DETECTION AND CLASSIFICATION OF GASTROINTESTINAL CANCER AND OTHER PATHOLOGIES THROUGH QUANTITATIVE ANALYSIS OF OPTICAL COHERENCE TOMOGRAPHY DATA AND GONIOPHOTOMETRY

Dr Iakovos Amygdalos

Department of Surgery and Cancer, Imperial College London,
St Mary's Hospital, London, UK

2014

A thesis submitted to Imperial College London for the degree of
Doctorate of Philosophy.

Supervisors

Professor George B. Hanna

Department of Surgery and Cancer, Imperial College London,
Academic Surgical Unit, St Mary's Hospital, London, UK

Professor Robert D. Goldin

Department of Medicine, Imperial College London
Department of Histopathology, St Mary's Hospital, London, UK

Abstract

The changes in light interaction between healthy and diseased tissues have been investigated as a potential diagnostic application. Here we attempt to differentiate between healthy and pathological gastrointestinal tissues using quantitative analysis of optical coherence tomography (OCT) data and goniophotometry.

A goniophotometer was constructed and calibrated using titanium oxide and microsphere phantoms. Measurements were carried out on human gastrointestinal tissue sections collected using the methodology described below. The anisotropy factor g was extracted from the scattering curves by fitting the Henyey-Greenstein function. Measurements on human samples were in the forward scattering range with g 0.6-0.7, in agreement with the literature.

Optical coherence tomography imaging was carried out on gastrointestinal tissues collected from patients undergoing elective surgery or endoscopy at St. Mary's Hospital, London. In total 146 patients were included. Data was processed using gradient analysis of signal attenuation and morphological analysis with kNN classification. Results were correlated with histological diagnoses. Gradient analysis results were statistically significant across most categories, showing particularly good differences in the gradient distributions between healthy and diseased oesophageal tissues. Morphological analysis and kNN classification produced sensitivity and specificity values for healthy oesophagus and cancer in surgical specimens reaching 100% / 97.87% and 99.99% / 99.91% respectively and high accuracy in detecting Barrett's oesophagus in endoscopic specimens, with sensitivity and specificity values of 99.80% and 99.02%. Results in rectal tissue where also noteworthy, with detection of dysplasia reaching a sensitivity and specificity of 99.55% / 96.01%.

Despite limitations in our work, we have shown that the detection of gastrointestinal pathologies using quantitative analysis of OCT data is a promising technique with good *ex vivo* results. Transferring the methodology to the *in vivo* domain holds a lot of potential as a future quick and reliable diagnostic technique.

Declaration

I hereby declare that I am the sole author of this thesis and that all work within it is my own. Any individuals who carried out any collaboration work are appropriately credited. I authorise the library of Imperial College London to lend this thesis to other institutions or individuals.

Collection and imaging of endoscopic gastrointestinal specimens was carried out by Shalil Solanki and Hussain Selmi, as part of their BSc projects. Methodological design, data processing, analysis and write-up of work on endoscopic gastrointestinal tissues for this thesis was carried out by the author.

Ethical approval was granted by the Royal Free Hospital & Medical School Research Ethics Committee. REC reference number: 09/H0720/105

March 2014

Dr Iakovos Amygdalos

Οἱ σπείροντες ἐν δάκρυσιν ἐν ἀγαλλιάσει θεριοῦσι.
-Ψαλ. 125,5

(They that sow in tears shall reap in joy.
-Psalm 126:5)

Dedication

To my parents, the giants on whose weary shoulders I humbly stand and to my sister, in whose achievements and thirst for academia I find inspiration. Thank you for your love, support and sacrifices.

To my friends, who stayed close to me even from thousands of miles away. Thank you for keeping me balanced, sane and grounded.

To the patients who, even in their darkest hour, did not hesitate to contribute. Your generosity is humbling.

Acknowledgements

Prof George B Hanna

Prof Robert D Goldin

Dr Daniel S Elson

Dr Pillar Beatriz Garcia Allende

Dr Yuki Nehikhare

Dr Shalil Solanki

Dr Hussain Selmi

Dr Hiruni Dhanapala

Dr Jonathan Hoare

Mr Justin Vale

Dr Jo Lloyd

Dr Marjorie Walker

The lab staff at St Mary's Hospital Department of Histopathology

Peer-reviewed publications and presentations from this thesis

Publications

Morphological analysis of optical coherence tomography images for automated classification of gastrointestinal tissues. Garcia-Allende PB, Amygdalos I, Dhanapala H, Goldin RD, Hanna GB, Elson DS. Biomed Opt Express. 2011 Oct 1;2(10):2821-36. Epub 2011 Sep 22 PMID: 22091441

Presentations

Morphological image analysis for classification of gastrointestinal tissues using optical coherence tomography. Garcia-Allende PB, Amygdalos I, Dhanapala H, Goldin RD, Hanna GB, Elson DS. Optical Coherence Tomography and Coherence Domain, Optical Methods in Biomedicine XVI, SPIE Photonics West 2012, San Francisco, USA (Jan 2012)

Detection of lower gastrointestinal pathologies through quantification of optical coherence tomography signal attenuation using a linear model. Selmi H, Amygdalos I, Solanki S, Garcia-Allende PB, Dhanapala H, Hoare J, Erry G, Elson DS, Goldin RD, Hanna GB. Society of Academic and Research Surgery (SARS) Annual Meeting 2012, University of Nottingham Medical School, Nottingham, UK (Jan 2012)

Detection of upper gastrointestinal pathologies through quantification of optical coherence tomography signal attenuation using a linear model. Solanki S, Amygdalos I, Selmi H, Garcia-Allende PB, Dhanapala H, Hoare J, Erry G, Elson DS, Goldin RD, Hanna GB. Society of Academic and Research Surgery (SARS) Annual Meeting 2012, University of Nottingham Medical School, Nottingham, UK (Jan 2012)

Detection of upper gastrointestinal pathologies with optical fibre probe fluorescence spectroscopy. Solanki S, Sauvage V, Amygdalos I, Selmi H, Dhanapala H, Hoare J, Goldin RD, Hanna GB, Elson DS. Society of Academic and Research Surgery (SARS) Annual Meeting 2012, University of Nottingham Medical School, Nottingham, UK (Jan 2012)

Optical fibre probe reflectance spectroscopy as a technique to diagnose lesions in the lower gastrointestinal tract. Selmi H, Sauvage V, Amygdalos I, Solanki S, Dhanapala H, Hoare J, Goldin RD, Hanna GB, Elson DS. Society of Academic and Research Surgery (SARS) Annual Meeting 2012, University of Nottingham Medical School, Nottingham, UK (Jan 2012)

Automated classification of gastrointestinal tissues using optical coherence tomography. Garcia Allende PB, Amygdalos I, Dhanapala H, Erry G, Goldin RD, Hanna GB, Elson DS. Advances in Optics for Biotechnology, Medicine and Surgery XII, Florida, USA (June 2011)

The optical properties of healthy versus cancerous urological tissue. Nehikhare O, Amygdalos I, Erry G, Tomlins P, Walker M, Vale J, Hanna GB. British Association of Urological Surgeons (BAUS) Section of Academic Urology, Society of Academic & Research Surgery (SARS) Conference, Royal College of Surgeons of Ireland, Dublin, Ireland (Jan 2010)

The optical properties of healthy vs. cancerous urological tissue. Nehikhare O, Amygdalos I, Tomlins P, Walker M, Lloyd J, Vale J, Hanna GB. Poster Presentation, 4th Antwerp Medical Students' Congress (AMSC), European Medical Students' Association (EMSA), Antwerp, Belgium (Sep 2010)

Table of Contents

List of Figures

List of Tables

List of Abbreviations

1 INTRODUCTION

1.1 The healthy gastrointestinal tract

1.1.1 The normal oesophagus and gastro-oesophageal junction

The oesophagus is a muscular tube, approximately 25cm long, which connects the pharynx to the stomach. It starts in the neck, descends through the thorax and pierces the diaphragm before joining the stomach at the gastro-oesophageal junction (GOJ). As with the rest of the gastrointestinal tract, its wall consists of multiple layers. These are the mucosa (itself split into epithelium, lamina propria and muscularis mucosae), submucosa, muscularis externa and adventitia [1, 2].

The oesophageal mucosa is a thick, smooth, greyish-pink layer. Its outermost layer is a non-keratinized, stratified squamous epithelium, which is between 300-500 μm thick and plays a protective role against thermal, chemical and mechanical injury [1, 3]. This is further divided into a basal proliferative layer, a parabasal layer and a flattened superficial layer. Epithelial cells are attached to the basal lamina, which in turn is attached to the lamina propria with the assistance of connective tissue papillae. The epithelium is constantly being regenerated, with cells moving from the basal layer to the superficial squamous layer over 2-3 weeks. During this process -which is considerably slower than in the stomach and intestine- the cells change shape to become progressively flatter, till they are desquamated. Below the epithelium is the lamina propria, containing mucosa-associated lymphoid tissue (MALT) and mucous glands. Cells within it secrete growth factors which regulate turnover and differentiation of overlying epithelial cells. Underneath the latter is the muscularis mucosae, a thin sheet of mainly longitudinal smooth muscle [1]. The submucosa is the strongest layer mechanically and contains larger blood vessels, nerves and more mucous glands [1, 3]. It loosely connects the mucosa to the muscularis externa and, thanks to its elastic fibres, assists in the closure of the oesophageal lumen during peristalsis. Finally, the muscularis externa is the outermost layer of the oesophagus and is composed of two layers of muscle, an outer longitudinal layer and an inner circular layer, which generate the peristaltic movements that propel food through the oesophagus. It is normally up to 300 μm thick [1].

The GOJ is a poorly defined anatomical area, where the distal oesophagus joins the proximal stomach (cardia). It can be considered as the last 1 cm of the oesophagus and first 1 cm of the stomach, though true anatomic location of the GOJ can vary between individuals. The GOJ

is histologically the transition point between the squamous oesophageal epithelium and the gastric columnar epithelium. The zone of transition is called the *Z line* [1, 3]. In many adults the *Z line* is actually proximal to the anatomical GOJ, with physiological columnar mucosa located in the distal oesophagus, which can be hard to distinguish from pathological columnar metaplasia in biopsies from the GOJ [3].

1.1.2 The normal stomach

The stomach is the widest part of the gastrointestinal tract and is situated in the upper abdomen. It is divided into the fundus, body, antrum and pylorus, which have different internal appearances and microstructures. As with the oesophagus, the stomach wall consists of mucosa, submucosa, muscularis externa and serosa [1].

The gastric mucosa is mostly reddish brown and, in the contracted stomach, is characterised by multiple folds or rugae. These are most marked along the greater curvature and towards the pylorus and disappear when the stomach is distended. The stomach epithelium is markedly different to that of the oesophagus and the transition from one to the other is abrupt and occurs at the cardiac orifice. There, the stratified oesophageal epithelium ends and is replaced by a simple columnar epithelium, consisting of mucous cells, which release gastric mucus from their apical ends. This forms a thick, protective and lubricant layer over the gastric lining. The latter is honeycombed by small, irregular gastric pits (approximately 200 μm in diameter), at the base of which several long, tubular gastric glands are located. These are divided into three groups – the cardiac, principal (in the body and fundus) and pyloric glands. The cardiac and pyloric glands consist predominantly of mucus-secreting cells. The principal glands have at least five cell types in their walls, namely chief, parietal, mucous neck, stem and neuroendocrine cells. They produce the digestive enzymes pepsin and lipase, gastric acid, intrinsic factor, mucus, gastrin, somatostatin and histamine. The lamina propria forms a connective tissue framework around the glands and contains lymphoid tissue, nerves and blood vessels. Underneath, the muscularis mucosae forms a thin layer of smooth muscle, whose contraction probably assists in the emptying of the glands. The gastric submucosa contains loose connective tissue with collagen bundles, elastin fibres and neurovascular bundles. It is surrounded by the muscularis externa, which is unique in the stomach, as it has three layers. From innermost outwards, these are the oblique, circular and longitudinal layers. They produce a churning movement which mixes

food with gastric secretions. The final layer is the serosa, which is an extension of the visceral peritoneum. It covers the entire stomach apart from the omental attachments at the lesser and greater curvatures and a small area near the diaphragm [1].

1.1.3 The normal large intestine

The large intestine is approximately 1.5m long, extending from the distal end of the ileum to the anus. Its principal function is the absorption of fluids and solutes [1].

The layered pattern described previously continues in the large intestine. The mucosa here is pale, smooth and raised into numerous folds in the colon. In the rectum it is thicker, darker and more vascular. The epithelium is predominantly lined by columnar and mucous (goblet) cells. These are also present in the intestinal glands (crypts), which also contain stem cells and neuroendocrine cells. The columnar cells are primarily responsible for absorption of gut contents and are characterised by apical microvilli, miniscule projections into the lumen which significantly increase the surface area of the intestine. The lamina propria and muscularis mucosae are similarly structured to the rest of the GI tract, as is the submucosa. The muscularis externa is quite different, as, with the exception of the appendix, the longitudinal muscle is aggregated into bands called taeniae coli. Between them, the longitudinal layer is much thinner. The circular fibres form a thicker layer in the rectum and constitute the internal anal sphincter in the anal canal. The serosa (visceral peritoneum) is variable in extent and forms small fat-filled appendices (epiploicae) along the colon [1].

1.2 Pre-malignant conditions and progression from healthy to cancer

1.2.1 Metaplasia, dysplasia and carcinoma in situ

Metaplasia is defined as the replacement of one mature cell type with another. This is almost always found in areas of damage, repair and regeneration, where tissues undergo changes to adapt to a environmental stimulus, but can also be physiological, as in the cervix. For example, gastro-oesophageal reflux damages the squamous epithelium of the oesophagus, which is replaced by columnar cells [2].

Dysplasia is characterised by loss in the uniformity of individual cells and disruption in their normal architectural organisation, i.e. disordered growth. It often occurs in metaplastic tissue and displays some of the characteristics of cancer, such as pleomorphism, hyperchromatism, high nuclear-to-cytoplasmic ratio and increased number of mitoses [2, 4]. However, the mitoses in dysplasia are almost always of a normal configuration, the difference with healthy tissue being that they occur at a higher frequency and in abnormal locations. For example, in dysplastic stratified epithelium, they are not limited to the basal layers, but instead appear throughout. *High grade dysplasia* exhibits more severe cytological and architectural changes than *low grade dysplasia* [2, 4].

When the entire thickness of the epithelium is severely dysplastic, but the lesion remains confined by the basement membrane, it is termed *carcinoma in situ*. This denotes a pre-invasive neoplasm, which may take years to breach the basement membrane and thus become invasive. Dysplasia does not necessarily progress to cancer and may in fact be reversible if not severe, however it frequently predates the development of cancer, for example in columnar-lined (Barrett's) oesophagus [2, 4].

1.2.2 Oesophagus and gastro-oesophageal junction

1.2.2.1 Oesophagitis

Causes of *oesophagitis* (inflammation of the oesophagus) include lacerations (such as *Mallory-Weiss tears*), chemical injuries, infections, gastro-oesophageal reflux disease (GORD) and eosinophilic infiltration (*Eosinophilic oesophagitis*). By far the most common is GORD, in

which chronic reflux of stomach contents into the GOJ and lower oesophagus causes inflammation, leading to morphological changes in the oesophageal epithelium [1-3]. As mentioned previously, the epithelium offers good protection against various types of injury to the oesophagus. However, with repeated exposure to the acidic, proteolytic contents of the stomach (and in extreme cases, biliary secretions from the duodenum), ulceration and fibrosis of the oesophageal wall may occur, accompanied by clinical symptoms such as pain and difficulty in swallowing [1, 2]. Histologically, hyperaemia may be the only visible change in early stages. In more advanced disease, eosinophils and neutrophils infiltrate the oesophageal mucosa and basal zone hyperplasia can be seen [2, 3]. Untreated, it can lead to oesophageal stricture development and *columnar-lined oesophagus* [1-3].

1.2.2.2 Columnar-lined (Barrett's) oesophagus

Columnar-lined or Barrett's oesophagus describes the epithelial metaplasia from squamous to columnar mucosa occurring in the oesophagus following long-term exposure to gastric contents [1-6]. It is present in 10% of GORD sufferers and is most common in white males between 40 and 60 years of age. Barrett's oesophagus is considered a pre-malignant condition, as it significantly increases the risk of developing oesophageal adenocarcinoma [1, 2, 5, 6]. Dysplasia is also diagnosed in 0.2%-2% of patients suffering from it [1, 2]. The risk of cancer development is approximately 3% for patients with columnar metaplasia, 18% for low-grade dysplasia and 34% in high-grade dysplasia [4].

The precise definition of Barrett's oesophagus has been subject of much debate in the international clinical and academic community. One of the main issues is whether intestinal metaplasia is absolutely necessary for the definition of Barrett's oesophagus or the presence of columnar metaplasia is enough [5]. Authorities in the United States require the presence of goblet cells, whereas the British Society of Gastroenterology does not [6].

Columnar-lined or Barrett's oesophagus is defined as the presence of endoscopically visible columnar epithelium within the oesophagus (proximal shifting of the squamo-columnar junction), which on biopsy is confirmed to have intestinal metaplasia (intestinalised epithelium or epithelium containing goblet cells) [2-4, 6]. Endoscopic detection of abnormal mucosa above the GOJ is necessary to ensure that metaplastic goblet cells from the cardia are not misdiagnosed for Barrett's oesophagus [2, 7]. Microscopically, *goblet cells*, which have

distinct mucous vacuoles that stain pale blue on haematoxylin & eosin staining (H&E), define intestinal metaplasia [2, 7] and their presence is necessary in order to diagnose Barrett's oesophagus. Odze et al (2005) describes non-intestinalised and intestinalised columnar-lined oesophagus, the latter having goblet cells [3].

The length of columnar-lined mucosa in the oesophagus is important, with 3 cm being the cut-off point between long and short segment Barrett's oesophagus [3, 6]. Intestinal metaplasia in the GOJ region may be ultra-short segment Barrett's oesophagus (defined as columnar metaplasia with goblet cells that is less than 1 cm in length) or chronic inflammation of the cardia with intestinal metaplasia. In both cases, GORD and *Helicobacter Pylori* infection are the major underlying factors [3].

1.2.3 Stomach

The gastric environment is strongly acidic (pH close to 1) which helps with digestion, but can also be very damaging to the stomach wall and other parts of the GI tract (such as the oesophagus, as discussed previously). A number of mechanisms have evolved to protect the gastric mucosa such as mucin secretion by foveolar cells, bicarbonate secretion by epithelial cells, a rich vascular supply and rapid epithelial regeneration. As a result, there is a thick, neutral pH mucous layer covering the mucosa, which prevents direct contact of food and acid with epithelial cells and any damage caused is repaired quickly thanks to epithelial regeneration, which replaces the entire gastric mucosa every 2-6 days [2].

Any factors which disrupt the normal protective mechanisms, exposing the gastric mucosa to the acidic environment, lead to inflammatory processes, which can progress from acute to chronic and have severe complications, such as development of malignancy. Common causes are non-steroidal anti-inflammatory drug (NSAID) use, *Helicobacter Pylori* infection, ingestion of harsh chemicals, excessive alcohol consumption and radio- or chemotherapy [2].

1.2.3.1 Acute gastritis

Acute gastritis is a transient mucosal inflammatory process which can be difficult to detect histologically, as the surface epithelium is intact and there is only some moderate oedema in the lamina propria and some vascular congestion. Scattered neutrophils may be present, which signify *active* inflammation (which may occur in both acute or chronic inflammation).

In more severe cases, haemorrhage and mucosal erosions may occur, where defects are created in the mucosa, down to the lamina propria, followed by pronounced neutrophilic infiltration and pus formation. Concurrent erosion and haemorrhage is termed *acute erosive haemorrhagic gastritis*, which can progress to ulceration [2].

1.2.3.2 Chronic gastritis and Helicobacter Pylori infection

In chronic gastritis there is long-term insult to the gastric mucosa, leading to chronic inflammation and more severe complications, such as peptic ulcer disease, intestinal metaplasia and dysplasia, which can progress to cancer [2].

