's inner surface, whereas the annotations consist of polygons encapsulating the whole bowel wall. Hence, the inner surface had to be extracted from the annotations. This inner surface  $S_{ano}$  was obtained from an annotated ROI  $A_{ano}$  in two steps: First, the boundary of  $A_{ano}$  was obtained by,

$$S_b = A_{ano} - (A_{ano} \ominus B), \quad (2.32)$$

where  $S_b$  is a binary image depicting the boundary voxels and  $\ominus$  represents the mathematical erosion operator applying a structuring element  $B$  consisting of a 18-connected neighborhood. Second, the angle between the gradient vectors from

the annotation and the level set segmentation was used to distinguish the inner from the outer surface by computing the inner product of the two:

$$S_a = \begin{cases} 1, & \langle \nabla \phi, \nabla A_{ano} \rangle > \frac{1}{4}, \\ 0, & \text{else} \end{cases} \quad (2.33)$$

where  $S_a$  is a binary image containing the voxels at which the two gradient vectors from the level set and the annotation point in the same direction. Finally, the inner surface  $S_{ano}$  of an annotation was extracted by taking the intersection of  $S_a$  and  $S_b$ ,

$$S_{ano} = S_a \cap S_b. \quad (2.34)$$

Figure 2.10 demonstrates this step.

Unfortunately, the inner contours from corresponding annotations of the two experts, i.e.  $S_{ano1}$ ,  $S_{ano2}$ , typically overlapped only partially. Particularly, there was variation at the border of mildly diseased regions. Therefore, we discarded the non-overlapping parts in order to determine an unbiased inter-observer variability in the truly overlapping areas.  $S_{ano1}$  was truncated by removing points that had a larger distance than  $\epsilon_{max} = 1cm$  to  $S_{ano2}$  and vice versa as indicated by

$$\begin{aligned} \hat{S}_{ano1} &= \{ \mathbf{x} \in S_{ano1} : d(\mathbf{x}, S_{ano2}) \leq \epsilon_{max} \}, \\ \hat{S}_{ano2} &= \{ \mathbf{x} \in S_{ano2} : d(\mathbf{x}, S_{ano1}) \leq \epsilon_{max} \}, \end{aligned} \quad (2.35)$$

where  $d$  is the smallest Euclidean distance between a point  $x$  to the other surface. The truncated surfaces are denoted by a hat (^). The distance between two contours,  $s_1$  and  $s_2$  was measured by

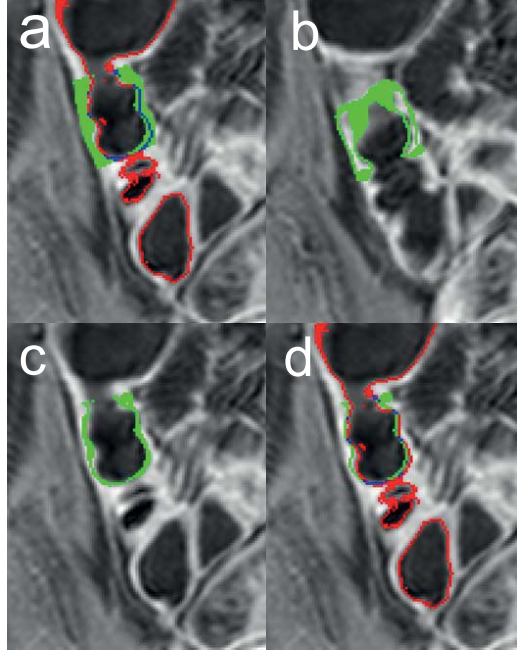


Figure 2.10. An illustration of the procedure to determine the inner-surface from an annotation. (a) A region with the level set surface in red, the annotation in green, and overlapping voxels in blue. (b) The inner and outer surface of the annotation in green. (c) Voxels in which the annotation's surface normal points in the same direction as the gradient of our level set function. (d) The annotated bowel wall's inner surface in green and our segmentation in red, with the overlap in blue.

$$D_{asym}(S_1, S_2) = \sqrt{\left(\frac{1}{N(S_1)} \sum_{i=1}^{N(S_1)} d(\mathbf{x}_i, S_2)\right)^2} \quad (2.36)$$

where  $N(S_1)$  is the number of voxels in  $S_1$ . The inter-observer variation ( $D_{sym}$ ) of the annotations was measured by

$$D_{sym}(\hat{S}_{ano1}, \hat{S}_{ano2}) = \min(D_{asym}(\hat{S}_{ano1}, \hat{S}_{ano2}), D_{asym}(\hat{S}_{ano2}, \hat{S}_{ano1})). \quad (2.37)$$

Similarly, the distance between the level set segmentation  $S_{seg} = \{\mathbf{x} \in \Omega: \phi(\mathbf{x}) = 0\}$  and the truncated surfaces  $\hat{S}_{ano1}$ ,  $\hat{S}_{ano2}$  was measured by  $D_{asym}(S_{seg}, \hat{S}_{ano1})$  and  $D_{asym}(S_{seg}, \hat{S}_{ano2})$  respectively. Observe that these distances are

asymmetric, because the level set segmentation always encapsulates a larger piece of bowel than the annotation. Wilcoxon's signed rank test was used to statistically assess differences between distances: a p-value smaller than 0.05 was considered to indicate a significant difference.

### **2.5.3.1 Qualitative outcomes on the test set**

Figure 2.11 shows how our algorithm coped with the challenges listen in the introduction (section 2.1). In Figure 2.11 heterogeneous bowel content, variation in the lumen diameter and large contrast variations between the lumen and the bowel wall can be clearly observed.

Figure 2.12 demonstrates the only example in which the level set segmentation clearly failed. In this case, the lumen contained merely fecal residue, which had a similar intensity as the bowel wall. Therefore, the algorithm confused bowel wall for lumen (it did not fail because of the constant luminal intensity!).

### **2.5.3.2 Quantitative outcomes on the test set**

Initially, we focused on the corresponding annotations, i.e. diseased regions identified and annotated by both experts. The mean distance ( $D_{sym}$ ) and corresponding standard deviation between the annotations of the experts was  $1.52 \pm 0.70\text{mm}$ . The mean distances and standard deviations of the level set segmentation to the experts ( $D_{asym}$ ) were  $1.44 \pm 0.55\text{mm}$  (annotator1) and  $1.48 \pm 0.68\text{mm}$  (annotator 2) respectively. Notice that all these distances are smaller than the largest voxel dimension, which was 2mm (out-of-plane) both for the retrospective and the prospective data.

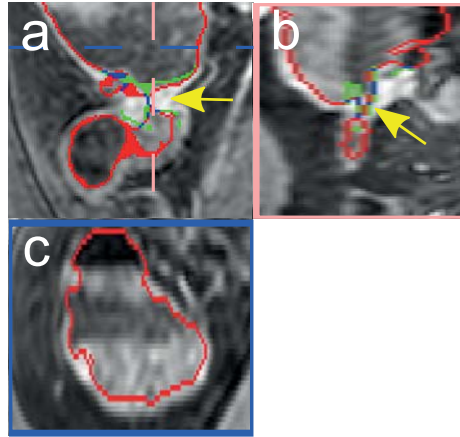


Figure 2.11. Cross-sections ((a) coronal, (b) sagittal, (c) axial) of a volume from the test set showing a diseased neo-terminal ileum. The red and green contours represent the level set segmentation and the inner-surface of the annotation made by expert 2. Overlapping parts are indicated in blue. Observe that there is a stenotic part (yellow arrow) and inhomogeneous bowel content. The distance between segmentation and annotation is 1,43mm.

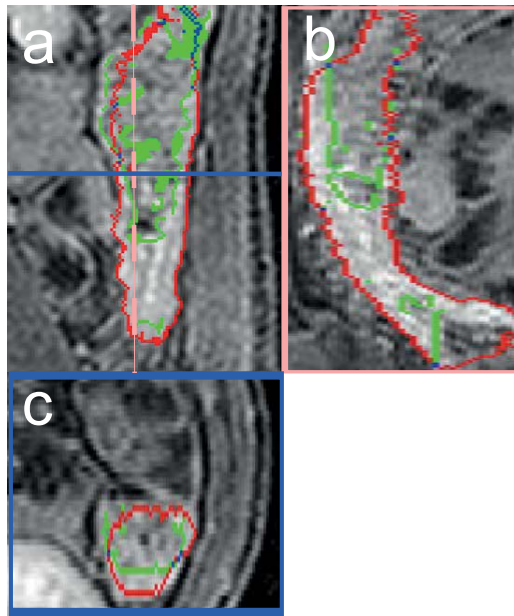


Figure 2.12. Cross-sections of a volume from the test set in which the level set algorithm (red) failed. The bowel lumen is filled with fecal residue only, having a similar intensity as the bowel wall. The distance between segmentation (red) and annotation (green) is 3.2mm.



Figure 2.13 (a) contains a Bland-Altman plot showing  $D_{asym}(S_{seg}, \hat{S}_{ano1})$  versus  $D_{asym}(\hat{S}_{ano1}, \hat{S}_{ano2})$  for the corresponding annotations. In words, it depicts the distance between the level set and the annotation of the first expert against the distance between second expert's annotation and first's. Similarly, Figure 2.13 (b) shows  $D_{asym}(S_{seg}, \hat{S}_{ano2})$  versus  $D_{asym}(\hat{S}_{ano1}, \hat{S}_{ano2})$  and Figure 2.14 shows  $D_{asym}(S_{seg}, \hat{S}_{ano1})$  versus  $D_{asym}(S_{seg}, \hat{S}_{ano2})$ .

Figure 2.14 illustrates that our algorithm approximates each of the annotators equally well. This can be seen as the points are rather randomly distributed around the origin of the y-axis ( $D_{asym}(S_{seg}, S_{ano1}) - D_{asym}(S_{seg}, S_{ano2}) = 0$ ). In other words, the algorithm is unbiased with respect to the annotators. What is more, as indicated in Table 2.4, none of the aforementioned distances are significant.

Mean distances to be compared	p-value
$D_{asym}(S_{ano1}, S_{ano2})$ versus $D_{asym}(S_{ano1}, S_{seg})$	0.71
$D_{asym}(S_{ano1}, S_{ano2})$ versus $D_{asym}(S_{ano2}, S_{seg})$	0.90
$D_{asym}(S_{ano1}, S_{seg})$ versus $D_{asym}(S_{ano2}, S_{seg})$	0.95

Table 2.4: Statistical comparison of distances

Finally, we assessed the performance of the level set segmentation against *all* annotations of both experts. Observe that the annotations matched only partly between the experts. Accordingly, we compared the level set to the full inner surfaces extracted from all the annotations (i.e.  $S_{ano1}, S_{ano2}$ ).

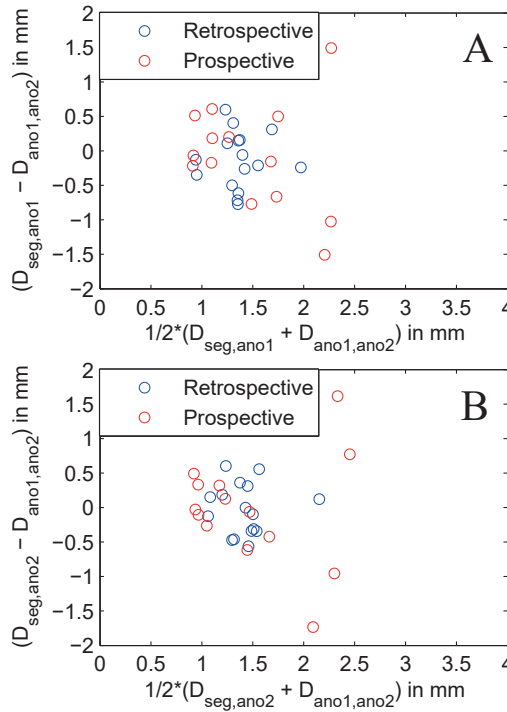


Figure 2.13. Bland-Altman plots of the inter-observer variability compared to the variability between the segmentation and the annotations made by experts. The blue and red data points denote respectively the annotations from the retrospective and prospective studies. (a) Our level set segmentation versus annotator 1. (b) Our level set segmentation versus annotator 2

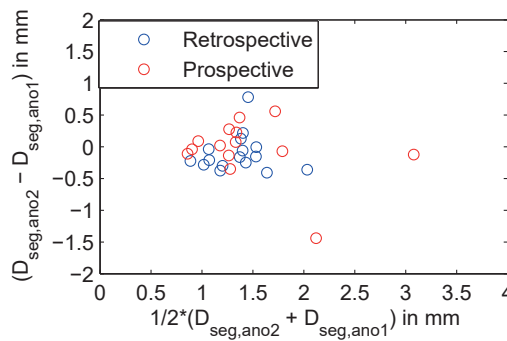


Figure 2.14. Bland-Altman plots of the distance between the level set and the annotation of the first expert against the distance between level set and the annotation by the second expert. The blue and red data points denote respectively the retrospective and prospective studies.

The mean distance and corresponding standard deviation from the level set segmentation to the first expert's annotation  $D_{asym}(S_{seg}, S_{ano1})$  was  $1.44 \pm 0.55$  mm. The mean distance and standard deviation from the level set to the second expert's annotation  $D_{asym}(S_{seg}, S_{ano2})$  was  $1.55 \pm 0.68$  mm. Figure 2.15 shows the histograms of the measured distances.

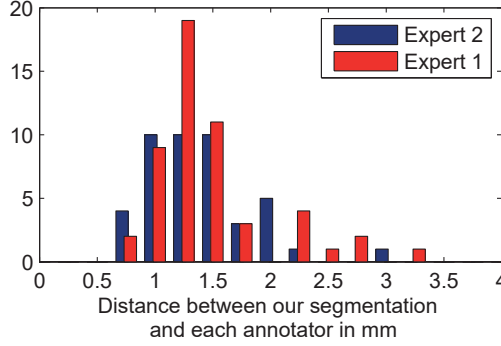


Figure 2.15. Histogram of distances between the level set segmentation and the first expert (in red) and the second expert (in blue).

#### 2.5.4 Robustness towards initialization

The variation in segmentation outcome due to differences in initialization was assessed for 15 randomly selected annotations. Initially, a centerline was drawn as described in Section IV.B and this procedure was repeated three months later for the 15 selected annotations. The technician performing the centerline tracking was blinded to the initial centerline. If the proposed method would be run on these centerlines the ensuing segmentations would overlap only partially due the differences in the starting- and endpoints. For a fair comparison, the end-points of the centerline were truncated so that the remaining points encompassed the same part of the bowel lumen. Next, the segmentations were computed from the modified centerlines. The distance between the two resulting segmentations and centerlines were both computed using (2.37) (see Table 2.5). The mean distance between the centerlines was  $1.83 \pm 0.9$  mm and the mean distance between segmentations was  $0.39 \pm 0.31$  mm. The measured distances between the centerlines are relatively small, because often a part of the lumen is partially occluded, which reduces the error. Observe that the differences in segmented

bowel lumen were much smaller than the initial difference between the centerlines. For the majority of surface points, the differences between the two segmentation results were below the voxel dimensions. To our opinion this signifies the robustness of our approach against realistic variations of the centerline.

ROI	1	2	3	4	5	6	7	8
Centerlines	1.08	1.64	1.25	1.83	1.23	1.91	0.91	1.87
Surfaces	0.51	0.37	0.17	0.18	0.59	0.25	0.09	0.19
ROI	9	10	11	12	13	14	15	
Centerlines	2.15	4.63	1.46	1.82	1.60	0.69	2.48	
Surfaces	0.26	0.24	0.29	0.82	0.05	0.56	0.50	

Table 2.5 The average distance between the two centerlines and the distance between the corresponding surfaces (in mm)

## 2.6 Discussion

We presented a new method to segment the bowel lumen from T1-weighted MR images using a level set approach. The technique worked for a heterogeneous content of the bowel lumen composed of three materials: fecal residue, water, and air as well as a diverse lumen exterior composed of three tissues with respectively dark, medium and bright intensities. Hence, it provided a natural extension to existing region-based active contour models. Particularly, it avoids splitting or merging multiple level set segmentations in a post-processing step.

Our algorithm was able to successfully segment the lumen in 60 out of 61 ROIs, spanning a wide variety of bowel geometries and material mixtures. The only case on which the method failed, concerned a bowel segment containing solely fecal residue where the contrast between wall and lumen was very low. The mean distance between our level set segmentation and the surface extracted from annotations by two experts was respectively  $1.43 \pm 0.55\text{mm}$  and  $1.48 \pm 0.68\text{mm}$ . This was comparable to the inter-observer variability between the experts' annotations:  $1.52 \pm 0.70\text{mm}$ . This inter-observer variability acts as a lower bound on the measured performance. Note that the voxel size was on average 1mm in-plane and 2mm out-of-plane, which implies that the inter-observer variability was

as accurate as one could reasonably expect. The experts independently annotated the diseased areas, which caused a discrepancy in the number of regions that were annotated. Therefore, the number of ROIs that was available for inter-observer comparison was lower than the total number of ‘unique’ ROIs. Furthermore, we needed to truncate some of the annotations for a fair inter-observer comparison. Although only the truncated annotations were used for the Bland-Altman plots, the truncation did not have any influence on the presented conclusions as the average distance between annotation and segmentation hardly changed in the comparison with all annotations.

In our work, we modeled slowly varying intensity fluctuations and abrupt intensity transitions in two separate terms: the smooth fluctuations in a term representing the mean local intensity and the abrupt ones in a term consisting of a weighted sum of mean material intensities. Alternatively, a combination of more sophisticated basis functions could be used (instead of the mean material intensities). Our future work could be to experiment with commonly used bases, such as a local polynomial basis [70] and Gabor wavelets.

This chapter showed how prior knowledge, such as the stratification pattern caused by the air-water-fecal layers inside the lumen and the distinction between bright (kidney, bladder and arteries), middle (muscle, the water in lumen) and dark (suppressed fat and air) anatomy in the bowel exterior, can be modeled by a neighborhood-variant data term. What is more, we asserted that the ratio between water and fecal residue was constant over a local ROI, which proved the case in our study. Still, we did observe that this ratio varied globally, over a patient. Overcoming this limitation might be another topic of future work, so that the method could be applied to the entire bowel.

However, our highest priority will now be to use the lumen segmentations to derive features related to Crohn’s disease activity. For instance, it eases the measurement of features such as bowel wall thickness. Future work will also definitely focus on automatic centerline extraction, which would further reduce the necessary interaction time.

## **Appendix**

The Appendix defines special conditions for transitions through which the level set should not propagate. The first transition through which the level set should

not propagate is the transition between air and bowel wall. Due to partial volume effects, the bowel wall voxels above air can have signal intensities in the range of water. In that case the level set tends to propagate through the wall. Initially, voxels above an air cavity are identified by

$$T_{aw}(\mathbf{y}) = \begin{cases} 1, & \frac{\langle \nabla \hat{C}_{air}(\mathbf{y}), \mathbf{n}_z \rangle}{|\nabla \hat{C}_{air}(\mathbf{y})|} > -\cos \theta_{max} \\ & \wedge |\nabla \hat{C}_{air}(\mathbf{y})| > \epsilon \\ 0, & \text{elsewhere} \end{cases}, \quad (2.38)$$

where  $T_{aw}(\mathbf{y})$  is a binary image depicting the invalid air to bowel wall transitions,  $\langle \rangle$  indicates the inner-product,  $\nabla()$  the gradient operator,  $\theta_{max}$  specifies the maximum angle of the gradient compared to  $\mathbf{n}_z$ , which is a unit vector oriented in the posterior direction for our coronal images, and  $\epsilon$  is a threshold to select the boundary voxels.

The first term of the condition in the top part of (2.38) constrains the direction of the gradient vector: an upward pointing normal likely corresponds to an air-bowel wall interface; if it deviates more than  $\theta_{max}$ , it most probably concerns a different transition. The second term constrains the influence of  $T_{aw}(\mathbf{y})$  to voxels located at transitions.  $T_{aw}(\mathbf{y})$  is used to stop the propagation of the active contour by setting  $C_{water}(\mathbf{y})$  and  $C_{fecal}(\mathbf{y})$  to zero and  $C_{air}(\mathbf{y})$  to one, when  $T_{aw}(\mathbf{y})$  is 1.

$$C_{air}(\mathbf{y}) = \hat{C}_{air}(\mathbf{y}) + T_{aw}(\mathbf{y}) \left( \hat{C}_{water}(\mathbf{y}) + \hat{C}_{fecal}(\mathbf{y}) \right). \quad (2.39)$$

By doing so,  $w(\mathbf{y})$  is equal to 0 and in turn  $\xi_i(\mathbf{x}, \mathbf{y})$  is zero. This makes that  $I(\mathbf{y}) - \xi_i(\mathbf{x}, \mathbf{y})$  deviates from zero, adding positive energy to (2.5). Therefore, the level set is encouraged to halt at the interface. Figure 2.16 (b) illustrates which voxels are selected through (2.38) in a typical situation.

The second transition through which the level set should not propagate is an apparent fecal residue to water transition. Due to partial volume effects, the intensity of healthy bowel wall may approach the intensity of fecal residue. Then, whenever the tissues surrounding the bowel have the same intensity as water, the algorithm would keep on propagating. In order to cope with this issue we adapted

the material fractions in a similar way as for the air to water transitions albeit involving slightly differing criteria. First, the gradient of the interface between water and fecal residue downwards, so that any such interface with a larger angle than  $\theta_{max}$  cannot be a true fecal residue to water transition. Accordingly, we define

$$T_{fw,1}(\mathbf{x}) = \begin{cases} 1, & \frac{\langle \nabla \hat{C}_{fecal}(\mathbf{x}), \mathbf{n}_z \rangle}{|\nabla \hat{C}_{fecal}(\mathbf{x})|} < \cos \theta_{max} \wedge |\nabla \hat{C}_{fecal}(\mathbf{x})| > \epsilon \\ 0, & elsewhere \end{cases}, \quad (2.40)$$

where  $T_{fw,1}(\mathbf{x})$  is a binary image depicting the invalid fecal residue to water transitions. Additionally, healthy bowel wall may be confused for fecal residue by our material fractions. Since healthy bowel is thin, these voxels can be detected by burring  $C_{fecal}(\mathbf{x})$  with a Gaussian kernel of size equal to the thickness of healthy bowel followed by a threshold. If the wall thickness is on the same order as the kernel size, the sides of the bowel wall move outwards (the full width at half maximum increases), the contrast drops and finally the whole structure disappears. For a thicker bowel wall, the location remains the same and therefore the dimensions of the structure remain unaffected. Therefore, thin bowel structures can be detected by

$$T_{fw,2}(\mathbf{x}) = \begin{cases} 1 & K_{\rho t} * \hat{C}_{fecal} \leq \tau_{thin} \\ 0 & elsewhere \end{cases}, \quad (2.41)$$

where  $T_{fw,2}(\mathbf{x})$  is a binary image that identifies the invalid fecal residue voxels,  $*$  depicts the convolution operator and  $K_{\rho t}(\mathbf{x})$  is a Gaussian kernel of size  $\rho t$  and  $\tau_{thin}$  is a threshold for classifying thin structures.

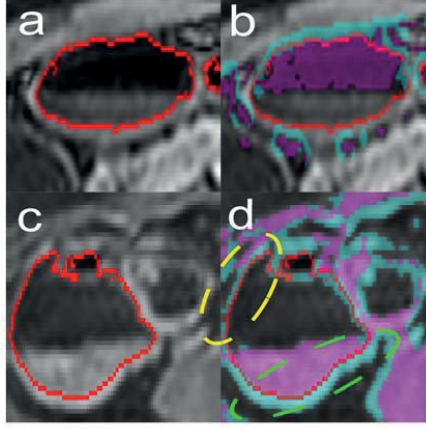


Figure 2.16. An illustration of the use of prior knowledge on the ordering of lumen materials. (a) An ROI from an axial slice depicting part of the transverse colon with the segmented bowel surface in red. (b) The voxels labelled as air have a magenta colour. The cyan color depicts voxels that were initially labelled as water or fecal residue. According to (2.38), these voxels were positioned at an invalid air-water interface, were therefore relabelled to air voxels. (c) An ROI from an axial slice depicting part of the ascending colon with the segmented bowel surface superimposed in red. (d) The regions labelled as fecal residue are shown in magenta. The cyan voxels were eliminated as invalid fecal residue to water transitions based on the criteria of (2.40) (e.g. those in the dashed green circle) and (2.41) (e.g. those in the dashed yellow circle).

Subsequently, (2.40) and (2.41) are combined into a single binary image  $T_{fw}(\mathbf{x})$  containing all voxels that were presumably erroneously classified as fecal residue

$$T_{fw}(\mathbf{x}) = T_{fw,1}(\mathbf{x}) \cup T_{fw,2}(\mathbf{x}). \quad (2.42)$$

Finally, the *corrected* material fractions of water  $C_{water}(\mathbf{x})$  and fecal residue  $C_{fecal}(\mathbf{x})$  are computed based on (2.38) and (2.42) resulting in

$$C_{fecal}(\mathbf{x}) = \hat{C}_{fecal}(\mathbf{x})(1 - T_{fw}(\mathbf{x})), \quad (2.43)$$



and

$$C_{water}(\mathbf{x}) = \hat{C}_{water}(\mathbf{x})(1 - T_{aw}) + \hat{C}_{fecal}(\mathbf{x})T_{fw}(\mathbf{x}). \quad (2.44)$$

Figure 2.16 illustrates how the aforementioned knowledge on the ordering of materials (transitions from air to bowel wall and from fecal residue to water) is used to correct estimates of the volume fractions and thereby avoid propagation of the level set from the lumen into the exterior.

Finally, a Gaussian kernel blurs the material fractions in order to represent the blurring effects caused by the PSF of the MRI scanner.

# 3 Semi-automatic bowel wall thickness measurement on MR enterography in patients with Crohn's disease

---

## Abstract

**Purpose.** To evaluate a semi-automatic method for the delineation of the bowel wall and the measurement of the wall thickness in patients with Crohn's disease.

**Methods.** Fifty-three patients with suspected or diagnosed Crohn's disease were selected. Two radiologists independently supervised the delineation of regions with active Crohn's disease on MRI, yielding manual annotations (Ano1, Ano2). Three observers manually measured the maximal bowel wall thickness of each annotated segment. An active contour segmentation approach semi-automatically delineated the bowel wall. For each active region, two segmentations (Seg1, Seg2) were obtained by independent observers, in which the maximum wall thickness was automatically determined. The overlap between (Seg1, Seg2) was compared with the overlap of (Ano1, Ano2) using Wilcoxon's signed rank test. The corresponding variances were compared using the Brown–Forsythe test. The Intraclass Correlation Coefficient (ICC) of semi-automatic thickness measurements was compared to the ICC of manual measurements through a likelihood-ratio test.

**Results.** Patient demographics: median age: 30, IQR: [25,38]; 33 female. The median overlap of the semi-automatic segmentations (Seg1 vs Seg2: 0.89) was significantly larger than the median overlap of the manual annotations (Ano1 vs Ano2: 0.72):  $p = 1.4 \cdot 10^{-5}$ . The variance in overlap of semi-automatic segmentations was significantly smaller than the variance in overlap of manual annotations ( $p = 1.1 \cdot 10^{-9}$ ). The ICC of semi-automatic measurement (0.88) was significantly higher than the ICC of manual measurement (0.45):  $p = 0.005$ .

**Conclusions.** The semi-automatic technique facilitates reproducible delineation of regions with active Crohn's disease. The semi-automatic thickness measurement sustains significantly improved interobserver agreement.

### 3.1 Introduction

Magnetic resonance imaging (MRI) facilitates evaluation of patients with Crohn's disease[71-73]. In particular, multiple MRI features are useful for assessment of Crohn's disease activity[14, 16, 28, 74]. Several disease-grading systems using these features have been validated against endoscopy and/or histopathology, notably the MaRIA, Clermont, London and CDMI scores,[14, 16, 75-77]. An important feature that is included in all the aforementioned disease activity scores is the maximal thickness of the bowel wall of a diseased segment.

Clinically, the bowel wall thickness in active Crohn's disease is *manually* measured using electronic calipers in the most thickened part of a segment. Generally, a manual measurement makes a method inherently subjective and limits the reproducibility of a feature. Accordingly, varying Intraclass Correlation Coefficients (ICC's) have been reported, reflecting this subjectivity[28, 78]. A method that (semi-) automatically performs the thickness measurement, might improve reproducibility. In turn, the reproducibility of the Crohn's disease scores could be enhanced.

A standard image processing approach to such a problem is to first (semi-)automatically delineate the region of interest ('segmentation'). Subsequently, a feature like the bowel wall thickness is derived from the segmented region. An image processing method that is frequently used for segmentation is the so-called active contour technique[41, 57, 58, 61]. Here, a very coarse outline of an object is initially created, which is then virtually deformed to yield the actual outline of the object of interest.

We have implemented an active contour approach to semi-automatically segment both the bowel wall's inner and outer surface taking into account the inhomogeneous bowel content, see appendix II. Subsequently, the single largest distance between the two surfaces is determined to yield the thickness measure. The method was integrated in the 3DNetSuite image postprocessing environment (Biotronics3D inc, London, United Kingdom).

The aim of the current study is to evaluate this semi-automatic method to segment the bowel wall and subsequently measure wall thickness. We hypothesize that our semi-automatic method significantly improves interobserver agreement compared to manual measurement.

## 3.2 Methods

### 3.2.1 Data

The data employed in this chapter were taken from two studies on Crohn's disease: (1) data from a prior, single center study referred to as retrospective data[69]; (2) data from a recently concluded multi-center study called the prospective data (publication in preparation).

Inclusion criteria for the retrospective study were: patients with histologically *proven* Crohn's disease,  $\geq 18$  years of age, undergoing MRI and ileocolonoscopy (within two weeks) as part of their clinical follow-up in a single tertiary center (the Academic Medical Center (AMC), Amsterdam, the Netherlands).

Inclusion criteria for the prospective study were: patients with *suspected or proven* Crohn's disease (based on clinical data, endoscopy or histopathology),  $\geq 18$  years of age, undergoing MRI and ileocolonoscopy within two weeks as part of their clinical follow-up in one of two tertiary centers (1. the Academic Medical Center, Amsterdam, the Netherlands or 2. The University College London Hospitals (UCLH), London, United Kingdom).

Exclusion criteria for both studies were: general contraindications for MRI (claustrophobia, pregnancy, renal insufficiency, pacemaker), an incomplete scan protocol or incomplete colonoscopy e.g. due to impassable strictures.

The local Medical Ethics Committee approved both studies. All patients gave written informed consent to usage of their data for future investigations.

The retrospective data was from all 27 patients consecutively included in the prior study at AMC between February 2009 and November 2010. Patients drank 1600ml of a hyperosmolar fluid (Mannitol, 2.5%, Baxter, Utrecht, The Netherlands) 1 hour before acquiring the MRI scans for optimal distension of the terminal ileum. MR imaging was performed on a 3.0T MRI scanner (Intera, Philips Healthcare, Best, The Netherlands). Imaging included amongst others a breath-hold contrast-enhanced T1-weighted spoiled gradient echo series with fat saturation. This was the sequence that was used for semi-automatic thickness measurement (see below).