The most common cause of chronic gastritis is *H. Pylori* infection [2, 4]. These bacilli are present in biopsies of almost all patients with duodenal ulcers and 90% of patients with chronic gastritis and gastric ulcers. In most patients the disease is limited to the antrum, though sometimes the cardia is involved and in some the infection progresses to the gastric body and fundus. In these cases there is multifocal mucosal atrophy, intestinal metaplasia and increased risk of gastric adenocarcinoma development. The presence of *H. Pylori* disrupts the protective mechanisms mentioned previously and the bacteria may even invade the mucosa [2].

Microscopically, biopsy specimens exhibit *H. Pylori* in the superficial mucus overlying epithelial cells, extending into the gastric pits in extreme cases. The distribution can be irregular, with areas of heavy colonization interweaved with almost bacteria-free patches. Neutrophils infiltrate the lamina propria, with some crossing the basement membrane, spreading into the epithelium and forming gastric pit abscesses [2, 3]. Plasma cells, lymphocytes and macrophages are also seen within the lamina propria. These *intra-epithelial neutrophils* and *sub-epithelial plasma cells* are characteristic of *H. Pylori* gastritis, which, when chronic, can lead to thickened rugal folds and mucosal atrophy [2, 3]. Furthermore, the development of mucosa-associated lymphoid tissue (MALT) aggregates may be induced, which may then transform into lymphoma [2].

H. Pylori infection is a major risk factor for carcinogenesis in the stomach [2, 8, 9], with more than 80% of gastric cancer being attributed to *H. Pylori* infection [9]. Studies have shown that *H. Pylori* infection is associated with a two-fold increase in risk of developing gastric cancer [9]. Chronic disease causes mucosal atrophy, with loss of parietal cell mass and intestinal

metaplasia, characterised by the presence of goblet cells. Moreover, the accumulation of free radicals and proliferative stimuli result in genetic alterations that lead to dysplasia and eventually cancer [2, 9]. This happens over a period of decades and affects 3% of infected patients [2].

1.2.3.3 Gastric polyps and adenomas

Polyps are projections from the gastric mucosa, 75% of which are caused by epithelial or stromal cell hyperplasia or inflammation, the rest being the result of ectopia or neoplasia. *Inflammatory* or *hyperplastic polyps* are common in chronic gastritis and if larger than 1.5 cm, they should be resected and examined for dysplasia. The majority, however, are under 1 cm in diameter and generally only exhibit signs of inflammation. Microscopically, they have irregular, cystically dilated and elongated foveolar glands, with an oedematous lamina propria [2].

Fundic gland polyps frequently occur in patients with familial adenomatous polyposis (FAP) and patients on proton pump inhibitor (PPI) therapy. Macroscopically they are similar to hyperplastic or inflammatory polyps, but are concentrated in the fundus and body. Microscopically, they are composed of cystically dilated, irregular glands lined with flattened parietal and chief cells, without any signs of inflammation [2].

Approximately 10% [2] of gastric polyps are *adenomas*, which are a form of gastric dysplasia [2, 4]. Their incidence is higher in older populations and in patients with FAP. They almost always occur on a background of chronic gastritis with atrophy and intestinal metaplasia. Carcinoma may be present in up to 30% of adenomas and is more frequent in lesions above 2 cm in diameter. Gastric adenomas are usually solitary lesions in the antrum and less than 2 cm in diameter. The majority are composed of intestinal-type columnar epithelium and they all have dysplasia which may be low or high grade, the latter showing more disruption in cell structure and organisation, including glandular budding, gland-within-gland formation and cribriform structures [2].

1.2.4 Large intestine

1.2.4.1 Polyps and adenomas

As described previously, polyps can occur anywhere along the GI tract, but are most frequently found in the colon. Most of them begin as *sessile* growths (i.e. without a stalk), which enlarge to create a stalk and become *pedunculated*. They can generally be classified as non-neoplastic, which include inflammatory, hamartomatous and hyperplastic polyps, or neoplastic, the most common of which is adenoma. The latter have the potential to turn cancerous [2].

Inflammatory polyps may form as a result of chronic injury and healing and exhibit lamina propria fibromuscular hyperplasia, mixed inflammatory infiltrates, mucosal erosion and epithelial hyperplasia [2, 10]. They are not associated with an increased risk of cancer development [10]. *Hamartomatous polyps* occur sporadically and generally as a result of genetic or acquired syndromes. *Hyperplastic polyps* are commonly diagnosed in patients between 60-80 years of age. They are smooth, nodular mucosal protrusions, typically less than 5 mm in diameter and are most frequently found in groups in the descending and sigmoid colon and rectum. They are composed of mature goblet and absorptive cells, whose delayed shedding leads to a crowed, serrated surface [2, 10]. They do not have the potential to become malignant, but must be distinguished from *sessile serrated adenomas*, which are histologically similar and can become cancerous. The latter are more commonly found in the ascending colon and lack any features of dysplasia [2, 10].

The most common and clinically significant *neoplastic polyps* are *colonic adenomas*, which are benign polyps and precursors to most colorectal adenocarcinomas. They are intraepithelial neoplasms, characterised by dysplasia [2, 10]. The majority do not progress to cancer, but screening and surveillance endoscopy are important in monitoring patients and diagnosing any changes to adenocarcinoma early. Typical adenomas range from 3 mm to 10 cm in diameter and can be sessile or pedunculated, with a velvety surface [2, 10]. Dysplasia is present, particularly at the surface of the adenoma and sometimes at the stalk. Epithelial cells fail to mature as they move up the surface layers and large nucleoli, eosinophilic cytoplasm and reduced goblet cells can be seen. Adenomas are classified as *tubular*, *tubulovillous* or *villous*, depending on whether they predominantly feature tubular glands or are covered by villi [2]. When dysplastic cells breach the basement membrane, the lesion is

defined as a *carcinoma in situ*, just as in the oesophagus and stomach. In this case however, the metastatic potential is very low, as there are no lymphatic channels in the colonic mucosa, which makes polypectomy a curative procedure [2, 11]. However, spread into the submucosa necessitates more aggressive treatment [2] and carries the diagnosis of invasive carcinoma [11]. *Familial adenomatous polyposis* causes the development of hundreds to thousands of adenomatous polyps, which will lead to adenocarcinoma if left untreated [2, 10-13].

Hereditary non-polyposis colorectal cancer (HNPCC) is also a cause of colorectal cancer [12, 13]. It is a rare autosomal dominant genetic syndrome which accounts for 3-5% of colorectal cancer. [13]

1.3 Epidemiology and pathophysiology of gastrointestinal cancer

1.3.1 Epidemiology

The worldwide annual incidence of gastric and colorectal cancer is 933,000 and 1,080,000 respectively [14, 15] and most patients present with advanced disease [16, 17]. An increase in the incidence of early gastric cancer has been reported, which is partly due to the increased availability of endoscopy, though the overall incidence in developed countries is falling [16]. The combined annual incidence of gastric and oesophageal cancer makes them the second most commonly diagnosed cancers worldwide [18]. In the UK, gastrointestinal (GI) cancers are amongst the most common malignancies, with new cases diagnosed in 2008 reaching 39,991 for colorectal, 8,173 for oesophageal and 7,610 for stomach cancer. This makes colorectal cancer the third most common in the UK, after breast and lung [13, 19] and about half of patients suffering from it die [12, 13].

1.3.1.1 Oesophageal cancer

The majority of oesophageal cancers are either adenocarcinomas, where the cells grow in a glandular pattern and produce mucin, or squamous cell carcinomas with intercellular bridges and keratin production [2, 4, 20]. Squamous cell carcinoma is more common worldwide, but the incidence of adenocarcinoma is rising in the West [2].

Oesophageal adenocarcinoma occurs most frequently in Caucasians and is 7 times more common in men than women [2, 21]. Its incidence varies 60-fold worldwide, with the highest being in Western countries, such as the United States and United Kingdom. The incidence has increased markedly in those countries since 1970 [2, 20, 21], from below 5% to 50% of all oesophageal cancers [2].

Oesophageal squamous cell carcinoma in the United States occurs mainly in men (4 time more common than women) over 45 years old and is nearly 6 times more common in African-Americans than Caucasians, most likely because of differences in alcohol and tobacco use, which are major risk factors. Other causes include caustic oesophageal injury, achalasia, frequent consumption of very hot beverages and previous radiotherapy (typically 10 years after exposure). The incidence of squamous cell carcinoma varies 180-fold

worldwide, with the highest incidences in Iran, central China, Hong Kong, Brazil and South Africa [2].

Oesophageal adenocarcinomas are generally late-presenting, resulting in an overall 5-year survival of less than 25% in most cases [20, 21]. In contrast, when detected early (carcinoma limited to mucosa or submucosa), the 5-year survival is closer to 80% [2]. The overall 5-year survival for squamous cell carcinoma of the oesophagus remains around 9%, even though increased endoscopic screening has led to earlier detection. In patients with superficial tumours, the 5-year survival rate rises to 75%. Lymph node involvement, which is common, is associated with a poor prognosis [2].

1.3.1.2   Gastric cancer

Gastric cancer incidence is very dependent on geography [2, 9]. It is 20 times more common in Japan, Chile, Costa Rica and Eastern Europe than in North America, northern Europe, Africa or Southeast Asia. In the former regions, mass endoscopic screening programs have raised the number of early gastric cancer cases to 35% of new diagnoses, whereas in low-incidence regions where such screening programs are not implemented, less than 20% of new diagnoses are early stage tumours [2]. The overall incidence of gastric cancer has dropped in Western countries [2, 9, 20], with gastric cancer accounting for less than 2.5% of cancer deaths in the United States, where it used to be the most common cause of cancer death in 1930. Many explanations have been put forward, including improvements in diet, such as reduced consumption of N-nitroso compounds, salt and smoked products or improvements in the management of *H. Pylori* infection and chronic gastritis [2, 9, 20]. Cancer of the cardia is on the rise, however [2, 20], probably because of the increased incidence of Barrett's oesophagus and distal oesophageal adenocarcinoma, which is morphologically and clinically similar to gastric cardia adenocarcinoma. Furthermore, the decrease in incidence has only been observed for intestinal-type adenocarcinoma [2].

The mean age of presentation of intestinal-type adenocarcinoma is 55 years, with a male-to-female ratio of 2:1. The majority of cases arise in high-risk areas and there is a natural progression from flat dysplasia and adenomas to cancer. On the contrary, diffusely infiltrating gastric cancer has a relatively uniform distribution around the world, a M:F ratio approaching unity and no known precursor lesions [2].

Gastric cancer is one of the leading causes of death worldwide [2, 9, 20, 22]. A major factor is that early symptoms are similar to chronic gastritis, which means that, in countries without screening programs, gastric cancers are generally diagnosed at later stages, which carry a worse prognosis [2, 9]. In the UK, gastric cancer is the fourth commonest cause of death from malignancy [22]. The worldwide 5-year survival rate is approximately 20% [9, 20], except in Japan where it reaches 60% [20], due to their rigorous screening programme which leads to earlier detection [23].

1.3.1.3   Colon and rectal cancer

Colon cancer is the most common cancer of the GI tract and is a major cause of morbidity and mortality worldwide [2, 10, 15]. There are 130,000 new cases and 55,000 deaths from colon cancer in the United States every year. This accounts for nearly 15% of all cancer related deaths, with only lung cancer ahead of it [2]. Fewer than 20% of cases occur before 50, with the highest incidence being between 60 and 70 years of age and males having a slightly higher rate than females. It is more prevalent in the United States, Canada, Australia and other developed countries [2, 12], with an up to 30 times smaller incidence in India, South America and Africa. In Japan and the United Kingdom rates are now at intermediate levels [2]. Colorectal cancer is thought to be caused by a combination of genetic and environmental factors, including hereditary syndromes, diet and lifestyle [12, 15].

Colorectal cancer is more common in males, with more than 80% of cases arising in those above 60 years of age [13] and a peak age of diagnosis at 70 years [12]. The lifetime risk of developing colorectal cancer is 5-10% by 80 years of age [12].

The Dukes system of classification is used to stage colorectal cancer. The corrected 5-year survival for patients with Dukes A or B is high (90%, 60-80% respectively), but only 35% for patients with Dukes C [13].

1.3.2 Pathophysiology

1.3.2.1   Oesophageal cancer

Oesophageal adenocarcinoma typically occurs in patients with Barrett's oesophagus and chronic GORD, with the presence of dysplasia further increasing the risk of cancer

development. It is usually found in the lower third of the oesophagus and may invade the gastric cardia [2, 4, 20]. Tumours may start as flat patches which can grow to large masses of 5 cm or more in diameter, or infiltrate diffusely or ulcerate and invade deeply. They most commonly form glands and produce mucin, often exhibiting intestinal-type morphology [2, 4]. Less frequently, diffusely infiltrative signet-ring cells can be seen. In many cases, Barrett's oesophagus can be seen around the tumour site [2].

Squamous cell carcinoma, on the other hand, presents in the middle third of the oesophagus in 50% of cases. It begins as a small, grey-white, plaque-like thickening called *squamous dysplasia*. Over a period of months or years, the tumour grows into a polypoid or exophytic mass which may obstruct the oesophagus, causing dysphagia or spread along the oesophageal wall in an ulcerative or diffusely infiltrative manner, causing thickening, rigidity and luminal narrowing. In extreme cases, it may invade surrounding structures such as the mediastinum, lungs, heart or great vessels such as the aorta. Most squamous cell carcinomas are moderately to well differentiated and spread to different lymph node groups according to location [2]. Well differentiated tumours exhibit cells with abundant eosinophilic cytoplasm, low mitotic count and keratinization, whereas poorly differentiated ones show scant cytoplasm, high mitotic activity and marked nuclear polymorphism [4]. Those in the upper third of the oesophagus spread to cervical lymph nodes, those in the middle third invade mediastinal, paratracheal and tracheobronchial lymph nodes, whereas those in the lower (abdominal) third metastasise to gastric and coeliac nodes. Additionally, intramural tumour nodules may be found away from the primary lesion, as cancer cells use the submucosal lymphatic network to spread along the oesophageal wall [2].

1.3.2.2 Gastric cancer

Over 90% of gastric cancers are adenocarcinomas and they are classified according to their location, macroscopic appearance and histology [2, 4, 9]. Most involve the gastric antrum and particularly the lesser curvature. There are two distinct histological types of gastric cancer according to the World Health Organization (WHO) and Lauren's classification system, intestinal and diffuse [2, 4, 9]. Intestinal-type adenocarcinomas tend to form bulky tumours composed of glandular structures and, although they can potentially penetrate the gastric wall, they generally tend to grow into exophytic masses or ulcerated tumours within the lumen. Mucin vacuoles are often present in the tumour cells and mucin may be present

in gland lumens. On the other hand, diffuse infiltrative tumours are composed of discohesive cells, with no visible glands and tend to infiltrate the stomach wall instead of forming discrete masses [2, 4, 9]. In some cases, large mucin vacuoles form, which push the nucleus to the periphery, giving the appearance of *signet-ring cells*. These are scattered in the stomach wall individually or in small groups, making them hard to distinguish at low magnification from inflammatory cells, such as macrophages. Both types of cancer can produce large mucin lakes and they often evoke a desmoplastic reaction that makes the gastric wall stiffer. In extreme cases, where the tumour has spread widely, the whole stomach may obtain a leather bottle appearance, which is called *linitis plastica* [2].

1.3.2.3 Colon and rectal cancer

The adenoma-carcinoma sequence accounts for as much as 80% of sporadic colon tumours and involves the accumulation of mutations that lead to the loss of normal cell regulatory mechanisms and development of cancer [2, 10, 12, 13]. The vast majority of colorectal cancers are adenocarcinomas [10-12], which affect the entire length of the colon equally, with similar histological properties [2]. However, tumours in the proximal colon tend to grow along one wall as polypoid, exophytic masses, which rarely cause obstruction, whereas in the distal colon lesions tend to be annular and their growth causes constrictions. Most tumours consist of tall columnar cells, which resemble the dysplastic characteristics of the adenomas they arose from, with fewer glands present in poorly differentiated tumours. They elicit a strong desmoplastic response, which causes the bowel wall to stiffen in affected areas and some may produce a lot of mucin that accumulates within it [2, 11]. Ten to fifteen per cent of adenocarcinomas are mucinous (defined as mucinous component > 50% [10, 11]) and they have an increased tendency to develop peritoneal metastases, invade adjacent organs and spread extensively to lymph nodes [10]. They are, thus, associated with a poorer prognosis, though TNM staging is still the main prognostic indicator, rather than tumour type [2, 10, 13]. Signet-ring cell carcinoma (composed of at least 50% signet ring cells [10, 11]) comprises less than 1% of colorectal cancers and occurs more often in individuals that are less than 50 years old and in ulcerative colitis sufferers [10].

1.4 Diagnosis of gastrointestinal cancer

1.4.1 Endoscopy

1.4.1.1 Oesophagogastroduodenoscopy

Endoscopy is the best method for diagnosis and surveillance of Barrett's oesophagus [5, 6] and is necessary for making a diagnosis of oesophageal cancer, as it allows direct visualisation of lesions and sampling of tissue for histological examination [21]. Single biopsies of suspicious lesions are 93% accurate, rising to 100% with the addition of several biopsies and brushings [21].

It is also the most specific and sensitive method for obtaining a definitive diagnosis of gastric cancer (other than laparotomy) [21]. Endoscopy with seven biopsies of suspicious gastric ulcers (at both margin and base) has a sensitivity of 98% for diagnosing gastric cancer [21].

1.4.1.2 Colonoscopy

Colonoscopy with histological examination of biopsies is the gold-standard investigation for colorectal polyps and cancer [11-13, 15, 24]. It can be used for therapeutic purposes too, such as local excision of polyps [12, 24].

1.4.2 Histopathology

Histological reporting confirms the diagnosis and provides information on tumour morphology and staging. This is important, as it guides further treatment (such as surgery, chemotherapy and radiotherapy) and provides information on the patient's prognosis, feedback on the quality of surgical resection and monitoring of the effects of neoadjuvant chemotherapy [10, 11, 25-27].

A number of stages are involved in the preparation of specimens for histological examination. Clinical information about the patient and specimen (such as operation type, tumour location, previous chemotherapy) is essential, as, in some specimens, the tumour site may not be macroscopically obvious, especially post-chemotherapy [25-27].

1.4.2.1 Specimen preparation

Mucosal biopsies should be fixed in formalin for at least 6 hours and no more than 3 days. Biopsies may be received free floating in formalin, or orientated by the endoscopist on filter paper, so that the mucosa is on top. Biopsies from multiple sites should be properly identified, ideally by being placed in multiple, clearly labelled pots by the endoscopist [7].

Polyps are treated in a similar way, with larger ones being sliced after fixation. The base or stalk should be inked to identify the resection margin. Slicing should be carried out in the axial plane and at least one block should contain both body and stalk [7, 10, 27].

Large specimens should be received unopened and ideally unfixed, allowing good orientation by the histopathologist before placing in formalin. The oesophagus should be opened along the anterior border, the stomach along the greater curve and the intestine along the antemesenteric border, unless there is a focal lesion in the way, in which case the incision should go around it. Luminal contents should be washed out with lukewarm water and without scrubbing, to avoid damaging the mucosa. Serial slices may be taken if the wall is thickened and with circumferential tumours, especially in the intestine. Margins should be inked when neoplasia may be present, including the circumferential margin [7, 11, 25-27]. The specimen should be pinned to a corkboard or stabilised in another suitable way and placed in sufficient formalin (10 times own volume) for at least 48 hours for large specimens. Paper wicks can be used to ensure formalin penetration in areas such as between the specimen and corkboard, or the lumen of unopened specimens [7, 25-27]. An alternative solution is to unpin specimens after 24 hours and let them float free in formalin [27]. Photography is not routine, but may be useful for meetings and post-fixation identification of non-obvious lesions [7, 25-27].

Oesophagectomy specimens contract as soon as they are excised and can lose a third of their natural length if fixed unpinned. It is recommended, therefore, that they are delivered fresh to the laboratory, as soon as possible after excision and pinned on corkboard before placement in formalin. The gastric part should be opened along the resection margin and the oesophagus opened along the anterior border or left unopened. If opened, the outer surface is painted, so that invasion of the serosa can be detected microscopically. The part containing the tumour is generally left unopened, in order to preserve the circumferential margin [25].

1.4.2.2 Tissue sampling and staining

Biopsies are embedded in their entirety and larger ones may be sliced first. The number and size of biopsies or biopsy fragments are recorded. Polyps are blocked whole or, if large, according to the way they have been sliced [7, 10, 27].

In large specimens, blocks are taken from the proximal and distal margins, the diseased area and lymph nodes. For tumours, larger blocks are commonly taken (*megablocks*), which can run longitudinally, transversely or both. Multiple blocks are normally taken from tumour sites, to allow examination of circumferential margins and the relationship to adjacent tissues (e.g. neighbouring Barrett's oesophagus) [25-27]. Random blocks from background mucosa are also taken and a careful search of neighbouring tissues (such as lesser and greater curve fat in gastrectomies) is carried out to block as many lymph nodes as possible [26, 27].

The specimens are stained with haematoxylin and eosin as standard [6, 7, 10], with other stains added for specific conditions. Alcian Blue stain helps identify goblet cells to confirm intestinal metaplasia and Giemsa, Cresyl Fast violet stains are used for the diagnosis of *H. Pylori* infection [3, 6, 7] . Sections may be taken singly or at different levels (e.g. 75 μm apart) [7].

1.4.2.3 Macroscopic reporting

When reporting on biopsies, the adequacy of the sample is commented upon and any useful clinical information or lack thereof also noted [7]. Maximal length and diameter of polyps should be recorded as well as a description of the external surface [10, 25, 27].

Amongst other things, the overall dimensions of the specimen and the position, distance from resection margins and macroscopic appearance of the tumour (e.g. ulcerated, polypoid) are recorded [4, 11, 25-27]. In colorectal tumours, perforation through the bowel wall is also recorded [11, 27].

1.4.2.4 Microscopic reporting

The main role of histological examination of a biopsy from a potentially malignant lesion is to confirm the diagnosis of adenocarcinoma or other diseases such as dysplasia. Comments can be made on the differentiation and classification of the tumour (e.g. intestinal-type or diffuse), but conclusions cannot be made about the bulk of the lesion without it being excised completely [26].

In polyps, tubulovillous classification applied to adenomas and degree of dysplasia recorded, as well as completeness of excision [7, 10]. Presence, size and distance from margins of any carcinoma is reported [7, 10, 27].

The histological type of tumour (adenocarcinoma, squamous cell carcinoma etc.), grade of differentiation, depth of invasion (including serosal involvement) and involvement of proximal, distal and circumferential resection margins are reported on. Vascular, lymphatic and perineural invasion are assessed and the number of involved and uninvolved lymph nodes in the specimen is determined. The presence of mucosal atrophy, metaplasia, dysplasia, *H. Pylori* (in the stomach) and the locations and extent of the areas involved are noted [10, 25, 26].

Fibrosis, mucin lakes and collections of keratin with apoptotic or no tumour cells signify the use of chemotherapy. The extent of tumour regression is described and the Mandard scale is used to standardise observations [25-27].

1.4.2.5 Benign vs malignant tumours

In the majority of cases a benign tumour can be confidently distinguished from a malignant one on the basis of morphology. Occasionally, the appearance of a neoplasm cannot confirm its biological behaviour, in which case techniques such as immunohistochemistry and molecular profiling are employed. In general, however, benign and malignant tumours are histologically diagnosed on the basis of differentiation, local invasion and metastasis [2].

A well differentiated tumour is one in which the neoplastic parenchymal cells closely resemble normal cells from the same tissue, both morphologically and functionally. In general, benign tumours are well differentiated. On the other hand, malignant growths

exhibit a wide range of differentiation, from completely undifferentiated (anaplastic) to, again, so well differentiated, that they are hard to distinguish from normal tissues, although this is unusual with gastro-intestinal tumours [2]. Poor differentiation carries a worse prognosis and increases the risk of lymphatic spread in early stage tumours [11]. Lack of differentiation is considered a hallmark of malignancy and is often associated with other morphological changes, such as *pleomorphism*, abnormal nuclear morphology, abnormal mitoses and loss of polarity. Pleomorphism refers to the variation in size and shape of cells and nuclei exhibited by tumour cells. Abnormalities in nuclear morphology include *hyperchromatism* (dark staining nuclei), clumping of chromatin, large nucleoli and a decrease in the nuclear-to-cytoplasm ratio (1:1 instead of the normal 1:4 or 1:6). Mitoses are not only much more numerous than normal tissues, but also exhibit abnormalities such as multipolar spindles and cells lose their normal polarity, growing in a disorganised way [2].

Benign lesions nearly always grow as cohesive, well organised units, which remain local and do not spread to neighbouring or distant sites. In fact, in well differentiated benign tumours the only characteristic that can identify them as neoplastic is their growth into a discrete mass, which is separate from the surrounding tissues. In contrast, malignant tumours invade and destroy surrounding tissues, especially when their rate of expansion is rapid. Slower-growing tumours may form some kind of margin, but even then small numbers of cells will infiltrate beyond that and into adjacent structures. Cancers do not recognise anatomical boundaries and their invasiveness is one of their most reliable diagnostic features [2].

Finally, the pathognomonic feature of malignancy is *metastasis*, which is the distant spread of tumour cells. This can be achieved in a number of ways, such as through the circulatory system (haematogenous spread), lymphatics and body cavities [2, 10, 12]. Haematogenous spread occurs more frequently through veins, as their thinner walls make them more penetrable than arteries and this commonly leads to liver and lung metastases, as those organs receive a large proportion of venous blood. The lymphatic system is the most common pathway for initial spread and the pattern of lymph node involvement reflects the natural lymphatic drainage of the area. Sometimes, local lymph nodes are bypassed (*skip metastasis*), but in most cases the pattern of spread is predictable and this aids in minimally invasive monitoring of tumour dissemination, by sampling *sentinel lymph nodes*. These are the first nodes in the region that receive lymph flow from the tumour and therefore have a high probability of containing tumour cells if the tumour has metastasised. Biopsies or

frozen sections from sentinel nodes are very useful in guiding therapy for breast and colon cancer. Finally, the seeding of body cavities occurs when carcinomas invade an open space, most commonly the peritoneum. Peritoneal spread is commonly diagnosed through staging laparoscopy, where biopsies from suspicious peritoneal lesions can be taken under direct view and then histologically examined [2].

A routine part of histological examination, therefore, is the assessment of local vascular and lymphatic invasion in the main specimen (e.g. a gastric biopsy or gastrectomy specimen), as well as spread to distant sites (e.g. lymph nodes or parts of other organs excised during the operation). More often than not, however, the diagnosis of distant metastasis is a clinical one (e.g. through imaging such as computed tomography) and can only be histologically confirmed if a specimen is available from that site.

1.4.2.6 Grading, staging and classification of tumours

In order to provide an accurate prognosis and to monitor effects of treatments, such as chemotherapy, it is important to determine the extent of tumour invasion and its aggressiveness. The *grading* of a cancer is based on the differentiation, number of mitoses and architectural features of the tumour. It varies for different types of cancer and is not always clinically useful, as the histological features do not always correlate well to the neoplasm's behaviour in vivo. *Staging* refers to the spread of the tumour and is based on the size of the primary lesion, lymph node spread and distant metastases. The major system in use is the TNM classification [2, 10]. For colorectal tumours, Dukes classification is also used [27].

Lauren's classification and the *WHO classification* are the most widely used systems for histologically classifying adenocarcinoma of the oesophagus, GOJ and stomach [4]. Lauren's classification separates adenocarcinomas into intestinal and diffuse types. The former exhibit a cohesive growth pattern and form glands, whereas the latter show little cohesion and little, if any, gland formation. A significant number of tumours exhibit both patterns and are termed "mixed" cancers [4]. The WHO classification splits adenocarcinoma into four subtypes: tubular, papillary, mucinous and signet ring. Tubular adenocarcinomas consist of tubular and acinar structures and papillary cancers have finger-like processes made of central fibrovascular stalks covered by neoplastic epithelium. Frequently, a tumour starts as

papillary and converts to tubular. When more than 50% of a tumour consists of extracellular mucin pools it is termed mucinous, whereas if those pools are intracellular, pushing the nuclei to one side and giving the appearance of a ring, the tumour is described as consisting of signet ring cells [4]. The WHO classification gives a more detailed description of the histology, whereas the Lauren classification is more relevant to the pathophysiology of the tumours. Accordingly, intestinal types are more common in older men, frequently arise from atrophic gastritis and have a better prognosis, whereas diffuse cancers are more common in women and do not seem to have a precursor lesion [4].

1.4.2.7   Histological effects of chemotherapy

Neoadjuvant chemotherapy can have a significant effect on tumour histology, sometimes causing such regression, that no tumour is macroscopically visible, being instead replaced by an area of scarring. In such cases, an extensive microscopic search must be carried out to find any residual tumour in areas of fibrosis. A reduction in the volume of malignant cells and an increase in the amount of stroma is seen. Mucinous metaplasia may be evident, as well as fibrosis and calcification of the stroma, with atypical fibroblasts present. The degree of fibrosis correlates with outcomes and can be quantified with the extent of tumour regression [11].

1.4.3 Limitations and missed diagnoses

As described above, endoscopy followed by histological examination of biopsies is the gold standard diagnostic process for gastrointestinal cancer. However, this method is not without limitations. A number of studies have shown that early and even sometimes advanced malignancies can be missed in conventional endoscopy [28].

The rate of missed cancers in the West has been reported to be between 6.7% - 14% on initial endoscopy [18]. Studies carried out in Japan report higher rates of 23.5-25.8% [18]. The discrepancies in missed diagnoses rates between East and West are, to a large extent, due to the different classification systems used for defining a lesion as malignant or not and differences in the morphology and progression of cancer in the two populations [18].

Amin et al (2002) investigated the delay between first endoscopy and histological diagnosis of cancer, excluding endoscopies which had been carried out more than 2 years before the

cancer diagnosis. They looked at 137 patients in Edinburgh with biopsy-proven gastric adenocarcinoma over a 5-year period. They found that in 8% of patients a repeat endoscopy was necessary within 4 weeks and in 14% of patients the diagnosis of cancer was missed and symptoms were attributed to benign pathology. The diagnosis for the latter group was eventually established after a median delay of 13 weeks (range 3-102). This false negative rate for first endoscopy was, therefore, 22% and could in fact be higher in reality, as only patients referred to surgery were included [22].

In a study of cancer patients in Finland, 4.6% had undergone upper gastrointestinal endoscopy where the cancer was undiagnosed (defined as endoscopy less than 3.5 years before the cancer was identified in a subsequent endoscopy or intervention). This led to a mean delay in diagnosis of 16.3 months [28]. The missed diagnoses were attributed to both lack of visible lesions on endoscopy (resulting in no biopsies being taken) and non-pathological biopsy specimens from sites which subsequently produced cancer biopsies [28].

A 5-year UK study of 181 patients with cancer found that 16% of cancers were not diagnosed on previous investigations (2 barium meals, 11 gastroscopies). This was attributed to experienced endoscopists failing to sample benign-looking lesions or inexperienced endoscopists missing high lesions [16].

A Japanese study looked at 8601 patients undergoing endoscopy, of which 3672 patients had a repeat examination within 3 years. From the latter, 0.9% of those diagnosed as normal, 2.2% of patients diagnosed with a non-malignant gastric ulcer and 17.2% of those originally diagnosed with gastric adenoma in the original endoscopy, were subsequently found to have gastric cancer on repeat examination [29].

Another area of concern is that of missed synchronous gastric neoplasms, where one lesion has been diagnosed but others are missed. This is particularly important when local resection techniques such as endoscopic mucosal resection are used to remove the malignancy, as any missed lesions will be left behind [30]. Lee et al. (2010) found that in 14 (27.5%) of 51 patients with multifocal synchronous gastric cancer, accessory lesions were missed by endoscopy and that small lesion size was significantly correlated with missed diagnosis.

There are many reasons for diagnostic errors during endoscopy. Gastric lesions, particularly ulcers, can look benign macroscopically, when in fact malignant. Use of proton pump inhibitors can exacerbate this effect and increase false-negative diagnoses [16, 22]. The experience of the endoscopist is crucial in such scenarios, where failure to biopsy can lead to missed cancer diagnoses [22]. In fact, multiple biopsies are recommended from both the rim and base of gastric ulcers, with 10 or more being a generally acceptable number [22].

The sensitivity and specificity of diagnosing a malignant lesion with endoscopy alone is 81% and 91.7% respectively, rising to 97% and 99.4% when combined with histology results [17]. This demonstrates the inability of endoscopy alone to adequately detect areas of disease, which is why non-targeted biopsies are essential. To maximise the diagnostic efficacy of endoscopy, the endoscopist needs to be suspicious enough to take biopsies from both potentially malignant lesions (e.g. ulcers) and seemingly benign areas, but also skilful enough to take the right number of biopsies from appropriate locations [17]. A lower threshold for biopsy taking and greater suspicion of seemingly benign lesions, along with a stricter protocol for repeat endoscopy would reduce the number of false-negatives and missed diagnoses [5, 22].

The development of light-based systems in medicine and surgery has generated interest in modelling light propagation through tissues and defining their optical properties, as this can lead to improvements in diagnostic and therapeutic applications [31-36]. Measurements of optical properties have been carried out on various tissues including skin, aorta, liver, muscle, stomach, oesophagus, colon, uterus, cervix and brain [35, 37-39].

The primary motivation in many cases has been to improve delivery of energy-based treatments such as photodynamic therapy (PDT). Knowledge of optical properties improves dosimetry and energy delivery, as the distribution of light in the tissue can be modelled [32-34, 40-44]. The therapeutic effect depends primarily on the amount of light absorbed and the penetration depth of light is dependent on optical tissue parameters such as μ_a, μ_s and g (defined in the next section) [42].

Others have investigated the potential of optical properties to distinguish between healthy and diseased tissue (most often cancer). The concept of an "optical biopsy", whereby diseased and malignant regions of surfaces such as skin and mucosa can be identified

quickly and accurately in vivo, is very attractive and a lot of research is focused on realising it through a number of methods, including simple imaging, optical property measurements or both [34, 35, 38, 39, 45-57]. Such a technology could significantly improve the efficiency of biopsies (by guiding mucosal sampling during endoscopy, thus reducing false negative rates) and help determine tumour margins in surgery [39, 47].

1.5 Optical Properties

1.5.1 Definitions

Optical properties are scientific values that describe how light interacts with different materials [37, 58]. They are obtained by using theories that describe light transport in tissue, to interpret measurements of observable quantities (e.g. reflection). The majority of optical properties reported for different tissues are based on a light propagation theory known as *radiative transport theory*, according to which, the radiance of light travelling into a certain direction is decreased by absorption and scattering, but increased by other light scattered in the same direction [37, 58]. Some of the main optical properties and other useful parameters are described here:

The *absorption and scattering coefficients* (μ_a and μ_s respectively) describe the probability per unit length of a photon's path, that the photon will be absorbed or conservatively scattered. Their units are (m^{-1}) [31, 40, 58, 59]. The *total attenuation coefficient* (μ_t) is equal to the sum of the absorption and scattering coefficients ($\mu_t = \mu_a + \mu_s$) [31, 37, 58]. The inverse, the *mean free path* (MFP), describes the mean distance a photon travels before a scattering or absorption event occurs [31, 40].

The *phase function p(s,s')* describes the probability of light coming from one direction (s') scattering into another (s) [37, 58, 59]. It is often characterised by a single parameter, called the average cosine of the phase function or the *anisotropy coefficient* (*g*) [37, 58, 59]. This factor is a measure of the asymmetry of scattering. When incoming light is scattered equally in all directions, this is called isotropic scattering and *g* has a value of 0. In forward and backward (i.e. anisotropic) scattering, *g* approaches 1 and –1 respectively [37, 44, 58, 59]. A number of equations to describe the phase function in tissues have been developed based on the *Henyey-Greenstein function*, such as the modified Henyey-Greenstein, Eddington and delta-Eddington, each with its own limitations [37, 58, 59].

Albedo (*a*) is the ratio of scattering to the sum of scattering and absorption. It ranges between 0 (no scattering) and 1 (no absorption). *Optical depth* or *optical thickness* (τ) is the product of tissue thickness and the sum of μ_a and μ_s. A value of 1 gives a probability of 37% (e^{-1}) that

light will travel at least that distance without an absorption or scattering event occurring [59].

The *effective attenuation coefficient* (μ_{eff}) is a complex interaction of μ_a, μ_s and g and is used in situations where measurement techniques cannot produce exact values of the three parameters. It is one of the most common optical properties being published [31], due to the popularity of certain methods (see later sections) used for tissue measurements, despite its limited ability to describe scattering and absorption compared to μ_a and μ_s [31].

1.5.2 Measurement techniques

A variety of techniques exist for measuring optical properties, which can be split into direct and indirect [37, 58, 59].

Direct techniques employ thin enough tissue samples to keep multiple scattering at negligible levels and utilise simple laws and equations, with no need for complicated modelling of light propagation. The sample thickness needs to be less than the MFP, so, for example, at 630nm tissue samples need to be less than 100µm thick [40]. Examples include measurements of unscattered transmission (which produce the total attenuation coefficient μ_t), effective attenuation measurements, e.g. with interstitial optical fibres (which give the effective attenuation coefficient μ_{eff}) and goniophotometric measurements (from which the single scattering phase function can be calculated) [37]. Goniophotometric measurements can in principle produce μ_s by integrating the scattering curve across all angles, but in practice this is difficult to achieve, as the signal strength at wider angles is weak compared to the incident beam [40].

Indirect techniques employ theoretical models of light scattering, which have inherent limitations. They can be classified as non-iterative, in which optical properties are directly related to experimentally measured parameters, and iterative, where the values of optical properties are estimated and optimised, until the calculated reflection and transmission match the measured values. The latter are more cumbersome but also more sophisticated [37]. As these techniques do not rely on optically thin samples, they are easier to implement, considering the practical difficulties of obtaining, preserving and mounting thin tissue samples. Also, some indirect methods may be applied in vivo, such as those measuring

diffuse reflectance. Diffuse transmittance is another measurement that is used in indirect methods. Both can be described in terms of their angular dependence and spatial variation. Integrating spheres, collimated detectors and optical fibres are commonly used to carry out these measurements. Optical fibres may be placed interstitially, to measure light distribution within tissues, in vitro and in vivo [40].

Iterative indirect methods include diffusion theory, adding-doubling models and Monte-Carlo simulations [58]. They usually include solutions for index-mismatched boundaries and/or multiple layers. Typically, they provide the absorption coefficient and reduced scattering coefficient based on total reflection and transmission. If a measurement of the unscattered transmission or phase function is available (which normally requires an additional technique to the one measuring reflection/transmission, e.g. goniophotometry), they can produce the scattering coefficient and anisotropy function as well [37].

As previously mentioned, the majority of optical properties reported for different tissues are based on a light propagation theory known as radiative transport theory [37]. This theory describes light propagation through a slab. In this theory, it is assumed that the medium is homogeneous, meaning that any variations in absorption or scattering coefficients must occur over distances much larger than the depth of the of the slab and that scattering particles are placed sufficiently far apart so as not to affect each other's scattering profiles (neither of which applies to dense, inhomogeneous media such as tissue). A further assumption is that the phase function can singularly describe the scattering characteristics of the whole medium (like an average of the scattering parameters of all individual particles in the medium) and that there are no light sources in the medium itself [59]. Furthermore, there are multiple sources of experimental errors that render values for optical properties inaccurate to varying degrees. These include: the physiological condition of the sample (hydration levels, temperature, surface smoothness, freshness), irradiation geometry, refractive index errors, orientation and aperture of sensing optical fibres, angular resolution of goniophotometric measurements and the limitations of theoretical models used to extract optical properties from experimental measurements. The dependence of optical properties on wavelength also makes it difficult to consolidate them, as many studies are carried out with different wavelengths [37, 40]. More and more sophisticated models and experimental techniques are being developed to counteract all the above effects.

1.5.3 Goniophotometry

A goniophotometer is a system designed to measure light intensity as a function of angle, giving the scattering phase function of samples such as biological tissue [31, 36, 40, 58-60]. It is a simple concept, which can, however, lead to increasingly sophisticated designs when attempting to maximise sensitivity, minimise background noise (whether electrical or optical), or take into account multiple factors such as refraction at mismatched boundaries, polarisation effects, source and detector profiles etc.