The prospective data was from 26 patients randomly selected from the prospective study data acquired at UCLH. This was done to have an approximately equal number of patients from two different medical centers. Patients were consecutively included in the prospective study from December 2011 until August 2014. The UCLH data was acquired with almost the same imaging protocol as the retrospective data. The most relevant difference concerned the patient preparation, which involved an additional ingestion of 800 ml Mannitol (2.5%) three to six hours prior to the examination to optimize the distension of the colon. MR imaging was also performed on a 3.0T MRI scanner (Ingenia, Philips Healthcare, Best, The Netherlands).

Detailed scan protocols are listed in Appendix I.

### **3.2.2 Annotations**

Two experienced abdominal radiologists (JS (>800 enterographies, 21 years), ST (>1600 enterographies, 13 years)) independently identified all regions they considered to represent active Crohn's disease. The presence of active Crohn's disease was based on all available MRI sequences. Two research fellows (JT respectively AM) independently annotated the data on behalf of the radiologists. Henceforth, these annotations are referred to as Ano1 and Ano2 respectively. Specifically each annotation was performed by successively drawing (2D) polygons in all slices including the diseased segment on the coronal contrast-enhanced T1 image. The stacks of 2D polygons constituted 3D volumes of active disease segments and served as references for the semi-automatic segmentations (see below). Annotations were considered to correspond (i.e. Ano1 and Ano2 detailing the same diseased segment) if they had at least 10% overlap.

The same two radiologists (JS, ST) and one research fellow (CP (>100 enterographies)) manually measured the maximal bowel wall thickness of each segment with disease activity, i.e. with an annotation Ano1 or Ano2 including segments with overlap between Ano1 and Ano2. The three observers were instructed to measure the bowel wall thickness of each segment, but no further guidance or instruction was given. In particular, observers were blinded to the measurements of each other. Henceforth these measurements will be referred to as Ob1, Ob2, Ob3.

### **3.2.3 Semi-automatic thickness measurement**

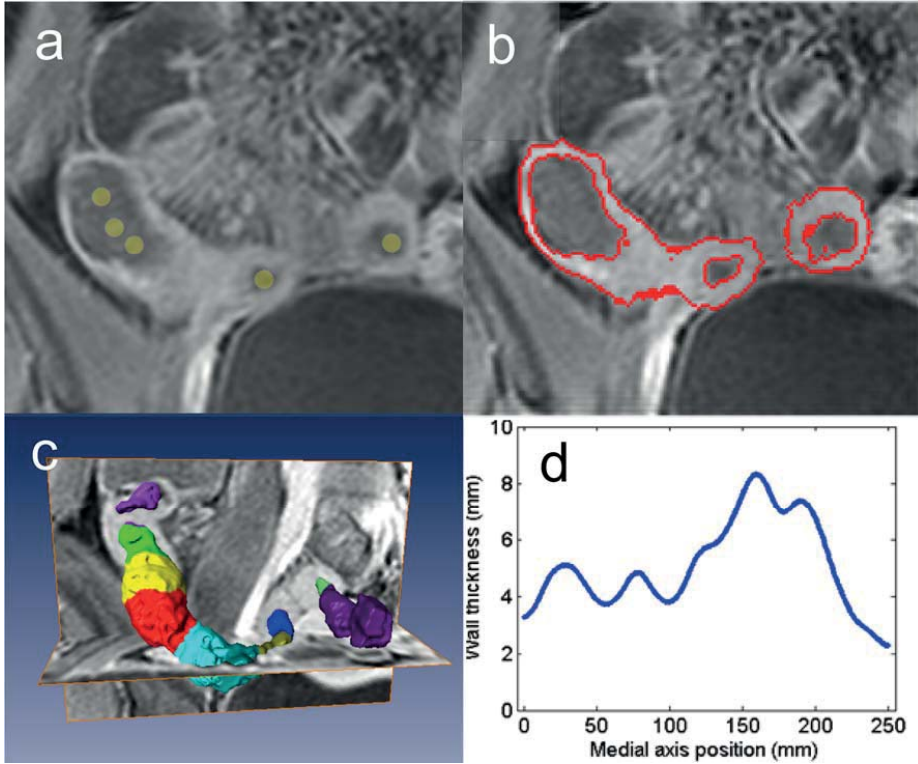


Figure 3.1: (a): initialization by placing a few points on the bowel's centerline. (b): example segmentation of the bowel wall indicated by a red line. (c): the average thickness determined over small patches along the circumference of the bowel; separate colors indicate such patches. (d): average patch thickness as a function of the position on the centerline. The maximum value represented the wall's maximal thickness, i.e. 8.2 mm in the example.

The method to measure the bowel wall thickness comprises four main steps:

1. Initialization
2. Identification of the bowel wall's inner surface
3. Identification of the bowel wall's outer surface
4. Thickness measurement

The method is initialized by manually placing a few points to indicate the bowel's centerline across a diseased section of the gastrointestinal tract (Figure 3.1 (a)). Subsequently, a small, virtual tube was constructed around this centerline that served as an initial model for the bowel wall's inner surface.

Next, the initial model was mathematically deformed, such that it delineated the transition from the bowel lumen to bowel wall as inferred from the MRI data. Therefore, we applied a well-described technique in image processing, referred to as active contour segmentation[44, 58, 61]. The technique takes into account the inhomogeneous content of the bowel lumen, see appendix II.

Subsequently, we fixed the delineated bowel wall's inner surface and applied a similar active contour segmentation to identify the wall's *outer* surface. This second segmentation was initialized by outwardly dilating the inner surface segmentation by 4 mm: approximately the average thickness of the healthy bowel wall[79]. This initial model was also mathematically deformed, such that it delineated the transition from the bowel wall to the adjacent tissues. The method was very similar to the one identifying the bowel wall's inner surface except for considering structures with varying signal intensity outside the bowel wall.

Figure 3.1(b) contains an example of the resulting segmentation of the bowel wall.

Finally, the distance was measured from each point on the segmented inner surface to the closest point on the segmented outer surface. These distances were averaged over small patches of 4 mm in length along the circumference of the bowel. The maximum average value was taken as the wall's maximal local thickness. Figure 3.1(c)(d) illustrates this procedure.

The mathematical details of the method are described in appendix II.

For the current study, in order to initiate the algorithm, two research fellows (RN, CP) independently indicated a centerline through the diseased segments delineated by annotations Ano1 and Ano2. This was done to allow subsequent comparison between the manual annotations and the semi-automatic segmentations.

Centerlines were drawn in the bowel lumen to completely cross the full extent of the annotated disease segments. Subsequently, the described algorithm yielded segmentations (Seg1 for research fellow 1, Seg2 for research fellow 2), and respective measures of the maximum bowel wall thickness (M1, M2).

Using this methodology, annotated regions Ano1 corresponded to semi-automatic segmentations Seg1 and the annotated regions Ano2 to the semi-automatic segmentations Seg2. Furthermore, in segments with overlapping Ano1 and Ano2

annotations, the manual measurements Ob1, Ob2, Ob3 corresponded to measurements (M1, M2). In annotated segments without overlap, Ob1, Ob2, Ob3 corresponded with only one measurement (either M1 or M2).

### 3.2.4 Evaluation measures

The performance of the semi-automatic segmentation procedure and the subsequent thickness measurement were separately evaluated by:

- Quantifying the *overlap* (correspondence) between the semi-automatic segmentations and manual annotations.
- Assessing the *distance* (i.e. the mismatch of the contours) between the semi-automatic segmentations and manual annotations
- *Visually grading the overlap* (correspondence) between the semi-automatic segmentations and manual annotations.
- *Correlating* the semi-automatic thickness measurements to the manual measurements.

A coefficient was determined reflecting the *overlap* or correspondence between the manual annotations and the semi-automatic segmentations (Figure 3.2(a)). In all cases, a semi-automatic segmentation covered a larger part of the bowel wall than the corresponding manual annotation. This was because the centerlines drawn by the research fellows extended beyond the manual annotations for some distance. The overlap coefficient was calculated as the percentage of volume of the manual annotations that did not overlap with the semi-automatic segmentation. This overlap measure is referred to as the semi-Dice coefficient, as it is essentially an asymmetric version of the Dice coefficient that is often applied in image processing research to measure overlap[80].

Furthermore, the mean shortest *distance* was determined from points on the manual delineations Ano1 and Ano2 to the semi-automatic segmentations Seg1 and Seg2 (Figure 3.2 (b)).



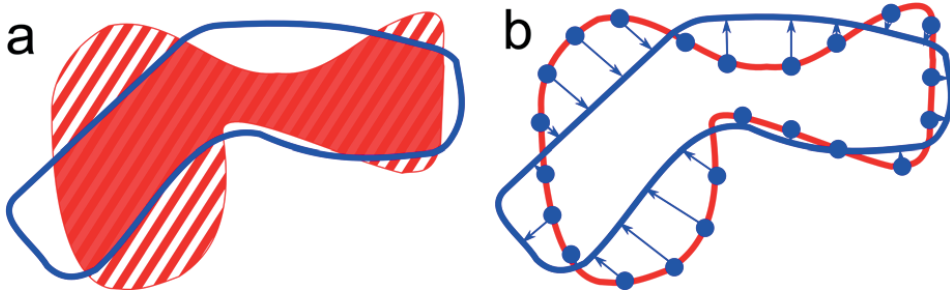


Figure 3.2: (a): the semi-Dice coefficient is defined as the volume of the annotation (A) not covered by the segmentation (S), i.e. the ratio of the striped volume to the dotted plus striped volume of the annotation. (b): the mean shortest distance between annotation (A) and Segmentation (S) is calculated by sampling points on A and averaging the distance of each such point to the closest position on S.

The research fellows who drew the paths also *visually graded* the accuracy of the semi-automatic segmentation on a four point Likert scale. The Likert scale reflected the percentage of overlap with the perceived lesion: 0 (<,50%] no overlap; 1: (50%,70%] poor overlap; 2: (70%,90%] moderate overlap; 3: (90%,<],complete overlap.

Finally, the intraclass *correlation* coefficient (ICC) between the manual thickness measurements Ob1, Ob2 and Ob3 and the thickness measurements derived from the semi-automatic segmentations was derived. ICC values were interpreted using the following criteria: <0.20, poor; >0.20≤0.40, fair; >0.40≤0.60, moderate; >0.60≤0.80, good; >0.80≤1.00, very good[81].

### 3.2.5 Statistical analysis

The overlap of the semi-automatic segmentations was statistically compared to the overlap of the manual annotations (Ano1, Ano2) using the Wilcoxon's signed rank test. The associated variances were statistically assessed by the Brown-Forsythe test. The same statistical tests were used to compare the mean shortest distances of the segmentations to the annotations.

The ICC's of (two) manual thickness measurements were statistically compared to the (two) semi-automatic measurements by means of a generalized Bland-Altman procedure. The overall ICC of the (three) manual measurements was statistically compared to the ICC of the (two) semi-automatic measurements through a likelihood-ratio test using a standard mixed model analysis.

A value of  $p < 0.05$  was considered statistically significant. All statistical analyses were performed with IBM SPSS Statistics version 22.0 for Microsoft Windows computers (SPSS, Chicago, III, USA).

### **3.3 Results**

The patient demographics were: median age: 30, IQR: [25,38]; 33/53 female. In the complete dataset 52/53 patients were identified with active Crohn's disease on MRI by either one of the radiologists or by both. In these patients there were 47 Ano1 annotations and 42 Ano2 annotations. Across both radiologists, there were 59 unique segments identified as active (i.e. annotated either as Ano1 or Ano2 alone, or by both), of which 30 corresponded, i.e. by having an overlap of more than 10%. The median overlap of these corresponding regions was 72%.

#### **3.3.1 Evaluation of the semi-automatic segmentations**

In Figure 3.3 the manual annotations (Ano1, Ano2) in the corresponding segments ( $n=30$ ) are compared with the semi-automatic segmentations Seg1 and Seg2. Figure 3.3(a) shows distributions of the semi-Dice coefficient (i.e. the overlap measure) and Figure 3.3(b) distributions of the mean shortest distances. The median semi-Dice coefficients were: Ano1 vs Seg1 = 0.87; Ano2 vs Seg2 = 0.76; Ano1 vs Ano2 = 0.72; Seg1 vs Seg2 = 0.89. The overlap of the semi-automatic segmentations (Seg1 vs Seg2) was significantly greater than the overlap of the two manual annotations (Ano1 vs Ano2):  $p = 1.4 \cdot 10^{-5}$ . Also, the variation in the overlap of the semi-automatic segmentations was significantly smaller than the variation in the overlap of the manual annotations ( $p = 1.1 \cdot 10^{-9}$ ). The median of the mean shortest distances were: Ano1 vs Seg1 = 1.31; Ano2 vs Seg2 = 1.28; Ano1 vs Ano2 = 1.07; Seg1 vs Seg2 = 0.64. The median of the distances between the semi-automatic segmentations (Seg1 vs Seg2) was significantly smaller than the median of mean shortest distance between the two manual annotations (Ano1 vs Ano2):  $p = 6.0 \cdot 10^{-6}$ . Also, the variation in the mean shortest distance of the semi-automatic segmentations was significantly smaller than the variation in the mean shortest distance of the manual annotations ( $p = 1.5 \cdot 10^{-9}$ ).

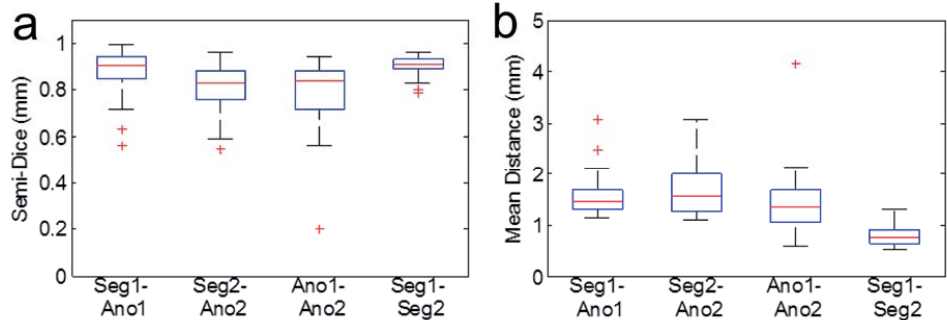


Figure 3.3. Comparison of the manual annotations Ano1 and Ano2 with the semi-automatic segmentations Seg1 and Seg2. Only corresponding regions are included, i.e. in which there was at least 10% overlap between the annotations Ano1 and Ano2 ( $n=30$ ). Each box plot shows the distribution of semi-Dice coefficients (a), respectively mean shortest distances (b) for a particular comparison (horizontally). The boxes display the median and 25th, respectively 75th percentiles of the data distribution; whiskers extend to 1.5 times the interquartile range; values outside these ranges are indicated as individual points.

Figure 3.4 is a bar chart summarizing the visual assessment of the semi-automatic segmentations, for all annotated regions. The distribution over the grades (0 to 3) for Seg1 was: 2, 1, 10 and 34, respectively ( $n=47$ ). The distribution for Seg2 over the grades was: 1, 1, 4 and 36, respectively ( $n=42$ ). The segmentations with grade 0 and grade 1 related to images with imaging artifacts ( $n=1$ ) and extensive fecal residue obscuring the bowel wall ( $n=2$ ), respectively.

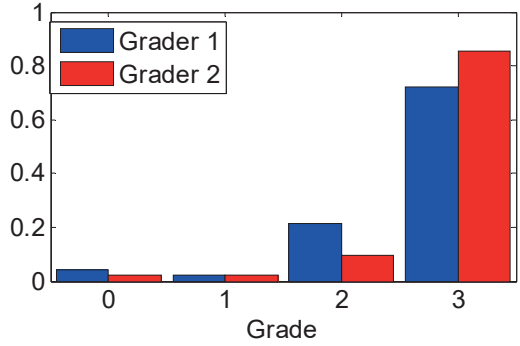


Figure 3.4. Bar chart collating the visual assessment of Seg1 and Seg2 by the research fellows initiating the segmentations. Horizontally are the Likert gradings: '0' no overlap; '1': 33% overlap; '2': 66% overlap; '3': complete overlap. Vertical are the fraction of segmentations in

a grading category (summing to 1 for both fellows). All annotations are included (n=47 for Seg1; n=42 for Seg2).

Figure 3.5 relates box plots of the semi-Dice coefficients of the segmentations to the grades given for all segmentations: Seg1 (a) respectively Seg2 (b). Additionally, Table 3.1 shows the medians of the semi-Dice coefficients, mean shortest distances and visual grades for all segmentations.

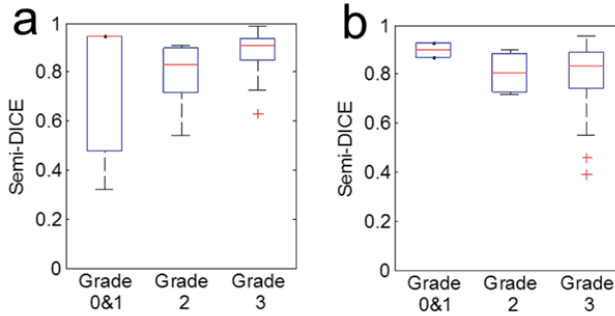


Figure 3.5. Box plots of semi-Dice coefficients as a function of the visual grades. (a): Seg1 (n=3 for Grade 0&1; n = 10 for Grade 2; n=34 for Grade 3); (b): Seg2 (n=2 for Grade 0&1; n = 4 for Grade 2; n=36 for Grade 3).

	Ano1-Seg1 (n=47)	Ano2-Seg2 (n=42)
Median semi-Dice (a.u.)	0.90 [0.82; 0.94]	0.84 [0.74; 0.90]
Median Distance (mm)	1.45; [1.29;1.63]	1.61 [1.36; 2.0]
Median Grade (a.u.)	3 [3;3]	3 [3;3]

Table 3.1: Median values and interquartile range of the semi-Dice Coefficient, mean shortest distance and visual grade comparing *all* annotated regions Ano1 and Ano2 to the segmentations Seg1 and Seg2.

### 3.3.2 Evaluation of the semi-automatic thickness measurement

Table 3.2 details the paired ICC coefficients of wall thickness measurements on the corresponding regions: Obs1, Obs2, Obs3, M1 *and* M2. Similarly, Table 3.3

and Table 3.4 shows the ICC's of the wall thickness measurements from *all* regions, i.e. depicting the ICC's for M1, respectively M2 separately. There is one semi-automatic measurement in the latter tables because there is no corresponding semi-automatic measurement for some regions.

The ICC's of the paired manual measurements on the overlapping segments varied from fair (lowest ICC: 0.34 (Obs1 vs Obs2)) to good (highest ICC: 0.60 (Obs2 vs Obs3)). The overall ICC of the manual measurement was moderate: 0.451. The ICC of the semi-automatic measurements on these segments was considered very good (0.88).

The ICC of the semi-automatic measurements was significantly higher than each ICC of the manual measurements. Particularly,  $p = 7.7 \cdot 10^{-5}$  for the comparison of the ICC of Obs2 vs Obs3 (0.60) to the ICC of M1 vs M2. Moreover, the overall ICC of the semi-automatic measurements was found to be significantly lower than the overall ICC of the manual measurement ( $p = 0.0054$ ).

	Obs1	Obs 2	Obs 3	M1	M2
Obs 1	1	0.342	0.516	0.542	0.453
Obs 2	0.342	1	0.603	0.617	0.529
Obs 3	0.516	0.603	1	0.737	0.738
M1	0.542	0.617	0.737	1	0.897
M2	0.453	0.529	0.738	0.897	1

Table 3.2: Intraclass correlation coefficients of manual thickness measurements Obs1, Obs2, and Obs3 and semi-automatic measurements M1,M2 for the corresponding regions (n=30).

	Obs 1	Obs2	Obs 3	M1
Obs 1	1	0.475	0.473	0.541
Obs 2	0.475	1	0.451	0.593
Obs 3	0.473	0.451	1	0.722
M1	0.541	0.593	0.722	1

Table 3.3: Intraclass correlation coefficients of manual thickness measurements Obs1, Obs2, and Obs3 and semi-automatic measurements M1 (n=47) for all segments with an annotation Ano1.

	Obs 1	Obs2	Obs 3	M2
Obs 1	1	0.325	0.584	0.485
Obs 2	0.325	1	0.658	0.545
Obs 3	0.584	0.658	1	0.702
M2	0.485	0.545	0.702	1

Table 3.4: Intraclass Correlation coefficients of manual thickness measurements Obs1, Obs2, and Obs3 and semi-automatic measurements M2 (n=42) for all segments with an annotation Ano2.

### 3.4 Discussion

This chapter evaluated a semi-automatic method to measure bowel wall thickness. The method consisted of four steps: (1) initialization by manually drawing a centerline, (2) segmentation of the bowel wall's inner surface, (3) segmentation of the wall's outer surface, and (4) measurement of the bowel wall thickness.

The eventual thickness measurement was fully dependent on the preceding segmentation step. In future research, the segmentation might also be at the basis to derive other features, e.g. the volume of the disease region. For these reasons, the method was separately evaluated regarding its performance (1) regarding the segmentation of a diseased part of the bowel wall and (2) to measure a bowel segment's wall thickness.

The overlap and distance between semi-automatic segmentations and the manual annotations were first determined on the corresponding segments (i.e. where the two independent manual annotations were made in the same segment of bowel). This allowed a direct comparison of the performance measures. The median overlap of the segmentations with these annotations was large: 0.87 (Seg1 vs Ano1) and 0.76 (Seg2 vs Ano2). Additionally, the median distance between segmentation and manual annotations was small: 1.31 mm (Seg1 vs Ano1) and 1.28 mm (Seg2 vs Ano2). What is more, the two semi-automatic segmentations

had significantly larger overlap and shorter distance to each other than the manual annotations. There was also a significantly smaller spread in overlap and distance between the segmentations than between the annotations. This signifies the good reproducibility of the semi-automatic segmentations.

The corresponding segments might be relatively 'easy' to segment because there was agreement between the annotators regarding the presence of active disease. However, the large majority of visual gradings indicated complete overlap of segmentation and annotation considering all regions (i.e. not only the corresponding ones). Only a few segmentations had poor (33%) to no overlap with the annotations: 3/47 for Seg1 and 2/42 for Seg2. The median semi-Dice, mean shortest distance and overlap grading (c.f. Table 3.1) further confirm the high accuracy of the segmentations on all regions. Figure 3.5 aimed to corroborate the relation between the visual gradings and the overlap measures. We refrained from statistically assessing the relation, because there are hardly any gradings in categories '0' & '1' and relatively few in category '2'.

Most importantly, the ICC of the semi-automatic thickness measurements was found to be significantly lower than the ICC of the manual measurement ( $p = 0.0054$ ). This demonstrates the limitation of performing manual thickness measurements and the benefit of semi-automatic measurement.

Previous studies have reported varying agreement in manual measurements of the bowel wall thickness[28, 78]. Part of our data is also included in[28]. In that study, the ICC was reported to be very good (0.87) for (two) experienced observers, while it was reported to be good (0.69) for a mixed pool of (four observers)[28]. On the other hand, a moderate ICC of 0.51 was reported in another study of 33 patients[78]. To our knowledge all the previous studies evaluated the thickness measurement on all bowel segments of patients in a study population. As such, these measurements probably included a large number of healthy segments, which could skew the ICC to higher values. Our thickness measurements were made on segments that were perceived as active by two experienced radiologists. In our opinion, the lower overall ICC than previously reported for our data (0.45 vs 0.69) shows the difficulty of objectively measuring the thickness of bowel segments with active disease on MRI.

Semi-automatic methods for segmentation have been widely employed for many other challenging problems in medical imaging such as lymph node detection[82], segmentation of skin lesions[83], tumor identification[84], and organ localization and segmentation[85]. These data illustrate the wide availability of techniques for (semi-) automatically segmenting abnormal regions. However, there is very limited work on (semi-) automatic segmentation of the bowel wall, particularly in relation to Crohn's disease. In our opinion this highlights the difficulty of the task. Related in part, Bhushan et al. [86] developed a motion correction and pharmacokinetic parameter estimation technique for identifying colorectal cancer using dynamic contrast enhanced MRI data. Furthermore, Schunk et al [87] analyzed MR images for their suitability in analyzing inflammatory bowel diseases, including Crohn's disease. However, they did not explore computational tasks, but instead focused on the clinical aspects. Recently, our group has developed a supervised learning framework for automatic detection and segmentation of Crohn's disease from abdominal MR images[40]. However, this method does not deliver a comprehensive segmentation of the colon wall, but identifies only small regions sized a several mm<sup>3</sup> affected by Crohn's disease.

Our work has several limitations. First, we focused on segments with active Crohn's disease on MRI only. We took this approach arguing that these segments are most relevant both for diagnosis and assessment of disease severity. Also, this approach concurs with clinical practice in which the MRI is often evaluated without knowledge of endoscopic results. As a consequence, there were segments presenting active disease on endoscopy that are not included in the measurements. Simultaneously, several of the segments (with active disease) in which the thickness measurement was done did have a normal appearance on the T1 images (scored for the prospective study). All these segments were graded with a '3' by our annotators (i.e. complete overlap of segmentation and annotation). We expect that including normal segments will only positively skew the ICC's.

Second, we did not correlate the thickness measurements to colonoscopic outcome measures. Conventionally, the thickness measurement is combined with other features in disease-grading systems, which are validated against the colonoscopic assessment [14, 16] [75, 76][77]. We consider such a validation outside the scope of our current chapter, but it is an important aspect in a



comprehensive analysis of the prospective data that is in progress (publication in preparation)

In conclusion, our data shows that a semi-automatic measurement technique facilitates a highly reproducible delineation of a region with active Crohn's disease. Furthermore, the semi-automatic thickness measurement achieves a significantly higher intraclass correlation than manual observers in active segments on MRI. As such, it may reduce the inter-observer variability of MRI grading systems for Crohn's disease.

## Appendix I: scan protocol

	Plane	Matrix	Slice (mm)	thickness FOV (mm <sup>2</sup> )	TR (ms)	TE (ms)	Flip angle
Balanced GE	Coronal	400x400	5	380x380	2,5	1,25	60
BTFE dynamic fat sat	Coronal	192x 192	10	380x380	2.0	1	45
T2-SSFSE	Coronal	400x400	4	380x380	6660	60	90
T2-SSFSE	Axial	528x528	4	400x400	759	119	90
T2-w SSFSE fat sat	Axial	320x320	7	380x302	1314	50	90
DCE sequence	Coronal	192x224x30	NA	380x439	2,9	1,8	15
3D T1-w SPGE fat sat	Coronal	200x240x90	NA	380x459	2.2	1.0	10
3D T1-w SPGE fat sat	Axial	384x384x90	NA	380x380	2.1	1.0	10

GE: gradient echo; BTFE: balanced turbo field-echo; SSFSE: single-shot fast spin echo; DCE: dynamic contrast enhanced; SPGE: spoiled gradient-echo; FOV: field of view; TR: repetition time; TE: echo time.

Table 3.5. Scan protocol applied to acquire the prospective data.

	Plane	Matrix	Slice (mm)	thickness FOV	TR (ms)	TE (ms)	Flip angle
T2-SSFSE	Coronal	256x256	4	400x400	516-758	65-118	90
T2-SSFSE	Axial	256x256	4	400x400	516-758	65-118	90

*Chapter 3: Semi-automatic bowel wall thickness measurements on MR enterography in patients with Crohn's disease*

T2-w SSFSE fat sat	Axial	288x288	7	375x300	1370-1450	70	90
DCE sequence	Coronal	144x144x14	NA	400x400x35	2,9	1,8	6
3D T1-w SPGE fat sat	Coronal	192x192x100	NA	400x400x200	1.87-2.19	1.0	10
3D T1-w SPGE fat sat	Axial	208x208x70	NA	400x400x140	1.87-2.19	1.0	10

SSFSE: single-shot fast spin echo; DCE: dynamic contrast enhanced; SPGE: spoiled gradient-echo; FOV: field of view; TR: repetition time; TE: echo time.

Table 3.6. Scan protocol applied to acquire the retrospective data.

## Appendix II: technical description of the segmentation algorithm

The bowel wall's *inner* surface was segmented using an active contours algorithm, which extends the well-known Local Binary fitting approach[58]. This algorithm was initialized by computing the signed distance transform to the boundary obtained by dilating a manually drawn centerline using a small structuring element of 3x3x3 voxels. A negative distance to the contour denoted the inside and a positive distance the outside of the lumen. This level set function evolved to optimize an energy function  $E_{totl}$ , which balanced a data term  $E_{dataI}$  and a shape term  $E_{shapel}$ ;

$$E_{totl} = E_{dataI} + E_{shapel} \quad (3.1)$$

The data term took into account that the lumen as well as the exterior could contain materials with varying intensity. Therefore, the intensity  $I(\mathbf{y})$  in a neighborhood around  $\mathbf{x}$  was modelled by a locally varying function  $\xi_i(\mathbf{x}, \mathbf{y})$ :

$$E_{dataI} = \int_{\Omega} \sum_{i=\{\text{lumen, background}\}} \int_{\Omega} K_{\rho}(\mathbf{x} - \mathbf{y}) \times \left( \frac{(I(\mathbf{y}) - \xi_i(\mathbf{x}, \mathbf{y}))^2}{2\eta_i^2(\mathbf{x})} + \log \eta_i(\mathbf{x}) \right) M_i(\phi_A(\mathbf{y})) d\mathbf{y} d\mathbf{x}, \quad (3.2)$$

In this equation the outer integral sums over the entire image domain  $\Omega \subset \mathbb{R}^3$ , so that  $\mathbf{x}$  represents an image coordinate. The summation is over the lumen and the background separately, while an indicator function ( $M_i(\mathbf{z})$ ) is applied that is one

if  $\mathbf{z} \in \Omega_i$  and zero otherwise. As such the inner integral represents a convolution sum that only takes into account terms from the region indexed by the indicator function. Furthermore,  $\mathbf{x} - \mathbf{y}$  is a neighbourhood coordinate and  $K_p(\mathbf{x} - \mathbf{y})$  a weight function (e.g. a Gaussian) imposing locality around voxel  $\mathbf{x}$ . Finally,  $\eta_i^2(\mathbf{x})$  is the residual w.r.t. the model (i.e. the local variance).

Essentially,  $\xi_i(\mathbf{x}, \mathbf{y})$  can take on an intensity that depends on the material encountered in  $\mathbf{y}$ . This can correspond to water, air or faecal material for the lumen. Similarly, dark, intermediate and bright intensities are assumed to occur in the background. As such, we were able to cope with the varying constituency on either side of the level set function.

The shape term served for regularization and consisted of the commonly used weighted minimal length term[44]:

$$E_{shape} = \int_{\Omega} g(\mathbf{x}, \phi_A) |\nabla \phi(\mathbf{x})| \delta_{\epsilon}(\phi_A(\mathbf{x})) d\mathbf{x}, \quad (3.3)$$

where  $\delta_{\epsilon}()$  represents a regularized Dirac function and  $g(\mathbf{x})$  is a weight function that is low whenever the gradient of the level set function is directed to the normal of the image gradient.