An optically thin sample of the material tested is illuminated by a thin beam of light and the scattered intensity is measured at different angles around the sample [31, 36, 60]. A typical configuration consists of a light source aimed at a sample (set at 90° to the axis of the incident beam), which is mounted at the centre of a rotating stage with an arm on which a detector is attached [31, 40, 58-61]. A photomultiplier is attached to the detector and readings of light intensity are taken at different angles around the sample, in a horizontal plane. The whole system may be manually operated or automated and computer-controlled [31, 40, 58-61].

In an ideal situation, light would enter and exit the sample without any other materials in the way. In practice however, especially with thin, floppy tissue samples, some kind of mounting system is used, such as a cuvette or set of glass slides to hold samples [31, 58]. This introduces internal reflections and refractive effects, because of the refractive index mismatch between the different materials [60]. In addition, light scattered at angles higher than the sample's critical angle undergoes total internal reflection and does not reach the detector [60].

To counteract these effects, some goniophotometers incorporate a tank filled with filtered water, in which the tissue sample is submerged. The tank is cylindrical, so that light exiting the tank is at normal angles to the tank wall and does not get refracted before reaching the detector [31, 36, 40, 61]. In some designs, both the detector and sample are in the water, thus eliminating the need for a cylindrical shape [59]. The principle behind using a water tank is that the refractive index of water (n=1.33) is much closer to glass (n=1.54) and tissue (n= approx. 1.45) than air (n=1), thus minimizing refraction effects both at the sample-glass and glass-water interfaces [60].

When optically thin samples are used (i.e. samples with a low concentration of scatterers), multiple scattering is minimised and single scattering approximations can be used to produce optical properties such as the scattering coefficient or phase function [36, 37, 40, 58, 59]. The physical thickness of optically thin samples depends on the concentration of scattering particles and the wavelength at which the goniophotometer operates. For example, it has been shown that, at 630nm human tissue samples need to be less than 100μm thick [31].

Another option is to mount the sample between a glass slide and glass hemisphere, with index matching material such as water or gel between the sample and glass surfaces. With this configuration, light travelling through the hemisphere will reach the glass-air boundary at right angles and therefore will not undergo further refraction before reaching the detector. This way one does not need a large water-filled tank, which may be cumbersome, and more importantly, if the goniophotometer is running at near-infrared wavelengths, which are of interest for the study of tissues, the absorption peaks of water at that wavelength are avoided [60].

Background noise is another source of error in goniophotometry, as the detected flux at angles is generally much smaller than that of the incident beam, especially in low-scattering samples. Measures to reduce it include shielding the system from outside light, painting surfaces black [59], placing apertures/slits in front of the source and/or detector [31, 59, 61] and normalising readings by splitting the beam before it hits the sample and using dual detectors and lock-in amplifiers to compare readings to source power in real time. The latter reduces background noise (even allowing operation in normal lighting conditions in some cases) and also corrects for variations in source power output [60, 61]. Polarisation is another factor affecting optical property measurements and some goniophotometers are designed to be polarisation-sensitive [40].

1.5.4 Phantoms

Phantoms are custom-made compounds designed to emulate the optical properties of biological tissues. Unlike biological tissues, they generally do not degrade over time and have stable optical properties, which can be calculated using light propagation theories. They can be used for a number of purposes, including: initial testing of system designs,

optimization of signal to noise ratio, performing routine quality control and comparing the performance of different systems [62].

In general terms, a phantom is made up of scattering and/or absorbing particles inside a neutral medium called the matrix. The choice of particles and matrix determines the mechanical and optical properties of a phantom and depends on the application the phantom is intended for. Typically more than 95% of a phantom is made up of its matrix, with only a very small proportion of its volume being taken up by scattering / absorbing particles [62].

Typical scatterers used in phantoms are lipid micro-particles, polymer micro-particles and white metal oxide powders. Lipid particles are biologically similar to the bi-lipid membranes of cells and organelles, which are thought to be the main scatterers in tissue. Polystyrene microspheres are manufactured under well-regulated conditions, allowing for precise control over their size and refractive index. Finally, titanium dioxide or aluminium oxide powders can also be obtained in well-controlled spherical formulations and have a high scattering coefficient. Titanium dioxide is one of the most common choices for scattering measurements, as it is a widely available pigment that is used in common white paint. Its main disadvantage is that it forms a suspension in most media, thus settling at the bottom when not stirred. Therefore, its commonly used with matrices that set, such as resin or agar [62].

The choice of matrix for the phantom has perhaps the highest impact on how it can be used. Polyester or epoxy resins are good for creating phantoms out of insoluble materials, such as titanium dioxide or polystyrene microspheres, as they prevent them from settling. In addition, their properties do not change over time. They are, however, not biologically compatible or compatible with organic chemicals. Their recommended use is for calibration, validation and comparison of systems [62].

Phantoms are commonly used to calibrate measurements methods and light propagation models, before transferring the technique to tissues. Some of the studies on gastrointestinal (GI) tissues described above utilized them [32, 39] and they have also been used to calibrate models that describe OCT signals in tissue.

1.5.5 Factors affecting optical properties of biological tissues and implications for clinical use

Light scattering in biological tissues is affected by the tissue's morphological and biochemical properties [31, 35, 36]. It is dependent on a multitude of factors, such as changes in refractive index between different structures or cell boundaries, the size and presence of intracellular organelles (such as the nucleus and mitochondria), the extracellular matrix or even larger structures such as capillaries [31, 44, 63]. The sizes of these structures can range from nanometres to millimetres and make it difficult to model the interaction of light with these tissues [31, 44, 64].

When considering the structure and contents of an individual cell, scattering will arise at a multitude of interfaces where mismatches in refractive index occur. The primary sources of scattering are expected to be organelles and their components [36, 63, 65]. Some authors argue that the cell itself is not as important for *in vivo* measurements, as surrounding structures or tissues will have a similar refractive index [36, 63], whereas others think the cell itself -and particularly its envelope- is an important scatterer [61].

Some organelles play a particularly important role in scattering due to their size and/or morphology. The nucleus, for example, is usually the largest organelle in the cell (3-10 µm) and increases in size relative to the cell in neoplastic processes [31, 36, 63, 65]. Mitochondria are also important scatterers due to their size (0.5 – 1.5 µm) and their unique folded membrane structure [36, 63]. Other organelles with multiple membrane layers and consequent refractive index fluctuations include the endoplasmic reticulum and Golgi apparatus [36, 63]. Lysosomes and peroxisomes do not contain complex structures, but are important scatterers due to their size (0.5 µm) [36, 65].

The differences in optical properties between healthy and diseased tissues such as cancer are hypothesised to mainly arise from changes in the size of scatterers within the cells of these tissues [35, 36, 65]. In particular, increases in nuclear size and nuclear-to-cytoplasmic ratio in tumour cells lead to increased light scattering [35, 36].

In general the scattering structures in tissues are modelled as spheres of varying sizes and the interaction of light with them is modelled using Mie theory, which simulates lights scattering from spheres of arbitrary sizes [36, 44]. A characteristic of Mie theory is that it

generally models forward scattering, the extent of which depends on the ratio of scatterer size to wavelength and the difference in refractive index between scatterers and the surrounding substrate [44]. The evidence shows that this describes the scattering properties of tissues well [44]. Small angle scattering is dominated by larger scatterers and large angle scattering by smaller size scatterers. Consequently, the useful information about larger structures such as nuclei and cells will be confined to very narrow angles (< 5°) [65].

The differences in size and shape of cells and their organelles between healthy and diseased tissues, such as cancer, lead to differences in their optical properties. This forms the basis of our work and the studies described in the literature.

1.5.6 Optical properties of gastrointestinal tissues

A number of studies have carried out measurements of optical properties on GI tissues. A selection of these studies is described below, organized by tissue type.

1.5.6.1 Oesophagus

Maier et al investigated the optical properties of oesophageal tumours intraoperatively in 6 patients (4 with squamous cell carcinoma and 2 with adenocarcinoma). Endoluminal illumination at 630nm was used, with the probe taking extraluminal measurements. The penetration depth and extinction coefficient of tumours in relation to their diameter was measured, allowing the treatment time for tumours of different diameters to be calculated [41]. No specific values of scattering properties were given, but the study demonstrated the potential usefulness of knowing penetration depth of light in oesophageal tumours in planning photodynamic therapy treatment.

Another study investigated the optical properties of healthy and cancerous oral, oesophageal and lung mucosa. Reflectance measurements were carried out during endoscopy in 9 patients, by placing a probe with two optical fibres (source and detector), set 2.5 millimetres (mm) apart, over areas deemed healthy or diseased. White light was used and spectroscopic measurements were carried out between wavelengths of 400 and 950 nanometres (nm). Scattering was found to be increased in normal tissue and the absorption coefficient was reduced in tumour tissue, but patient-to-patient variability was high [32].

In the next study, optical property measurements were carried out on excised samples of the gastro-oesophageal junction (GOJ), with the purpose of improving PDT planning. Tissue samples from excised surgical specimens were frozen, homogenized and placed in 200µm-thick cuvettes. An integrating sphere with a spectrometer was used and measurements were carried out between 300 and 1400nm. An inverse Monte Carlo simulation was used to produce values for μ_a , μ_s and g. Differences in μ_a and μ_s were demonstrated between stomach and oesophageal tissue and between normal stomach tissue and adenocarcinoma (52.3 % and 9.4% respectively). Squamous cell carcinoma was shown to have lower μ_a and μ_s compared to normal oesophagus. The anisotropy factor g remained constant [42].

1.5.6.2   Stomach

Spectroscopic measurements have been carried out on gastric mucosa too. A study employing inverse adding-doubling (IAD) and Monte Carlo techniques on transmittance and reflectance measurements of gastric mucosa samples showed that in the range between 400-2000nm the reduced scattering coefficient decreases smoothly with increasing wavelength [33]. This is useful to know when measuring optical properties and planning photodynamic therapy at different wavelengths.

Monte Carlo simulations were also applied in another study, where a probe with 11 optical fibres was used (a source fibre and 10 detection fibres at varying distances from the source fibre), connected to an imaging spectrometer, to obtain values for the absorption and scattering coefficients of normal gastric mucosa. Measurements of diffuse reflectance were carried out in the antrum and fundus of 35 patients, through the working channel of an endoscope. Peaks in μ_a were found at 541nm and 576nm, the absorption range of haemoglobin. The scattering coefficient was found to be lower in the antrum than the fundus over nearly the entire spectrum, reflecting the differences in cellular organization between those two regions [39].

This was also shown in the next study, which employed a multi-fibre endoscopic probe to carry out reflectance measurements on 51 patients undergoing gastroscopy for symptoms of dyspepsia. Biopsies were taken from the sites of reflectance measurements to correlate histological findings to optical properties. Spectroscopy was carried out in the range of 480-950nm and the data was processed with a Monte Carlo simulation and Henyey-Greenstein

function to extract μ_a and μ_s. The scattering coefficient of areas of antral gastritis was found to be significantly lower than normal controls, particularly between 600nm and 700nm. Findings were similar in the fundus. In contrast, the absorption coefficient was not significantly different between normal and pathological samples for either site. The scattering coefficient also varied with different levels of inflammatory infiltrate, providing therefore a way to grade different levels of gastritis. The overall method showed a diagnostic sensitivity of 66%, 80% and 100% in patients with mild, moderate or marked inflammation respectively [47].

1.5.6.3   Bowel

Marchesini et al. in 1994 demonstrated that optical properties of GI tissues could be measured quickly and easily. A spectrometer was used with an integrating sphere, to measure reflectance and transmittance on excised human colon samples consisting of mucosa, muscularis mucosa and submucosa, in the of range of 300-800nm. The absorption and reduced scattering coefficient were extracted using the diffusion approximation model [43].

The next two studies, from Wei et al., investigated the differences in optical properties of healthy and adenomatous colon tissue. In the first, tissues were taken from 13 human colons (26 samples in total, 13 healthy, 13 adenomatous), frozen and microtomed to produce slices with a mean thickness of 40µm. A double integrating sphere technique was used to carry out measurements of diffuse reflectance and transmittance and collimated transmittance, from which absorption, scattering and anisotropy coefficients were calculated at five different wavelengths. The IAD algorithm was used to extract the optical properties of the samples. Variation of absorption and scattering coefficients for all tissues with wavelength was noted, as expected. Additionally, both the absorption and scattering coefficients were significantly higher in all adenomatous tissues compared to normal tissues while the anisotropy factor showed differences without reaching statistical significance [34]. In the second study, the methodology was similar. Twelve healthy and 12 adenomatous samples were used and a double integrating sphere technique was employed again to measure the diffuse transmittance and reflectance and the collimated transmittance. Monte Carlo modelling found optical properties to be different between normal and adenomatous tissues. This was wavelength dependent, with the highest difference in absorption and

scattering coefficients between normal and adenomatous mucosa occurring at 780nm. For the muscle layer, the maximum difference in absorption coefficient also lay at 780nm whereas for the scattering coefficient it lay at 850nm [46].

1.6 Optical Coherence Tomography
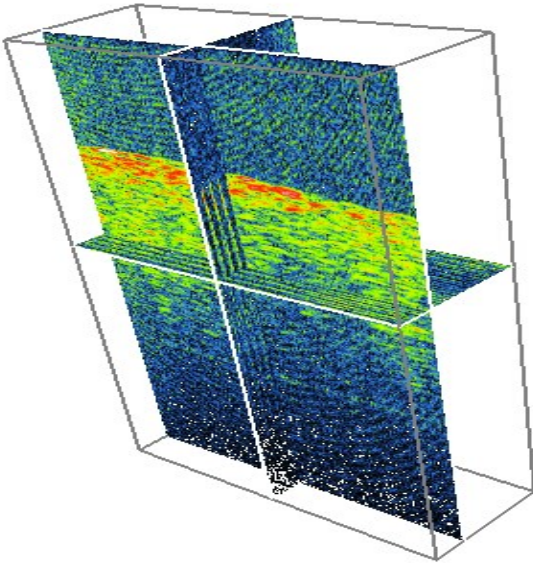
Figure 1.1: Three-dimensional representation of OCT imaging, demonstrating coronal, saggital and axial views

1.6.1 Principles of OCT

Optical coherence tomography produces high-resolution cross-sectional images of tissue in vivo. It uses infrared light rather than acoustic energy and has a ten-fold higher resolution than does high-frequency ultrasound, though the maximum depth of optical coherence tomography is lower than with ultrasound imaging [6, 66].

Axial (Z) Scanning

Reference Mirror

$E_R(t), P_R(\omega)$

Collimation Lens

Lateral (X or Y) Scanning

Beam Splitter

$E_S(t), P_S(\omega)$

Low coherence light source

Objective Lens

Beam Reducer

Photo detector

Sample under test

Computer & Display

Filtering   Demod. / processsing   ADC
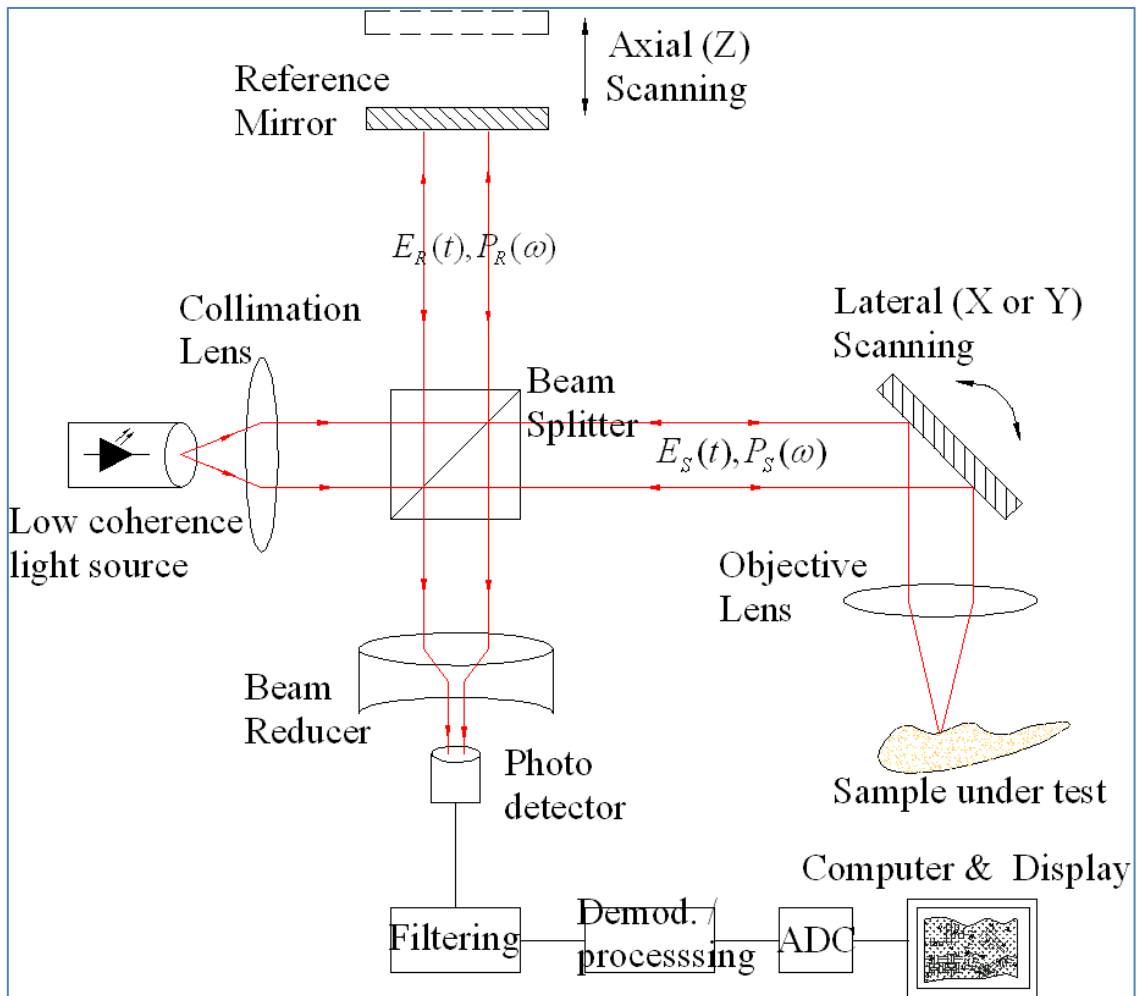
Figure 1.2: Diagram of typical OCT system configuration

It is a non–invasive imaging technique, based on split beam interferometry. A laser beam is split into a reference and a sample path, which travel equal distances and are reflected by a mirror and sample respectively, then recombined inside an interferometer. An interference pattern is obtained, with peaks corresponding to sharp refractive index changes within a sample, e.g. anatomical layers in a tissue sample [45, 48-51, 53, 67]. The interference pattern produced at one point in the sample provides a single depth scan, also called an a-scan. By scanning the beam across a sample, multiple a-scans can be taken, producing a b-scan. Typically a two-dimensional cross-sectional image may consist of 500 depth scans over 5mm. Multiple b-scans form a c-scan, a three-dimensional image [45, 48, 55, 67].
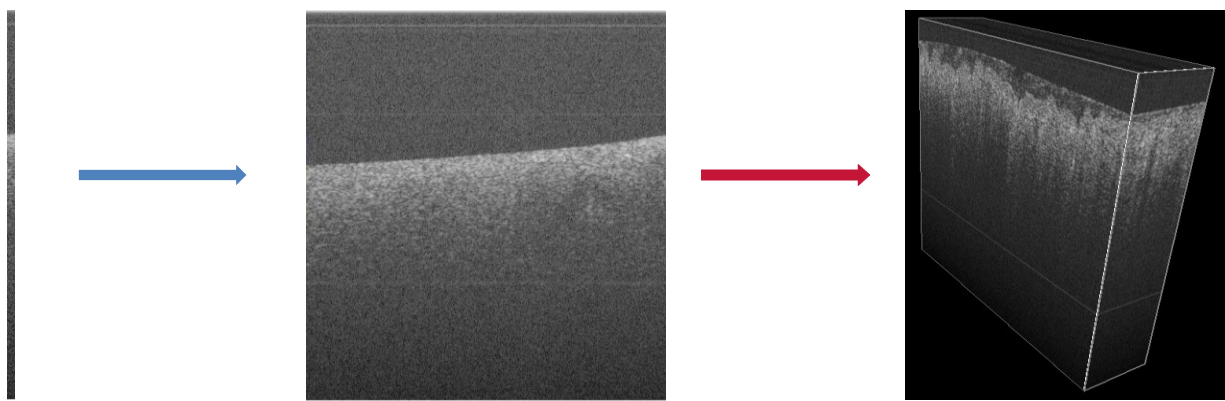
Figure 1.3: Diagram demonstrating the relationship between a single axial scan (A-scan), a two-dimensional image (B-scan) and a three-dimensional stack of images (C-scan)

The depth of penetration of the OCT signal (and therefore depth of OCT images) is limited by scattering in deeper layers, so that typical depth scans are 1-3mm at wavelengths of 800-1300nm [45, 50]. The axial (depth) resolution of the OCT system is limited by the coherence length, a statistical property of the light source, which is inversely proportional to the bandwidth (the distribution of wavelengths in the spectrum of the light source). The transverse resolution is limited by the optical focusing characteristics of the system [48, 49]. Depending on system characteristics, a very high resolution can be obtained, comparable to histology, in the region of 0.5-20µm. This is high enough to identify microscopic features such as villi, glands, crypts, lymphatic aggregates and blood vessels [48-51, 53]. A typical OCT configuration is shown in Figure 1.2.

1.6.2 Clinical use of optical coherence tomography in gastrointestinal tissues

Optical coherence tomography is an attractive technology because of its ease of use, rapid function and relatively low cost [45, 66]. It can also be adapted to work in vivo, runs in situ and in real time, is portable and can operate through air (unlike ultrasound), without contrast agents [48]. OCT has been used to image eye, GI, heart, skin and nervous system tissues [48, 49, 51, 66]. It has also been used to image cartilage and evaluate its degeneration in osteoarthritis and for intravascular, dental and neural imaging [45].

Work has been carried out on OCT with the aim of using it as an adjunct to histopathology or even a future replacement. A common aim of many projects around the world is the development of "optical biopsies", where histological diagnosis and tumour margin detection can be carried out in vivo, during endoscopy or surgery, with no need for sample excision. With its current characteristics, OCT could be used to guide biopsies and reduce sampling error rates, or to identify the invasion depth of tumours (T-staging) and monitor response to therapy [48-50]. It may bring histopathology and endoscopy together into one technology, especially with technological improvements to resolution and image acquisition rate [50].
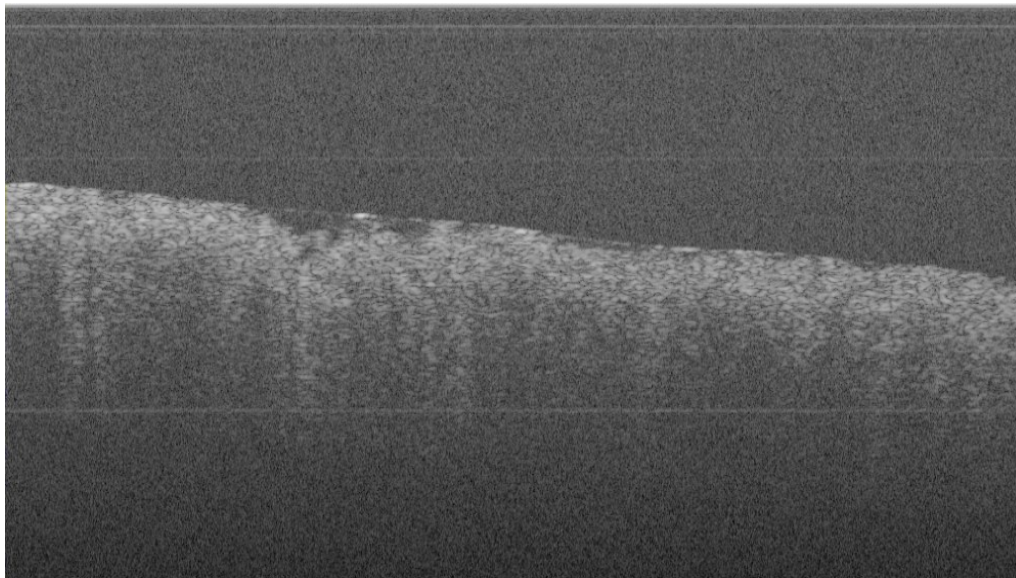
Figure 1.4: Optical coherence tomography image of normal colon

The main focus of clinical applications of OCT has been the identification of changes in tissue structure using high-resolution images, akin to "in-vivo histology". In gastrointestinal tissues, OCT provides a resolution near the cellular level that could detect pre-malignant changes such as Barrett's oesophagus, which may go undetected with conventional endoscopy.

It has been shown that endoscopic OCT can visualise the 5-layer structure of the normal oesophagus [49, 68] and distinguish it from the glandular epithelium of the stomach [49, 68], as well as from pathologies such as intestinal metaplasia [69] Barrett's dysplasia [49, 68], adenocarcinoma [49, 68] or squamous cell carcinoma [49]. Colon tissues have also been studied using the same principles and differences in tissue architecture have been used to distinguish between healthy colon [48, 51, 70, 71], dysplasia [71], ulcerative colitis [48, 51, 70] and cancer. The same has been demonstrated on ex-vivo specimens too [48, 72].

In all of these cases, the diagnostic ability has been based on detecting the changes in structure occurring with various pathologies, such as the change from squamous to glandular epithelium in Barrett's or the total loss of architecture in neoplasms [68]. Some authors have employed scoring systems to classify OCT images [51], whereas others have also employed computer algorithms to look for differences in image brightness which would correlate to differences in scattering between healthy and diseased tissues [71]. One study looked at the thickness of the mucosa as well as the effect of probe pressure on images, particularly the visualisation of deeper tissues [72].

All of the above authors focus on the acquisition of high resolution images to aid in detection of pathologies as early as possible and in vivo. Limitations include the subjectivity of assessing these images and the significant intra and inter-observer variation. For example, the sensitivity and specificity of OCT for Barrett's have been shown to range from 68% - 88% and 70% - 82% respectively [49, 73, 74]. Furthermore, qualitative analysis relies on training the observers. To overcome these limitations studies are now focusing on quantitative analysis of OCT data.

1.6.3 Quantitative analysis of optical coherence tomography data

As mentioned above, OCT is an attractive technology for tissue imaging because of its ease of use, speed and high resolution, amongst other things. However, the OCT signal and consequently image quality are affected by the scattering properties of tissues. To counteract this, early work on further processing of OCT images focused on improving image quality by modelling the propagation of the OCT sample beam through tissue. The aim was to improve aspects such as image contrast and signal-to-noise ratio (SNR) [52, 75].

In 2000, Thrane et al. described a novel model based on the extended Huygens-Fresnel principle (EHF) [52]. The latter describes light propagation through a medium with a random spatial variation in the refractive index [76]. The model treated optical distortions as random effects and assumed that the propagation of light towards an area of refractive index mismatch (such as a tissue layer), the reflection from that area and the propagation back to the surface were all statistically independent of each other. Backscattering from the reflection plane was considered diffuse and split into scattered and unscattered parts.

The authors demonstrated that it worked in both the single and multiple scattering regimes and showed that by accounting for an effect termed the "shower curtain effect" a stronger OCT signal can be obtained. The "shower curtain effect" referred to the fact, that, if a scattering medium is placed between a light source and an observer, the position of the scattering medium in relation to the two points (source & observer) affects how the observer sees the light source. For example, a person standing right behind a shower curtain cannot see a distant observer, but can be seen by them. Furthermore, a parameter called the heterodyne efficiency factor $\Psi(z)$ was defined, which described the reduction in the signal-to-noise ratio (SNR) caused by scattering in tissue. The authors showed that inclusion or exclusion of the shower curtain effect in their modelling had a significant effect on the way heterodyne efficiency factor changed with depth. The theory was tested against experimental measurements carried out on aqueous suspensions of latex microspheres and the study concluded that the model helped maximize the signal obtained from OCT systems where the lens is not in direct contact with the material being scanned and the backscattering is diffuse, such as in biological tissues [52].

1.6.3.1 Optical properties and signal attenuation

Following on from the work described above, the same authors expanded their model to extract information about the optical properties of the tissues being imaged, which could be used for diagnostic purposes [75]. The attractive features of OCT could now be transferred to the domain of optical property measurement and an extra dimension could be added to the pure imaging-based diagnostic capabilities of OCT, by possibly identifying pre-malignant and early malignant changes, where layer organization is preserved, thus rendering them undetectable by imaging alone [75].

In 2004, Levitz et al. modified the above model to extract the scattering coefficient and effective anisotropy factor g_{eff}, by fitting the depth profile of coherent backscattered light in an a-scan. Their extraction algorithm worked by selecting a region of interest from the raw OCT image, which is made up of multiple a-scans. These were averaged and, after linear smoothing was applied, an averaged curve was produced, a section of which was chosen to fit the model to. Pixels near the surface of the sample, where reflection is specular, were excluded from the fitting routine [53].

The user then entered initial guesses of the scattering coefficient and root-mean-square scattering angle (θ_{rms}) and the model fitted the data to a signal vs. depth curve, returning values for the scattering coefficient, θ_{rms} and error estimates. The effective anisotropy factor was then defined as $g_{eff} = \cos\theta_{rms}$. Typically, this was an overestimation of g. An evaluation study was carried out on 6 epoxy resin phantoms with three distinct scattering coefficients (from set titanium dioxide concentrations) and two absorption coefficients (from photocopying machine toner concentrations), using an OCT system running at 1300nm. Values for g_{eff} produced by the model overlapped with those calculated by Mie theory. The scattering coefficient was overestimated by the algorithm, but showed a step-wise increase between phantoms. An ex-vivo study was also carried out on human atherosclerotic lesions. Fourteen aortic segments, obtained within four hours of autopsy, were imaged with OCT and extracted optical properties were compared to histological examination of the samples. Qualitative assessment of OCT images showed good correlation with histology, allowing for differentiation between normal and atherosclerotic samples. In terms of optical properties, it was shown that normal samples had a g_{eff} between 0.95 and 1 in 90% of cases, whereas more than 50% of atherosclerotic samples were below 0.95. Similarly, about 95% of normal

samples had μ_s values between 15mm^{-1} and 39mm^{-1} while in about 60% of atherosclerotic samples it was below 15mm^{-1} [53].

Another single scattering model was developed by van der Meer et al. and used to extract optical properties of healthy arteries and atherosclerotic plaque. Thirteen human carotid artery specimens were imaged with OCT running at 800nm with a bandwidth of 120nm. An algorithm was developed based on the single scattering approximation (exponential decay of OCT signal with depth) that was fitted to the average of 50-100 adjacent a-scans, producing the attenuation coefficient. The model was first tested on a set of bi-layer phantoms (consisting of Intralipid solutions of different concentrations over a plastic scattering layer) The OCT signal for the second layer in the phantom studies fitted very well with the algorithm for all concentrations of Intralipid (the first layer) and the determined values of μ_t compared well to those expected (10mm^{-1}). Similar results were obtained for the vascular tissue, where the algorithm fitted µt well for all areas of interest. It was shown that the attenuation of OCT light was highest in thrombus (11.2 ± 2.3mm^{-1}) and the lowest in lipid-rich regions (3.2 ± 1.1mm^{-1}), with intermediate values for the intima (5.5 ± 1.2mm^{-1}). The attenuation coefficients of both diffuse intimal and lipid-rich tissue were significantly different to other plaque components. The authors found discrepancies in their values compared to those of Levitz et al. (2004), which they put down to differences in temperature of samples used for the two studies, the wavelength and bandwidth of the OCT systems and the models used. They argue that the single scattering model, however, is valid for attenuation coefficients up to 17mm^{-1} for single layered models and that their results in phantoms show their model is valid for layered media too. They conclude that the study demonstrated the proof of principle of using a single-scattering model to extract optical properties of healthy and diseased arterial tissue, but more systematic testing is required to establish its reliability in larger scattering depths and its accuracy in distinguishing healthy and diseased areas [55].

Xu et al. (2008) used an OCT microscope to measure the optical properties of individual arterial wall and plaque layers, by imaging transverse sections of arteries, as well as taking luminal measurements. The OCT system used for transverse measurements was custom made, running at 1320nm with 7.3µm axial resolution in air (5.4µm in tissue) and 25µm lateral resolution. A standard OCT system was used for luminal imaging, attached to an optical fibre catheter probe. Seventy-two formalin-fixed coronary artery specimens were

obtained post-mortem. Cross-sections were produced by cutting the arteries at various points of interest, which were imaged at normal angles to the artery axis. Calibration was carried out on a phantom composed of 0.03% 0.2μm polystyrene microspheres. The effect of mounting angle on measurements was also investigated, as it was hypothesized that the direction of the OCT beam through an anisotropic medium such as the arterial wall would affect the optical properties measured. The region in OCT images that was fitted to the model started 50μm below the specular reflection peak (i.e. the sample surface) and ended were the signal was attenuated to 37% (1/e). This region was chosen because that is where single scattering approximation can be applied. The optical properties of the samples showed little angular dependence, whereas they varied significantly when imaged from the lumen side compared to transverse measurements. Backscattering and attenuation decreased from lipid to fibrous to calcified plaque. Limitations of this study included the fact that the samples were formalin fixed and post-mortem and that a single scattering model was used. It was pointed out that measurements were taken within 1 MFP, rendering the model valid, but it was acknowledged that, due to plaque geometry, it is not always possible to fit to single-scattering regions and more sophisticated multi-scattering models may be better suited [57].

Tomlins et al. (2010) quantitatively analysed OCT images from 19 oral mucosa specimens (15 normal, 4 dysplastic), looking for differences between healthy and dysplastic tissue. Linear gradient analysis of the OCT signal was applied to a selected region of interest in each B-scan near the surface, where single scattering dominates. A colourmap of each specimen was also produced, indicating which regions were dysplastic and which normal. Using a manually selected gradient value as the threshold, they achieved a sensitivity and specificity of 90% [77]. A study using a similar methodology was carried out on 19 formalin-fixed specimens (13 fibroepithelial polyps, 2 with mild and 4 with moderate/severe dysplasia) from the same group. A statistically significant difference was demonstrated between normal/mild and moderate/severe groups [77].

1.6.3.2 Monte Carlo

A Monte Carlo (MC) model of the OCT signal has also been described. MC modelling treats light as photon packets, without using wave equations. Therefore, it could be argued that it is unsuitable to describe OCT, a process that is based on the wave properties of light.

However, when modelling the OCT signal in terms of its intensity only, without referring to the field or phase, MC modelling is valid [76]. MC modelling works by calculating the statistical probability of individual photons being scattered in certain directions and of those photons contributing to the OCT signal. Individual scattering events for each photon are treated as independent of past or future scattering events. An initial guess of the medium's scattering properties is required to calculate the probabilities of scattering events and then these are averaged out and fitted to measured data to provide definitive values of the scattering properties. The model in this study was designed to simulate the heterodyne efficiency factor of an OCT signal (i.e. the attenuation of the OCT signal with depth, as described above). It was compared against experimental measurements on phantoms and against the EHF model developed by Thrane and good agreement was shown. The authors concluded that MC modelling is thus suitable for describing the OCT signal [76].

Monte Carlo modelling was again used in another study to describe the OCT signal, with the aim of investigating contributions of different orders of scattering to the OCT image and separating the effects of single and multiple scattering and of diffusion and non-diffusion components. The model described light in terms of photons, which were split into categories. Those that had gone through multiple scattering carried distorted information about the depth of the scatterers and were called multiply-scattered photons (MSP). In contrast, least-scattered photons (LSP) carried reliable information on scattering depth. Photons that exceeded a defined number of scattering events were deemed to be part of the diffusion component of the OCT signal and were labelled diffuse component photons (DCP). The algorithm was tested on OCT images of a thumb. It was shown that LSP primarily form the image of upper skin layers, up to optical depths of 1mm, whereas MSP make a blurry contribution to the whole image. The diffusion component contributed mainly to the imaging of deep layers, also producing a blurred image, but only of the lower layers. The non-diffusion component produced images of layers up to 1.4mm optical depth [78].

1.6.3.3   Refractive Index

Another study by van der Meer looked at effect of temperature on optical properties. Here, carotid samples and a phantom made from dairy butter were imaged with OCT and their optical properties extracted under controlled temperature conditions (monitored with a

thermocouple) between 18oC and 37oC. The OCT system was again at 800nm with an axial resolution of 3.5µm and a lateral resolution of 7µm. The index of refraction was measured for each sample by placing on a reflective surface that could be seen on the OCT image and using a formula that incorporates the refractive index of the surrounding medium (air) and the difference in path lengths inside and outside the sample. The algorithm described in the previous study (van der Meer et al. 2005) was optimized to take advantage of the refractive index information. The refractive indexes of intima (1.352 ± 0.002), media (1.382 ± 0.006), calcifications (1.63 ± 0.05), lipid (1.42 ± 0.04) and lipid phantom (1.417 ± 0.009) were produced at 37oC and were shown to decrease when increasing the temperature from 18oC to 37oC, with the highest difference in the lipid phantom. The attenuation coefficient for arterial and plaque segments could be determined, but not reliably for lipid pools. A decrease in μ_t was observed with increasing temperature. The study demonstrated the importance of refractive index in obtaining accurate measurements of optical properties, as the decrease in refractive index (from 18 oC to 37oC) was followed by a decrease in μ_t. The effect of temperature on refractive index and μ_t was demonstrated and the results of this study suggested that fatty tissues should be measured at 37oC [56].
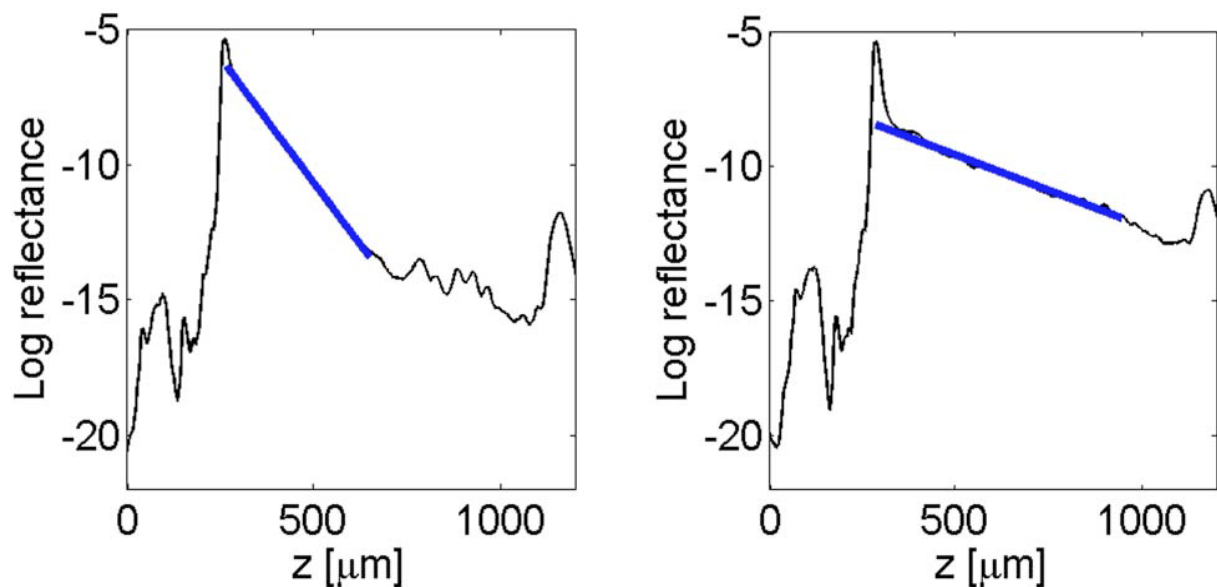
Figure 1.5: Linear fitting of OCT signal attenuation. More scattering sample (higher attenuation) on the left, less scattering sample on the right

The studies above demonstrate the potential of OCT as an optical property measurement system, but also show a number of limitations. First of all, different groups used different methodologies, including models, OCT systems and wavelengths, phantoms and calibration

methods. This means their results cannot be readily compared. Additionally, there was considerable variability in results within each group, with optical property values being given as ranges. Many studies were designed as proof-of-principle demonstrations, without systematic calibration and long-term data gathering. The sample numbers were low and there were differences in the physiological conditions of tissues between different studies. The models were not perfect with deviations from experimental values shown, which were attributed to experimental technique or errors in calibration procedures. Furthermore, some were based on single and some on multiple scattering approximations, some on EHF and some on Monte Carlo simulations. Finally, none of the studies could provide definitive values for optical properties or a systematic assessment of their ability to distinguish between healthy and diseased tissues.