The total energy function was optimized iteratively by a two-step process: (1)  $\xi_i(\mathbf{x}, \mathbf{y})$  and  $\eta_i(\mathbf{x})$  were updated by minimizing (3.2) for a fixed  $\phi_A$ ; (2)  $\phi_A(\mathbf{x})$  was adjusted using a gradient descent approach.

This optimization resulted in a segmentation of the inner surface of the bowel wall, which was fixated. Subsequently, a similar active contour approach was taken to coarsely segment the *outer* surface of the bowel wall with a second level set function. This second level set function was initialized by outwardly dilating the final segmentation of the inner surface by 4 mm: approximately the average thickness of the healthy bowel wall[79]. The ensuing boundary again served to generate a level set function  $\phi_B(\mathbf{x})$  in which a positive distance to the contour denoted the inside (wall) and a negative distance background. The level set again evolved to optimize an energy function:  $E_{totO} = E_{dataO} + E_{shapeO}$ .

The data term of the second energy function was very similar to the one segmenting the bowel lumen except for that the bowel wall was assumed to

contain only one component:  $\xi_{wall}(\mathbf{x}, \mathbf{y}) = \mu_{wall}(\mathbf{x})$ . Therefore, the indicator function  $M_{wall}(\mathbf{z})$  was adjusted such that it only took into account terms corresponding to the wall (and discard terms if  $\mathbf{z}$  were in the lumen).

The shape term  $E_{shapeO}$  consisted of three terms;

$$E_{shapeO} = \lambda_1 E_{regularO} + \lambda_2 E_{thickO} + \lambda_3 E_{crossO} \quad (3.4)$$

The first term in (3.4) was the same minimal length term as applied to the inner surface segmentation ( $E_{regularO} = E_{ShapeI}(\phi_B)$ ). The second term targeted to keep the outer surface close to the average wall thickness:

$$E_{thickO} = \int_{\Omega} (\phi_A(\mathbf{x}) \delta_{\epsilon}(\phi_B(\mathbf{x}) - \mu_{average})^2 d\mathbf{x}, \quad (3.5)$$

which returns the squared deviation from the average bowel wall thickness integrated along the outer surface.

The third term of (3.4) merely prevented the inner and the outer surface representations from crossing each other:

$$E_{cross}(\phi_A, \phi_B) = \int_{\Omega} H(\phi_A(\mathbf{x})) H(\phi_B(\mathbf{x})) d\mathbf{x} . \quad (3.6)$$

in which  $H$  represents a heavy side function. Essentially, equation (3.6) returns a (positive) number if the outer surface crosses the inner surface and is zero otherwise. Setting  $\lambda_3$  to a very large value precludes such a crossing.

The energy function for the bowel wall's outer surface segmentation was optimized following the same strategy for the inner surface segmentation. The two resulting level sets together yielded an accurate delineation of the bowel wall.

## 4 Semi-automatic assessment of the small bowel and colon in Crohn's disease patients using MRI (the VIGOR++ project)

---

### *Abstract*

**Background:** MRI disease activity scores show promise for evaluation of Crohn's disease (CD), although varying degrees of reproducibility have been reported. Potentially, these can be improved by use of software assisted semi-automated measurements. The aim of this study was to develop and validate a predictive MRI activity score for ileocolonic CD activity based on software assisted semi-automatic measurement of MRI features.

**Methods:** 120 patients (66 female, median age 35) patients with suspected or known CD were prospectively recruited from two centers to undergo consecutive MRI and ileocolonoscopy. An MRI based disease activity score (the "VIGOR" score) was developed based on subjective radiologist observations and semi-automatic measurements of bowel wall thickness, excess volume and contrast enhancement (A1) using a retrospective cohort of 27 patients with known CD and the Crohn's Disease Endoscopic Index of Severity (CDEIS) as reference standard. A second score was developed based on only the subjective radiologist observations. For validation, both scores were applied to the prospective dataset, along with two existing MRI activity scores (MaRIA and the London score). Interobserver agreement was evaluated using the intraclass correlation coefficient (ICC).

**Results:** The VIGOR score ( $20.5 \times A1 + 0.2 \times \text{Excess volume} + 2 \times \text{mural T2}$ ) and the subjective model had comparable correlation to CDEIS as the MaRIA and London score (Ob1/2,  $r=0.57/0.60$ ,  $0.42/0.56$ ,  $0.40/0.48$  and  $0.39/0.50$ , respectively). The VIGOR score had a significantly higher ICC compared to the other activity scores ( $0.81$  vs.  $0.43\text{--}0.55$ ,  $p<0.001$ ).

**Conclusions:** The VIGOR score achieves comparable accuracy to conventional MRI activity scores, but with significantly improved reproducibility, favouring use for therapy evaluation and monitoring of disease activity

## **4.1 Introduction**

Crohn's disease is an inflammatory bowel disease, which most commonly occurs in the small bowel and colon, but can manifest itself throughout the gastrointestinal tract. Magnetic resonance imaging (MRI) is increasingly used for diagnosis and phenotyping of Crohn's disease, because it is safe, non-invasive and has high accuracy for evaluation of enteric disease and extramural complications[1]. Furthermore, multiple MRI features such as wall thickness and bowel wall signal have been validated as biomarkers of Crohn's disease activity, demonstrating good correlation with endoscopic and histopathologic inflammation grading[14, 16].

Several MRI disease activity scores were subsequently developed to address the *relative* usefulness of the MRI features. Such activity scores incorporate a variety of features and have also been externally validated[14, 16, 28, 75]. Currently, these are slowly disseminating in clinical practice. The Magnetic Resonance Index of Activity (MaRIA) for example has been developed using the CDEIS and incorporates quantitative measurement of relative bowel wall contrast enhancement along with qualitative evaluation of the presence of mural ulceration and abnormal T2 signal [2]. Other indices, such as the London score also rely on qualitative grading of various features by reporting radiologists [3,5]. Before widespread adoption for evaluating disease activity and therapeutic monitoring, MRI activity scores must be proven accurate across the spectrum of disease severity, and reproducible between radiologists. The current literature, however, reports variable interobserver agreement for many features used in MRI activity scores[18, 28]. Moreover, although MRI shows high accuracy for severe disease activity (91% accuracy), diagnostic performance drops considerably for mild disease or disease in remission (62% accuracy)[19].

One potential solution to the current limitations of MRI is to reduce the observer variability of activity scores by incorporating novel software solutions, which can automatically extract relevant features from the MRI data, reducing the risk of observer bias. For example, a recent study by Naziroglu *et al* described an image processing method that was able to semi-automatically delineate regions of active disease on MRI based on automatic segmentation and is able to provide

measurements of bowel wall thickness (Chapter 3). The algorithm requires minimal user input and semi-automatic measurements showed high correlation with radiologist measurement, while improving interobserver agreement (Chapter 3). In another study by Li et al, motion correction in free-breathing dynamic contrast enhanced (DCE)-MRI was used to increase the reliability of dynamic enhancement features[88], another important component of disease activity scores. These features were developed under the VIGOR++ project, which received funding from the European Union's Seventh Framework Programme (FP7).

We hypothesise that a scoring system combining semi-automatic computer-aided measurements with conventional radiologist scoring of MRI features could potentially improve accuracy and reproducibility in comparison to existing MRI scores.

The aim of this study was to develop and validate a predictive MRI model for ileocolonic CD activity incorporating novel computer-aided semi-automatic measurement of MRI features using an ileocolonoscopy standard of reference, and to compare performance with existing MRI activity scores

## **4.2 Methods**

### **4.2.1 Retrospective cohort**

For development of the scoring system, an independent cohort was used, consisting of 27 patients with known Crohn's disease undergoing MR enterography (MRE) and ileocolonoscopy within four weeks. Prior to MRE, a standardized small bowel preparation was used consisting of 4 hours fasting and 1600 mL 2.5% Mannitol solution ingested over 1 hour before the scan. This cohort has previously been reported[28]. Three patients were excluded from the original cohort, because no informed consent could be obtained for future research.

### **4.2.2 Prospective cohort**

Between October 2011 and September 2014, consecutive patients  $\geq 18$  years with suspected or known Crohn's disease and scheduled for ileocolonoscopy were recruited from two European tertiary referral centers for inflammatory bowel

disease (1. Academic Medical Center (AMC), Amsterdam, the Netherlands, and 2. University College London Hospital (UCLH), London, United Kingdom). All included patients underwent MRE and ileocolonoscopy within two weeks, while clinical disease activity was assessed using the Harvey-Bradshaw Index (HBI) at the time of MRI[89].

Patient exclusion criteria were contraindications to MRI (e.g. pacemakers, claustrophobia), failure to comply with the oral contrast protocol (see below), a gap of more than two weeks between MRI and ileocolonoscopy and incomplete or insufficient bowel cleansing precluding accurate mucosal assessment. Ethical permission was obtained from both institutions' medical ethics committee and written informed consent was obtained from all patients.

#### **4.2.3 MRI protocol**

In the prospective cohort, patients fasted for at least 4 hours before the examination and were instructed to drink a total of 2400 mL 2.5% Mannitol solution (Baxter, Utrecht, the Netherlands) split in two doses: 800 mL (3 hours prior to MRI) and 1600 mL (1 hour prior to MRI), to achieve distension of both colonic and small bowel segments. MRI examinations were performed on a 3.0 T MRI unit (Ingenia/Achieva; Philips Healthcare, Best, the Netherlands) in the supine position using a phased-array body coil. The MRI protocol used in both centers is outlined in Table 4.1. The dynamic contrast enhanced (DCE) sequence consisted of 300 consecutive volumetric acquisitions at a temporal resolution of 1.2 seconds/volume. Intravenous gadolinium contrast was administered 60 seconds after the start of the dynamic sequence using the standard contrast agent in the participating centers (Gadovist 1.0 mmol/L, Bayer Schering Pharma, Berlin, Germany; Dotarem 0.5 mmol/L, Guerbet, Paris, France). Coronal and axial contrast-enhanced 3D T1-weighted spoiled gradient-echo (SPGE) images were then acquired in the delayed phase (approximately 7 minutes after contrast injection). To reduce bowel peristalsis, three separate doses of 10 mg intravenous butylscopolamine bromide (Buscopan, Boehringer Ingelheim) were given during the examination. DCE sequences were registered using the method described Li *et al.*[88].



	Plane	Slice thickness (mm)	FOV	TR (ms)	TE (ms)	Flip angle
Balanced GE	Coronal	5	380x380	2.5	1.25	60
BTFE dynamic	Coronal	10	380x380	2-2.1	1	45
T2-SSFSE	Coronal	4	380x380	628-660	60	90
T2-SSFSE	Axial	4	400x400	759	119	90
T2-w SSFSE fat saturation	Axial	7	380x380	967-1314	50	90
DCE sequence	Coronal	2.5	380x380-439	2.9	1.8	15
3D T1-w SPGE fat saturation	Coronal	2	380x380-459	2.2-2.4	1.0-1.1	10
3D T1-w SPGE fat saturation	Axial	2	380x380	2.1-2.3	1.0-1.1	10

BTFE, balanced turbo field-echo; DCE, dynamic contrast enhanced; FOV, field of view; GE, gradient echo; SPGE, spoiled gradient-echo; SSFSE, single-shot fast spin echo; TE, echo time; TR, repetition time.

Table 4.1. Protocol for MRI acquisition

#### 4.2.4 Image analysis

Scans from the retrospective cohort were individually evaluated by four observers (C.Y.N., D.P., J.S., J.M.) resulting in four evaluations per dataset[28]. MRI examinations from the prospective cohort were evaluated using online viewer software (3Dnet Suite, Biotronics3D, London, UK) by two pairs of observers (Ob1: C.Y.N, J.S.; Ob2. D.P, S.T.), with extensive experience in MRE (>1100, >800, >500 and >1500 examinations, respectively). The first pair of observers was from AMC, the second pair from UCLH. Each MRI dataset was independently evaluated by one observer from both pairs, resulting in two evaluations per dataset. Observers were blinded to each other's findings, general patient information and clinical data.

	Grading categories			
MRI Features	0	1	2	3

Mural T2 signal	Normal bowel wall	Minor increase in T2 signal (dark gray)	Moderate increase in T2 signal (light gray)	Marked increase in T2 signal (areas of white)
Perimural T2 signal	Normal mesentery	Increase in mesenteric signal; no fluid	Small fluid rim ( $\leq 2$ mm)	Larger fluid rim ( $> 2$ mm)
T1 enhancement	Normal bowel wall	Minor enhancement; significantly less than nearby vessels	Moderate enhancement; somewhat less than nearby vessels	Marked enhancement; approaches intensity of nearby vessels
T1 pattern	Not applicable	Homogeneous	Mucosal	Layered
Length of disease	0 cm	$> 0\text{--}5$ cm	$> 5\text{--}15$ cm	$> 15$ cm
Wall thickness	1–3 mm	$> 3\text{--}5$ mm	$> 5\text{--}7$ mm	$> 7$ mm
Wall thickness in mm <sup>a</sup>				
RCE				
Comb sign	Absent	Present		
Edema	Absent	Present		
Ulcers	Absent	Present		
Abscess□	Absent	Present		
Fistulas	Absent	Present		
Pseudopolyps	Absent	Present		

<sup>a</sup> Measured using electronic calipers

MRI=magnetic resonance imaging, RCE=relative contrast enhancement

Table 4.2. MRI features and grading categories

Overall scan quality was graded on a scale from 0 (non-diagnostic images) to 3 (diagnostic images without artefacts). Subsequently, the bowel was divided in five segments to be evaluated separately: the terminal ileum (most distal 20 cm of the small bowel), ascending colon, transverse colon, descending/sigmoid colon and rectum. Segment distension, defined as the percentage of adequately distended bowel for diagnostic evaluation, was graded from 0 to 4 ( $< 20\%$ ,  $20\text{--}40\%$ ,  $40\text{--}60\%$ ,  $60\text{--}80\%$ ,  $> 80\%$ ). All segments were evaluated using the MRI features in table 2. Segmental MaRIA and London score were also calculated for all segments[14, 16], as detailed in Appendix 1. The relative contrast enhancement (RCE) of the bowel wall (incorporated in MaRIA) was calculated for each

segment using the method described by Rimola *et al.*[14], while omitting the noise correction term (see Appendix 1). Active segments were defined for each reader as having  $>0$  grade on at least one manual disease feature on MRI (Table 4.2).

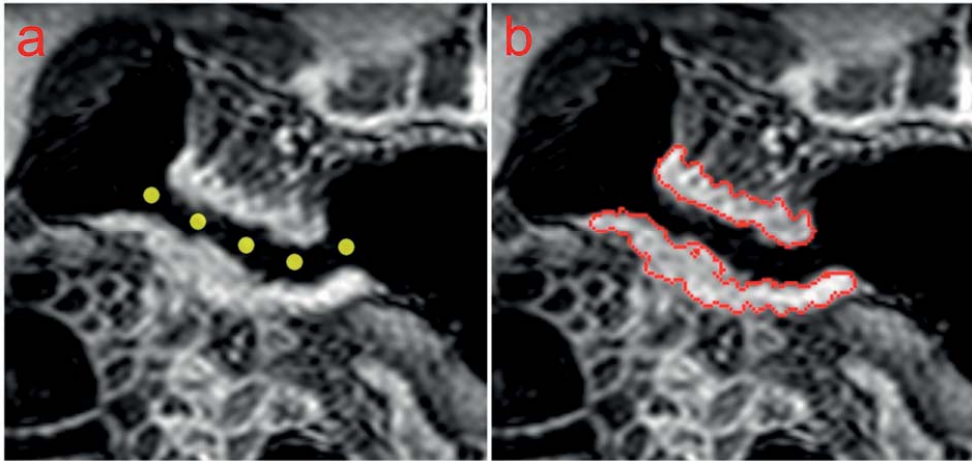


Figure 4.1 (A) Placement of centerline points in the lumen of an affected transverse colon segment. A few centerline points are placed in the middle of the lumen in one or more slices. (B) The delineation of the inner and outer bowel wall surfaces is visualized by a red line. Presently this is shown on a coronal slice, but it can be visualized in a similar way in reconstructed sagittal or transversal planes.

#### 4.2.5 Semi-automatic measurements

The bowel's centerline was indicated for all MRI active segments by manually placing a number of widely spaced points within the lumen of the bowel through the diseased region on the post-contrast coronal T1-weighted SPGE sequence (Figure 4.1A). Subsequently, the inner and outer bowel wall surfaces of the affected bowel wall were automatically delineated using the method described in Chapter 3 (Figure 4.1B). From this delineation the following automatic bowel wall thickness (ABWT) features were obtained: maximum bowel wall thickness (mm), mean bowel wall thickness (mm) and excess bowel wall volume ( $\text{mm}^3$ ). The excess bowel wall volume was defined as the volume of the delineated region exceeding normal thickness. This normal thickness was determined by semi-

automatic measurement of the mean bowel wall thickness in 25 randomly selected healthy segments (0 CDEIS and no disease on MRI) from the retrospective cohort.

Each delineation of the diseased region was also used as a 3D region of interest on DCE images to extract the A1 feature, describing the initial slope of increase of the enhancement curve[88].

#### **4.2.6 Reference standard**

For the prospective cohort, ileocolonoscopy was performed within two weeks of MRI using a standard endoscope (model CF-160L, Olympus) by either a gastroenterologist or a senior resident in gastroenterology under direct supervision of a gastroenterologist. The endoscopist used the Crohn's Disease Endoscopic Index of Severity (CDEIS) to evaluate endoscopic disease[10]. CDEIS scores were calculated for each segment and stenosis scores were added to the corresponding segmental scores. The endoscopist was blinded to findings on MRI, except for cases where a balloon-dilatation procedure was indicated. In these cases, the length of stenosis on MRI was used to determine the feasibility of balloon-dilatation.

#### **4.2.7 Model development**

The model development was limited to the manual and semi-automatic features that were measured in at least 10% of active segments, as inclusion of low incident features can introduce unwanted biases, especially for features only associated with very severe disease. Furthermore, features with high intercorrelation to another feature ( $r \geq 0.95$ ) were excluded, in which case the feature with the highest correlation to CDEIS was retained.

Active segments (as identified by at least one of the four observers) from the *retrospective* cohort were included for model development, performed using a previously described exhaustive search method for biomarker discovery[90]. In summary, this method evaluates *all* possible combinations of MRI features (Table 4.2) as candidate models for predicting CDEIS. The rank correlation to CDEIS of each such model in the retrospective data was then determined using a 50-fold bootstrap cross-validation[91]. Eventually, this procedure delivers a best performing model, which has the highest correlation to CDEIS. However, the

method generates a whole *ensemble* of models with correlation coefficients that do not differ significantly from this best performing model (based on paired Student's t-tests with Bonferroni correction for multiple testing[92]). As such, a pragmatic choice needs to be made for selecting the optimal model from the *ensemble*. To that end, we selected the features occurring in more than 50% of models in the ensemble. From this selection, a generalized linear regression model was formed, after which features with non-significant regression coefficients were removed from the model.

The procedure was repeated twice, with and without the semi-automatic features included as candidate features. The chosen model including semi-automatic features was termed the "VIGOR score". The alternate model without semi-automatic features and including only radiologist evaluations was used as a comparative reference for the VIGOR score and termed the "radiologist score".

#### **4.2.8 Validation**

The developed VIGOR score and radiologist score were tested by correlating derived segmental MRI scores to CDEIS on the independent, prospective dataset. This was done separately for both observers. In the same way, the segmental MaRIA and London scores were correlated to CDEIS. Segments with missing model features were excluded, to provide a fair comparison between different models. Interobserver agreement was calculated for all overlapping segments (i.e. deemed active by both observers).

#### **4.2.9 Statistical analysis**

Active endoscopic disease was defined as a CDEIS equal to or higher than 3[93]. Correlation of activity scores to CDEIS was done using Spearman rank correlation. Interpretation of Spearman's correlation coefficient was done as follows: 0–0.20, very weak;  $\geq 0.20$ –0.40, weak;  $\geq 0.40$ –0.60, moderately;  $\geq 0.60$ –0.80, strong;  $\geq 0.80$ –1.00, very strong. Interobserver agreement was evaluated using weighted kappa values for ordinal data and intraclass correlation coefficients (ICC) for continuous data, using the following criteria for interpretation: 0–0.20, poor; 0.21–0.40, fair; 0.41–0.60, moderate; 0.61–0.80, good; 0.81–1.00, very good[81]. Spearman correlation coefficients were

compared using the Steiger Z-test for overlapping dependent correlations[94]. ICC's were compared by calculating the variance through bootstrapping, after which paired Student's t-tests were performed.

We considered a P-value of  $< 0.05$  to indicate a statistically significant difference. Linear regression modeling and validation was implemented with R Statistical language (v3.1.2, Austria, Vienna). Descriptive statistics were analysed using SPSS 22 for Mac (SPSS, Chicago, Ill).

## **4.3 Results**

### **4.3.1 Retrospective cohort**

In the retrospective cohort, 27 known Crohn's disease patients were evaluated. Seven segments were excluded, due to hemicolectomy (n=2) and poor visibility of the rectum (n=5). Furthermore, 3 segments were excluded because of poor colonic distension: one ascending colon and two rectal segments. Of the remaining 125 segments, 49 were described as active by at least one observer (18 terminal ileum, 16 ascending colon, 4 transverse colon, 7 descending/sigmoid colon and 4 rectum segments). The following automatic features were not available in a number of active segments: A1 due to failed registration (n=5), A1 due to the segment being outside the DCE field-of-view (n=14; 1 terminal ileum, 13 colon/rectum) and ABWT due to no/failed segmentation (n=3). Eventually 27 segments were used for development of the VIGOR score; all 49 active segments were used for development of the radiologist score.

### **4.3.2 Prospective cohort**

A total of 171 patients were prospectively recruited (96 AMC, 75 UCLH). Of these, 51 patients were excluded for the following reasons:  $> 14$  days between MRI and colonoscopy (n=7), failure to comply with the oral contrast protocol (n=6), examination performed on a 1.5-T scanner (n=13), cancelled or aborted ileocolonoscopy (n=5), incomplete MRI protocol (n=12; e.g. missing sequences and incomplete DCE), failed DCE registration (n=6), insufficient bowel cleansing (n=1) and severe language barrier (n=1). The final prospective study cohort consisted of 120 patients (74 AMC, 46 UCLH), for which demographics and clinical characteristics are provided in table 3. Eighteen patients with suspected

CD eventually received a different final diagnosis: irritable bowel syndrome (n=8), ulcerative colitis (n=2), diverticulosis (n=2), undefined IBD (n=1), intestinal malrotation (n=1), previous infection (n=1), unclear disease (n=1) and no disease (n=2). These patients were included in the study cohort as they fulfilled the study inclusion criteria, although they did not show signs of active disease on MRI. As such, the activity scores were not evaluated in these patients.

<b>Total no. of patients</b>	<b>120</b>
Female, n (%)	66 (55)
Age at MRI (years), median (IQR)	35 (24–44)
Previous surgery, n (%)	43 (36)
Concomitant treatments	
Anti-TNF antibodies, n (%)	31 (26)
Steroids, n. (%) of patients	22 (18)
Thiopurines, no. (%)	19 (16)
5-ASA, no. (%) of patients	23 (19)
Methotrexate, no. (%)	8 (7)
CRP (mg/L), median (IQR)	3.8 (1–11)
HBI, median (IQR)	5 (2–8)
Montreal classification (n=103)	
Age at diagnosis (years), median (IQR)	22 (16–27)
Disease location	
L1 ileal, n (%)	41 (40)
L2 colonic, n (%)	14 (14)
L3 ileocolonic, n (%)	48 (47)
L4 upper GI tract involvement, n (%)	4 (4)
Disease behaviour	
B1 inflammatory	53 (52)
B2 stricturing	35 (34)
B3 penetrating	15 (15)
Perianal involvement, n (%)	23 (22)

5-ASA, 5-acetylsalicylic acid; CDEIS, Crohn's disease Endoscopic Index of Severity; CRP, C-reactive protein; GI, gastrointestinal; HBI, Harvey-Bradshaw Index; IQR, interquartile range; MRE, magnetic resonance enterography; TNF, tumour necrosis factor.

Table 4.3 Clinical characteristics of the prospective cohort

A total of 600 bowel segments were evaluated by two observers (Ob1, Ob2). Mean scan image quality (0–3) was 2.2 (SD: 0.6). Mean distension values for terminal ileum and colon were both 3.4 (SD: 0.7). Ob1 and Ob2 identified 96 and 103 active segments on MRI, respectively. After exclusions, a respective total of 69 and 68 active segments were included in the analysis (Figure 4.2).

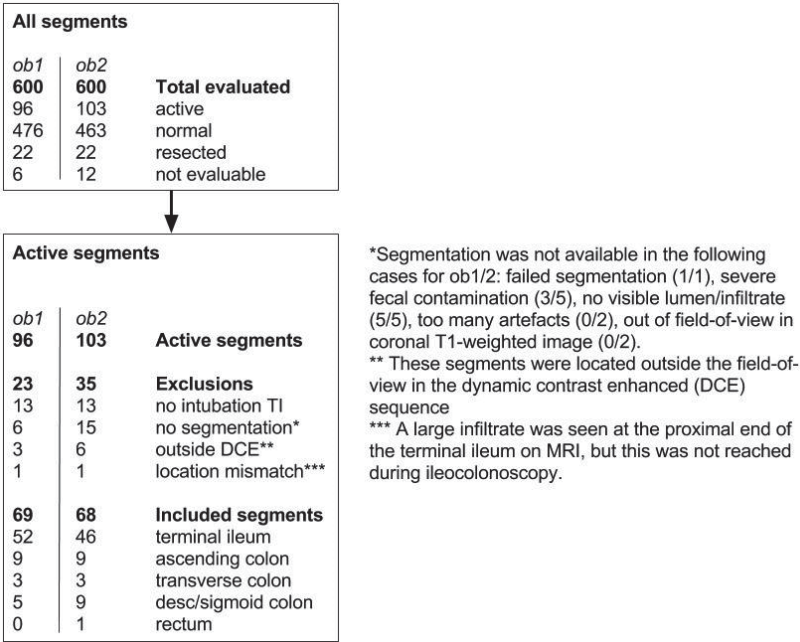


Figure 4.2. Flow diagram detailing selection of segments for inclusion in model development

Thirty-five segments with endoscopic disease (median CDEIS: 11, IQR: 8–15; 4 terminal ileum, 31 colon/rectum) were not identified as active by any observer on MRI. Compared to MRI true-positive segments (median CDEIS: 17, IQR: 11–23), CDEIS in these false-negative segments was significantly lower compared to true-positive segments (median CDEIS: 11, IQR: 8–15;  $p < 0.001$ ). Of the false-negative segments, 4/35 and 2/35 cases had a  $\text{MaRIA} \geq 7$  for Ob1 and Ob2,



respectively. These cases did not overlap and were not identified as active as none had a grade >0 on any individual, manual disease feature.

### **4.3.3 Model development**

The following features were included in the model development: mural T2 signal, perimural T2 signal, T1 enhancement, enhancement pattern, bowel wall thickness in millimetres, comb sign, length, RCE, ABWT max, ABWT excess volume and A1. The following features – fistula, abscess, pseudopolyps and ulcers – were excluded because of low incidence in the retrospective dataset (<10% of segments). ABWT mean was excluded as it was highly intercorrelated with ABWT max ( $r=0.98-1.00$ ), while the latter produced a higher correlation to CDEIS.

The search produced 697 models with no significant difference to the best performing model. The feature distribution for these models is shown in Figure 4.3. The top four features above the pre-specified 50% model prevalence cut-off were selected for the VIGOR score – A1, ABWT excess volume, mural T2 signal, perimural T2 signal and comb sign, of which A1, ABWT excess volume and mural T2 remained significant in a generalized linear regression model. On the retrospective dataset this produced the following parameterized model:

$$\text{VIGOR score} = 20.5 \times \text{A1} + 0.2 \times \text{ABWT excess volume} + 2 \times \text{mural T2} \quad (4.1)$$

Restricting the search to the manual features produced 20 models with no significant difference to the best performing model, of which the following six features were above the pre-specified 50% model prevalence cut-off – length, RCE, mural T2 signal, enhancement T1, pattern and comb sign. Of these, length, RCE, mural T2 signal and comb sign remained significant in a generalized linear regression model. The optimal model based on these features was:

$$\begin{aligned} \text{Radiologist score} = & 5.3 \times \text{length} + 0.06 \times \text{RCE} \\ & + 2.5 \times \text{mural T2} + 6.3 \times \text{comb sign} \end{aligned} \quad (4.2)$$

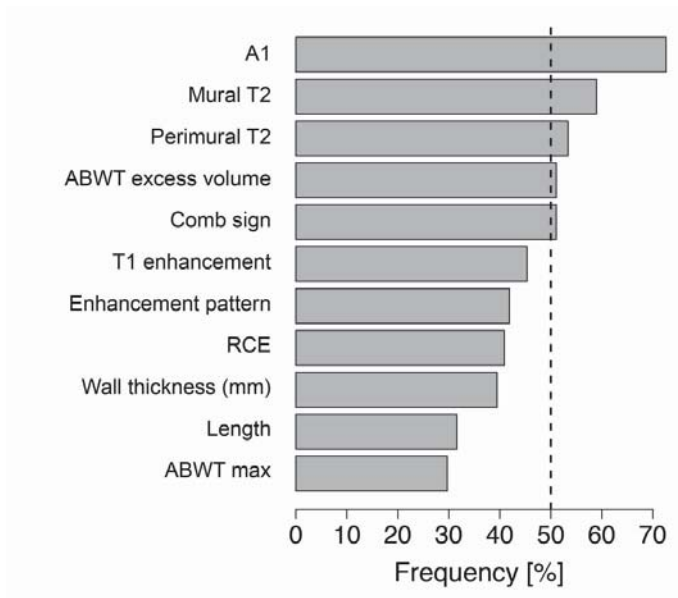


Figure 4.3. Feature distribution in the top 697 models based on manual and semi-automatic features with 50% model prevalence cut-off

#### 4.3.4 Model validation and comparison

Correlations to CDEIS for each observer and interobserver agreement are presented in Figure 4.4. The VIGOR score showed moderate-to-strong correlations to CDEIS (Ob1/2:  $r=0.57/0.60$ ). Weak-to-moderate correlations to CDEIS were seen for MaRIA ( $r=0.40/0.48$ ), while the London score ( $r=0.39/0.50$ ) and the radiologist score ( $r=0.42/0.56$ ) showed moderate correlations to CDEIS. There were no significant differences between the VIGOR score and the other activity scores in the correlation to CDEIS ( $p=0.09-0.68$ ). The VIGOR score showed a very good ICC (0.81), while other activity scores showed moderate ICC's (MaRIA, 0.43, London, 0.46; radiologist score, 0.55). Comparison of ICC's indicated significant differences between the VIGOR score and the MaRIA, London score as well as the radiologist score ( $p<0.001$ ). Correlation to CDEIS and interobserver agreement for individual MRI features can be found in Appendix 2.

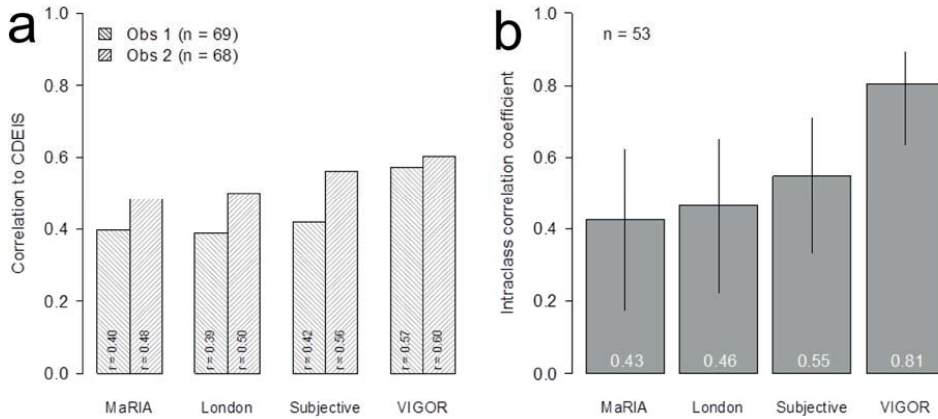


Figure 4.4 Bar plot showing activity scores Spearman correlation to CDEIS per observer and intraclass correlation coefficients (ICC's) in overlapping segments.

#### 4.3.5 Analysis on all evaluated segments

When all evaluated segments of patients diagnosed with CD (n=102) were included in the analysis, i.e. also including those which were normal on MRI (n=468), the median correlation of manual features increased from 0.23 to 0.47 for Ob1 and from 0.34 to 0.49 for Ob2, while interobserver agreement (ICC/kappa) increased from 0.33 to 0.59. For the MaRIA, London and manual activity scores, a slight increase in correlation to CDEIS was seen (0.09 for both observers), while the median of ICC's increased from 0.46 to 0.75.

#### 4.4 Discussion

In this large development and validation study, evidence was provided for a new MRI CD activity scoring system incorporating both manual observations and computer-aided semi-automatic features. Significantly improved reproducibility was demonstrated in comparison to existing manual activity scores, such as the MaRIA, London score and our derived radiologist score. The VIGOR score showed the highest accuracy against the endoscopic standard of reference, although the difference with other activity scores was not significant.

Compared to other studies, we found relatively lower correlations with CDEIS, while interobserver agreement for individual features and activity scores were also below that currently reported in the literature[28, 75, 95]. However an important difference in our methodology to previous studies comparing MRI and clinical activity scores is that we only included segments identified as active on MRI. This was deliberate as large numbers of normal segments tend to skew correlations and can result in over-optimistic conclusions. Furthermore, this is how MRI is used in clinical practice as standalone test. Indeed, when we repeated our analysis with normal segments included, we observed improvements in the correlation of radiologist scores to CDEIS and better interobserver agreement.

The large number of false-negative segments in our study ( $n=35$ ) were in most cases below the MaRIA cut-off for active disease ( $\geq 7$ ), and thus unlikely to be caused by perceptual error. Furthermore, these segments represented less severe endoscopic disease compared to true-positive segments.

In clinical trials, the need for objective measures of disease activity to evaluate response to therapy has led to the use of validated endoscopic scores, such as the CDEIS or SES-CD[10, 96]. Currently, MRI activity scores are being investigated for use as outcome measures in clinical trials, to provide a less invasive alternative to endoscopy with the additional benefit of extraluminal evaluation[12]. For this, we found suitable characteristics for the newly developed semi-automatic model, such as promising correlation to CDEIS and crucially high interobserver agreement. However, properties for therapeutic monitoring should be investigated using longitudinal studies, as was done by Ordas *et al* for evaluation of MaRIA[95]. Clearly, for MRI scores to be used in multicenter studies, a high level of reproducibility between radiologists is imperative. Furthermore, cut-off values should be determined based on commonly used outcomes, such as mucosal healing[97].

The diagnostic capabilities of semi-automatic MRI features should be further investigated. Of particular interest is the A1 feature, which does not rely on visually perceptive characteristics, and might bring additional value to detect disease activity. Concurrent steps will have to be made in automation of the

current segmentation technique, to provide full gastrointestinal tract evaluation without the need for guidance by the radiologist.

In our study we tested linear regression models, since these models are easy to apply, interpretable and understandable, and take advantage of the fact that most MRI features have positive linear relationship to Crohn's disease severity. Since it was computationally feasible, we were able to perform an exhaustive search over *all* potential models in order to develop the VIGOR score. This provides an advantage over heuristic approaches, which can produce suboptimal models. Such an exhaustive search has already shown promising potential in various fields of biomarker discovery[98, 99]. Furthermore, the VIGOR score was developed on a combination of manual *and* semi-automatic features to fully employ the capabilities of MRI for disease evaluation.

In the development and validation studies by Rimola *et al.*, a rectal enema was used to provide distension of the colon[14, 75]. In our study, we have shown that similar results can be achieved using an oral preparation with an additional 800 mL of Mannitol solution 3 hours prior to the exam. Indeed reader grading suggested colonic distension was generally good. Bowel preparation for the retrospective cohort did not include specific colon preparation. A few segments with poor colonic and rectal distension were removed from the retrospective group (n=3) as these might have introduced bias to the developed model.

Our study has several limitations. Firstly, our analysis was performed only on segments identified as active on MRI, albeit using a low threshold for active disease (one manual disease feature graded  $>0$ ). As such, we did not test whether use of semi-automatic MRI features could improve sensitivity for active endoscopic disease, which is “occult” based on conventional manual MRI features. It is interesting that the best performing metric was based on the slope of enhancement (A1) after intravenous gadolinium, something that is difficult to appreciate using the human eye. It could be that such enhancement parameters are also abnormal in diseased segments otherwise missed by conventional MRI. Clearly, considering the difficulty in unassisted detection of mild disease activity[19], evaluation of diagnostic capabilities should be a priority of future studies.

The DCE sequence used in our retrospective cohort used a smaller field of view compared to the sequence used in the prospective cohort, which limited the number of A1 data for model development. Because the field was positioned on the terminal ileum, the excluded segments were mainly colonic and rectum segments (16 out of 17 cases). Furthermore, in the prospective cohort, six patients were excluded due to failed DCE registration and respectively 9 and 21 segments were excluded due to missing semi-automatic features. These segments often showed relatively poor distension or fecal residue (figure 2), revealing the current limitations of semi-automatic features in segments with sub-optimal preparation.

Currently, steps are being taken to optimise the time-efficiency of semi-automatic MRI measurements and to provide full integration in viewer software. Clearly, these aspects are essential for clinical applicability, which requires easy to use techniques.

In conclusion, the use of semi-automatic features for assessment of patients with CD provides improved reproducibility over conventional activity scores, favouring its use for therapy evaluation and monitoring of disease activity. Accurate and reproducible MRI scores could improve the physician's trust in these scores to make consistent and effective treatment decisions.

## Appendix 1

Calculation of the London score, the Magnetic Resonance Index of Activity (MaRIA) and the relative contrast enhancement (RCE) using bowel wall signal intensity (SI) measured in a region of interest:

*London score*

$$= 1.79 + 1.34 \times \text{Wall thickness} + 0.94 \times \text{mural T2 signal}$$

$$\text{MaRIA} = 1.5 \times \text{Wall thickness (mm)} + 0.02 \times \text{RCE} + 5 \times \text{edema} + 10 \times \text{ulceration}$$

$$\text{RCE} = \frac{\text{SI postcontrast} - \text{SI precontrast}}{\text{SI precontrast}}$$

\* RCE calculation did not include a noise correction term, as was used by Rimola *et al* [2], since inconsistent noise measurements were observed in our data, yielding arbitrary RCE values.

## Appendix 2: Prospective cohort: individual MRI features

Correlations with CDEIS and interobserver agreement for individual subjective and semi-automatic MRI features can be found in table 4. For Ob1, weak to moderate correlations were seen for mural T2, T1 enhancement, bowel wall thickness, RCE and all semi-automatic features, with A1 and excess volume showing the highest correlation to CDEIS ( $r=0.51$ ,  $p<0.001$ ). For Ob2, weak to moderate correlations were seen for all subjective and semi-automatic features, with A1 again showing the highest correlation to CDEIS ( $r=0.55$ ,  $p<0.001$ ). Very good interobserver agreement was seen for ABWT mean, ABWT excess volume and the A1 feature, while good agreement was seen for RCE and ABWT features and moderate agreement was seen for enhancement pattern and length. Perimural T2, T1 enhancement, bowel wall thickness, comb sign and ulcers showed fair agreement, while mural T2 and edema showed poor agreement. The low agreement for mural T2 can be partially explained by a large number of one grade

differences (25/53), while only a low number of two or three grade differences were seen (11/53).

MRI features	Observer 1 n=69		Observer 2 n=68		Interobserver n=53	agreement*
	r	p Value	r	p Value	Coefficient (95% CI)	
Mural T2 signal	<b>0.29</b>	<b>0.03</b>	<b>0.44</b>	<b>&lt;0.001</b>	0.05 (0.00–0.18)	
Perimural T2 signal	0.17	0.15	<b>0.36</b>	<b>&lt;0.01</b>	0.34 (0.10–0.58)	
T1 Enhancement	0.15	0.22	<b>0.42</b>	<b>&lt;0.001</b>	0.31 (0.11–0.52)	
Enhancement pattern	<b>0.25</b>	<b>0.04</b>	<b>0.39</b>	<b>0.001</b>	0.48 (0.26–0.69)	
Wall thickness (mm)	<b>0.42</b>	<b>&lt;0.001</b>	<b>0.44</b>	<b>&lt;0.001</b>	0.30 (0.01–0.54)	
Length	0.23	0.06	<b>0.28</b>	<b>0.02</b>	0.52 (0.34–0.71)	
RCE	<b>0.28</b>	<b>0.02</b>	<b>0.32</b>	<b>&lt;0.01</b>	0.64 (0.44–0.77)	
Comb sign	0.23	0.06	<b>0.26</b>	<b>0.03</b>	0.35 (0.10–0.61)	
Edema	0.17	0.16	<b>0.30</b>	<b>0.01</b>	0.14 (0.00–0.40)	
Ulcers	0.13	0.30	<b>0.26</b>	<b>0.03</b>	0.25 (0.02–0.48)	
Automated features						
Maximum wall thickness	<b>0.30</b>	<b>0.01</b>	<b>0.36</b>	<b>&lt;0.01</b>	0.76 (0.61–0.85)	
Mean wall thickness	<b>0.26</b>	<b>0.03</b>	<b>0.34</b>	<b>&lt;0.01</b>	0.87 (0.78–0.92)	
Excess volume	<b>0.51</b>	<b>&lt;0.001</b>	<b>0.37</b>	<b>&lt;0.01</b>	0.81 (0.68–0.88)	
A1	<b>0.51</b>	<b>&lt;0.001</b>	<b>0.55</b>	<b>&lt;0.001</b>	0.82 (0.70–0.90)	

\* Inter-rater reliability values are intraclass correlation coefficients for continuous variables, linear weighted kappa values for ordinal and binary variables

DWI, diffusion weighted imaging; MRI, magnetic resonance imaging; r, spearman correlation coefficient; RCE, relative contrast enhancement.

Table 4.4 Correlations between individual MRI features and CDEIS in the terminal ileum and interobserver agreement of individual MRI features





# 5 Simulation of scanner- and patient-specific low-dose CT imaging from existing CT images

---

## *Abstract*

**Purpose:** Simulating low-dose Computed Tomography (CT) facilitates in-silico studies into the required dose for a diagnostic task. Conventionally, low-dose CT images are created by adding noise to the recorded projection data. However, this is not always achievable in practice as the raw data are simply not available. This chapter aims to present a new method for simulating patient-specific, low-dose CT images without the need of the original projection data. The methods assume fan beam imaging and image reconstruction by parallel-beam filtered-backprojection merely to proof the principles of our method and to show that representative noise distributions are obtained.

**Methods:** The low-dose CT simulation method included the following steps: (1) computation of a virtual sinogram from a high dose CT image by means of the radon transform; (2) simulation of an 'reduced'-dose sinogram with appropriate amounts of noise; (3) subtraction of the high-dose virtual sinogram from the reduced-dose sinogram; (4) reconstruction of a noise volume via filtered back-projection; (5) addition of the noise image to the original high-dose image. The required scanner-specific parameters were retrieved from calibration images of a water cylinder. The apodization window was obtained from the noise power spectrum (NPS) in a small region of interest in the center of those images. Furthermore, the bowtie filter attenuation characteristics were derived from the pixel variance. Finally, the X-ray tube output parameter (reflecting the photon flux) and the detector read-out noise were computed from the pixel variance at various exposure levels. The low-dose simulation method was evaluated by comparing the noise characteristics in simulated images with experimentally acquired data.

**Results:** We found that the models used to recover the scanner specific parameters fitted accurately to the calibration data. What is more, the retrieved apodization window of the reconstruction filter, the bowtie filter, the X-ray tube output parameter and the detector read-out noise were comparable to values reported in literature. Finally, the simulated low-dose images accurately reproduced the noise characteristics in experimentally acquired low-dose-volumes.

**Conclusion:** The developed methods truthfully simulate low-dose CT imaging, without requiring projection data. The scanner-specific parameters can be estimated from experimentally acquired calibration data. The new methodology could aid in further optimizing CT protocols by facilitating in-silico studies on dose dependency of low contrast object detectability.

## **5.1 Introduction**

Since the early experiments by Sir Godfrey Hounsfield and Allan McLeod Cormack in the 1950's and 1960's, computed tomography (CT) has established itself as one of the most important medical imaging modalities[4]. In fact, the number of CT examinations is still increasing[23]. An important disadvantage of CT, however, is the exposure to ionizing radiation that is inherent to the technique. Accordingly, it is common practice to keep the radiation dose as low as reasonable achievable (ALARA). Unfortunately, lowering the dose yields a lower signal-to-noise ratio and thus a poorer image quality which may hamper subsequent diagnosis. Optimization of the dose/quality trade-off is a far from trivial problem as one cannot simply expose subjects to a range of radiation doses for ethical reasons. Therefore, a lower-dose CT image is usually simulated by adding noise to the underlying projection data, i.e. the sinogram[24-26, 100-103]. Subsequently, the lower-dose image is reconstructed from these noisy projections using the scanner's software. However, this approach is not always achievable in practice as the projection data are often simply not available. This chapter studies a method to generate low-dose CT images based on existing image data. Therefore, we introduce new methodology to determine key system parameters such as the reconstruction kernel, bowtie filter, the X-ray tube output and the read-out noise by a simple calibration procedure. These system parameters determine the noise properties of the simulated low dose CT-images. Furthermore, retrospective investigation of the influence of low-dose imaging might be permitted if one could generate such data directly from existing images. Obviously, this requires that the simulation process complies with the physics of image formation to produce reliable lower-dose CT volumes.

### **5.1.1 Related Work**

Mayo et al[25] and Frush et al. [102] were among the first to simulate low-dose CT images. They added Gaussian noise to the projection data, after which the images were generated by means of the scanner's reconstruction software. Any such approach assumes that the number of photons hitting the detector is large. However, when only a low number of photons is detected, the properties of the noise in the sinograms become much more complex. Then, the readout noise

becomes significant and the measured signal is best described by compound Poisson statistics[104-106]. Still, many low-dose CT-simulators have merely added Gaussian noise to the raw projection data[26, 101, 107, 108]. Zabic et al [109] extended the noise model to correctly reflect the noise (co)variances under photon-starvation conditions and appropriately simulate detector noise artifacts. Furthermore, Wang et al. combined the raw data acquired at two tube-voltages, which allowed also simulating adjustments in the tube-voltage[110]. Similarly, Wang et al [111] present a method for generating simulated low-dose cone-beam CT (CBCT) preview images. Essentially, correlated noise is injected into the original projections after which images are reconstructed using both conventional filtered backprojection (FBP) and an iterative, model-based image reconstruction method (MBIR).

Simultaneously, the need for meaningful characterization of image noise beyond that offered by pixel standard deviation became increasingly important[26, 101, 102, 112]. Boedeker et al. [113] and Faulkner et al. [114] proposed to use the NPS and the noise equivalent quanta (NEQ) to describe the noise properties in CT images, whereas Joemai et al. [101] used the NPS and variance to validate their low-dose CT model.

The work that was done to describe the NPS of CT images also yielded techniques to estimate the reconstruction kernel. This proved very valuable, since manufacturers are often reluctant to disclose the kernels. Other scanner-specific parameters, such as the bowtie filter and the readout noise, were derived from the projection data [26] [106][104]. To the best of our knowledge the bow tie filter was never derived from actual image data.

Initially, Wang et al. [110] and Kim and Kim [115] presented preliminary work on simulating low-dose CT scans from the reconstructed images. Wang et al aimed to develop a simulation technique based on image data such that it produced similar results as a method using projection data. Kim and Kim [116] presented a comprehensive, image-based framework for reduced-dose CT simulation. The key characteristics of the CT system are estimated in this work based on several measurements of a tapered, cylindrical phantom: the reconstruction filter, noise parameters and the photon flux of the X-ray tube. Subsequently, reduced dose CT

noise images are generated by means of a synthesized sinogram. The noise equivalent quanta (NEQ) is a key parameter that is used to determine the system parameters. Essentially, it reflects the (squared) SNR in a CT image, measured in the noise image of a uniform object. Kim and Kim adopt a linear relation between NEQ per detector element and the NEQ per image to specify the amount of noise that has to be added. This relation was derived (amongst others) by Wagner [117] and Hanson [118] assuming that the attenuation at the projection angles is uniform.

### **5.1.2 Objective**

This chapter presents a new framework to simulate lower-dose CT imaging from existing CT images without the original projection data. We take a different approach to image-based low dose CT simulation compared to the chapter by Kim et al[116]. Particularly, our method to estimate the system parameters relies on the variance in signal intensity reflecting the noise level. While doing so we do not need to assume that the noise properties of the projections are uniform. A further novelty of our work is the estimation of the bowtie filter.

Our simulation method first creates a virtual sinogram from a high-dose CT image, which is processed to yield one corresponding to a lower dose. Subsequently, the high-dose virtual sinogram is subtracted from the lower-dose sinogram. The resulting noise sinogram is used to reconstruct a noise volume via filtered back-projection. Addition of the noise volume to the original high-dose image results in the lower-dose image.

By forward modeling the entire acquisition process, a spatially varying noise distribution is generated. The noise characteristics depend on the actual object that is being imaged as well as the aforementioned acquisition parameters. The entire approach is validated by means of the NPS derived from separate CT images of a water barrel and an anthropomorphic, pelvic phantom.

The methods assumes that a virtual X-ray source produces monochromatic photons with energy equal to the effective energy of the original, polychromatic X-ray tube (as in Refs.[115, 119]). Furthermore, in-plane, fan beam imaging and

image reconstruction by filtered-backprojection is assumed to prove the principles of our method. We consider the extension of our methods to cone beam imaging and reconstruction still an obvious next step of our work.

The chapter is organized as follows. Section 5.2 describes the low-dose simulation method. Subsequently, Section 5.3 goes into how several system parameters can be estimated from CT images. Section 5.4 lists the experiments that are done in order to measure the parameters 5.4.2 and to validate the model 5.4.3. The outcome is discussed in Section 5.5.

## **5.2 Lower-dose CT simulator**

The lower-dose CT simulator consists of nine steps (see Figure 5.1):

1. An attenuation image  $\mu_{\text{high}}$  is constructed from a high-dose CT image  $J_{\text{high}}$ ;
2. A virtual sinogram  $R_{\text{high}}$  is generated from  $\mu_{\text{high}}$  by means of the Radon transform[2, 120]. Note that computing the radon transform requires interpolation, hence this virtual sinogram is slightly more blurred than the true sinogram ;
3. A virtual, noiseless measurement of the number of detected photons  $N_{\text{det, high}}$  is generated from  $R_{\text{high}}$ ;
4. A virtual, noisy measurement  $N_{\text{red}}$  at reduced dosed is created with appropriate amounts of Poisson and Gaussian noise — reflecting the quantum and readout noise components [4] as well as the noise already present in the original high-dose image;
5. A virtual, 'reduced'-dose sinogram  $R_{\text{red}}$  is computed from  $N_{\text{red}}$ . This ensures that all smoothing effects that are inherent to the discrete radon transform (step 2) and the discrete inverse radon transform (step 7) do not impose additional blurring to the object being imaged;
6. A noise sinogram  $R_{\text{noise}}$  is obtained by subtracting the virtual sinogram  $R_{\text{high}}$  from the reduced-dose sinogram  $R_{\text{red}}$ ;
7. A noise attenuation image  $\mu_{\text{noise}}$  is reconstructed by means of the inverse Radon transform ---via filtered back-projection (FBP)--- from the noise sinogram  $R_{\text{noise}}$ ;
8. A noise image  $J_{\text{noise}}$  is constructed from  $\mu_{\text{noise}}$ ;
9. A low-dose attenuation image  $J_{\text{low}}$  is formed by adding  $J_{\text{noise}}$  ---which contains noise only--- to the original high-dose CT image  $J_{\text{high}}$  ;

In the next subsections, we will detail steps 1--9.

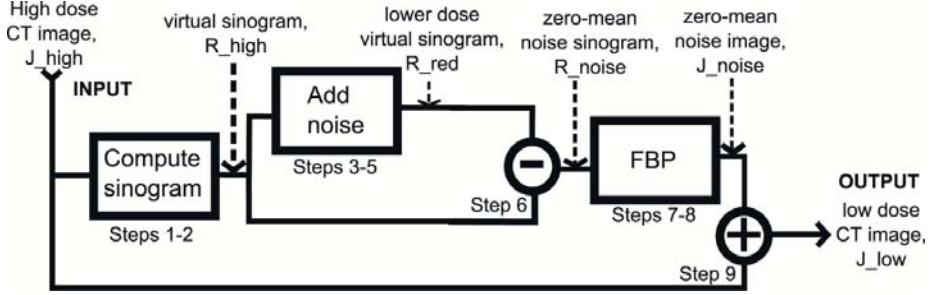


Figure 5.1: Schematic overview of the low-dose CT simulator

### 5.2.1 The virtual sinogram (Steps 1-2)

The attenuation of X-ray radiation on a path from source to detector is described in a discretized form of Lambert-Beer's law by:

$$N_{\text{det}} = N_0 e^{-\sum_{q=0}^{q_{\text{end}}} \mu(q\Delta s)\Delta s}, \quad (5.1)$$

in which  $N_{\text{det}}$  denotes the number of photons that hit the detector,  $N_0$  the number of photons that would hit the detector in case no object is present,  $\Delta s$  the step size,  $q$  the position on the path,  $q_{\text{end}}\Delta s$  the end-position of the path and  $\mu(q\Delta s)$  represents the (position dependent) attenuation coefficient, also called attenuation image. Henceforth, the summation in the exponent of this equation will be called an attenuation projection,  $p$ .

The attenuation image  $\mu_{\text{high}}$  is calculated from the input, high-dose image  $J_{\text{high}}$  in step 1 via:

$$\mu_{\text{high}} = \frac{\mu_{\text{water}} J_{\text{high}}}{1000} + \mu_{\text{water}}, \quad (5.2)$$

in which we use for  $\mu_{\text{water}}$  the attenuation coefficient of water at the effective energy of an X-ray tube of 80 keV[119, 121].

The aggregate of such attenuation projections in a parallel-beam scanner geometry can be approximated from the high-dose attenuation image using the Radon transform:



$$P_{\text{high}}(k, l) = \sum_{n=1}^{N_n} \sum_{m=1}^{N_m} \mu_{\text{high}}(n, m) \delta(n \cos(k\Delta\theta) + m \sin(k\Delta\theta) - l\Delta t) \Delta x \Delta y, \quad (5.3)$$

where  $\delta(\cdot)$  is the delta function,  $k\Delta\theta$  the gantry angle,  $l\Delta t$  the position of a point on a straight line intersecting the isocenter,  $(n, m)$  the pixel coordinates, and  $\Delta x$  and  $\Delta y$  the pixel sizes.

For a fan-beam geometry, however, Eq. (5.3) is not valid. Therefore, the projections  $P_{\text{high}}(k, l)$  need to be reordered in such a way that they correspond to the projections in fan-beam geometry. In the latter geometry, let  $i\Delta\gamma$  represent the angular position on the detector ring (fan angle) and  $j\Delta\beta$  the offset angle from the line through the source and isocenter (gantry angle). The relation between a parallel projection  $P_{\text{high}}(k, l)$  and a fan-beam projection  $R_{\text{high}}(i, j)$  can be derived to be [120]:

$$R_{\text{high}}\left(\frac{1}{\Delta\beta}\left[k\Delta\theta - \arcsin\left(\frac{l\Delta t}{D_{\text{si}}}\right)\right], \frac{1}{\Delta\gamma}\arcsin\left(\frac{l\Delta t}{D_{\text{si}}}\right)\right) = P_{\text{high}}(k, l), \quad (5.4)$$

where  $D_{\text{si}}$  is the (X-ray) source to isocenter distance.  $R_{\text{high}}(i, j)$  should be considered an approximate, *virtual* sinogram, particularly since the noise is strongly reduced due to all the interpolations and averaging involved in its calculation. Eqs. (5.3) and (5.4) together form step 2. The associated, virtual transmission  $T_{\text{high}}(i, j)$  is calculated by:

$$T_{\text{high}}(i, j) = e^{-R_{\text{high}}(i, j)}. \quad (5.5)$$

## 5.2.2 Adding noise to the virtual sinogram

### 5.2.2.1 Theory (Step 3)

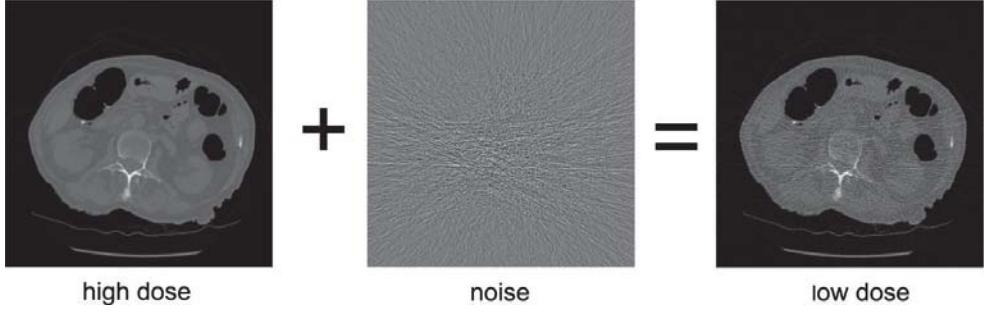


Figure 5.2. A low-dose CT image  $J_{\text{low}}$  (right) is simulated by adding a patient-specific (zero-mean) noise image  $J_{\text{noise}}$  (middle) to the original high-dose image  $J_{\text{high}}$  (left).

Above, we introduced the idea to create a low-dose attenuation image by adding a space-variant object-dependent noise image to the high-dose attenuation image (see Figure 5.2). As such, the pixel variance of the low-dose image  $\text{var}[\mu_{\text{low}}(n, m)]$  is given by

$$\text{var}[\mu_{\text{low}}(n, m)] = \text{var}[\mu_{\text{high}}(n, m)] + \text{var}[\mu_{\text{noise}}(n, m)], \quad (5.6)$$

where  $\text{var}[\mu_{\text{high}}(n, m)]$  and  $\text{var}[\mu_{\text{noise}}(n, m)]$  are the pixel variances of the high-dose image and the noise image, respectively.

As a result of filtered back-projection, *any* attenuation image  $\mu(n, m)$  is a weighted sum of attenuation projections  $R(i, j)$ :

$$\mu(n, m) = \left( \frac{\pi \Delta t}{M} \right) \sum_{j=1}^M \sum_{i=-N}^{N-1} c_{\text{tot}}(n, m, i, j) R(i, j), \quad (5.7)$$

in which  $c_{\text{tot}}(n, m, i, j)$  represents the coefficients of the reconstruction filter (including all the interpolation steps),  $2N$  the number of detector elements and  $M$  the number of gantry angles composing a full revolution.

Since the noise in the projections is assumed to be independent, the pixel variance equals

$$\text{var}[\mu(n, m)] = \left( \frac{\pi \Delta t}{M} \right)^2 \sum_{j=1}^M \sum_{i=-N}^{N-1} c_{\text{tot}}(n, m, i, j)^2 \text{var}[R(i, j)]. \quad (5.8)$$

According to Eq. (5.8), the correct noise characteristics are created when each attenuation projection contains an appropriate amount of noise. Thus, the following condition should be satisfied:

$$\text{var}[R_{\text{low}}(i, j)] = \text{var}[R_{\text{high}}(i, j)] + \text{var}[R_{\text{noise}}(i, j)], \quad (5.9)$$

in which  $\text{var}[R_{\text{high}}(i, j)]$  is the variance in the high-dose sinogram,  $\text{var}[R_{\text{low}}(i, j)]$  the variance in a sinogram acquired at the specified dose and  $\text{var}[R_{\text{noise}}(i, j)]$  the variance in the noise sinogram.

In general, the variance of an attenuation projection  $\text{var}[R(i, j)]$  can be approximated by a first order Taylor series approximation [106, 122, 123] as:

$$\text{var}[R(i, j)] \approx \frac{1}{N_{\text{det}}(i, j)} + \frac{\sigma_e^2}{N_{\text{det}}^2(i, j)}. \quad (5.10)$$

Eq. (5.10) consists of two terms. The first reflects the quantum (photon) noise, which obeys a Poisson distribution. The second term represents the readout noise, which is modeled by zero-mean Gaussian noise  $G(0, \sigma_e^2)$  with variance  $\sigma_e^2$ . The number of detected photons  $N_{\text{det}}(i, j)$  is given by

$$N_{\text{det}}(i, j) = N_0(i)T(i, j), \quad (5.11)$$

where  $T(i, j)$  is the transmission at fan angle  $i\Delta\gamma$  and gantry angle  $j\Delta\beta$ .  $N_0(i)$  is a function of the fan angle  $i\Delta\gamma$  due to the bowtie filter. The bowtie filter is incorporated by means of its transmission coefficients as a function of fan angle. Furthermore,  $N_0$  can be defined as a function of protocol- and scanner-dependent variables:

$$N_0(i) = \frac{Kwd_{\text{fan}}I\tau}{M}T_B(i), \quad (5.12)$$

in which  $w$  is the collimation (width of the fan beam),  $d_{\text{fan}}$  the detector size in the fan angle direction at the isocenter,  $\tau$  the rotation time,  $I$  the tube current,  $T_B(i)$  the transmission of the bowtie filter at fan angle  $i\Delta\gamma$ , and  $K$  a constant reflecting the X-ray tube output in photons/(mAs.mm<sup>2</sup>). The parameters  $K$  and  $T_B(i)$  are scanner-specific and need to be estimated using calibration scans (if not known a-priori), for which a procedure is detailed in Section 5.4.2. The other parameters in

Eq. (5.12) can be typically retrieved from the literature [4, 124] or are included in the DICOM-header. Note that the radiation dose is steered via the exposure  $I\tau$ . Essentially, step 3 consists of calculating  $N_{det}$  using Eqs. (5.5), (5.11) and (5.12).