1.7 Aims and overview of thesis

The aim of this work is to investigate the effect of the differences in optical properties between healthy and pathological gastrointestinal tissues on OCT imaging and the potential of using that effect as a clinical diagnostic test for cancer and other gastrointestinal diseases, particularly precancerous states such as Barrett's oesophagus and dysplasia.

To achieve this, we will be carrying out quantitative analysis of OCT data using two techniques. The first one is gradient analysis of OCT signal attenuation, which has been previously applied on other tissues. The second one is a novel technique called morphological analysis, which has not been applied to OCT images before. The aim will be to develop a methodology that can demonstrate differences between healthy and diseased tissue and can accurately predict the histological diagnosis of new specimens.

As described in Section 1.5.5, the main factors affecting optical properties in tissues are scatterer size and density. Scatterer size affects the anisotropy coefficient g and scatterer density primarily affects the scattering coefficient, as well as the measured value of g when multiple scattering occurs. It has been shown that there is a link between the scattering coefficient and the degradation of the OCT signal (more signal loss in highly scattering tissues). The scattering coefficient describes how much light is lost due to scattering, but does not provide any information about the direction in which the light is lost. This is described by the anisotropy coefficient.

 Information about the direction in which light moves in tissues would be useful in improving models of OCT signal transmission and could improve the diagnostic accuracy of techniques depending on OCT signal analysis. For that reason, a goniophotometer was also constructed with the aim of measuring the anisotropy coefficient g of healthy and diseased gastrointestinal tissues.

A secondary aim will be to try and extract information about the anisotropy coefficient from the OCT signal, something which has been little investigated in the literature. Measurements of the anisotropy coefficient using OCT would then be validated against those carried out with goniophotometry -which is the gold standard- on the same tissues.

2 GONIOPHOTOMETRY

2.1 Construction and initial calibration of a goniophotometer

2.1.1 Background

As described in the introduction, a goniophotometer is a system that measures the intensity of scattered light at different angles around a sample illuminated by a thin light beam. Through these measurements the anisotropy factor g of a material or tissue can be calculated. In this section we describe the construction of a prototype goniophotometer and initial testing on epoxy phantoms.

2.1.2 Aims

To build a prototype goniophotometer providing consistent scattering curves from samples with minimal noise or error.

To develop a data processing methodology for extracting g from the scattering curves.

To test the system on a set of epoxy phantoms with predicted values of g.

2.1.3 Methods

A goniophotometer was constructed with a light source consisting of a 1310nm-wavelength laser beam, produced by a laser diode connected to an optical fibre and collimation package. The wavelength was chosen to match that of the Thorlabs OCT system, so that wavelength-dependent optical properties obtained through the two methods could be compared. At the centre of the system lies a rotation stage, which allows for manual or motorized motion. A custom-made top plate with an arm for the detector is attached to it. The rotation stage is computer-controlled, via serial port, allowing for accurate positioning with a 0.001° minimum step size. A gap in the centre allows for easy mounting of stationary objects at the centre of rotation. The detector is a Germanium sensor, which works in the near-infrared and infrared regions. It is connected to an amplifier, which gives an arbitrary reading of intensity, with a dynamic range of 5 orders of magnitude (10^4 to 10^9). The amplifier is connected to a Keithley multi-meter, which in turn is connected to the same computer as the rotating stage, via a GPIB interface. This allows for readings to be recorded and stored automatically on the computer. The sample mounting system places the front surface of the

sample at the centre of rotation and holds the sample at 90° to the laser beam. In order to hold floppy samples such as tissue or thin slices of phantoms a system whereby the sample is held against a lens was designed. Both a single and double lens configuration were tested, as described below. To reduce background noise, the goniophotometer has been shielded by an enclosure, made of a metal framework covered by black cardboard. The cardboard reflectance at 1310nm was found to be less than 5% by spectrometry. This reduces both external light and internal reflections. The enclosure is further shielded by a black, non-reflective sheet, which is draped over it. Measurements are run in the dark. Custom-made software is used to control the goniophotometer, running in a Windows (Microsoft Corporation, USA) environment on a personal computer. The software allows the user to define the number of readings to be taken and the measurement angles and provides an output of the data in a comma-separated value (.csv) file format. Figure 2.2 demonstrates the configuration of the goniometer.

A set of phantoms was used for the initial testing and calibration of the goniometer. These were made by mixing epoxy resin and titanium oxide powder at varying concentrations and letting the mixture set. Slices of different thickness could then be cut from phantoms of different densities, allowing measurements to be taken from samples of varying scattering properties. A set of uncut phantoms is shown in Figure 2.1.

Figure 2.1: Epoxy phantoms of increasing density from left to right.

The production process for the epoxy phantoms was designed to produce standard-sized phantoms of various densities. For the purposes of our study we needed phantoms with scattering coefficients varying from high to very low, approaching -or in the domain of- single scattering. Two methods can be used to produce phantoms with very small scattering coefficients, the first one being to make very dilute phantoms and the second one to make

very thin phantoms. The first method is theoretically easier, but difficult in practice, as there are inherent limitations in the manufacturing technique. These include clumping of the titanium oxide powder and difficulties in producing perfectly homogeneous phantoms. With very dilute mixtures the problem would be exacerbated, as there could be large areas of the phantom with no powder at all. Furthermore, biological tissues have set scattering properties which would require very thin, microtomed slices to enter the single-scattering regime. Consequently, the goniometer sample mount was designed for thin slices, which meant the phantoms had to be thin too. Additionally, in order to ensure light is scattered forwards through any given sample and does not escape from the sides, thin, wide sections must be used.

Taking the above into account, it was decided to obtain thin slices of phantoms of varying densities for the purposes of our experiments. As microtomes are routinely used to cut thin slices of tissues including bone, this was the first option that was explored. Samples were sent to a specialist lab in Germany to be microtomed using different types of blades and microtomes. Sections between 10-50 microns were produced and sent back for testing in our system. Though they were fairly durable and of the desired thickness, they were creased, had cracks in some areas and were not of the same width throughout. The main issue from the lab report was that the material was too hard and suffered from built-in tension, which caused the sections to crease and shrink as soon as they were cleaved off the main block. In addition to that, the whole process was cumbersome and expensive, so a compromise was made.

Instead of microtoming the phantoms, new ones of varying densities were made. These were then cut using a diamond saw and polished to produce slices of varying thickness. These were then used to carry out calibrations and initial testing, as well as a pilot study to determine the anisotropy coefficient g.

Alignment and repeatability of the goniophotometer at this stage was deemed good, as the curves produced are generally symmetrical and overlap well for measurements of the same phantoms (Figure 2.3). However, as can be seen in the same figure, a noise pattern became evident during measurements on the epoxy phantoms.
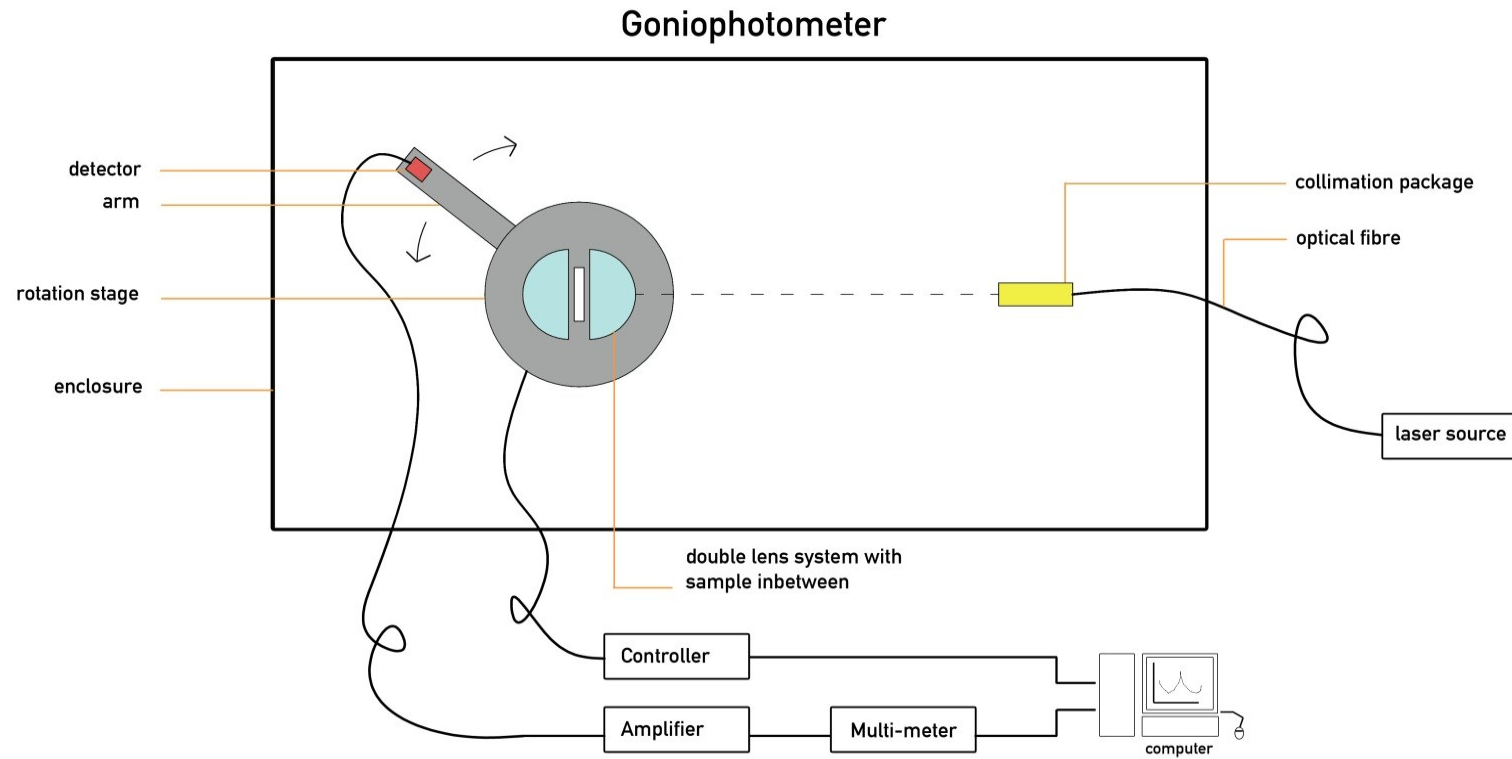
Figure 2.2: Goniophotometer setup. The rotating stage and detector can be seen on the left, with the laser source on the right, all inside the light-proof enclosure. The system pictured here featured a double-lens mount, which was changed later as described in the main text

This pattern appeared when the rotation step size was reduced to below 1°, at 0.5° and less. To establish whether the new double-lens system was to blame, the old clamp system was used for another set of measurements. Measurements on the same phantom at decreasing step sizes confirmed that the pattern was present and that curves overlapped with very good agreement, showing it was a real, consistent effect that had gone unnoticed before because of averaging over larger step sizes (Figure 2.3).

By taking measurements with the detector cap on it was demonstrated that the noise pattern was not optical in nature, but electrical and likely originating from the rotating stage, because of its periodicity and the fact that a relatively large current is used to drive the motor. Measures were taken to reduce electrical noise by moving the detector cable away from the rotating stage and motor control cable (suspending it from the top of the enclosure) and also reducing the current to the motor by choosing a slower motion profile. This led to a decrease in overall noise levels.

A lens-based mounting system was introduced to hold tissue samples and thin phantom slices. Taking into account the effect of total internal reflection as discussed above and in the literature, it was decided to go for a system where the sample is held against a semi-cylindrical lens. This would have all the benefits discussed earlier, but would also be easier to align, since only the horizontal and rotational alignment would matter, in contrast to a hemisphere which would have to be perfectly aligned in the vertical direction.

Eventually, instead of cylindrical lens, a slightly flatter one was chosen, due to two factors: the difficulty and cost of procurement of a custom-made lens and a calculation of the refractive properties of the two lenses which showed that the flatter one was better for our experiments. As shown in Figure 2.4, when taking into account the refractive index of tissue and the glass the lens is made of, our configuration is closer to the ideal line, causing less deviation between measured and true angles compared to a perfect half-cylinder.
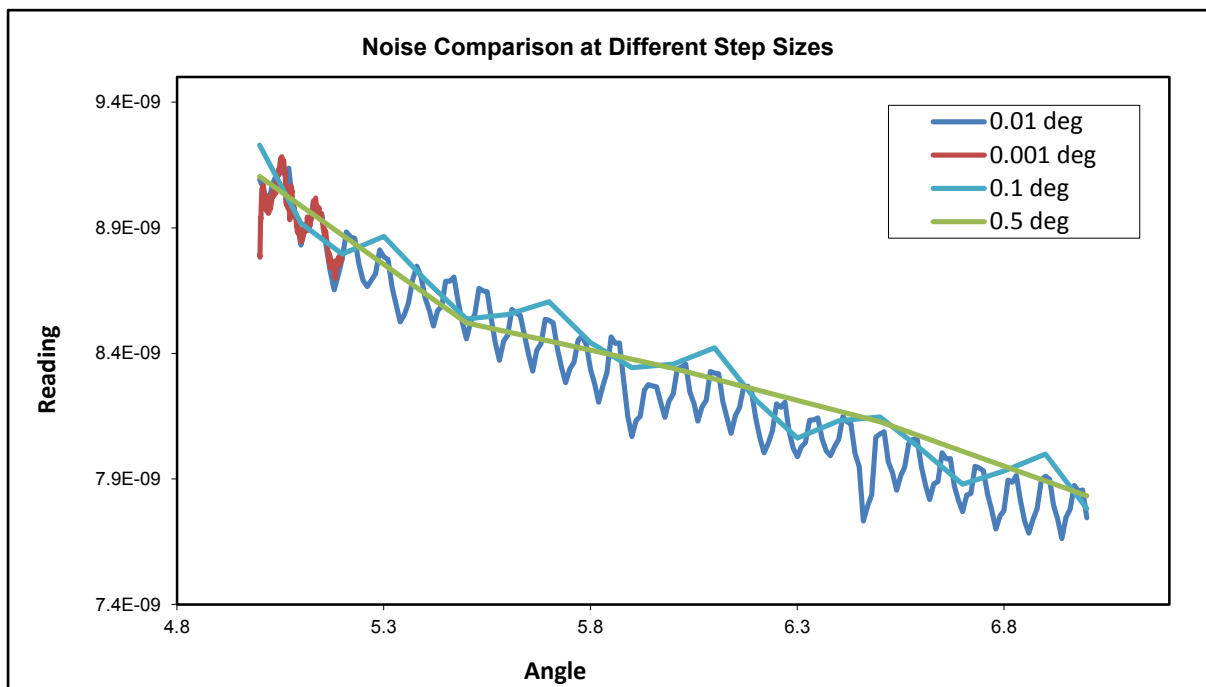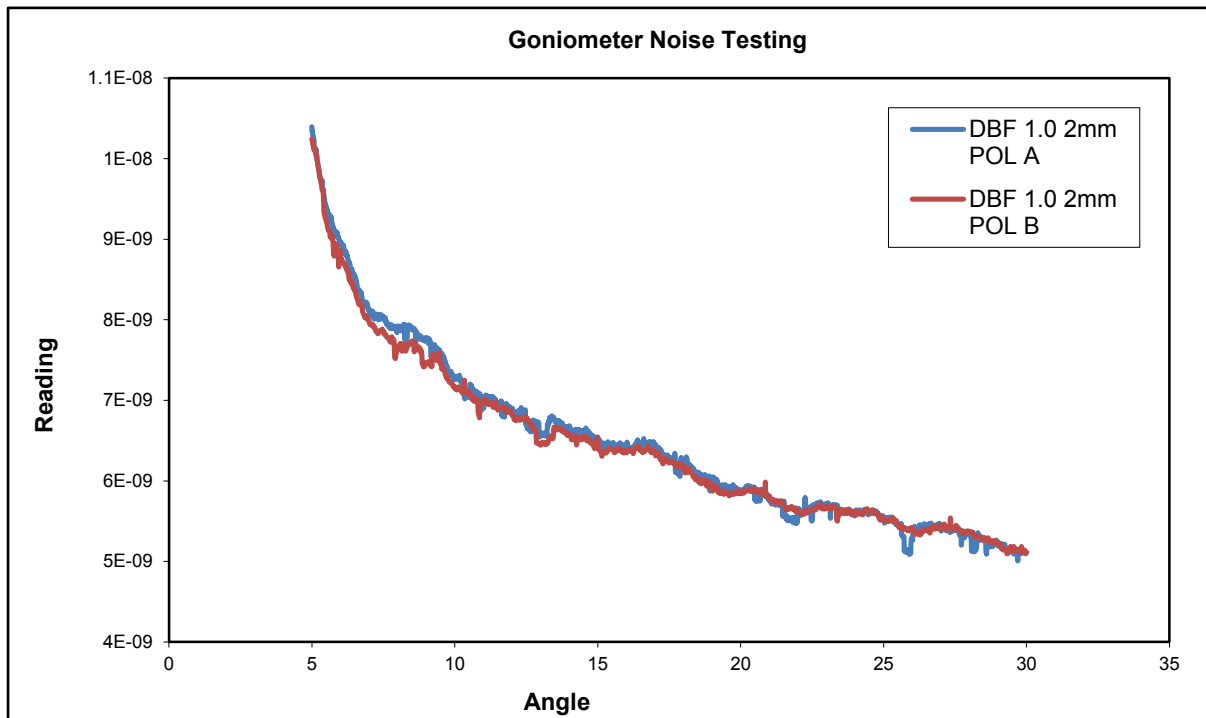
Figure 2.3: Goniometer noise testing. The top graph shows good repeatability when measuring the same phantoms, however there is a noise pattern, which was further investigated by repeating measurements of the same phantom at different step sizes (bottom graph).

First a single lens was used, with a glass slide held against it and then a double lens system. Ultrasound gel was used as an index-matching medium between samples and lens surfaces. The arrangement of the new mounting system means that the "useful data" of forward scatter is compressed by the lens into the –60 to 60 degrees region and a decompression algorithm needs to be applied to give the correct shape for the curves, as described in the literature [59, 60]. Measurements from the single lens configuration are shown in Figure 2.5. Four superimposed measurements of 4mm 0.0625mg/g phantom are shown, demonstrating generally good agreement between runs. A double-lens system was then tried, with the sample held between the flat surfaces of the two lenses. This was to try to apply the improvement in data gathering of front scatter, to backscattered light too. Two consecutive measurements on a 2mm 0.0625mg/g polished phantom with the double lens system are shown in the same figure.