### **5.2.2.2 Model Implementation (Steps 4-6)**

This section describes how  $R_{red}(i, j)$  is created in such a way that Eq. (5.9) is fulfilled. Note that the noise sinogram  $R_{noise}(i, j)$  is obtained via  $R_{noise}(i, j) = R_{red}(i, j) - R_{high}(i, j)$ . Here,  $R_{high}(i, j)$  is the virtual sinogram (obtained via Eq. (5.4)) and not the real sinogram associated with  $\mu_{high}(i, j)$ . Due to the interpolations and averaging in the calculation of the discrete radon transform, the noise of  $R_{high}(i, j)$  is assumed to be negligible, so that  $\text{var}[R_{noise}(i, j)] \approx \text{var}[R_{red}(i, j)]$ .

Eq. (5.10) indicates that the quantum noise and the readout noise can be added to yield the total noise. Accordingly,  $R_{red}(i, j)$  is calculated in two steps.

First, only the quantum noise inherent to the detected number of photons is simulated. Essentially, this is implemented by drawing samples from a Poisson distributions with expectation value  $N_{red, q}(i, j)$ .  $N_{red, q}(i, j)$  is calculated from Eq. (5.11). Substituting Eqs. (5.10), (5.11) and (5.12) into Eq. (5.9) yields an explicit relation between  $I_{low}$ , chosen a-priori,  $I_{high}$ , given by the high-dose image, and  $I_{red}$ , which is the tube current that yields the correct amount of noise to be used for creating the sinogram:

$$I_{red} = \frac{I_{high} I_{low}}{I_{high} - I_{low}} \quad (5.13)$$

in which  $I_{high}$  and  $I_{low}$  are the tube currents of the high-dose and the low-dose image to be simulated. Note that this equation essentially compensates for quantum noise already present in the high-dose image.

Second, readout noise is simulated by repeatedly drawing samples  $N_{red, r}(i, j)$  from a Gaussian distribution with mean zero and variance  $\sigma_{red}^2$ . Given that  $N_{red, r}$  is calculated as described previously,  $\sigma_{red}^2$  is computed by:

$$\frac{\sigma_{red}^2}{N_{red, q}^2} = \frac{\sigma_e^2}{N_{det, low}^2} - \frac{\sigma_e^2}{N_{det, high}^2}. \quad (5.14)$$

This equation is simplified by substituting Eqs. (5.11), (5.12) and (5.13) into Eq. (5.14) and dropping all redundant terms to yield:

$$\sigma_{\text{red}}^2 = \sigma_e^2 \frac{I_{\text{high}}^2 - I_{\text{low}}^2}{(I_{\text{high}} - I_{\text{low}})^2}. \quad (5.15)$$

As before, this equation compensates for Gaussian noise already present in the high-dose data. The virtual noisy measurement (step 4) is found by adding the read-out noise to the Poisson process

$$N_{\text{red}}(i, j) = N_{\text{red,q}}(i, j) + N_{\text{red,r}}(i, j). \quad (5.16)$$

We adjust  $N_{\text{red}}$  and set it to one, whenever the equation delivers a number that is smaller than one (photon starvation), which corresponds to a very small  $N_{\text{red}}(i, j)/N_0(i, j)$  as  $N_0 > 1$ . Hence, the virtual 'reduced' dose sinogram (step 5) becomes

$$R_{\text{red}}(i, j) = -\ln(N_{\text{red}}(i, j) / N_0(i, j)), \quad (5.17)$$

which yields the noise sinogram (step 6)

$$R_{\text{noise}}(i, j) = R_{\text{red}}(i, j) - R_{\text{high}}(i, j). \quad (5.18)$$

### 5.2.3 Reconstructing the image from the noisy sinogram (Steps 7-9)

In step 7, we opt to reconstruct  $\mu(i, j)$  in a parallel-beam geometry, because it is computationally much less expensive and because our scanner uses such a reconstruction approach[120]. Therefore, the fan-beam projections  $R_{\text{noise}}(i, j)$  need to be reordered into parallel-beam projections  $P_{\text{noise}}(k, l)$  by inverting Eq.(5.4) :

$$P_{\text{noise}}\left(\frac{1}{\Delta\theta}(j\Delta\beta + i\Delta\gamma), \frac{1}{\Delta t}D_{\text{si}}\sin(i\Delta\gamma)\right) = R_{\text{noise}}(i, j), \quad (5.19)$$

Just as before, uniformly sampled parallel projections  $P(k, l)$  are obtained by linear interpolation of  $R(i, j)$ . Subsequently, filtered back-projection is used to construct a noise image via:

$$\mu_{\text{noise}}[n, m] = \left( \frac{\pi \Delta t}{M} \right) \sum_{k=1}^{M_{\text{par}}} \sum_{l=-N_{\text{par}}}^{N_{\text{par}}-1} P_{\text{noise}}(k, l) \cdot c(n \Delta x \cos(k \Delta \theta) + m \Delta y \sin(k \Delta \theta) - l \Delta t), \quad (5.20)$$

in which  $c$  are the reconstruction filter coefficients, and  $M_{\text{par}}$  and  $N_{\text{par}}$  the number of gantry angles and detector elements in the parallel beam geometry. Note that  $c$  only contains the coefficients of the apodized ramp filter while  $c_{\text{tot}}(i, j, n, m)$  also incorporates the necessary interpolation steps. Clearly, the reconstruction filter coefficients are scanner- and protocol-specific. A calibration procedure for obtaining the filter coefficients is described below. Next,  $\mu_{\text{noise}}(n, m)$  was scaled to Hounsfield units in step 8 by:

$$J_{\text{noise}}(n, m) = 1000 \left( \frac{\mu_{\text{noise}}(n, m) - \mu_{\text{water}}}{\mu_{\text{water}}} \right). \quad (5.21)$$

Finally, the low-dose image can be obtained in step 9 as:

$$J_{\text{low}}(n, m) = J_{\text{high}}(n, m) + J_{\text{noise}}(n, m). \quad (5.22)$$

### 5.3 Parameter estimation

This section describes how the required scanner- and scan-specific parameters may be computed from calibration scans:

1. The reconstruction filter coefficients in  $c(l)$  as well as  $c_{\text{tot}}(n, m, i, j)$
2. The bowtie filter transmission  $T_B(i)$
3. The X-ray tube output parameter  $K$
4. The readout noise variance  $\sigma_e^2$

#### 5.3.1 The reconstruction filter coefficients

Eq. (5.7) can be rewritten in matrix notation in which a system matrix  $C_{\text{tot}}$  contains the filter coefficients  $c_{\text{tot}}(n, m, i, j)$ :

$$\boldsymbol{\mu} = \mathbf{C}_{\text{tot}} \mathbf{R}, \quad (5.23)$$

where  $\mu$  is a  $N_{pix}$  vector (the reconstructed image),  $C_{tot}$  is  $N_{pix} \times M2N$  matrix, and  $R$  is a  $M2N$  vector holding the attenuation projections.  $N_{pix}$  is the number of pixels in the image to be reconstructed.

In our implementation  $C_{tot}$  actually consists of a series of matrix multiplications, each representing a different processing step:

$$C_{tot} = C_{back} C_{filt} C_{fan2par}. \quad (5.24)$$

Here,  $C_{fan2par}$  is a  $M_{par}2N_{par} \times M2N$  matrix that implements the fan-beam to parallel-beam transformation;  $C_{filt}$  is a  $M_{par}2N_{par} \times M_{par}2N_{par}$  Toeplitz matrix performing the high-pass filtering and  $C_{back}$  is a  $N_{pix} \times M_{par}2N_{par}$  matrix that incorporates the back-projection step. Observe that  $C_{fan2par}$  and  $C_{back}$  can be derived from the beam geometry of the simulation. The matrix  $C_{filt}$  has the filter coefficients  $c(l)$  in its rows. Essentially,  $c(l)$  is the  $M_{par}2N_{par}$  'core' backprojection filter that needs to be estimated in order to compute  $c_{tot}(n, m, i, j)$ .

Conventionally, the reconstruction filter  $c_{tot}(n, m, i, j)$  is derived from the NPS in a region of interest (ROI)[113, 114, 116, 125]. If all fan-beam projections  $R(i, j)$  used to reconstruct the ROI contain white noise (and aliasing is negligible), the NPS becomes radially symmetric[125]. This is approximately the case in the center of a water cylinder that is placed in the center of the scanner. The pixels inside such a ROI are reconstructed by the projections located at the center of the detector array. The expected number of detected photons is constant, since the water cylinder is locally approximately flat, hence the noise level in each of these projections is the same. Therefore, the noise is approximately white as the amount of cross-talk is negligible[4].

Here, the NPS, which is the Fourier transform of the autocorrelation function, is radially symmetric and given by:

$$NPS(\omega_r) = H_{tot}^2(\omega_r) S(\omega_r) \quad (5.25)$$

where  $\omega_r = \sqrt{\omega_x^2 + \omega_y^2}$ ,  $\omega_x$  and  $\omega_y$  are the frequencies in Cartesian coordinates,  $H_{tot}(\omega_r)$  the modulation transfer function of the scanner (see below), and  $S(\omega_r)$  the NPS of the projections. The assumption that  $R(i, j)$  contains white noise makes that the NPS becomes [113, 125]

$$S(\omega_r) \approx \frac{1}{\omega_r}, \quad (5.26)$$

Eq. (5.26) is valid when the sampling of the gantry angles of  $R(i, j)$  is uniform and sufficiently dense.

$H_{\text{tot}}$  is modeled to consist of three elements, namely one apodized ramp filter and two interpolation filters. The apodized ramp filter  $H_{\text{filter}}$  ensures a mathematically correct reconstruction up to the cut-off frequency of the apodization filter. Furthermore, one interpolation filter  $H_{\text{fan2par}}$  represents the transformation of the fan-beam rays to uniformly sampled parallel-beam rays, and the other one  $H_{\text{back}}$  reflects the interpolation along the path during the back-projection. Consequently, when aliasing is ignored,  $H_{\text{tot}}$  is given by:

$$H_{\text{tot}}(\omega_r) = H_{\text{filter}}(\omega_r) H_{\text{back}}(\omega_r) H_{\text{fan2par}}(\omega_r), \quad (5.27)$$

where  $H_{\text{filter}}$  is the Fourier transform of  $c(l)$ , the 'core' backprojection filter, which can be further decomposed into

$$H_{\text{filter}}(\omega_r) = \omega_r H_{\text{apo}}(\omega_r), \quad (5.28)$$

with  $H_{\text{apo}}$  a cut-off window designed to avoid ringing artifacts near large tissue transitions and to suppress noise in the image. The goal is to determine the shape of  $H_{\text{apo}}$ , which we approximate by:

$$H_{\text{apo}}(\omega_r) = \frac{\left( a + b \cos\left(\frac{\pi \omega_r}{f_{\text{par}}}\right) \right)}{a + b}, \quad (5.29)$$

where  $a$  and  $b$  are two filter parameters. For some values of  $a$  and  $b$ ,  $H_{\text{apo}}(\omega_r)$  is equal to cut-off windows found in literature[126]. The NPS is estimated in a ROI using the periodogram, which is an estimate of the NPS and is defined by,

$$\text{NPS}(\omega_x, \omega_y) = \sum_{i=1}^N \frac{|\mathcal{F}\{J_{\text{noise}}(x, y)\}|^2}{N} \quad (5.30)$$



in which  $N$  is the number of images used to estimate the NPS,  $J_{\text{noise}}$  is a zero-mean noise image (e.g. obtained by subtracting repeated images of the same slice) and  $\mathcal{F}\{\}$  symbolizes the Fourier transform. We assume that only linear interpolation is used and therefore Eq. (5.25) becomes (after filling in the previous equations):

$$\text{NPS}(\omega_r) = \omega_r \left( \frac{a + b \cos\left(\frac{\pi \omega_r}{f_{\text{par}}}\right)}{a + b} \right)^2 \text{sinc}^4\left(\frac{\omega_r}{f_{\text{fan}}}\right) \text{sinc}^4\left(\frac{\omega_r}{f_{\text{par}}}\right) \quad (5.31)$$

in which  $f_{\text{fan}}$  is the Nyquist frequency of the detector array with detectors of size  $d_{\text{fan}}$  at the isocenter and  $f_{\text{par}}$  is the Nyquist frequency of the rebinned detectors of size  $d_{\text{par}}$ .

Essentially, the parameters  $a$  and  $b$  are estimated by fitting the model described in Eq. (5.31) to the NPS measured in a ROI in the center of a water cylinder placed in the center of the scanner. For that we use the Levenberg-Maquardt optimization algorithm. Thereafter,  $c(l) = \mathcal{F}^{-1}\{H_{\text{apo}}(\omega_r)\}$  and  $c_{\text{tot}}(n, m, i, j)$  is obtained via Eq. (5.24).

### 5.3.2 The bowtie filter

We will now demonstrate how the bowtie filter transmission can be estimated in a least-squares sense from the pixel variance measurements in a phantom. This is an additional novelty of our work, which is required if projection data cannot be obtained.

Assuming that phantom images are acquired at a dose that is high enough to ignore the electronic noise, then the pixel variance  $\text{var}[\mu(n, m)]$  can be rewritten by substituting Eqs. (5.10), (5.11) and (5.12) into Eq. (5.8):

$$\text{var}[\mu(n, m)] = \left( \frac{M^2 \pi^2 \Delta t^2}{I w d_{\text{fan}} K \tau} \right) \sum_{j=1}^M \sum_{i=-N}^{N-1} \frac{c_{\text{tot}}(n, m, i, j)^2}{T_B(i) T(i, j)}. \quad (5.32)$$

Apart from  $K$  and  $T_B(i)$ , the other parameters in this equation are known ( $c_{\text{tot}}$  may be estimated following the procedure from the previous section and  $T$  is obtained via Eq. (5.5)). Writing this equation in matrix notation yields:

$$\text{var}[\boldsymbol{\mu}] = A \mathbf{C}_{\text{tot}}^2 \frac{1}{\mathbf{T}_B \cdot \mathbf{T}}, \quad (5.33)$$

where  $\text{var}[\boldsymbol{\mu}]$  is an  $N_{\text{pix}}$  vector (with  $N_{\text{pix}}$  again the number of pixels);  $A = \frac{M2\pi2\Delta t2}{lwd_{\text{fan}}K\tau}$  is a scalar;  $(\cdot)^2$  reflects element-wise square (of the  $N_I \times M2N$  matrix  $\mathbf{C}_{\text{tot}}$ , see above);  $\frac{1}{(\cdot)}$  corresponds to an element-wise division;  $(\cdot)(\cdot)$  corresponds to an element-wise multiplication;  $\mathbf{T}_B$  is an  $M2N$  vector representing the bowtie filter transmission, and  $\mathbf{T}$  is an  $M2N$  vector containing the transmission values.

Eq. (5.33) can be rewritten by putting the elements of  $\mathbf{T}$  into a  $M2N \times M2N$  diagonal matrix  $\mathbf{D}_T$  and by using the property that  $\mathbf{T}_B$  does not depend on the gantry angle into:

$$\text{var}[\boldsymbol{\mu}] = \mathbf{C}_{\text{tot}}^2 \frac{A}{\mathbf{D}_T \mathbf{D}_M \mathbf{T}_B} \quad (5.34)$$

in which  $\mathbf{D}_M$  is a  $M2N \times 2N$  matrix that replicates the bowtie filter for all gantry angles and  $\mathbf{T}_B$  is now a  $2N$  vector representing the transmission of the bowtie filter. In fact  $\mathbf{D}_M$  consists of  $M$  'stacked' ( $2N \times 2N$ ) unit matrices.

Clearly, Eq. (5.34) is a linear equation that might be solved analytically. Unfortunately, the system matrix and the number of parameters are very large, and therefore such a purely analytic approach is computationally very expensive. Therefore, we opt to model the bowtie filter's transmission by:

$$T_B(i) = o_{\text{low}} + (1 - o_{\text{low}}) \left( \sum_{q=0}^4 a_q \cos\left(\frac{\pi q i}{N}\right) \right)^2. \quad (5.35)$$

The summation in the equation represents a truncated fourth order Fourier series. Only cosines are used as the bowtie filter is a symmetric function and the sum is squared to ensure that the transmission is not negative. Furthermore, we assume that the transmission in the center of the bowtie filter is one ( $T_B(0) = 1$ ), which is

imposed by constraining  $a_0 = 1 - \sum_{q=1}^4 a_q$ . Thereby, we assume that the transmission also has a lower limit (as the bowtie filter itself has a finite thickness). To that end  $o_{\text{low}}$  is a constant representing the minimal transmission. Furthermore, the weighted-sum construction ensures that the transmission can indeed converge to 1.

We assert that  $T_B$  is a monotonically decreasing function to both sides of the filter. Therefore, we devised the following simple penalty term:

$$P(\mathbf{a}) = \sum_{i=0}^{N-1} \frac{\partial T_B(i, \mathbf{a})}{\partial i} u\left(\frac{\partial T_B(i, \mathbf{a})}{\partial i}\right), \quad (5.36)$$

with  $u(\cdot)$  the Heavyside function and  $\mathbf{a}$  the 4 parameter vector of the model (c.f. Eq. (5.35)). Essentially, Eq. (5.36) sums all positive derivative values over half the filter (which is symmetric by definition).

Finally, the filter parameters are estimated by solving:

$$\mathbf{a}, \mathbf{A} = \arg \min_{\mathbf{a}, \mathbf{A}} \left\{ \sum_{k=1}^{N_{\text{pix}}} \chi_k(\mathbf{a}, \mathbf{A}, k) + \beta P(\mathbf{a}) \right\}, \quad (5.37)$$

with

$$\chi_k(\mathbf{a}, \mathbf{A}, k) = \left( \text{var}[\mu_k]_{\text{experimental}} - \text{var}[\mu_k, \mathbf{a}, \mathbf{A}]_{\text{model}} \right)^2, \quad (5.38)$$

where  $A$  is the scaling constant from Eq. (5.33) which is essentially a gain factor that needs to be simultaneously estimated and  $\beta$  the weight of the penalty term

The required pixel variances to solve this equation will be obtained from a central region in images of a water cylinder that is repeatedly scanned (see below).

### 5.3.3 The X-ray tube output parameter and the readout noise level

Finally, we will demonstrate how the X-ray tube output parameter  $K$  and the variance of the readout noise  $\sigma_e^2$  can be estimated from the pixel variance  $\text{var}[\mu(n, m)]$  measured at different tube currents. Therefore, Eq. (5.8) will be simplified in such a way that  $K$  and  $\sigma_e^2$  can be derived from a linear fit.

First, substituting Eq. (5.10) into Eq. (5.8) yields:

$$\text{var}[\mu(n, m)] = \left( \frac{\pi \Delta t}{M} \right)^2 \sum_{j=1}^M \sum_{i=-N}^{N-1} c_{\text{tot}}(n, m, i, j)^2 \left( \frac{1}{N_{\text{det}}(i, j)} + \frac{\sigma_e^2}{N_{\text{det}}(i, j)^2} \right) \quad (5.39)$$

Subsequently, substituting Eqs. (5.11) and (5.12) into (5.39) and reshuffling the variables yields

$$\text{var}[\mu(n, m)] = \left( \frac{\pi^2 \Delta t^2}{IM^2 w K \tau} \right) \sum_{j=1}^M \sum_{i=-N}^{N-1} \frac{c_{\text{tot}}(n, m, i, j)^2}{T_B(i) T(i, j)} + \left( \frac{\pi \sigma_e \Delta t}{IKM w \tau} \right)^2 \sum_{j=1}^M \sum_{i=-N}^{N-1} \frac{c_{\text{tot}}(n, m, i, j)^2}{T_B(i)^2 T(i, j)^2}. \quad (5.40)$$

Consequently, the next relation emerges when the exposure  $I\tau$ , the X-ray tube output  $K$ , and the variance of the readout noise  $\sigma_e^2$  are separated from all other variables of Eq. (5.40):

$$\text{var}[\mu(n, m)] = \frac{C_p(n, m)}{KI\tau} + \frac{C_e(n, m)}{K^2 I^2 \tau^2} \sigma_e^2, \quad (5.41)$$

in which  $C_p(n, m)$  and  $C_e(n, m)$  are given by

$$C_p(n, m) = \left( \frac{\pi^2 \Delta t^2}{M^2 w} \right) \sum_{j=1}^M \sum_{i=-N}^{N-1} \frac{c_{\text{tot}}(n, m, i, j)^2}{T_B(i) T(i, j)}, \quad (5.42)$$

and

$$C_e(n, m) = \left( \frac{\pi^2 \Delta t^2}{w^2 M^2} \right) \sum_{j=1}^M \sum_{i=-N}^{N-1} \frac{c_{\text{tot}}(n, m, i, j)^2}{T_B(i)^2 T(i, j)^2}. \quad (5.43)$$

$C_p(n, m)$  and  $C_e(n, m)$  can be calculated using the reconstruction filter weights  $c_{\text{tot}}$  (from Section 5.3.1), the bowtie filter (from Section 5.3.2), and the transmission (calculated via Eq. (5.5)). Eq. (5.41) is ill-posed whenever the variance is measured at a single exposure. Therefore, acquisitions are obtained at multiple exposures. Next, the model is fitted in a least-squares sense using the following non-linear model,

$$\{K, \sigma_e^2\} = \arg \min_{K, \sigma_e^2} \left\{ \sum_{k=1}^{N_{\text{pix}}} \sum_{i=1}^{N_I} \chi_{k,i}(K, \sigma_e^2) \right\}, \quad (5.44)$$

with

$$\chi_{k,i}(K, \sigma_e^2) = \left( \text{var}[\mu_{k,i}]_{\text{measured}} - \text{var}[\mu_{k,i} | K, \sigma_e^2]_{\text{model}} \right)^2, \quad (5.45)$$

in which the model is expressed by Eq. (5.41),  $N_{\text{pix}}$  and  $N_I$  define the number of pixels and exposures respectively. The initial parameters are obtained by first solving a simpler problem that emerges when Eq. (5.41) is approximated by two linear equations. If the tube current is very high, the contribution of the read-out noise to the pixel variance is often neglected. In this case, the second term in the right hand side of Eq. (5.41) is ignored, hence  $K$  can be approximated from :

$$\text{var}[\mu(n, m)] \approx \frac{C_p(n, m)}{I \tau K}. \quad (5.46)$$

With  $K$  known,  $\sigma_e^2$  can be estimated using new images that are acquired at a lower dose level. Reshuffling Eq. (5.41) gives:

$$\sigma_e^2 = \frac{K^2 \tau^2 I^2}{C_e(n, m)} \left( \text{var}[\mu(n, m)] - \frac{C_p(n, m)}{K I \tau} \right). \quad (5.47)$$

These estimates for  $K$  and  $\sigma_e^2$  are the initial parameters for minimizing Eq. (5.44).

## 5.4 Results

### 5.4.1 Measurement data

CT images of a water cylinder 34 cm in diameter and an anthropomorphic pelvic phantom were acquired on a Philips Brilliance 64 CT scanner at the Academic Medical Center in Amsterdam, The Netherlands. A modified CT colon protocol was used, since the intended application is CT colonography. The modifications only concerned the tube current, which was adjusted to control the dose level and the acquisition mode, which was sequential for the water cylinder (i.e. imaging the exact same plane) and the pelvic phantom. Table 5.1 and Table 5.2 list the

parameter settings. Note that the scan protocol parameters are controlled by the user whereas the geometry parameters are scanner dependent.

Phantom type	water cylinder	pelvic phantom	
Acquisition mode	sequential	sequential	
Kernel	B	B	
X-ray tube Voltage (kV)	120	120	
Slice Thickness: $w$ (mm)	0.68	0.68	
Collimation	40 x 0.625	64 x 0.625	
Matrix	512 X 512	512 X 512	
Diameter field of measurement:	500	500	
Set type	Calibration/ Training	Test	Test
Field of view	350	350	350
Pixel sizes: $d_{\text{pix}}$ (mm)	0.68	0.68	0.68
Exposure: $I\tau$ (mAs)	250, 120, 60, 30	210, 170, 120, 85, 60, 30, 20, 15	80, 40,15
Number of rotations	25 (except for	13	128
Copies per rotations	40	40	1
Number of slices	1000 (except	520	128

Table 5.1 Scan protocol parameters.

Number of detector rings	64
Source to isocenter distance, $D_{\text{si}}$ (mm)	570
Source to detector distance, $D_{\text{sd}}$ (mm)	1040

Field of measurement: $D_{\text{FOM}}$ (mm)	500
Number of detectors, $N_{\text{det}}$	672
Detector size, $d_{\text{det}}$ (mm)	1.41
Detector size at iso-center, $d_{\text{fan}}$ (mm)	$\sim 0.77$
Sampling after rebinning, $d_{\text{par}}$ (mm)	$d_{\text{fan}}/2$
Number of gantry angles per revolution, $M$	1160

Table 5.2 Scanner's geometry parameters.

CT images of the water cylinder were used to estimate the unknown, scanner-specific parameters:  $H_{\text{apo}}$ ,  $T_B(i)$ ,  $K$  and  $\sigma_e^2$ . Here, the settings listed under 'Calibration/Training' (Table 5.1) were used. Subsequently, separate images of the water cylinder and images of the pelvic phantom were used to validate the low-dose simulation model (settings listed under 'Test'). Therefore, simulated and measured noise characteristics were compared by means of the pixel variances and the NPS.

## 5.4.2 Parameter estimation

### 5.4.2.1 The reconstruction filter coefficients

The volumes emanating from successive X-ray tube rotations at 250mAs were pairwise subtracted to yield 2000 zero-mean noise images (i.e. corresponding slices from successive rotations). As such the attenuation profile of water cylinder is implicitly corrected. The NPS was computed in a small rectangular ROI consisting of 64x64 pixels in the center of the images. Figure 5.3 shows the result.

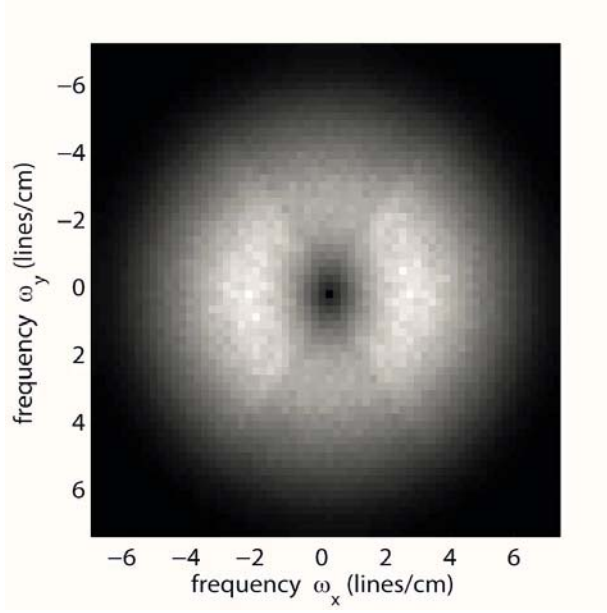


Figure 5.3. The NPS calculated from a 64x64 ROI in the center of a 34cm water cylinder. The NPS was normalized such that it ranged from 0 to 1

Subsequently, the parameters of the apodization window  $H_{\text{apo}}$  were estimated by fitting our model (Eq. (5.31)) to the NPS. The left plot of Figure 5.4 demonstrates how well the model fits the data, while the right plot shows in blue the apodization window which is used in the remainder of the chapter (Eq. (5.29)). In the same plot, the algorithmic transfer function,  $H_{\text{alg}}$ , which is defined by  $H_{\text{alg}}(\omega) = (H_{\text{apo}}(\omega_r)H_{\text{back}}(\omega_r)H_{\text{fan2par}}(\omega_r))^2$  and represents the total apodization, is depicted in pink.



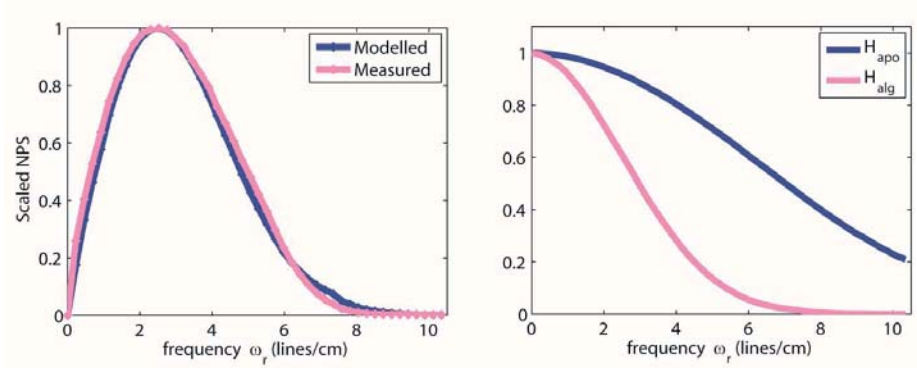


Figure 5.4. Left: the (angular averaged) radial NPS calculated from the images of a water cylinder (pink) and the fitted model, i.e. Eq. (5.31) (blue). Right: the ensuing apodization window  $H_{apo}$  (blue, Eq. (5.29)) and the algorithmic transfer function,  $H_{alg}$ .

#### 5.4.2.2 The bowtie filter

The 2000 zero-mean noise images from the previous step were used to calculate pixel variances (see Figure 5.5). Subsequently, only those variances were retained up to 154 mm from the center, i.e. inside the water cylinder. As such, artifacts at the boundary of the cylinder and problems due to signal clipping outside the cylinder are avoided. Furthermore, samples up to 10 pixels (7mm) from the center were discarded, since the variance of the central pixels cannot be estimated with sufficient precision. Samples were collected from within the remaining region along 80 evenly distributed radial lines drawn outward from the center (see Figure 5.5). The bowtie filter parameters  $a$  were estimated for each of the 80 radial segments separately (via Eq. (5.37)), after which the associated bowtie filter transmissions were averaged. We took this approach for computational reasons since the matrices in Eq. (5.33) are extremely large. The initial parameter setting for every estimation was  $a_{init} = 0.5, 0, 0, 0$ , which corresponds to a single cosine (see below).  $\sigma_{low}$  was set to 0.15, which is comparable to the minimum found in other scanners[26, 104].

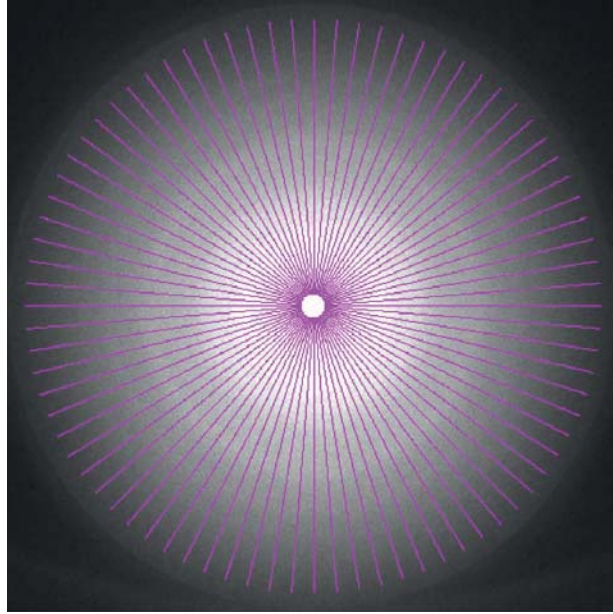


Figure 5.5. Variance image and all of the radial segments. The bowtie filter was calculated for each such segment, i.e. for each separate line on either side of the center.

Figure 5.6 shows the estimated bowtie transmission. The left plot gives the profile of the mean bowtie transmission, its 95% confidence interval (all in blue) as well as the initialization (pink). Notice that the bowtie transmission was estimated more precisely for the central detectors than for the ones at the periphery (reflected in the smaller confidence interval). Particularly, the variance is large for the range of  $|\text{DetectorID}| > 200$ , which corresponds to the edges of the water cylinder.

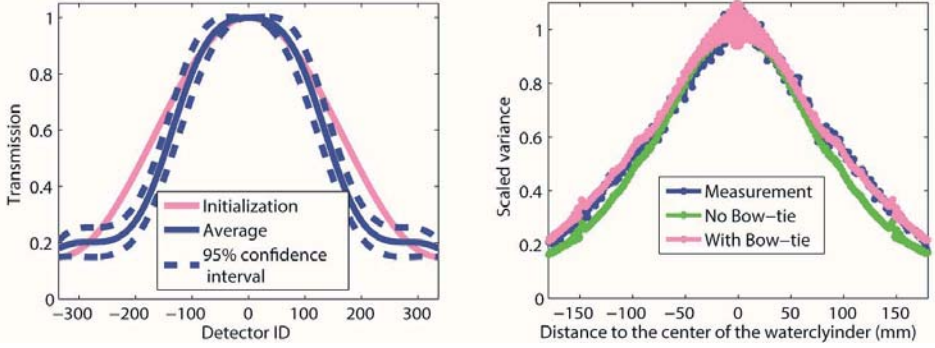


Figure 5.6. left: The mean estimated bowtie filter (blue solid), the 95% confidence intervals (blue dashed) and the bowtie used as initialization (pink). right: the values of the pixels measured (blue) on the diagonal of the variance image depicted in Figure 5.5, the analytically computed variance with (pink) and without (green) bowtie filter. The variance was scaled by dividing the original by the mean.

The right plot displays the variance measured (blue) along one of the radial segments depicted in Figure 5.5. Additionally, the variance as a function of position was computed by means of Eq. (5.34) with (pink) and without the bowtie filter (green). This shows that adding the bowtie filter to the model, enabled it to describe the measured variance more accurately.

The shape of the transmission profile was similar to previously estimated bowtie filters of CT-scanners[26, 104].

#### 5.4.2.3 The X-ray tube output parameter and the readout noise level

The variance per pixel was estimated from the water cylinder images for each exposure level in the training set. As specified in Section 5.4.2.2, the analysis was restricted to pixels positioned within the 7 mm (10 pixels)  $-154$  mm distance interval from the center.

Figure 5.7 shows the ratio ( $\rho$ ) of measured variances at 250 and 120 mAs as a function of the distance to the image center. Neglecting the read-out noise, it follows from Eq. (5.46) that

$$\rho = \frac{\text{var}[\mu(r)]_{I_{\text{low}}}}{\text{var}[\mu(r)]_{I_{\text{high}}}} = \frac{I_{\text{high}}}{I_{\text{low}}} = \frac{250}{120} = 2.08.$$

Figure 5.7 essentially shows that this assumption was only approximately valid at these dose levels as a paired  $t$ -test showed that the pink line varied significantly from the blue one ( $p < 0.05$ ). Nonetheless, the difference was only 2%.

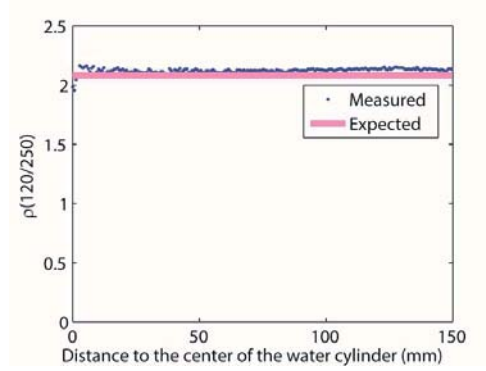


Figure 5.7. The ratio  $\rho$  between pixel variances at 120 mAs and 250 mAs as a function of distance to the center of the image. Measurements are depicted in blue, the expected ratio  $250/120 = 2.08$  is indicated in pink.

Next, Figure 5.8 (left) depicts estimates of  $K$  obtained at different exposure levels.  $K$  was estimated using Eq. (5.46) by fitting a line through all data points, after which the slope of the line corresponds to  $1/(I\tau K)$ . Clearly, the estimation of  $K$  stabilizes at higher exposure levels and increasingly deviates from stability as the exposure level decreases. We attribute this to the increasing importance of the read-out noise due to which Eq. (5.46) is not valid anymore. Henceforth, the estimated value for  $K$  at 250 mAs is used as an initialization to compute the initial value of  $\sigma_e^2$ .

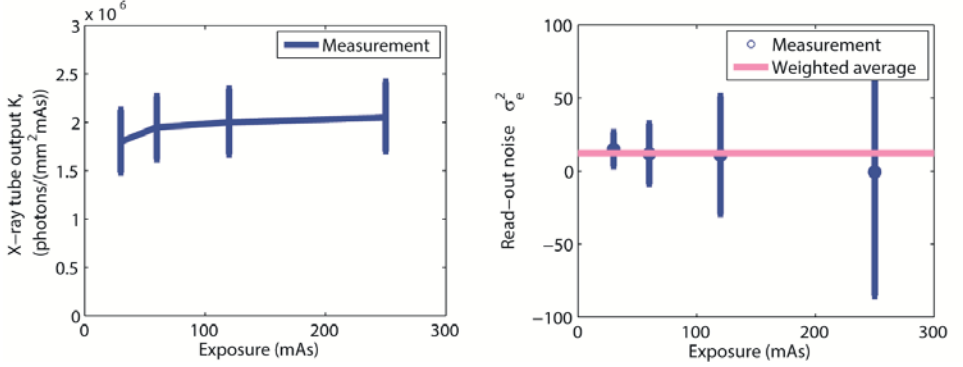


Figure 5.8. left:  $K$  as a function of the exposure. right:  $\sigma_e^2$  as a function of the exposure. The error-bars indicate 2.5-97.5 percentile of the estimation  $K$  and the  $\pm 1$  times the standard deviation of the estimation  $\sigma_e^2$ .

Figure 5.8 (right) displays the read-out noise and uncertainty as a function of exposure in the training data. Table 5.3 lists the estimated values of  $K$  and  $\sigma_e^2$  at each exposure level. The weighted average of estimates of  $\sigma_e^2$  over all exposure levels is used to initialize the minimization procedure described by Eq. (5.44). Here, the weights were inversely proportional to the variances in the estimations of  $\sigma_e^2$ .

The read-out noise  $\sigma_e^2$  was computed using Eq. (5.47), again based on the 7 mm (10 pixels)--154 mm distance interval from the center. Figure 5.9 shows histograms of estimated  $\sigma_e^2$  values at 120 mAs and 30 mAs. Clearly, the distribution is wider at 120 mAs (reflecting less precise estimation), because the total noise is dominated by the Poisson component. Note that the variance can take on negative values due to the subtraction involved in Eq. (5.47). Simultaneously, observe that the mean is larger than zero in each case.

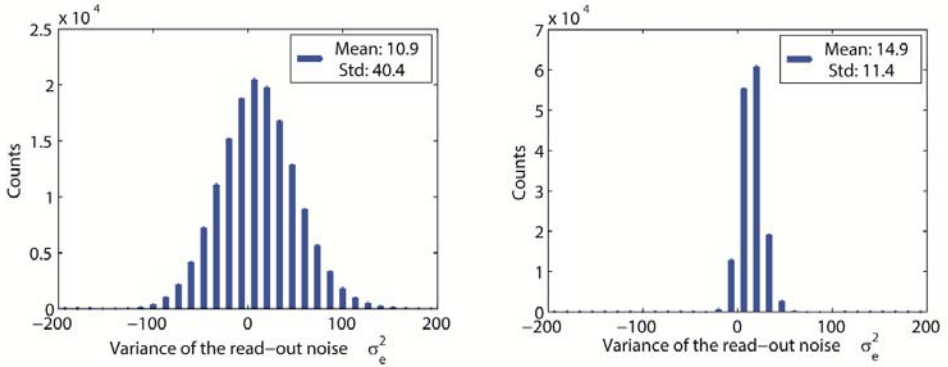


Figure 5.9. Histograms of  $\sigma_e^2$  estimated in the pixels of the water cylinder at 120 mAs (left) and 30 mAs (right).

Exposure (mAs)	250	120	60	30
$K(\times 10^6)$	2.08	2.03	1.98	1.84
$\sigma_e^2$	-0.6	10.9	11.8	14.9

Table 5.3 Estimates for  $K$  and  $\sigma_e^2$  at each exposure level.  $\sigma_e^2$  is computed using  $K$  estimated at 250 mAs.  $K$  is expressed in the number of photons/(mm<sup>2</sup>. mAs.revolution).

Finally, Figure 5.10 displays a contourplot depicting the density distribution as a function of the pixel-variances and  $C_p$  within the 7 mm (10 pixels) --154 mm ROI of the 250 mAs images. Notice how the pink line (solid), described by Eq. (5.46), is nicely in the center and its 95% confidence interval (dashed) delineates the point cloud. The density of points is much higher in the lower left than the upper right of the figure because more measurement points originate from the periphery of the water cylinder, which have a lower variance. The final values of  $K$  and  $\sigma_e^2$  were estimated using Eq. (5.44) Eq. (5.45) and the weighted mean of the values in Table 5.3 as an initialization. They were found to be  $2.17 \times 10^6$  photons/(mm<sup>2</sup>. mAs) and 23.3(average of the values at the lowest two exposures), respectively. Note that  $K$  and  $\sigma_e^2$  are higher as all data were combined than for each tube current separately. These final values will be used in the remaining experiments.



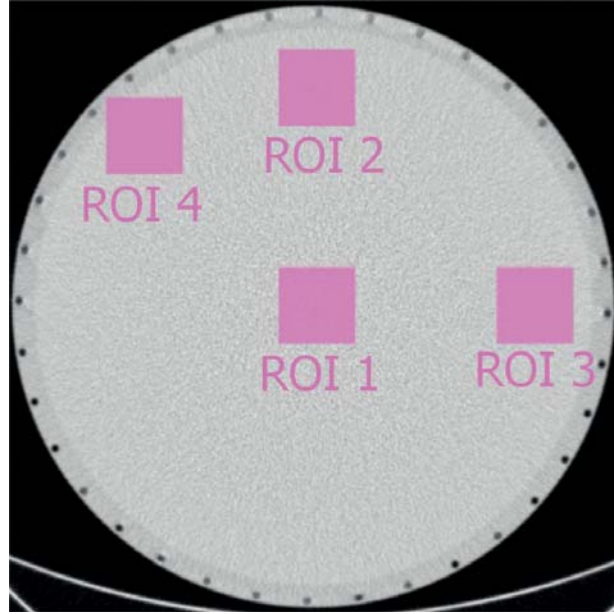


Figure 5.11: Image depicting four ROIs that were used to validate the low-dose CT simulation.

Furthermore, Figure 5.12 shows the standard deviation as a function of the distance to the center at different exposure levels: 60 mAs (lower three) and 15 mAs (upper three), respectively. The simulations at 60mAs (bottom three curves) were obtained using images scanned at 250mAs (bottom), 120mAs (middle) and 85mAs (top), while the simulations at 15mAs (top three curves) used images obtained at 250mAs (bottom), 60mAs (middle) and 21mAs (top). The pink and blue lines correspond to the angular averaged experimental and the simulated data.

The figure shows that the differences between simulations and experiments are small. Only in the center of the 15 mAs simulations small differences of maximally 5% are noticeable. Therefore, the noise in the virtual sinogram can be neglected at the studied dose levels. Notice that the deviation becomes larger in simulations from increasingly higher doses. This is because increasingly more noise needs to be added. We attribute the (almost) absence of noise to the averaging that occurs in generating the virtual sinogram (i.e. the radon transform). This is different from methods based on raw projection data: such methods



generate noise based on data that already contains noise, whereas the noise in our data is severely suppressed.

Additionally, Table 5.4 gives the relative root-mean-squared difference  $\epsilon_\sigma$  between the simulated and experimental noise levels as a function of the exposure:

$$\epsilon_\sigma = \sqrt{\frac{1}{NM} \sum_{n=1}^M \sum_{m=1}^M \frac{(\sigma_{\text{simu}}(n, m) - \sigma_{\text{exp}}(n, m))^2}{\sigma_{\text{simu}}^2(n, m)}}, \quad (5.48)$$

in which  $\sigma_{\text{simu}}(n, m)$  and  $\sigma_{\text{exp}}(n, m)$  are the simulated and experimentally acquired standard deviations at pixels  $(n, m)$  and  $N$  and  $M$  indicate the number of pixels on both axes.

Figure 5.12 and Table 5.4 demonstrate that the simulated noise level closely approximates the experimental noise level in the water cylinder.

Exposure (mAs)	210	170	120	85	60	42	30	21	15
$\epsilon_\sigma(10^{-2})$	1.9	1.4	1.5	1.5	1.4	2.1	2.3	2.3	2.5

Table 5.4: Relative RMS difference  $\epsilon_\sigma$  between simulated and experimentally acquired noise levels as a function of the exposure.

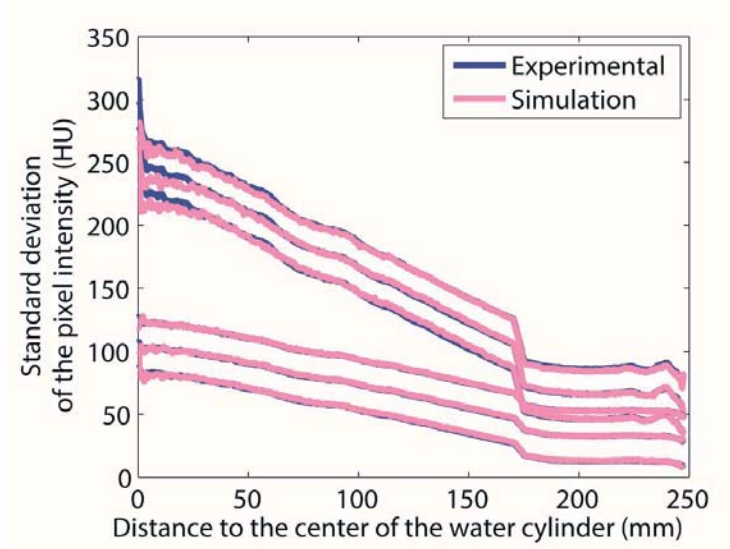


Figure 5.12: The standard deviation of the noise in a 34cm water cylinder as a function of the distance to the center of the water cylinder for simulations of 15mAs data (top three lines) from images obtained at (from top to bottom) 21, 60 and 250 mAs and for simulations of 60mAs data from images obtained at (from top to bottom) 85, 120 and 250 mAs. An offset of  $\pm 20$  HU was added to separate the three curves for display purposes. The pink lines resulted from experiments, the blue lines from our simulations.

Figure 5.13 shows contour plots of the NPS calculated from ROIs depicted in Figure 5.11. The blue and pink lines correspond to the experimental and the simulated NPS, respectively. Table 5.5 gives the relative root-mean-squared difference  $\epsilon_{\text{NPS}}$  between the experimental ( $\text{NPS}_{\text{exp}}$ ) and simulated NPS ( $\text{NPS}_{\text{simu}}$ ) for each ROI, i.e.:

$$\epsilon_{\text{NPS}} = \sqrt{\frac{\sum_{n=1}^M \sum_{m=1}^M (\text{NPS}_{\text{simu}}(\omega_n, \omega_m) - \text{NPS}_{\text{exp}}(\omega_n, \omega_m))^2}{\sum_{n=1}^M \sum_{m=1}^M \text{NPS}_{\text{simu}}^2(\omega_n, \omega_m)}}. \quad (5.49)$$

Figure 5.13 and Table 5.5 signify that the shape of the 2D noise power spectra from the simulations closely approximates those encountered experimentally.

ROI number	1	2	3	4
------------	---	---	---	---

$\epsilon_{NPS}$

0.16	0.16	0.12	0.14
------	------	------	------

Table 5.5 Relative RMS difference  $\epsilon_{NPS}$  between measured and simulated NPS for the ROIs depicted in Figure 5.11

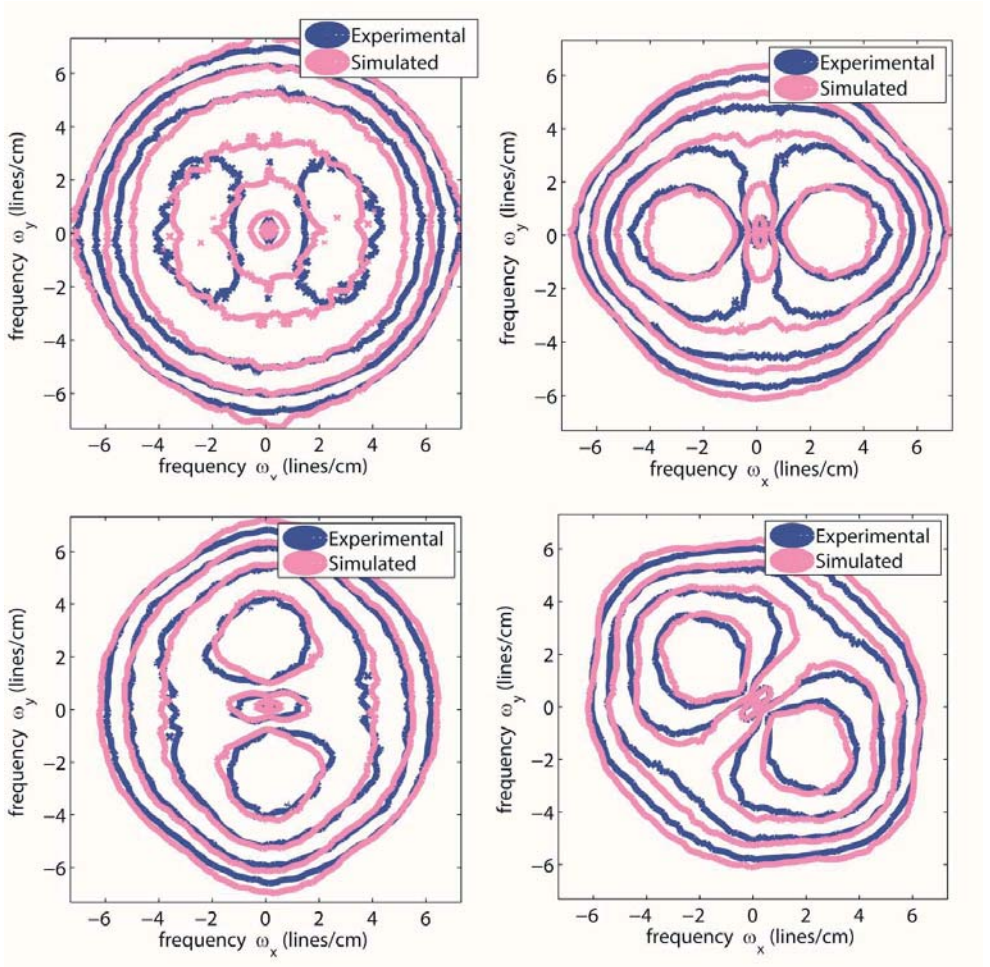


Figure 5.13: Contour plots of the experimental (blue) and the simulated (pink) NPS at 15 mAs. The iso-contours depict the same value for the pink and blue curves. Each line corresponds to an increase of factor 2 while moving inwards; the actual values for the outermost iso-contours are: 74  $\text{Hu}^2\text{cm}^2$  (topleft, ROI1), 35  $\text{Hu}^2\text{cm}^2$  (top right, ROI 2), 35  $\text{Hu}^2\text{cm}^2$  (left, ROI 3) and 32  $\text{Hu}^2\text{cm}^2$  (right, ROI 4) These ROI's were indicated in Figure 5.11.

#### **5.4.3.2 Pelvic phantom**

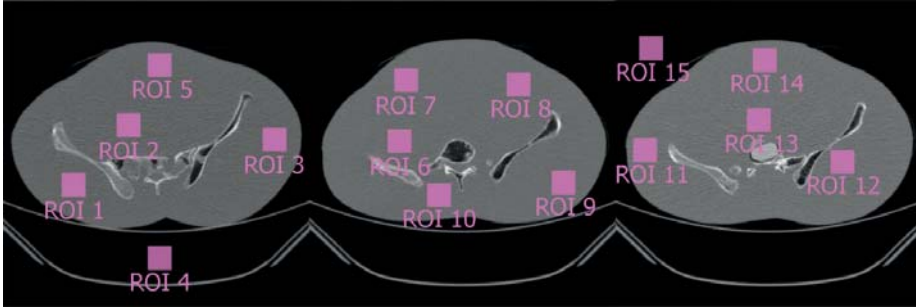


Figure 5.14: Three slices of the pelvic phantom imaged at 80 mAs in which ROIs were selected for the assessment of the low-dose simulation method.

Fifteen ROIs were selected in which the noise properties were analyzed, see Figure 5.14. Each ROI was composed of 41 x 41 pixels. The standard deviation in each ROI was determined for both the experimental scans and the simulated scans at 80, 40 and 15 mAs. For the simulations at 40 and 15 mAs, the scans at 80 mAs were used as the high-dose image. A validation at 80 mAs was possible by computing the noise properties directly from the zero-mean noise image  $\mu_{\text{noise}}$  that was simulated assuming the original image was acquired at infinite dose. For each slice and exposure level, 64 simulations were created to take variations in noise realizations into account. On average, the standard deviation of the simulated images deviated 5.3%, 2.4% and 4.3% from the standard deviation of the experimental images for the experimental acquired scans at 80, 40 and 15 mAs, which was within the 95% confidence interval of the estimated standard deviation.

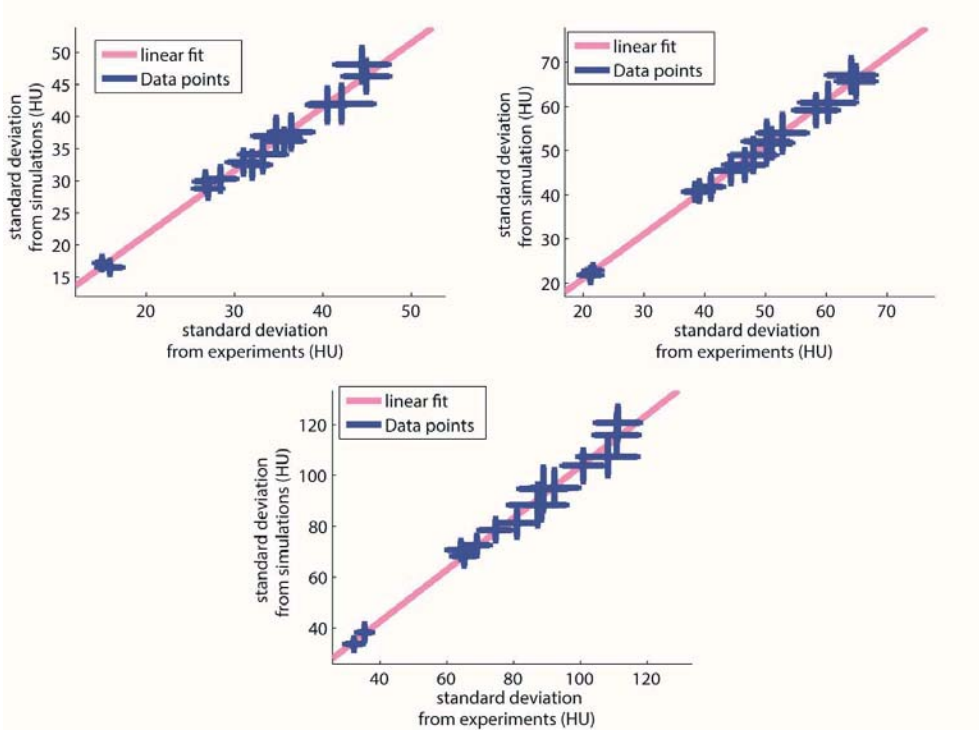


Figure 5.15: Standard deviations obtained from the simulated images as a function of the ones obtained from the experimentally acquired data at 80 mAs (top left), 40 mAs (top right) and 15 mAs (bottom). Each point indicates the average standard deviation in one of the ROIs depicted in Figure 5.14. The bars depict the 95% confidence interval of the estimation. The pink line represents the linear fit, the parameters of the fit are listed in Table 5.6.

Exposure (mAs)	slope	offset	correlation
80	0.995	1.71	0.994
40	1.006	0.94	0.998
15	1.018	1.75	0.994

Table 5.6 The parameters and the correlation of the linear fit illustrated in Figure 5.15

Figure 5.15 and Table 5.6 show that the simulated scans closely approximated the noise strength in the experimental scans. Nonetheless, the difference between the simulations and experimental scans was larger for the measurements based on the pelvic phantom than the ones based on the water cylinder. We attributed this to

the presence of bone-like structures, that cause beam hardening, which was not taken into account in our method. Furthermore, the NPS was computed in four arbitrarily selected ROIs from experimental and simulated scans at 15mAs, see Figure 5.16. Once more, the figures show how well the simulation technique approximates the experimentally acquired low-dose images.

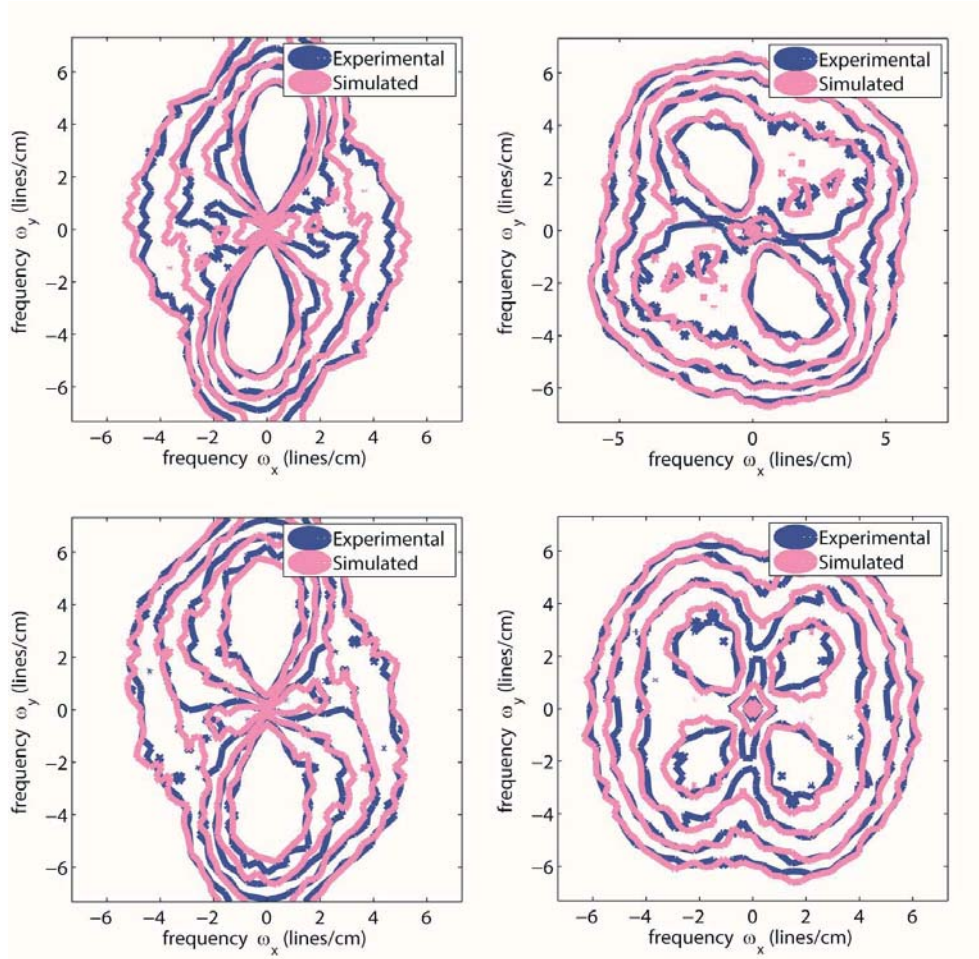


Figure 5.16: Contour plots of the experimental (blue) and the simulated (pink) NPS from ROIs 3, 7, 12, 14 of the pelvic phantom (depicted in Figure 5.14) at 15 mAs. The iso-contours depict the same value for the pink and blue curves. Each line corresponds to an increase of factor 2 while moving inwards; the actual values for the outermost iso-contours are: 14



Hu<sup>2</sup>cm<sup>2</sup> (top left, ROI 3), 8 Hu<sup>2</sup>cm<sup>2</sup> (top right, ROI 7), 18 Hu<sup>2</sup>cm<sup>2</sup> (left, ROI 12) and 7 Hu<sup>2</sup>cm<sup>2</sup> (right, ROI 14,)

## **5.5 Discussion**

We presented a novel method to simulate patient-specific, low-dose CT images from existing high-dose images assuming reconstruction by parallel beam filtered backprojection. Scanner-specific parameters i.e. the apodization window of the reconstruction filter, the bowtie filter, the X-ray tube output parameter and the read-out noise were estimated using calibration images of a water cylinder. Therefore, new estimators were developed that used reconstructed images and did not require projection data. The low-dose simulation was evaluated by comparing the noise characteristics of simulated low-dose images with experimentally acquired low-dose images.

One of the main strengths of our chapter is that it allows scientists to simulate low-dose CT in a well-documented and reproducible manner while not being dependent on the availability of raw projections as well as scanner- and scanning-parameters. Note that it is not our purpose to improve on the simulation methods that are based on the raw-projection data. Instead, the aim is to simulate low dose CT image from the higher dose images, which is useful whenever the raw projection data is not shared by vendors. Additionally, it is relevant for retrospective CT-studies in which the raw data is usually not stored. Moreover, it is important for the developers of image processing algorithms: it gives them a realistic tool to study the robustness of their techniques under noisy circumstances.