Measurements were then tried at 0.5º increments, which produced a noisy pattern. It was hypothesized that the pattern was due to inter-reflections between the lens surfaces and internal reflections within each lens. This was confirmed by repeating the measurements with a single lens (without a sample), whereupon the pattern disappeared. After some shielding was added around the lens areas that were not at the measurement plane, the pattern was removed (Figure 2.6). Looking at the curve, there is a sharp drop, as expected, between around 60 and 110 degrees on both sides, due to "compression" of the light at wide angles by the lenses and the fact that little light is scattered around the sides of the sample.
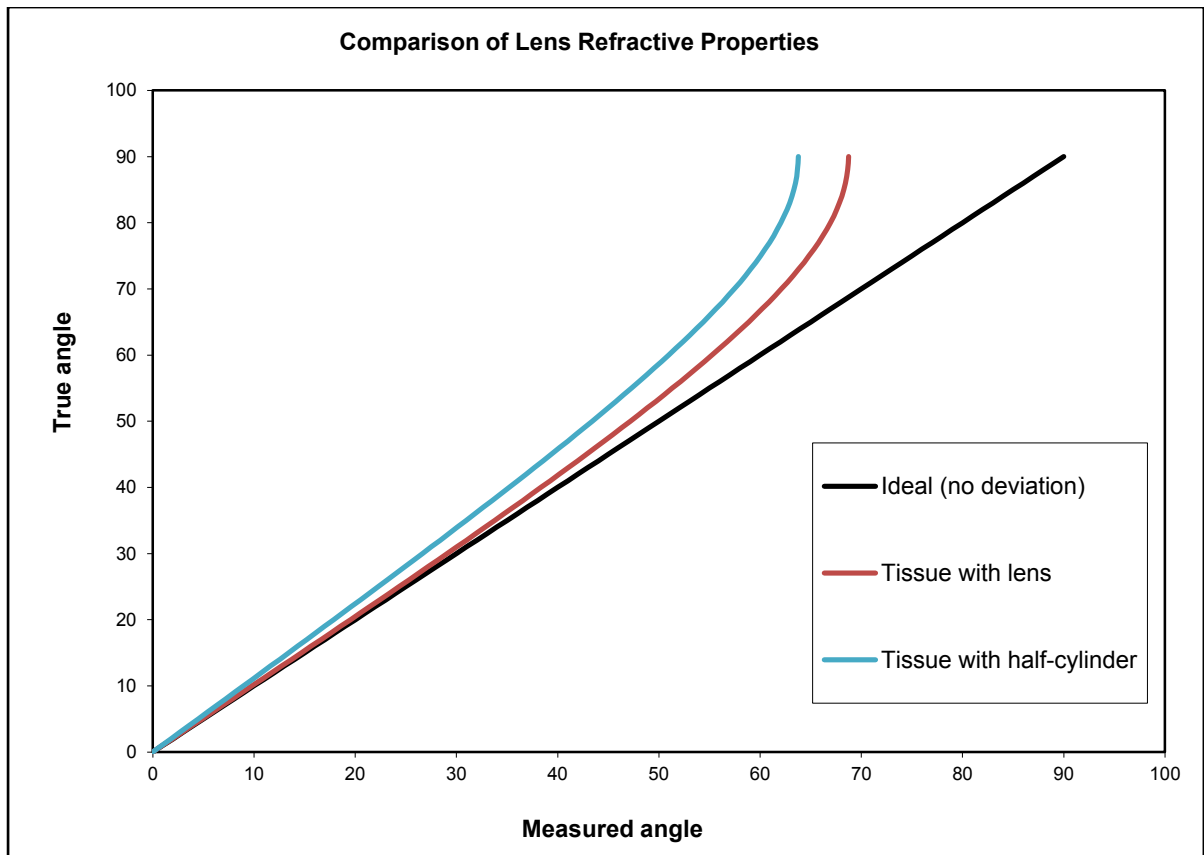
Figure 2.4: Simulation of refractive properties of a half-cylinder mounting system versus the lens used in our setup. The differences between measured and actual light exit angles are shown.
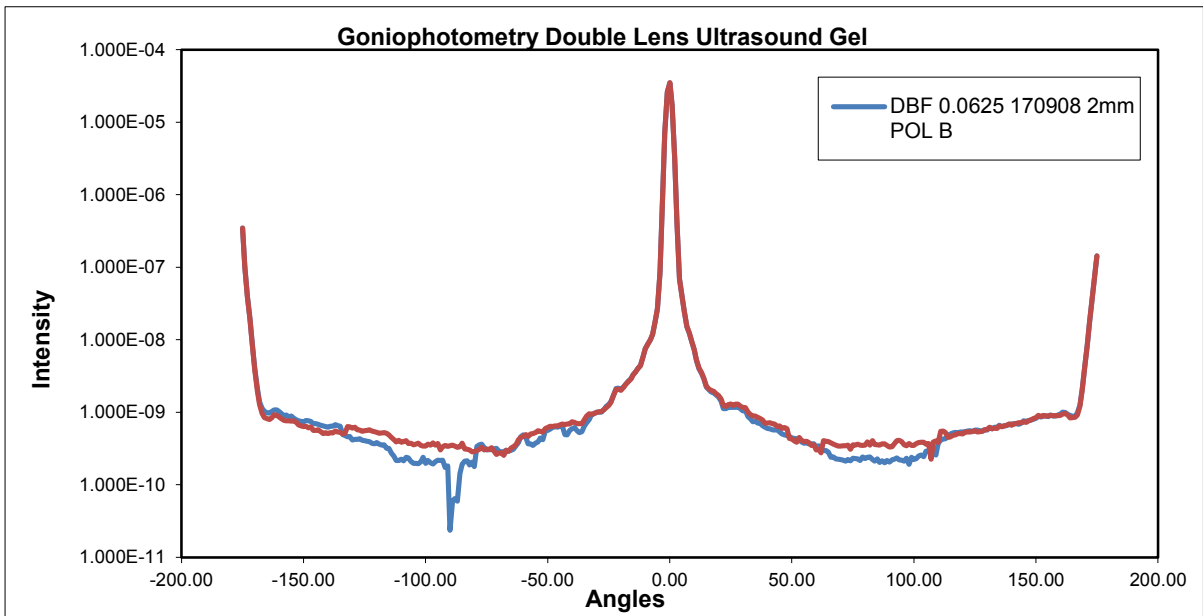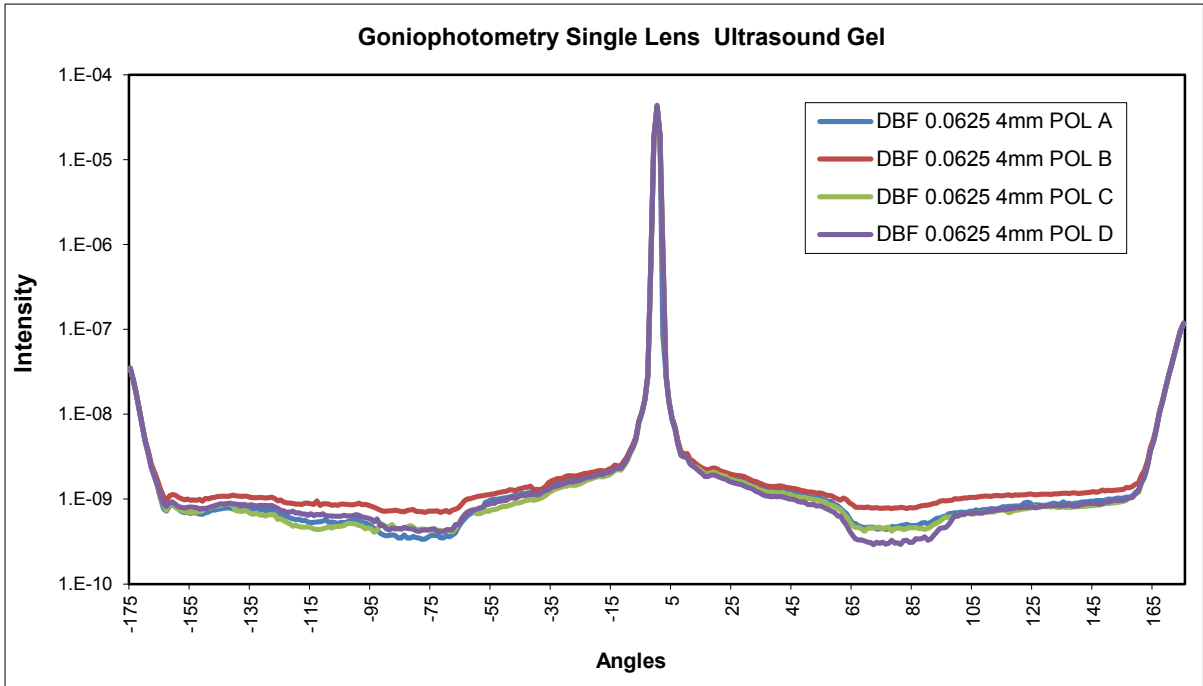
Figure 2.5: Comparison of single lens (top graph) vs double lens (bottom graph) mounting system for the goniometer.
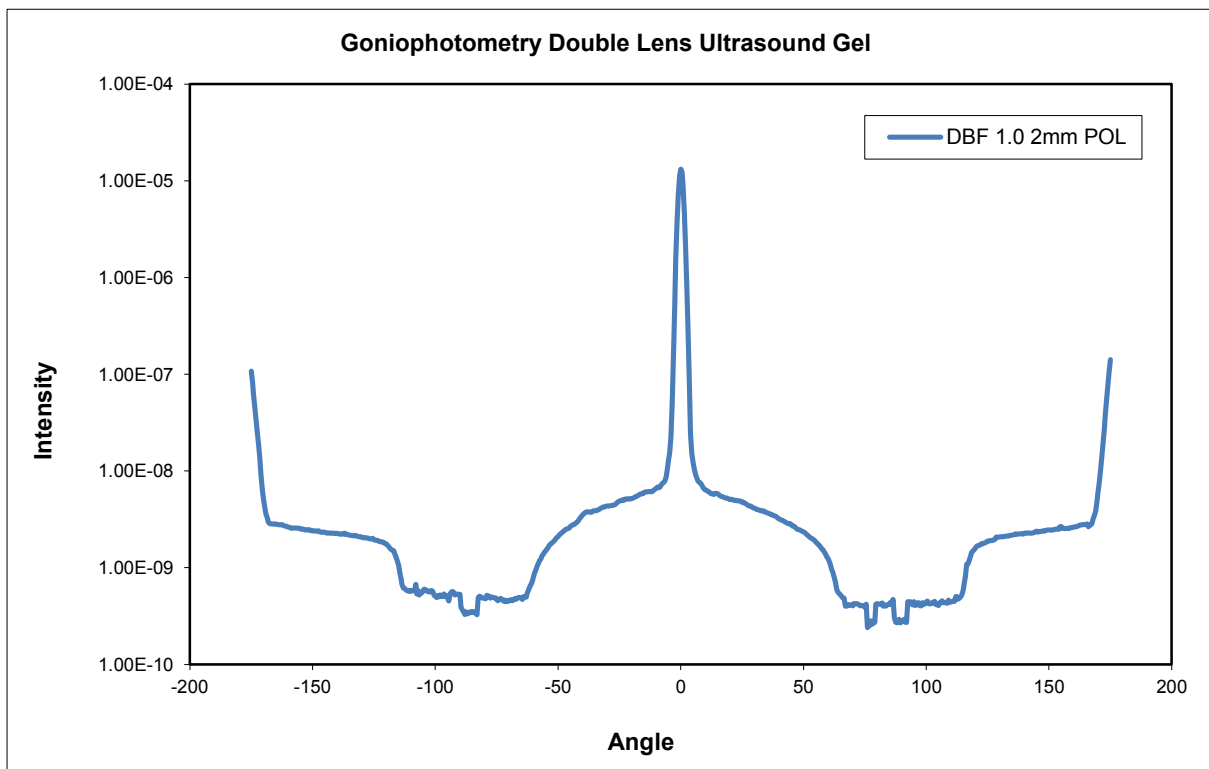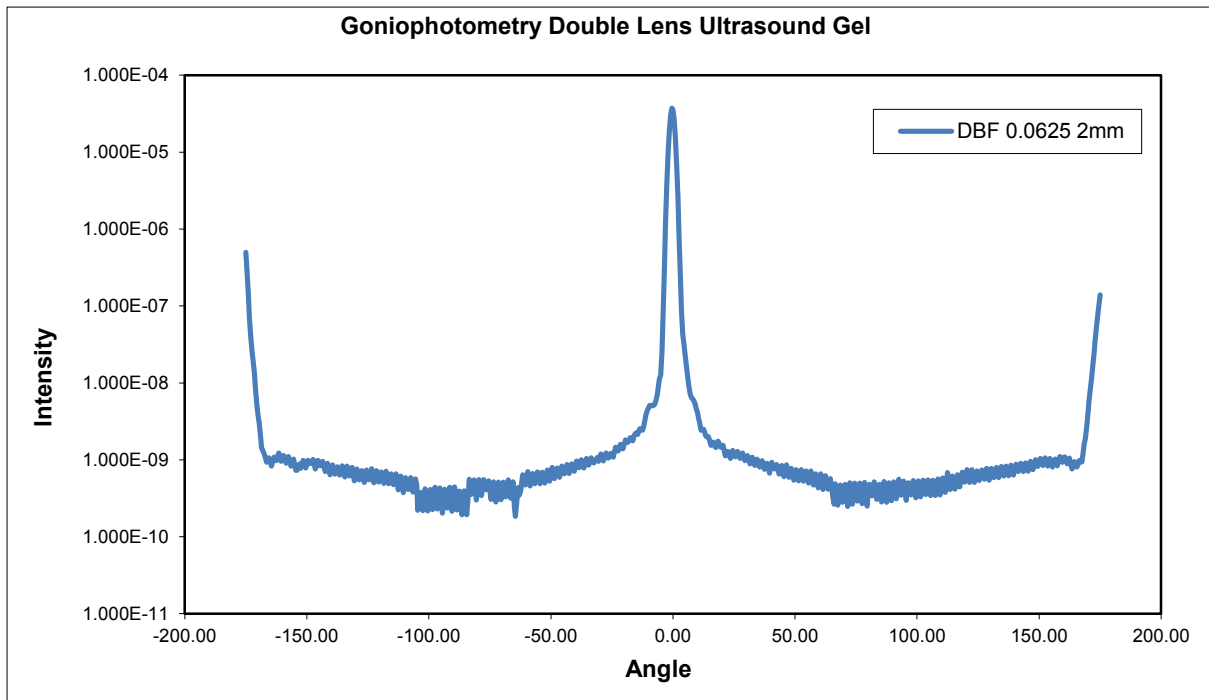
Figure 2.6: Measurements with double lens system before and after extra shielding was added. The noisy pattern (top graph) has been removed after the extra shielding.

The system was then used to conduct a pilot study to investigate the effect of sample thickness on the measurement of optical properties, specifically the anisotropy factor *g*. The aim of the study was to model the error in *g* values caused by measurements on multiple thicknesses of the same material, across a range of densities, so that it can be predicted and corrected for in future experiments. A secondary aim of the pilot study was to develop and test measurement and data processing protocols and identify limitations in our technique.

It was hypothesized that thicker samples will cause underestimation of the scattering function, as multiple scattering would lead to a more diffuse and isotropic scattering pattern, giving values of *g* that are lower than predicted (which in forward scattering materials, such as tissue, approaches 1). Values of *g* should progressively increase from thicker to thinner samples.

A phantom of 0.0625mg/g density was cut and polished into four slices, of 8, 4, 2 and 1mm thickness. Goniophotometry was carried out with the system described above, using the double lens system and ultrasound gel as the index-matching medium. The Henyey-Greenstein function was fitted to the scattering curves to give values of *g* for each thickness. The equation is described below:

$$p(\theta) = \frac{(1-g^2)}{(1+g^2-2g\cos\theta)^{\frac{3}{2}}} \qquad (1)$$

The useful data from each curve was in the range of –60 to 60 degrees (because of the lens effect), so that section was used for phase function fitting. The data was processed using Microsoft Excel software (© Microsoft Corporation, USA). Correction for lens effects was applied and the two sides of the curve were averaged to minimise error. As the refractive index of resin is close to glass, the angle deviation was not large and the corrected curve did not look significantly different. The steps for the 8mm slice are shown in Figure 2.7.

Whilst fitting the Henyey-Greenstein equation it became clear that the fit and resulting value of *g* was heavily dependent on the area of the curve used for the fitting. The difference in fitting the Henyey–Greenstein function to different regions of the scattering curve is shown in Figure 2.8.
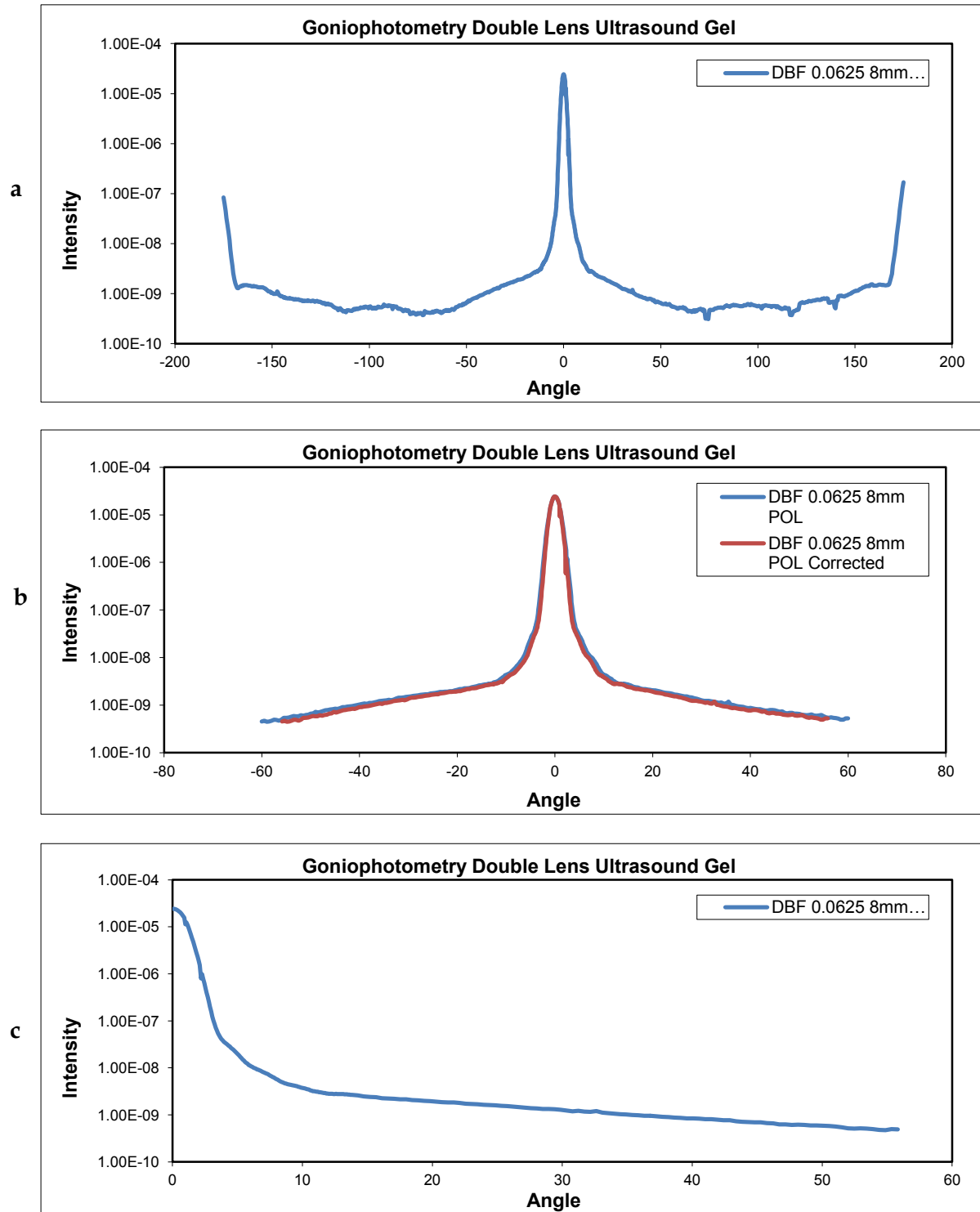
Figure 2.7: Processing of goniophotometry data - generation of corrected average curve. a) Original curve b) Selected area of interest between –60º and 60º with corrected angles superimposed (red) c) Averaged area of interest.
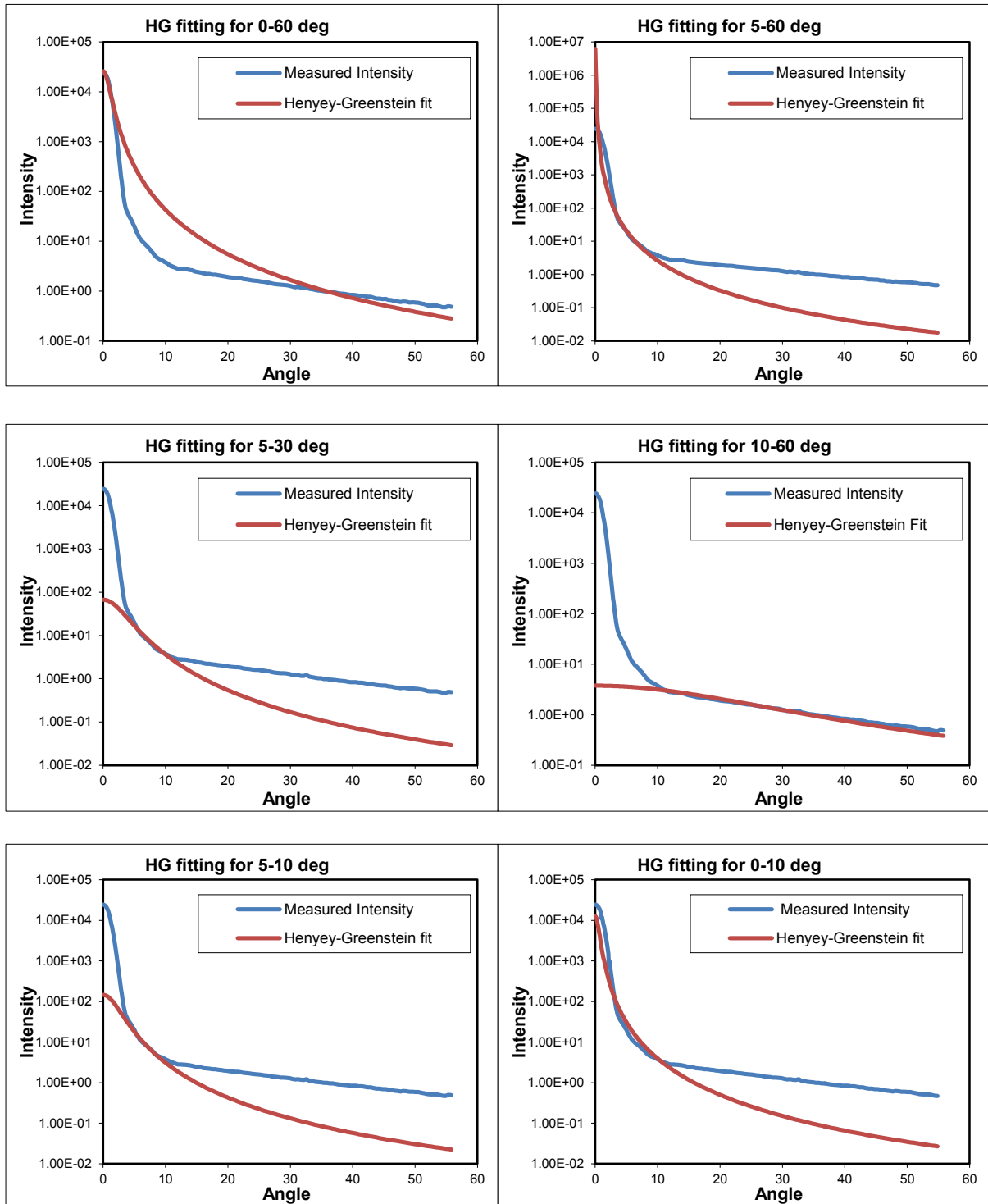
Figure 2.8: Fitting of Henyey-Greenstein function to different regions of interest in the scatter curve. From top left to bottom right: 0-60 degrees, 5-60 degrees, 5-30 degrees, 10-60 degrees, 5-10 degrees and 0-10 degrees.