We demonstrated that the models used to recover the scanner-specific parameters accurately described the calibration data. The estimated reconstruction filter corresponded well to smooth reconstruction kernels found earlier[113, 115, 116]. We recognize, however, that a more general model such as an higher order cosine series may be necessary to describe the very sharp kernels. The shape of the bowtie filter estimated by us closely resembles the bowtie filter used in a Siemens scanner[26, 104, 127]. Furthermore, we estimated the X-ray tube output parameter at  $K = 2.17 \cdot 10^6$  photons/mm<sup>2</sup>. mAs and the read-out noise variance at  $\sigma_e^2 = 23.2$ . These estimated values were of the same order of magnitude as

reported previously: Massoumzadeh et al. [26] estimated  $2.7 \cdot 10^6$  photons/mm<sup>2</sup>.mAs for  $K$  and values from 40 through 200 for  $\sigma_e^2$ ; Faulkner et al. [114] estimated  $4.15 \cdot 10^6$  and  $6.46 \cdot 10^6$  photons/mm<sup>2</sup>.mAs for  $K$ . Ma et al. [106] found 10 for  $\sigma_e^2$ . Finally, we showed that the simulated low-dose images accurately reproduced the noise in experimental low-dose volumes.

A limitation of our approach is in the assumption that monochromatic photons are produced by a virtual X-ray source. As such, our method does not take beam hardening effects into account, which likely cause the encountered deviations in the noise characteristics between simulated and experimental scans of the pelvic phantom. At the same time, the differences between the simulated and experimental data were relatively small even in the presence of bony structures and at relatively low exposures of 15 mAs. Essentially, a polychromatic approach would require a spectral dependency in our framework, particularly concerning the x-ray tube output  $N_0$  and the calculation of attenuation projections. We consider this an important topic for further research.

Additionally, a limitation is that our method does not take tube current modulation into account. The tube current modulation essentially adjusts the tube current to the part of the body being imaged and the size of the patient. A variation per slice can be simply incorporated in our method by adjusting  $N_0$  to the actually used tube current  $I_{\text{high}}$  which may be stored in the DICOM-header. However, it might not be easy to recover complexer variations of  $N_0$ , e.g. as a function of the gantry angle.

Our method requires a full field of view which is the case for many protocols. Clearly, one could always prospectively reconstruct full field images to satisfy this requirement. In our model, the detector noise was approximated by zero-mean Gaussian noise, which is commonly used in literature[26, 106, 110, 115, 128], but strictly speaking is not entirely correct[109].

The techniques to extract the system parameters assumed in-plane, fan beam imaging and filtered backprojection for tomographic reconstruction. This should be considered a calibration step. The simulation method also focused on in-plane fan-beam imaging and image reconstruction by filtered backprojection, particularly to demonstrate the feasibility of simulating low dose CT scanning from higher dose images. As such, the effects of interslice noise correlations



associated with helical scanning were not considered. We hypothesize that these correlations are small, but might not be negligible. An obvious next step will be to simulate cone-beam imaging and backprojection and fully integrate helical scanning in the model. This would require a fairly straightforward adaptation of the methods to generate the virtual sinogram and the technique to perform filtered backprojection. Our methods are not applicable to iterative reconstruction methods as such methods violate crucial model assumptions. This is perhaps the largest limitation of our work.

In summary, the developed methods truthfully simulated low-dose CT imaging without requiring projection data. This new technology might facilitate large-scale studies into the diagnostic accuracy for lower CT dose. In turn, it could aid in further reducing the radiation risks of CT-examinations.

## 6 Conclusion

---

The first part of this thesis focussed on the VIGOR++ project, whose purpose was to develop a novel tool to grade Crohn's disease severity based on MRI. Specifically, three essential problems were solved: 1. (semi-)automated delineation of the bowel wall; 2. precise extraction of features from this delineation; 3. development of a reproducible disease scoring system through a comprehensive feature search and validation of this system by comparing it to state-of-the-art systems based on manually measured features. The second part of the thesis focused on developing a realistic low-dose CT-image simulator from reconstructed high-dose CT images.

## 6.1 Conclusions

In chapter 2, a new method was presented to segment the bowel lumen from T1-weighted MR images using a level set approach. The technique can cope with a heterogeneous content of the bowel lumen composed of three materials: fecal residue, water, and air as well as a diverse lumen exterior composed of three tissues with respectively dark, medium and bright intensities. Hence, it provides a natural extension to existing region-based active contour models. Particularly, it avoids splitting or merging multiple level set segmentations in a post-processing step. Our algorithm was able to successfully segment the lumen in 60 out of 61 ROIs, spanning a wide variety of bowel geometries and material mixtures. The mean distance between our level set segmentation and the surface extracted from annotations by two experts was  $1.43 \pm 0.55\text{mm}$  and  $1.48 \pm 0.68\text{mm}$ , respectively. This was comparable to the inter-observer variability between the experts' annotations:  $1.52 \pm 0.70\text{mm}$ . The lumen segmentation was used to derive features related to Crohn's disease activity in chapter 3.

In chapter 3, a semi-automatic method to measure the bowel wall thickness was evaluated. The method consists of four steps: (1) initialization by manually drawing a centerline; (2) segmentation of the bowel wall's inner surface; (3) segmentation of the wall's outer surface; (4) measurement of the bowel wall thickness. The method was validated using the examinations of 56 consecutive patients with suspected or known Crohn's disease undergoing 3T MR enterography. In each of these images two annotators independently annotated all regions with active Crohn's disease (Ano1, Ano2). Furthermore, three observers (Ob1, Ob2, Ob3) manually measured the maximal bowel wall thickness of each

segment annotated by Ano1 or Ano2. The method was separately evaluated regarding its ability (1) to segment a diseased part of the bowel wall and (2) to measure a segment's thickness. The median overlap of the semi-automatic segmentations (Seg1 vs Seg2: 89%) was significantly larger than the median overlap of the manual annotations (Ano1 vs Ano2: 72%):  $p = 1.4 \cdot 10^{-5}$ . The variance in overlap of the semi-automatic segmentations was significantly smaller than the variance in the overlap of the manual annotations ( $p = 1.1 \cdot 10^{-9}$ ). The intraclass correlation coefficient (ICC) of the semi-automatic measurement (0.88) was significantly higher than the ICC of the manual measurement (0.32-0.65):  $p = 0.005$ . Therefore, our data showed that a semi-automated measurement technique facilitated a highly reproducible delineation of a region with active Crohn's disease. Furthermore, the semi-automatic thickness measurement sustained a significantly higher ICC than manual observers in active segments on MRI. As such, it may reduce the inter-observer variability of an MRI grading systems for Crohn's disease.

In chapter 4, a method was developed for scoring disease activity based on manually and automatically determined MRI features using ileocolonoscopy as the reference standard. Patient data (both MRI and colonoscopy) from a retrospective study (27 patients) was used to develop a scoring system, the VIGOR model, based on semi-automatically determined features. Similarly, a system was developed incorporating only manually determined features: the manual model. A total of 120 patients (66 females, median age 35) were prospectively included for validation. The validation of the VIGOR model showed comparable grading accuracy to conventional scoring systems, such as the MaRIA, London score and the manual model. Simultaneously, a significantly higher ICC was seen for the VIGOR model (0.81 vs 0.43–0.55). Contrary to other studies, we found relatively low grading accuracy and interobserver agreement for individual features and scoring systems, which was partially explained by the inclusion of only active segments. In conclusion, the use of semi-automatic features for assessment of patients with CD provides improved reproducibility over conventional scoring systems, favoring its use for therapy evaluation and monitoring of disease activity. Accurate and reproducible MRI scores could improve the physician's trust in these scores to make consistent and effective

treatment decisions. Additional benefits in diagnostic capabilities remain to be investigated in future studies.

In Chapter 5, a novel method was presented to simulate patient-specific, low-dose CT images from existing high-dose images assuming reconstruction by filtered backprojection. Scanner-specific parameters such as the apodization window of the reconstruction filter, the bowtie filter, the X-ray tube output parameter and the read-out noise were estimated using calibration images of a water cylinder. Therefore, new estimators were developed that used reconstructed images and did not require projection data. The low-dose simulation was evaluated by comparing the noise characteristics of simulated low-dose images with experimentally acquired low-dose images. The method might facilitate large-scale in-silico studies into the diagnostic accuracy for lower CT dose. In turn, it could aid in further reducing the radiation risks of CT-examinations.

## **6.2 Future work on MRI-based assessment of Crohn's disease severity**

An outlook of future work on the assessment of Crohn's disease severity can be divided into two parts: advances in methodology and progress towards clinical application of newly developed techniques.

The methods presented in this thesis could be extended in several ways so that they become more robust or able to measure more salient features (and thus become more sensitive). For instance, the bowel wall segmentation algorithm models slowly varying intensity fluctuations and abrupt intensity transitions in two separate terms: the smooth fluctuations in a term representing the mean local intensity and the abrupt ones in a term consisting of a weighted sum of mean material intensities. Alternatively, a combination of more sophisticated basis functions could be used (instead of the mean material intensities). Future work could be to experiment with commonly used bases, such as a local polynomial basis [70] and Gabor wavelets. A local polynomial basis could be used to cope with strong B0-field inhomogeneity while retaining a large integration scale. Alternatively, introducing Gabor wavelets might enable the modeling of ridge like structures such as thin pieces of bowel wall. Finally, a modified nearest mean classifier estimated the material fractions. More sophisticated, hybrid methods such as estimating the material fractions through an Expectation Maximization

(EM) approach could also increase the robustness of the presented method. Clearly, this framework might be extended to other problems besides the segmentation of the bowel wall.

Incorporating additional anatomical knowledge may increase the robustness of the segmentation algorithm even further. Currently, shape information is modelled by either the minimum length term (or a weighted one steered by image information[44]) or by learning a complex shape distribution from training shapes (and/or use a Parzen density estimation to fill the gaps between the training shapes)[129-131]. The first option incorporates a very limited amount of prior knowledge while the second option requires an extensive training set to grasp all degrees of shape variation, which in the case of the bowel is unavailable (or at least very costly to obtain). A novel approach would be to model the bowel by simple lower dimensional shapes such as an irregular tube or locally penalize certain curvatures. Lately, work has been published in which complex shapes were represented by wavelet decomposition[132]. The coefficients learned as such acted to regularize the segmentation process.

Feature driven segmentation would also be of interest, in which case segmentation and feature extraction is done simultaneously. For instance, in chapter 3, the outer-wall segmentation step might be regularized by a thickness constraint. The stratification of the bowel wall (or presence of edema) might also be modelled implicitly into the segmentation algorithm, so that the actual measurement of the stratification pattern improves the segmentation and vice-versa.

Another aspect is to further minimize the interaction, which is on the order of seconds for well-distended and short pieces of diseased bowel up to a few minutes for collapsed and highly folded ones. The interaction could be reduced by automatically extracting the centerline. Automatically drawing a centerline is very challenging as the lumen shape is very complex and is sometimes not identifiable due to strictures. Automatically detecting whether disease activity is present and its approximate location would make the analysis more efficient in daily practice.

The MRI features showed a large variation among observers, and understanding the origin of these variations may help improving the automated measurements in scoring disease activity. For instance, how should the thickness be measured (both as a stand-alone feature and as a variable in a scoring system)? Should a single

feature, multiple features or even a thickness distribution describe the bowel wall thickness? Such studies require segmentations, which are now available due to the work in Chapter 2 and 3. Chapter 4 has shown that the volume is an important feature and investigating more descriptors of volume could lead to even stronger scoring systems. We hypothesize that volumetric and surface-based features more closely correlate to colonoscopic features than point based ones, as for instance the CDEIS scores the inflammation based on the size the affected surface.

Although improvements can be made by enhancing existing and developing novel features it is unlikely MRI can perfectly predict CDEIS. This is because MRI and colonoscopy visualize different aspects of the disease (severity). Furthermore, recent work showed that different degrees of inflammation frequently occur in a single surgical specimen with both inflammation and fibrosis present in a single voxel[133].

Additionally, the VIGOR++ score or the automated features may also be validated against a histopathological score in the future. Histopathology assesses transmural disease activity, straightforwardly and locally. Semi-automated features such as the bowel wall thickness and the A1-feature can be measured locally as well, although finding exact correspondence might be a challenging issue. A validation using the histologic scale may enhance the understanding of how disease activity should be expressed at a segment scale.

### **6.3 Future work on simulation of low-dose CT**

Chapter 5 has demonstrated that realistic, low-dose CT-images can be produced from high-dose images. Such a realistic low-dose model for state-of-the-art clinical CT scanners needs to be further developed. The necessary extensions should include: a cone-beam helical acquisition geometry, a poly-energetic X-ray spectrum, object dependent attenuation, and state-of-the-art iterative reconstruction algorithms. Nonetheless, for most scanners that are in use today, the approximation that are done in our simulation method are still valid. Therefore, this new technology might facilitate large-scale studies into the diagnostic accuracy with lower CT dose. In effect, it could aid in further reducing the radiation risks of CT-examinations.

# References

---

- [1] J. Panes, Y. Bouhnik and W. Reinisch, "Imaging techniques for assessment of inflammatory bowel disease: Joint ECCO and ESGAR evidence-based consensus guidelines," *J Crohns Colitis*, vol. 7, p. 556-585, 2013.
- [2] J. Prince and J. Links, *Medical Imaging Systems*, Pearson Prentice Hall, 2006.
- [3] D. Bruining, G. Bhatnagar, J. Rimola, S. Taylor, E. Zimmermann and J. Fletcher, "CT and MR enterography in Crohn's disease: current and future applications," *Abdominal Imaging*, vol. 40, pp. 965-974, 2015.
- [4] W. Kalender, *Computed Tomography*, Publicis MCD Verlag, 2000.
- [5] B. Allen, M. Baker, D. Einstein and e. al, "Effect of altering automatic exposure control settings and quality reference mAs on radiation dose, image quality and diagnostic efficacy in MDCT enterography for active inflammatory Crohn's disease," *Am J Roentgenol*, no. 195, pp. 89-100, 2010.
- [6] A. Kambadakone, P. Prakash, P. Hahn and D. Sahani, "Low dose CT examinations in Crohn's disease: impact on image quality, diagnostic performance, and radiation dose.," *Am J Roentgenol*, no. 195, pp. 78-88, 2010.
- [7] M. Kappelman, K. Moore, J. Allen and S. Cook, "Recent Trends in the Prevalence of Crohn's Disease and Ulcerative Colitis in a Commercially Insured US Population," *Dig Dis Sci*, vol. 58, no. 2, p. 519-525, 2013.
- [8] D. Baumgart and W. Sandborn, "Crohn's disease," *Lancet*, vol. 6736, p. 1-16, 2012.
- [9] W. Best, J. Beckte, J. Singleton and F. Kern, "Development of a Crohn's disease activity index. National Cooperative Crohn's Disease Study," *Gastroenterology*, no. 70, pp. 439-444, 1976.
- [10] J. Mary and M. R., "Development and validation of an endoscopic index of the severity for Crohn's disease: a prospective multicentre study. Groupe d'Etudes Therapeutique des Affections Inflammatoires du Tube Digestif (GETAID)," *Gut*, no. 30, pp. 983-989, 1989.
- [11] G. D'Haens, K. Geboes, M. Peeters, F. Baert, F. Penninckx and P. Rutgeerts, "Early lesions of recurrent Crohn's disease caused by infusion of intestinal contents in excluded lieum," *Gastroenterology*, no. 114, pp. 262-267, 1998.
- [12] J. Benitez, M. Meuwis, C. Reenaers, C. Kemseke, P. Meunie and E. Louis, "Role of endoscopy, cross-sectional imaging and biomarkers in Crohn's disease monitoring," *Gut*, no. 62, pp. 1806-1816, 2013.
- [13] K. Horsthuis, S. Bipat, R. Bennink and J. Stoker, "Inflammatory bowel disease diagnosed with US, MR, scintigraphy, and CT: metaanalysis of prospective studies," *Radiology*, vol. 247, p. 64-79, 2008.
- [14] J. Rimola, S. Rodriguez, O. García-Bosch, I. Ordás, E. Ayala and M. Aceituno, "Magnetic resonance for assessment of disease activity and severity in ileocolonic Crohn's disease," *Gut*, vol. 58, pp. 1113-20, 2009.
- [15] J. Rimola, I. Ordas, S. Rodriguez and J. Panes, "Colonic Crohn's disease: value of magnetic resonance colonography for detection and quantification of disease activity," *Abdominal Imaging*, vol. 35, pp. 422-427, 2010.
- [16] M. Steward, S. Punwani, I. Proctor, Y. Adjei-Gyamfi, F. Chatterjee, S. Bloom, M. Novelli, S. Halligan, M. Rodriguez- Justo and S. Taylor, "Non-perforating small bowel Crohn's disease assessed by MRI enterography: Derivation and histopathological validation of an MR-based activity index," *European Journal of Radiology*, vol. 81, pp. 2080-2088, 2012.
- [17] J. Tielbeek, J. Makanyanga, S. Bipat, D. Pendse, C. Nio, F. Vos, S. Taylor and J. Stoker, "Grading Crohn's disease activity with MRI: Interobserver variability of MRI features, MRI scoring of severity and correlation with Crohn's Disease Endoscopic Index of Severity," *American Journal of Roentgenology*, no. 201, pp. 1220-1228, 2013.
- [18] M. Ziech, S. Bipat, J. Roelofs, C. Nio, B. Mearadji and S. Van Doorn, "Retrospective comparison of magnetic resonance imaging features and histopathology in Crohn's disease patients," *Eur J Radiol*, vol. 80, pp. 299-305, 2011.
- [19] K. Horsthuis, S. Bipat, P. Stokkers and J. Stoker, "Magnetic resonance imaging for evaluation of disease activity in Crohn's disease: a systematic review," *Eur Radiol*, vol. 19, pp. 1450-60, 2009.



- [20] J. Tielbeek, F. Vos and J. Stoker, "A computer-assisted model for detection of MRI signs of Crohn's disease activity: future or fiction?," *Abdominal Imaging*, vol. 37, no. 6, pp. 967-973, 2012.
- [21] E. Haacke, R. Brown, M. Thompson and R. Venkatesan, *Magnetic Resonance Imaging: Physical Principles and Sequence Design*, New-York: Wiley-Liss, 1999.
- [22] M. Ziech, P. Bossuyt, A. Laghi, T. Lauenstein, S. Taylor and J. Stoker, "Grading luminal Crohn's disease: which MRI features are considered as important," *European Journal of Radiology*, vol. 81, pp. e467-472, 2012.
- [23] OECD, "OECD health data (www.oecd.org).," The Organisation for Economic Cooperation and Development, 2012.
- [24] R. Gelder, H. Venema, J. Florie, C. Nio, I. Serlie, M. Schutter, J. van Rijn, F. Vos, A. Glas, P. Bossuyt, J. Bartelsman, J. Lameris and J. Stoker, "CT colonography: Feasibility of substantial dose reduction - comparison of medium to very low doses in identical patients," *Radiology*, no. 232, pp. 611-620, 2004.
- [25] J. Mayo, K. Whittal, A. Leung, T. Hartman, C. Park, S. Primack, G. Chambers, M. Linkeman, T. Toth and S. Fox, "Simulated dose reduction in conventional chest CT: validation study," *Radiology*, no. 202, pp. 453-457, 1997.
- [26] P. Massoumzadeh, S. Don, C. Hildebolt, K. Bae and B. Whiting, "Validation CT dose-reduction simulation," *Medical Physics*, vol. 36, pp. 174-189, 2009.
- [27] Z. Li, D. Mahapatra, J. Tielbeek, J. Stoker, L. van Vliet and F. Vos, "Image Registration Based on Autocorrelation of Local Structure," *Transactions on Medical Imaging*, vol. 35, no. 1, pp. 63-75, 2016.
- [28] J. Tielbeek, J. Makanyanga, S. Bipat, D. Pendse, C. Nio, F. Vos, S. Taylor and J. Stoker, "Grading Crohn Disease Activity With MRI: Interobserver Variability of MRI Features, MRI Scoring of Severity, and Correlation With Crohn Disease Endoscopic Index of Severity," *American Journal of Roentgenology*, vol. 201, no. 6, pp. 1220-1228, 2013.
- [29] J. Tielbeek, F. Vos and J. Stoker, "A computer-assisted model for detection of MRI signs of Crohn's Disease activity: future or fiction?," *Abdominal Imaging*, vol. 37, no. 6, pp. 967-973, 2012.
- [30] F. Vos, J. Tielbeek, R. Naziroglu, Z. Li, P. Schueffler, D. Mahapatra, A. Wiebel, C. Lavini, J. Buhman, H.-C. Hege, J. Stoker and L. van Vliet, "Computational Modeling for Assessment of IBD: to be or not to be?," in *Annual International Conference of the IEEE Engineering in Medicine and Biology*, 2012.
- [31] J. Nappi and H. Yoshida, "Feature-guided analysis for reduction of false positives in CAD of polyps for computed tomographic colonography," *Medical Physics*, vol. 30, pp. 1592-1601, 2003.
- [32] V. van Ravesteijn, T. Boellaard, M. van der Paardt, I. Serlie, M. de Haan, J. Stoker, L. van Vliet and F. Vos, "Electronic cleansing for 24-H limited bowel preparation CT colonography using principle curvature flow," *IEEE Transactions on Biomedical Engineering*, vol. 60, no. 11, pp. 3036-3045, 2013.
- [33] M. Franaszek, R. Summers, P. Pickhardt and J. Choi, "Hybrid Segmentation of Colon Filled With Air and Opacified Fluid for CT Colonography," *IEEE Transactions on medical imaging*, vol. 25, no. 3, pp. 358-368, 2006.
- [34] C. Van Wijk, V. Van Ravesteijn, F. Vos and L. Van Vliet, "Detection and Segmentation of Colonic Polyps on Implicit Isosurfaces by Second Principle Curvature Flow," *IEEE Transactions on Medical Imaging*, vol. 29, no. 3, pp. 688-698, 2010.
- [35] S. Grigurescu, S. Nevo, M. Liedenbaum, R. Truyen, J. Stoker, L. van Vliet and F. Vos, "Automated detection and segmentation of large lesions and masses in CT colonography," *IEEE Transactions on Biomedical Engineering*, vol. 57, no. 3, pp. 675-684, 2010.
- [36] I. Serlie, F. Vos, R. Truyen, F. Post, J. Stoker and L. van Vliet, "Electronic cleansing for computed tomography (CT) colonography using a scale invariant three material model," *IEEE Transactions on Biomedical Engineering*, vol. 57, no. 6, pp. 1306-1317, 2010.
- [37] E. Thornton, M. Morrin and J. Yee, "Current Status of MR Colonography," *RadioGraphics*, vol. 30, pp. 201-218, 2010.
- [38] W. Zhang, J. Liu, J. L. A. Yao, T. Nguyen, S. Wank, W. Nowinski and R. Summers, "Mesenteric Vasculature-Guided Small Bowel Segmentation on 3-D CT," *IEEE Transactions on Medical Imaging*, vol. 32, no. 11, pp. 2006-2021, 2013.
- [39] D. Mahapatra, P. Schuffler, J. Tielbeek, J. Makanyanga, J. Stoker, S. Taylor, F. Vos and J. Buhmann, "Active Learning Based Segmentation of Crohn's Disease using Principles of Visual Saliency," in *IEEE International Symposium for Biomedical Imaging*, 2014.

- [40] D. Mahapatra, P. Schuffler, J. Tielbeek, J. Makanyanga, J. Stoker, S. Taylor, F. Vos and J. Buhmann, "Automatic Detection and Segmentation of Crohn's Disease Tissues from Abdominal MRI," *IEEE Transactions on Medical Imaging*, vol. 32, no. 12, pp. 2332-2348, 2013.
- [41] M. Kass, A. Witkin and D. Terzopoulos, "Snakes: Active contour models," *International Journal on Computer Vision*, vol. 1, pp. 321-331, 1988.
- [42] S. Osher and J. Sethian, "Fronts propagating with curvature dependent speed: Algorithms based on Hamilton-Jacobian formulations," *Journal of Computational Physics*, vol. 79, pp. 12-49, 1988.
- [43] D. Lesage, E. Angelini, I. Bloch and G. Funka-Lea, "A review of 3D vessel lumen segmentation techniques: Models, features and extraction schemes," *Medical Image analysis*, vol. 13, no. 2009, pp. 819-845, 2003.
- [44] V. Caselles, R. Kimmel and G. Sapiro, "Geodesic Active Contours," *International Journal of Computer Vision*, vol. 22, no. 1, pp. 61-79, 1997.
- [45] R. Kimmel and A. Bruckstein, "Regularized laplacian zero crossings as optimal edge integrators," *International Journal of Computer Vision*, vol. 53, no. 3, pp. 225-243, 2003.
- [46] M. Holtzman-Gazit, R. Kimmel, N. Peled and D. Goldsher, "Segmentation of thin structures in volumetric medical images," *IEEE Transactions on Image Processing*, vol. 15, no. 2, pp. 354-363, 2006.
- [47] A. Vasilevsky and K. Siddiqi, "Flux Maximizing Geometric Flows," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 24, no. 12, pp. 1565-1578, 2002.
- [48] L. Vese and T. Chan, "A multiphase Level Set Framework for Image Segmentation using the Mumford Shah Model," *International Journal of Computer Vision*, vol. 50, no. 3, pp. 271-293, 2002.
- [49] T. Chan and L. Vese, "Active contours without edges," *IEEE Transactions on Image Processing*, vol. 10, no. 2, pp. 266-277, 2001.
- [50] T. Brox and D. Cremers, "On Local Region Models and a Statistical Interpretation of the Piecewise Smooth Mumford-Shah Functional," *International Journal of Computer Vision*, vol. 84, pp. 184-193, 2009.
- [51] D. Cremers, M. Rousson and R. Deriche, "A Review of Statistical Approaches to Level Set Segmentation: Integrating Color, Texture, Motion and Shape," *International Journal of Computer Vision*, vol. 72, no. 2, pp. 195-215, 2007.
- [52] S. Geman and D. Geman, "Stochastic relaxation, Gibbs distributions and the Bayesian restoration of images," *Transactions on Pattern Analysis and Machine Intelligence*, vol. 4, no. 11, pp. 721 - 741, 1984.
- [53] Y. Leclerc, "Constructing simple stable description for image partitioning," *International Journal for Computer Vision*, vol. 3, no. 1, pp. 73-102, 1989.
- [54] S. Zhu and A. Yuille, "Region competition: Unifying snakes, region growing and Bayes/MDL for multiband image segmentation," *IEEE Transactions on Pattern Analysis and Machine Learning*, vol. 18, no. 9, pp. 884-900, 1996.
- [55] C. Samson, L. Blanc-Feraud, G. Aubert and J. Zerubia, "A variational model for image classification and restoration," *Transaction on Pattern Analysis and Machine Learning*, vol. 22, no. 5, pp. 460-472, 2000.
- [56] N. Paragion and R. Deriche, "Geodesic active regions and level set methods for supervised texture segmentation," *International Journal of Computer Vision*, vol. 46, no. 3, pp. 223-247, 2002.
- [57] D. Mumford and J. Shah, "Optimal Approximations by Piecewise Smooth Functions and Associated Variational Problems," *Communications on Pure and Applied Mathematics*, vol. 42, pp. 577-658, 1989.
- [58] C. Li, C. Kao, J. Gore and Z. Ding, "Implicit Active Contours Driven by Local Binary Fitting Energy," in *Computer Vision and Pattern Recognition CVPR*, 2007.
- [59] C. Li, C. Kao, J. Gore and Z. Ding, "Minimization of Region Scalable Fitting Energy for Image Segmentation," *IEEE Transactions on Image Processing*, vol. 17, no. 10, pp. 1940-1949, 2008.
- [60] S. Lankton and A. Tannenbaum, "Localizing Region-Based Active Contours," *IEEE Transaction on Image Processing*, vol. 17, no. 10, pp. 2029-2039, 2008.
- [61] L. Wang, L. He, A. Mishra and C. Li, "Active Contours Driven by Local Gaussian Distribution Fitting Energy," *Signal Processing*, vol. 86, pp. 2345-2447, 2009.

- [62] C. Lee, Robust image segmentation using active contours a level set approach, Raleigh: North Carolina State University, 2005.
- [63] M. Heiler and C. Schnoerr, "Natural image statistics for natural image segmentation," in *IEEE International Conference on Computer Vision*, 2003.
- [64] J. Kim, J. Fisher, A. Yezzi, M. Cetin and A. Willsky, "A nonparametric statistical method for image segmentation using information theory and curve evolution," *IEEE Transaction Image Processing*, vol. 14, pp. 1486-1502, 2005.
- [65] O. Michailovich, Y. Rathi and A. Tannenbaum, "Image segmentation using active contours driven by the Bhattacharya gradient flow," *Transactions on Image Processing*, vol. 16, no. 11, pp. 2787-2801, 2007.
- [66] K. Ni, X. Bresson and T. Chan, "Local Histogram Based Segmentation Using the Wasserstein Distance," *International journal on Computer Vision*, vol. 84, pp. 97-111, 2009.
- [67] M. Rajchl, J. Yuan, J. White, E. Ukwatta, J. Stirrat, C. Nambaksh, F. Li and T. Peters, "Interactive hierarchical-flow segmentation of scar tissue from late-enhancement cardiac MR images," *IEEE Transactions on Medical Imaging*, vol. 33, pp. 159-172, 2014.
- [68] M. Rousson and N. Paragios, "Prior Knowledge, Level Set Representations and Visual Grouping," *International Journal of Computer Vision*, vol. 76, pp. 231-243, 2008.
- [69] M. Ziech, C. Lavini, M. Caan, C. Nio, P. Stokkers, S. Bipat, C. Ponsioen, A. Nederveen and J. Stoker, "Dynamic contrast-enhanced MRI in patients with luminal Crohn's disease," *Eur J Radiol*, vol. 81, no. 11, p. 3019-3027, 2012.
- [70] C. Wijk, R. Truyen, R. van Gelder, L. van Vliet and F. Vos, "On Normalized Convolution to Measure Curvature Features for Automatic Polyp Detection," in *Medical Image Computing and Computer-Assisted Intervention MICCAI 2004*, Rennes, 2004.
- [71] N. Gourtsoyiannis, J. Grammatikakis, G. Papamastorakis, J. Koutroumbakis, P. Prassopoulos and M. Rousomoustakaki, "Imaging of small intestinal Crohn's disease: comparison between MR enteroclysis and conventional enteroclysis," *European Radiology*, vol. 16, no. 9, pp. 1915-25, 2006.
- [72] C. Tillack, J. Seiderer, S. Brand, B. Goke, M. Reiser and C. Schaefer, "Correlation of magnetic resonance enteroclysis (MRE) and wireless capsule endoscopy (CE) in the diagnosis of small bowel lesions in Crohn's disease," *Inflammatory bowel disease*, vol. 14, no. 9, pp. 1219-28, 2008.
- [73] J. Albert, F. Martiny, A. Krummenerl, K. Stock, J. Lesske and C. Gobel, "Diagnosis of small bowel Crohn's disease: a prospective comparison of capsule endoscopy with magnetic resonance imaging and fluoroscopic enteroclysis," *Gut*, vol. 54, no. 12, pp. 1721-7, 2005.
- [74] M. Ziech, P. Bossuyt, A. Laghi, T. Lauenstein, S. Taylor and J. Stoker, "Grading luminal Crohn's disease: which MRI features are considered as important?," *European journal of radiology*, vol. 81, no. 4, pp. 467-72, 2012.
- [75] J. Rimola, I. Ordas, S. Rodriguez, O. Garcia-Bosch, M. Aceituno and J. Llach, "Magnetic resonance imaging for evaluation of Crohn's disease: validation of parameters of severity and quantitative index of activity," *Inflammatory bowel disease*, vol. 37, no. 5, pp. 1759-68, 2011.
- [76] A. Buisson, A. Joubert, P. Montoriol, D. Da Ines, C. Hordonneau and B. Pereira, "Diffusion weighted magnetic resonance imaging for detecting and assessing ileal inflammation in Crohn's disease," *Alimentary pharmacology & therapeutics*, vol. 37, no. 5, pp. 537-45, 2013.
- [77] C. Hordonneau, A. Buisson, J. Scanzi, F. Goutorbe, B. Pereira and C. Borderon, "Diffusion-weighted magnetic resonance imaging in ileocolonic Crohn's disease: Validation of quantitative index of activity," *The American journal of gastroenterology*, vol. 109, no. 1, pp. 89-98, 2013.
- [78] H. Siddiki, J. Fidler and J. Fletcher, "Prospective comparison of state-of-the-art MR enterography and CT enterography in small-bowel Crohn's disease," *AJR*, vol. 193, pp. 113-121, 2009.
- [79] M. Ziech and J. Stoker, "MRI of the small bowel: Enterography," in *MRI of the Gastrointestinal Tract*, Springer-Verlag, 2010, pp. 117-134.
- [80] L. Dice, "Measures of the amount of ecologic association between species," *Ecology*, vol. 26, no. 3, pp. 297-302, 1945.
- [81] J. Landis and G. Koch, "The measurement of observer agreement for categorical data," *Biometrics*, vol. 33, no. 1, pp. 159-74, 1977.

- [82] A. Barbu, M. Suehling, X. Xu, D. Liu, S. Zhou and D. Comaniciu, "Automatic detection and segmentation of lymph nodes for ct data," *IEEE Transactions on Medical Imaging*, vol. 31, no. 2, pp. 240-250, 2012.
- [83] N. Pereyra, N. Dobigeon, H. Batatia and J. Tournet, "Segmentation of skin lesions in 2-d and 3-d ultrasound images using spatially coherent generalized rayleigh mixture model," *IEEE Transactions on Medical Imaging*, vol. 31, no. 8, pp. 1509-1520, 2012.
- [84] Y. Song, W. Cai, J. Kim and D. Feng, "A multistage discriminative model for tumor and lymph node detection in thoracic images," *IEEE Transactions on Medical Imaging*, vol. 31, no. 5, pp. 1061-1075, 2012.
- [85] A. Criminisi, D. Robertson, E. Konukoglu, J. Shotton, S. Pathak, S. White and K. Siddiqui, "Regression forests for efficient anatomy detection and localization in computed tomography scans," *Medical Image Analysis*, vol. 17, no. 8, pp. 1293-1303, 2013.
- [86] M. Bushan, J. Schnabel, L. Risser, M. Heinrich, J. Brady and M. Jenkinson, "Motion correction and parameter estimation in dcemri sequences: Application to colorectal cancer.," in *MICCAI*, Toronto, 2011.
- [87] K. Schunk, "Small bowel magnetic resonance imaging for inflammatory bowel disease," *Top. Magn Reson Imaging*, vol. 13, no. 6, pp. 409-425, 2002.
- [88] Z. Li, J. Tielbeek, M. Caan, C. Puylaert, M. Ziech, C. Nio, J. Stoker, L. van Vliet and F. Vos, "Expiration-Phase Template-Based Motion Correction of Free-Breathing Abdominal Dynamic Contrast Enhanced MRI," *IEEE Transactions on Biomedical Engineering*, vol. 62, pp. 1215-25, 2015.
- [89] R. Harvey and J. Bradshaw, "A simple index of Crohn's-disease activity," *Lancet*, p. 514, 1980.
- [90] P. Schuffler, D. Mahapatra, J. Tielbeek, F. Vos, J. Makanyanga, D. Pendse, C. Nio, J. Stoker, S. Taylor and J. Buhman, "A model development pipeline for Crohn's disease severity assessment from magnetic resonance images," *Abdom Imaging Comput Clin Appl*, vol. 8198, pp. 1-10, 2013.
- [91] T. Hastie, R. Tibshirani and J. Friedman, The elements of statistical learning, Springer, 2001.
- [92] H. Abdi, "The bonferroni and Sidak Corrections for Multiple Comparisons," 2007.
- [93] M. Daperno, F. Castiglione, L. de Ridder and et.al., "Results of the 2nd part Schientific Workshop of the ECCO (II): Measures and markers of prediction to achieve, detect, and monitor intestinal healing in Inflammatory Bowel Disease," *J. Crohn's Colitis*, vol. 5, pp. 484-98, 2011.
- [94] J. Steiger, "Tests for comparing elements of correlation matrix," *Psychol. Bull.*, vol. 87, pp. 245-51, 1980.
- [95] I. Ordas, J. Rimola, S. Rodriguez and e. al., "Accuracy of magnetic resonance enterography in assessing response to therapy and mucosal healing in patients with Crohn's," *Gastroenterology*, vol. 146, pp. 374-82, 2014.
- [96] M. Daporno, G. D' Haens, G. van Assche and et.al, "Development and validation of a new simplified endoscopic activity score for Crohn's disease: The SES-CD," *Gastrointest. Endosc.*, vol. 60, pp. 505-12, 2004.
- [97] G. Mazzuoli, F. Guglielmi, E. Antonelli and e. al, "Definition and evaluation of mucosal healing in clinical practice," *Dig. Liver. Dis.*, vol. 45, pp. 969-77, 2013.
- [98] I. Cima, R. Schiess, P. Wild and e. al, "Cancer genetics-guided discovery of serum biomarkers signatures for diagnosis and prognosis of prostate cancer," *Proc Natl Acad Sci USA*, vol. 108, pp. 3342-7, 2011.
- [99] S. Surinova, M. Choi, S. Tao and e. al, "Prediction of colorectal cancer diagnosis based on circulating plasma proteins," *EMBO Mol Med*, pp. 1166-78, 2015.
- [100] J. Florie, R. E. Gelder, M. P. Schutter, A. van Randen, H. W. Venema, S. de Jager, V. P. M. van der Hulst, A. Prent, S. Bipat, P. M. M. Bossuyt, L. C. and Baak and J. Stoker, "Feasibility study of computed tomography colonography using limited bowel preparation at normal and low-dose levels study," *Eur. Radiol.*, vol. 17, no. 12, pp. 3112-22, 2007.
- [101] R. M. S. Joemai, J. Geleijns and W. J. H. Veldkamp, "Development and validation of a low dose simulator for computed tomography," *Eur. Radiol.*, vol. 20, no. 4, pp. 958-66, 2010.
- [102] D. P. Frush, C. C. Slack, C. L. Hollingsworth, G. S. Bisset, L. F. Donnelly, J. Hsieh, T. Lavin-Wensell and J. R. Mayo, "Computer-simulated radiation dose reduction for abdominal multidetector {CT} of pediatric patients," *Am. J. Roentgenol.*, vol. 179, no. 5, pp. 1107-13, 2002.

- [103] O. Amir, D. Brunstein and A. Altman, "Dose optimization tool," in *Medical Imaging 2003: Visualization, Image-Guided Procedures and Display*, 2003.
- [104] B. R. Whiting, P. Massoumzadeh, O. A. Earl, J. A. O'Sullivan, D. L. Snyder and J. F. Williamson, "Properties of preprocessed sinogram data in {X}-ray computed tomography," *Med. Phys.*, vol. 33, no. 9, pp. 3290-303, 2006.
- [105] K. Yang, A. L. C. Kwan, S. Y. Huang, N. J. Packard and J. M. Boone, "Noise power properties of a cone-beam {CT} system for breast cancer detection," *Med. Phys.*, vol. 35, no. 12, pp. 5317-27, 2008.
- [106] J. Ma, Z. Liang, Y. Fan, Y. Liu, J. Huang, W. Chen and H. Lu, "Variance analysis of X-ray CT sinograms in the presence of electronic noise background," *Med. Phys.*, vol. 39, no. 7, pp. 4051-65, 2012.
- [107] A. S. Wang and N. J. Pelc, "Synthetic CT: Simulating low dose single and dual energy protocols from a dual energy scan," *Med. Phys.*, vol. 38, pp. 5551-5562, 2011.
- [108] A. S. Wang and N. J. Pelc, "Synthetic CT: simulating arbitrary low dose single and dual energy protocols," in *Medical Imaging 2011: Physics of Medical Imaging, Proc. SPIE 7961*, 2011.
- [109] S. Zabic, Q. Wang, T. Morton and K. M. Brown, "A low dose simulation tool for CT systems with energy integrating detectors," *Med. Phys.*, p. 031102, 2013.
- [110] A. S. Wang, S. Feng and N. J. Pelc, "Image-based synthetic CT: simulating arbitrary low dose single and dual energy protocols from dual energy images," in *Medical Imaging 2012: Physics of Medical Imaging, Proc. SPIE 8313*, 2012.
- [111] A. S. Wang, J. Webster Stayman, Y. Otake, S. Vohgt, G. Kleinszig, A. Jay Khanna, G. L. Gallia and J. H. Siewerdsen, "Low-dose preview for patient-specific, task-specific technique selection in cone-beam CT," *Medical Physics*, vol. 41, no. 7, p. 071915, 2014
- [112] A. Britten, M. Crotty, A. Kiremidjian, H. Grundy and E. Adam, "The addition of computer simulated noise to investigate radiation dose and image quality in images with spatial correlation of statistical noise: an example application to X-ray CT of the brain," *Br. J. Radiol.*, vol. 77, no. 916, pp. 323-8, 2004.
- [113] K. Boedeker, V. Cooper and M. McNitt-Gray, "Application of the noise power spectrum in modern diagnostic MDCT: part I. Measurement of noise power spectra and noise equivalent quanta," *Phys. Med. Biol.*, vol. 52, no. 14, pp. 4027-46, 2007.
- [114] K. Faulkner and B. Moores, "Analysis of X-ray computed tomography images using the noise power spectrum and autocorrelation function," *Phys. Med. Biol.*, vol. 29, no. 11, pp. 1343-52, 1984.
- [115] C. W. Kim and J. H. Kim, "Application of CT Simulation Technique for Virtual Ultra-Low-Dose Trial in CT Colonography," in *Abdominal Imaging, Computational and Clinical Applications, 4th International Workshop*, Nice, France, 2012.
- [116] C. Kim and J. Kim, "Realistic simulation of reduced-dose CT with noise modeling and sinogram synthesis," *Med. Phys.*, vol. 41, no. 1, p. 011901, 2014.
- [117] R. Wagner, D. Brown and M. Pastel, "Application of information theory to the assessment of computed tomography," *Med. Phys.*, vol. 6, no. 2, pp. 83-94, 1979.
- [118] K. Hanson, "Detectability in computed tomographic images," *Med. Phys.*, vol. 6, no. 5, pp. 441-451, 1979.
- [119] A. Wunderlich and F. Noo, "Image covariance and lesion detectability in direct fan-beam X-ray computed tomography," *Phys. Med. Biol.*, vol. 53, no. 10, pp. 2471-93, 2008.
- [120] A. Kak and M. Stanley, *Principles of Computerized Tomographic Imaging*, IEEE Press, 1988.
- [121] J. H. Hubbel and S. M. Seltzer, "Tables of {X}-Ray Mass Attenuation Coefficients and Mass Energy-Absorption Coefficients (version 1.4)," 2004. [Online]. Available: <http://physics.nist.gov/xaamdi>, [Accessed: 2013, March 3].
- [122] A. Papoulis, *Probability, random variables and stochastic processes*, McGraw-Hill, 1984.
- [123] A. Macoviski, *Medical Imaging Systems*, Prentice-Hall, 1982.
- [124] ImPACTGroup, "Comparitive specifications: 64-slice CT-scanners," 2009.
- [125] M. F. Kijewski and P. F. Judy, "The noise power spectrum of CT-images," *Phys. Med. Biol.*, vol. 32, no. 5, pp. 565-75, 1987.

- [126] S. J. Riederer, N. J. Pelc and D. A. Chesler, "The noise power spectrum in computed X-ray Tomography," *Phys. Med. Biol.*, vol. 23, no. 3, pp. 446-54, 1978.
- [127] J. Wang, H. Lu, Z. Liang, D. Eremina, G. Zhang, S. Wang, J. Chen and J. Manzione, "An experimental study on the noise properties of X-ray CT sinogram data in  $\{R\}$  adon space," *Phys. Med. Biol.*, vol. 53, no. 12, pp. 3327-41, 2008.
- [128] T. M. Benson and B. K. B. de Man, "Synthetic CT noise emulation in the raw data domain," in *Synthetic CT noise emulation in the raw data domain*, 2010.
- [129] J. Kim, M. Cetin and A. Willsky, "Nonparametric shape priors for active contour-based image segmentation," *Signal Processing*, pp. 3021-3044, 2007.
- [130] M. Rousson and N. Paragios, "Prior Knowledge, Level Set Representations and Visul Grouping," *International Journal of Computer Vision*, vol. 76, pp. 231-243, 2008.
- [131] S. Dambreville, Y. Rathi and A. Tannenbaum, "A Framework for Image Segmentation Using Shape Models and Kernel Space Shape Priors," *Transactions on Pattern Analysis and Machine Intelligence*, pp. 1385-1399, 2008.
- [132] Y. Gao, B. Corn, D. Schifter and A. Tannenbaum, "Multiscale 3D representation and segmentation with applications to hippocampal/caudate extraction from brain MRI," *Medical Image Analysis*, pp. 374-385, 2012.
- [133] J. Rimola and e. al., "Characterization of Inflammation and Fibrosis in Crohn's Disease Lesions by Magnetic Resonance Imaging," *The American Journal of Gastroenterology*, pp. 432-440, 2015.

# Summary

---

Cross-sectional medical imaging techniques have become indispensable in assessing abnormalities in the anatomy and physiology of the abdomen in clinical practice. This thesis addresses two specific challenges. The first part of the thesis investigates the application of MRI and image processing techniques to assess Crohn's disease activity. The second part presents a realistic low-dose CT-image simulator to facilitate in-silico dose optimization based on available "high-dose" data sets.

Grading Crohn's disease activity is important for monitoring disease activity and treatment planning. Crohn's disease is a chronic inflammatory disease of the gastrointestinal tract with a relatively high incidence in the Western world, affecting 700 thousand individuals in Europe alone. The disease is characterized by a chronic relapsing and remitting course, i.e. periods of exacerbations are alternated by episodes of diminished disease activity. Accordingly, the mere presence of the disease must be distinguished from active disease, which can occur at varying levels of severity and each level requiring a different treatment. Ideally, a disease activity score should be objective, reproducible, quantifiable, non-invasive and comprehensive. Currently, imperfect clinical, endoscopic and histopathology scores such as CDAI, CDEIS and D'Haens index are used to assess the disease activity. The aim of the VIGOR++ project was to develop a novel grading scheme based on multi-sequential MRI volumes.

Segmentation of the bowel's lumen surface is a crucial step to automatically assess Crohn's disease activity based on magnetic resonance (MR) images. This is challenging due to MR signal variations, heterogeneous lumen content, the presence of stenosis, and the diversity of the surrounding tissues. We presented a region-based active contour algorithm that can handle these challenges by incorporating prior knowledge about properties of the lumen content (e.g. the stratified appearance) and knowledge of the adjacent anatomy in a space-variant regional mixture model. This information was added to the model by estimating material fractions inside each voxel. A level set representation employing a gradient-descent scheme was used to obtain the actual segmentation. The method was tested on 61 regions from 59 patients presenting Crohn's disease activity. It was shown that the difference between the automated segmentations and manual



annotations is comparable to the difference between the manual annotations themselves.

The lumen surface acted as a starting point for bowel wall segmentation and subsequent thickness measurements. The method consisted of four steps: (1) initialization by manually drawing a centerline; (2) segmentation of the bowel wall's inner surface; (3) segmentation of the wall's outer surface; (4) measurement of the bowel wall thickness. The algorithm was run twice starting from two centerlines. This yielded two segmentation results of the bowel wall and two measures of its thickness. The method was separately evaluated regarding its ability (1) to segment a diseased part of the bowel wall and (2) to measure a segment's thickness. The overlap of the semi-automatic segmentations was significantly larger than the overlap of the manual annotations. The intraclass correlation coefficient (ICC) of semi-automatic measurements was significantly higher than the ICC of manual measurements. As such, the semi-automated measurement technique facilitates a highly reproducible delineation of regions with active Crohn's disease. Furthermore, the semi-automatic thickness measurement sustains a significantly higher intraclass correlation than manual observers.

In VIGOR++, a method was developed for scoring disease activity based on manual and automated MRI features using ileocolonoscopy as the reference standard. Patient data (both MRI and colonoscopy) from a retrospective study was used to conduct an exhaustive search to select the best scoring system. Next, this system was validated based on an independent test set. The so-called VIGOR++ system was compared with other state-of-the-art scoring systems such as the MaRIA-score and the London score. The VIGOR++ score was shown to have a considerably higher inter-observer agreement than both the MaRIA and the London score.

The second part of the thesis is dedicated to abdominal CT. A low-dose Computed Tomography (CT) simulator was presented, which facilitated in-silico studies into the required dose for a diagnostic task. Conventionally, low-dose CT images are simulated by adding noise to the recorded projection data. However, this is not always achievable in practice as the raw projection data are simply not available. I aimed to present a new method for simulating patient-specific, low-dose CT images without the need of the original projection data. The required scanner-specific parameters such as the apodization window, the bow-tie attenuation



coefficients, the X-ray tube output parameter (reflecting the photon flux) and the detector read-out noise are retrieved from calibration images of a water cylinder. The low-dose simulation method was evaluated by comparing the noise characteristics in simulated images with experimentally acquired data. It was demonstrated that the simulated low-dose images accurately reproduced the noise characteristics of experimentally acquired low-dose-volumes. Hence, the new methodology could aid in further optimizing CT protocols by facilitating in-silico studies on dose dependency of low-contrast object detectability.

# Samenvatting

---

Cross-sectionale medische beeldbewerkingstechnieken zijn onmisbaar geworden in de klinische praktijk voor de beoordeling van abnormaliteiten in de anatomie en fysiologie van het abdomen. Dit proefschrift behandelt twee specifieke problemen in dit domein. Het eerste deel van het proefschrift onderzoekt de toepassing van MRI en beeldbewerkingstechnieken om de ziekte van Crohn activiteit te beoordelen. Het tweede deel presenteert een realistische lage dosis CT-beeld simulator die ons in staat stelt dosis optimalisatie experimenten uit te voeren op oude hoge dosis beelden.

Het graderen van de ziekte van Crohn activiteit is belangrijk voor het monitoren van de ziekte voortgang en het bepalen van de behandelstrategie. De ziekte van Crohn is een chronische ontstekingsziekte in het maag-darmkanaal met een relatief hoge incidentie in de Westerse wereld. In Europa alleen al lijden zo'n 700 duizend individuen aan deze ziekte. De ziekte kenmerkt zich door onregelmatige afwisselingen tussen actieve en minder actieve fasen. Hierom is het belangrijk om naast het bestaan van de ziekte ook de activiteit te bepalen. Deze kan worden opgedeeld in verschillende score niveaus en elk score niveau heeft een ander behandelplan. Idealiter is zo'n ziekte activiteit score objectief, reproduceerbaar, kwantitatief, niet-invasief en allesomvattend. Op dit moment worden er helaas imperfecte score systemen gebruikt, zoals CDAI, CDEIS en D'Haens, die zijn gebaseerd op o.a. endoscopie en histopathologie. Het doel van het VIGOR++ project was het ontwikkelen van een scoring systeem uit multi-sequentiële MRI-volumes.

De segmentatie van het darm binnen-oppervlak was een cruciale stap in het automatisch bepalen van de ziekte van Crohn activiteit uit MRI beelden. Dit was een complex probleem door willekeurige variaties in het MRI signaal, heterogeniteit van het darmlumen, de aanwezigheid van stenose en de verscheidenheid van de omringende weefsels. In dit proefschrift werd een nieuw "region based active contours" algoritme gepresenteerd dat met dergelijke complexiteit kon omgaan door gebruik te maken van voorkennis van de inhoud van het darmlumen (bijv. de gelaagdheid van de inhoud) en de omliggende anatomie, door dit te modelleren in een plaats-afhankelijk regionaal mixture model. Hierin werd voor elke potentiële lumen voxel bij voorbaat een schatting gemaakt van de materiaal fractie. Het algoritme maakte gebruik van een "level

set” representatie van het darmoppervlak, daarnaast werd de gradiënt afdaling methode gebruikt om de uiteindelijke oplossing te vinden. De methode was getest op 61 darmsegmenten van 59 patiënten met de diagnose ziekte van Crohn. Hierin werd aangetoond dat het verschil tussen de verkregen segmentaties en de handmatige annotaties vergelijkbaar was met het verschil tussen de annotaties onderling.

De segmentatie van het lumen oppervlak diende als startpunt voor het verkrijgen van de darmwand segmentatie en de daar uitvolgende diktemetingen. De methode bestond uit vier stappen: (1) Initialisatie door een handmatig getekende pad; (2) Segmentatie van het darm binnen oppervlak; (3) Segmentatie van het darm buiten oppervlak; (4) De meting van de dikte van de darmwand. Twee onderzoeksassistenten tekenden (elk begeleid door een andere radioloog) onafhankelijk van elkaar paden en vervolgens werd voor elk pad het algoritme uitgevoerd. Hierdoor werden voor veel segmenten twee segmentaties verkregen en dus ook twee diktemetingen. De methode werd vervolgens apart gevalideerd aan de hand van de kwaliteit van de segmentaties en de kwaliteit van de diktemetingen. De gevonden overlap tussen de segmentaties was significant hoger dan de overlap tussen de handmatige annotaties. Tevens was de intraclass correlatiecoëfficiënt (ICC) van de semiautomatische diktemetingen significant groter dan de ICC van de handmatige metingen. Als zodanig stelde deze semiautomatische methode de onderzoekers in staat zeer reproduceerbare omlijnningen en diktemetingen te verkrijgen van zieke stukken darmwand.

In VIGOR++ werd een methode ontwikkeld om de ziekte activiteit te scoren aan de hand van handmatig en automatisch verkregen MRI features en gebruikmakend van ileocolonoscopie als referentie standaard. Patiënt data (zowel MRI als colonoscopie) van een retrospectieve cohort werd gebruikt om de beste features te selecteren voor het model. Vervolgens werd het model getest op een onafhankelijke test-set. Het zogenaamde VIGOR++ systeem werd vergeleken met andere zeer recentelijk ontwikkelde scoringssystemen zoals de MaRIA-score en de London score. Hieruit volgde dat de VIGOR++ score een veel hogere interobserver overeenkomst had dan zowel de MaRIA als de London score.

Het tweede deel van het proefschrift was gewijd aan abdominale CT. Hierin werd een lage dosis CT simulator gepresenteerd, die het mogelijk maakte om studies uit te voeren waarin de minimale stralingsdosis bepaald zou kunnen worden voor een specifieke diagnostische taak. Het was gebruikelijk om lage dosis CT-beelden

te simuleren door ruis toe te voegen aan de ruwe projectie data. Echter was dit niet altijd mogelijk in de praktijk omdat de ruwe projectie data simpelweg niet beschikbaar was. In dit proefschrift werd een nieuwe methode geïntroduceerd om patiënt specifieke lage dosis CT beelden te simuleren zonder daarvoor de ruwe projectiedata te gebruiken. De vereiste scanner-specifieke parameters zoals het afkapvenster van het reconstructiefilter, het bowtie filter, de fotonen flux afkomstig van de röntgenbron en de detector ruis werden verkregen uit kalibratie beelden van een watercilinder. De lage dosis CT simulator was gevalideerd door de ruiseigenschappen in de gesimuleerde beelden te vergelijken met die in experimenteel verkregen data. Hierin werd gedemonstreerd dat de gesimuleerde lage dosis beelden accuraat de ruiseigenschappen nabootsten van de experimenteel verkregen beelden. Hierdoor zou de stralingsdosis van bestaande CT protocollen verder verlaagd kunnen worden en mogelijk geoptimaliseerd.

# Dankwoord

---

Het is een surrealistisch gevoel om te realiseren dat ik dan eindelijk ben toegekomen aan het dankwoord. Op zo'n moment realiseer ik me dat ik het proefschrift niet zou hebben kunnen volbrengen als ik niet voldoende gesteund werd door de mensen om me heen. Ondanks dat ik zeer waarschijnlijk een hoop personen onrecht aan doe door slechts een klein aantal personen persoonlijk te bedanken, wil ik toch een poging wagen.

Uiteraard zou ik mijn begeleiders Frans, Lucas en Jaap willen bedanken voor de kans die ze me gegeven hebben en ook voor hun tomeloze geduld en vertrouwen, zelfs op de momenten dat dit bij mij zelf ontbrak. Jaap, je klinische input was onmisbaar, het is goed dat je me er regelmatig op wees dat ik niet met een strikt wiskundig probleem bezig was. Lucas, je hebt me geleerd dat de essentie van een model ligt in zijn onderliggende aannames en niet in de beschrijving van het model zelf. Dit dwong mij kritisch te leren kijken naar de op het oog simpele maar cruciale “waarheden”. De simpele vragen bleken vaak de meest moeilijk te beantwoorden, ik neem deze wijsheid in het vervolg niet alleen ter harte in beeldanalyse problemen maar in het leven algemeen. Frans, ik ben je dankbaar voor de juiste motiverende woorden zodra deze nodig waren. Je vertelde me regelmatig om elke dag kleine stappen te maken (in het schrijven/segmenteren), dan zouden dat er vanzelf veel worden. Had ik maar beter geluisterd in dat aspect! Ik heb een bewondering voor hoe je theorieën en modellen weet te reduceren tot de kern en deze uit te leggen aan minder technisch onderlegde projectgenoten en studenten. Dit proces observeren was zeer leerzaam voor mij.

Dank aan mijn VIGOR collega's en in het bijzonder Laurence en Rado van 3Dnet medical. De vele avonduren die we samen op teamviewer doorbrachten om de project deadlines te halen, hebben een vaste gebruiker van me gemaakt. En speciale dank aan Jeroen van Schie. Het is best bijzonder, hoe je elkaar kunt gaan waarderen, ondanks onze diametraal verschillende persoonlijkheden. De herinneringen aan onze tijd als kamergenoten tijdens de werkoverleggen in Londen toveren nog steeds een lach op mijn gezicht. Als laatste een speciale dank aan Carl voor het geven van de nodige klinische inzichten. Ik heb overigens genoten van de vele inhoudelijke Skype gesprekken over hoe ziekte van Crohn features het beste gemeten kunnen worden.

Een zeer grote hoeveelheid dank gaat uit naar mijn (oud)collega's van de Quantitative Imaging Group, ik ben blij de kans te hebben gehad te werken met een zeer diverse maar vooral ook leuke open groep collega's. Ik zou over elk van jullie wel iets moois kunnen schrijven, wederom zal ik me moeten inhouden. Vincent, je was mijn gids toen ik net bij de groep kwam kijken, dank je dat ik over je schouder mee mocht kijken. Robert, we kennen elkaar onze hele Delftsche tijd al, het is fijn dat er tijdens het proefschrift een aantal bijzondere herinneringen zijn toegevoegd. Een speciale dank gaat uit naar Zhang, mijn kamergenoot gedurende de eerste jaren. Het ontroerde me toen je zei dat je de afgelopen jaren meer tijd met mij door hebt gebracht dan met je echtgenote. Hopelijk heb je dat het laatste jaar kunnen inhalen. Ik waardeer het dat je in die tijd een echt klankbord voor me was en het was mooi om te zien hoe we gegroeid zijn in die periode. Juan Pedro, dank je voor het met verve opvullen van het gat dat Zhang heeft achtergelaten. Verder zou ik de rest van (oud) QI expliciet willen bedanken, dus Milos, Lennard, Qiaole, Alex, Rahil, Hui, Sanneke, Ganesh, Martijn G, Joor, Jiangfei, Thijs, Robert, Ronald, Mandy, Annelies, Nadya, Mojtaba, Thijs, Lena, Jelena, Carlas, Jeroen H, Kedir, Anna, Jeroen S, Tian, Jelle, Tom, Babak, Willem, Gyllion, Emile, Leon, Jos, Hamidreza, Martijn B, Mohammed, Richard, Bernd, Sjoerd, Jeroen K, Ted, Piet, Ron en Jaap C dank jullie voor jullie steun en de tijd samen.

Daarnaast zou ik VOC-bier willen bedanken voor de in- en ontspanning op zondag en in het bijzonder het gezelschap onnavolgbaar, zonder jullie was ik waarschijnlijk nooit van Rotterdam gaan houden. Ook zou ik mijn clubgenoten willen bedanken voor hun steun. Als ik ergens mee zat, kon ik altijd bij ieder van jullie mijn ei kwijt. Een extra bedankje voor mijn huisgenoten in de PhD periode Hidde, Pieter en Alwin. Het is waar dat je elkaar pas echt leert kennen als je samenwoont en vooral in zo'n vormende periode als net na het studeren. Het hieruit voortgekomen collectief Nachtdier is een mooie verrassing aan het einde van mijn proefschrift.

En Carlijn, dank je voor je aanwezigheid en je emotionele steun. Je hebt de rusteloosheid in me weggenomen, waardoor ik me weer heb leren focussen. Petros, broertje van me, op jou kon ik altijd rekenen voor elke type ondersteuning. Je bent trouwens de enige van wie ik wist dat je me flink de waarheid durfde te vertellen. Daarvoor wil ik je in het bijzonder bedanken! En als laatste wil ik mijn ouders Hiwet en Dimitri bedanken, jullie zijn het fundament waar ik altijd op kon

terugvallen. Het vertrouwen dat jullie in mij hebben is inspirerend te noemen.  
Zonder jullie was dit boekje er nooit geweest.

# Curriculum Vitae

---

Robiel Naziroglu was born in Utrecht, the Netherlands, on December 11<sup>th</sup> 1986. In 2005, he obtained his Gymnasium diploma at Christelijk Lyceum Zeist and started to study applied physics in Delft. He earned his BSc in Applied physics in 2009 and his MSc in Biomedical Engineering in 2011.

In 2011, he started as a PhD involved in the VIGOR++ project at the Quantitative Imaging group of Delft University of Technology on the topic of *Analysis methods for abdominal imaging: quantifying Crohn's disease activity*. He was supervised by Prof. dr. ir. Lucas J. van Vliet and dr. Frans M. Vos.



---