2.1.4 Results

The measured values for *g* for each phantom thickness are shown below.

Thickness (mm)	g values per curve area fitted (degrees)							
	0-10	0-60	5-10	5-30	5-60	10-30	10-60	30-60
8	0.988	0.979	0.95	0.931	0.998	0.613	0.613	0.544
4	0.99	0.989	0.953	0.926	0.994	0.501	0.478	0.465
2	0.985	0.985	0.994	0.998	0.998	0.51	0.395	0.256
1	0.978	0.986	0.991	0.999	0.999	0.485	0.409	0.249

Table 2.1: Measured anisotropy values (*g*) for different thicknesses of a 0.0625mg/g phantom. The values obtained from fitting the Henyey-Greenstein function to different areas of the scattering curve are shown.

2.1.5 Discussion

Different values are given when fitting different areas of the curve to the Henyey-Greenstein function as is clearly shown. When fitting near the axis, *g* ranges from 0.926 to 0.999, whereas towards the periphery the range is 0.249 to 0.613. From the results it is clear that the pilot study did not confirm our hypothesis. There was no ascending trend of *g* values from thicker to thinner samples; in fact the opposite was shown when looking at *g* values from fitting at the periphery. As expected, there was significant variation in results when fitting the Henyey-Greenstein function to different parts of the curve, which emphasizes the importance of developing a concise data processing protocol, where the best area from the scattering curve in terms of quality and consistency of results is chosen.

A number of reasons could be to blame for the results of this pilot. The phantom used in this study was the clearest one in our range, which would make any scattering hard to pick up, let alone the difference between thicker and thinner slices. In fact, it was almost as clear as the lenses used for the mounting system and its possible that the change in distance between the front and back lens might have had a bigger effect on the scattering curves than the change in optical thickness of the sample. Measurements with a single lens system on denser samples could be used to test this theory. Another reason could be that our data processing methodology is flawed. It is possible that we are not applying the Henyey-Greenstein function correctly and/or that a different model might be more suitable, either for all phantoms or for this particular density. It may be that we have reached the single scattering regime with this phantom density and simpler equations for the scattering function are more

suitable here. The next step is to modify the goniophotometer and improve our optical property extraction technique, before carrying out a larger study with a range of phantoms.

2.2 Goniophotometry of microsphere phantoms

2.2.1 Background

In the previous section the process of building a goniophotometer and preliminary testing on epoxy phantoms was described. In this section we aim to improve upon the previous design and on the methodology used to extract the anisotropy factor from the raw data.

Furthermore, a different type of phantom is used in this section, consisting of microsphere suspensions, which better resemble human tissues and whose optical properties can be calculated and controlled more accurately. Two sets of microsphere phantoms were used in this study, each consisting of batches of different microsphere sizes. To get a useful range of phantoms, the scattering coefficients, attenuation coefficients and g values need to be comparable to those of tissues and ideally cover a range from below to above the properties of real tissue.

From the literature, the scattering coefficient of oesophageal and gastric tissue ranges from 20.5mm^{-1} to 12.2 mm^{-1} [42]. The stock suspensions were already significantly above that, so a number of dilutions were carried out to produce a series of phantoms with scattering coefficients ranging from above those of tissue to very small values, which, at 1mm path length, would be in the single scattering regime. The latter is important in our experiments as, in theory, the anisotropy coefficient is determined from the size of the scatterers and not affected by their concentration. In experiments, however, the concentration affects the type of scattering (i.e. single vs multiple), therefore affecting measurements of anisotropy, giving an underestimation of the true value of g at large optical thicknesses, with results improving at lower concentrations [44, 59, 79, 80].

2.2.2 Aims

To improve on our previous goniophotometer design in order to reduce errors in data collection.

To improve on our data processing methodology in order to measure g more accurately and consistently.

To test our new system and methodology on a new set of phantoms, which more closely resemble the properties of tissues.

2.2.3 Methods

Improvements were carried out on the goniophotometer, which included the installation of a more sensitive, fully automatic detector and amplifier console, a beam trap and improved shielding to reduce background light levels. A beam trap is essentially an opening in the goniophotometer wall, directly opposite the laser source, which diverts the laser beam into a closed tube, away from the area of measurements, when the detector is not at the centre. This prevents light bouncing back off the goniophotometer wall and affecting measurements.

The sample mount was also improved. The back lens was removed, as tissues are mainly forward scattering, which simplified the mounting process and removed a lot of noise from measurements. Furthermore, the mount was placed on rotation and translation platforms, which allowed for precise alignment of the sample to the laser beam and detector (Figure 2.10). The laser source remained the same, but an aperture was placed in front of it, in order to reduce beam divergence and decrease the spot size on the sample and to minimise background light levels. The new design, shown in Figure 2.9, works faster, has lower noise levels and higher sensitivity than before (Figure 2.11).

The new goniophotometer design was tested on two sets of phantoms consisting of aqueous suspensions of polystyrene microspheres. These are commonly used in the literature for equipment testing and calibration, as their optical properties can be calculated using Mie theory. Values for g and μ_s can be given for different sphere materials, sizes and concentrations. The phantoms are made from polystyrene microspheres in pure water, with 0.1% SDS surfactant and 0.05% NaN_3 antimicrobial solution (Bangs Laboratories, IN, USA).
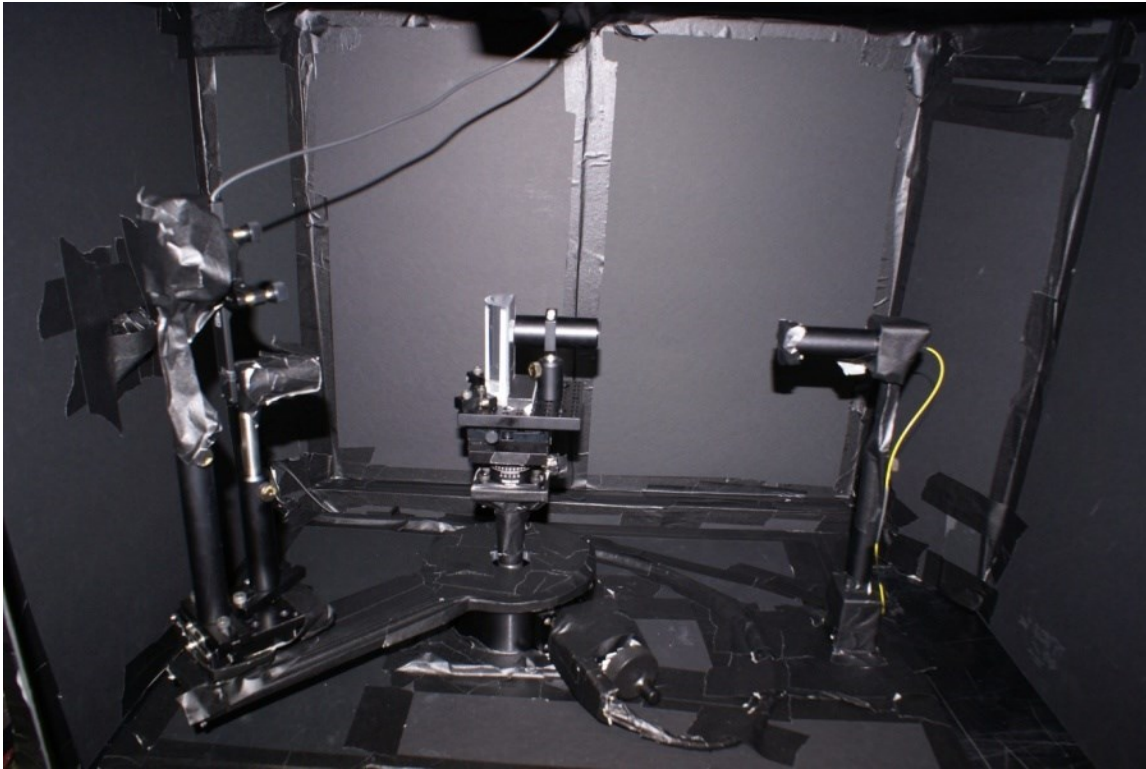
Figure 2.9: View of the goniophotometer with the front cover of the enclosure open. The detector, sample mount and laser source can be seen from left to right.

The first set of microsphere phantoms consisted of spheres with diameters of 0.8µm and 1.92µm. A dilution series was made to produce phantoms of different concentrations and hence scattering coefficients (see Appendix 6.2.3). The microspheres arrived as suspensions at defined concentrations. After vigorous agitation to homogenise the suspensions, pre-calculated volumes were aspirated using a pipette and added to pure water, aiming for a final volume of 2ml for each phantom. This was to allow for repeat measurements at each concentration, as the cell (described below) used for measurements could hold approximately 400µL. As a rule of thumb, the minimum starting volume drawn from each suspension was never below 200µL, to stay well within the accuracy of the pipette and thus minimise errors. Therefore, at certain key concentrations multiple batches were produced, to allow further sampling for the most dilute phantoms, instead of attempting to aspirate tiny volumes from the original dense suspensions.
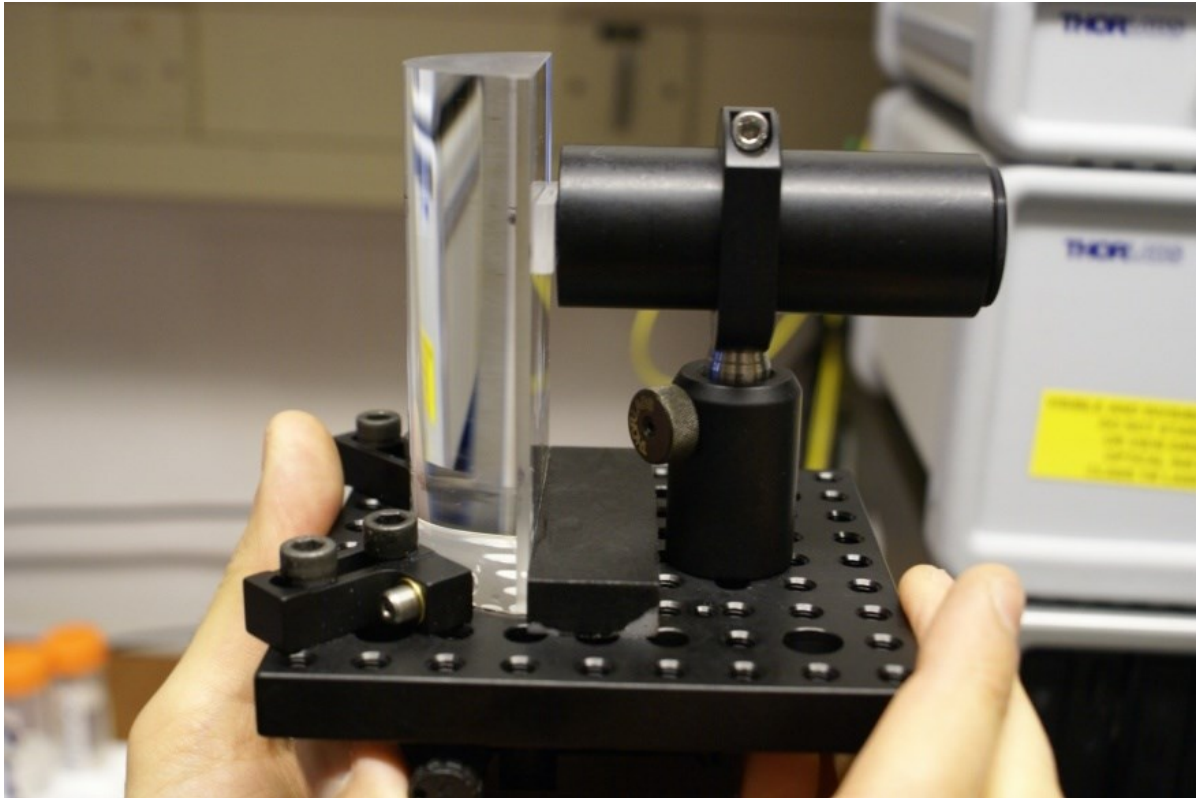
Figure 2.10: A close-up of the mounting system. The sample (in this case, the cell used for aqueous sample measurements, described later) is held against the front lens by a hollow tube, through which the laser beam passes.

The predicted g values for the dilutions were calculated using Scott Prahl's online Mie calculator (http://omlc.ogi.edu/calc/mie_calc.html), with the refractive index of polystyrene being defined as 1.57 (manufacturer's information) and the refractive index of water as 1.32235 [81]. The microsphere suspensions were placed inside a 1mm thick glass cell (Starna, UK) and glycerol was used as the index-matching agent between the cell and the lens. The cell is shown in Figure 2.11. After each measurement the cell contents were discarded and the cell was rinsed and dried to avoid contamination of each sample from the previous ones.

Average light intensity measurements (from 10 readings at a time) were taken from -90 to 90 degrees, in steps of 0.5 degrees. The data was processed in the same way as that of the epoxy resin phantoms (described in Section 2.1.3) and the Henyey-Greenstein function was fitted to them to give values of g. The results are laid out in Section 2.2.4.
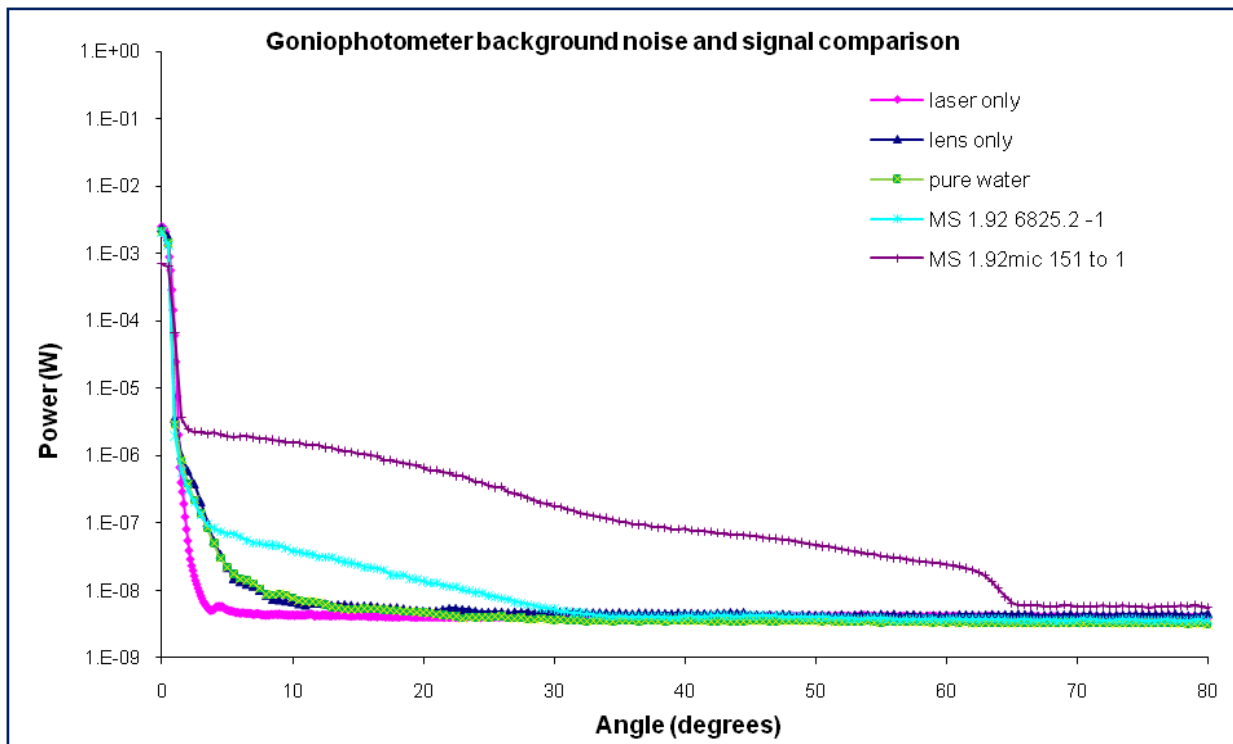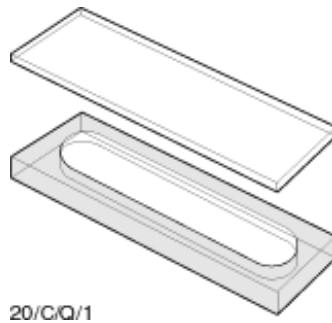
Figure 2.11: Top: the 1mm-thick cell used for goniophotometric measurements of liquids. Bottom: comparison of background noise levels and measurements from two different concentrations of aqueous suspensions of 1.92 μm diameter microspheres.

The second set of phantoms was designed to test the goniometer on wider range of anisotropy factors, namely aiming for g = 0.6, 0.7, 0.8 and 0.9. The closest available from the manufacturer's catalogue (Bangs Laboratories) were chosen, which were 0.58μm, 0.69μm, 0.96μm and 1.73μm microspheres, with predicted g values of 0.597, 0.704, 0.792 and 0.902 respectively.

A dilution series was created for each microsphere size, with an improvement on the previous methodology being the preparation of phantoms at set optical thicknesses, in order to make comparisons easier. As mentioned previously, the optical thickness is the product of

the sample thickness and the total attenuation coefficient, so in this case it is equal to the attenuation coefficient, as the cell used for the measurements has a 1mm path length. To create a suitable range of optical thicknesses a 10micron slice of tissue with a scattering coefficient of 12mm^{-1} was defined as the gold standard. The optical thickness of such a slice of tissue would be 0.01mm x 12 = 0.12 and our range of phantoms had optical thicknesses above and below that range.

Goniometric measurements were carried out in the same way as with the first set of microsphere phantoms, producing .csv files. However, the data this time was imported into and processed using Matlab (Mathworks, MA, USA). The code used can be found in the Appendix 6.2.1 and 6.2.2.

Each measurement curve was first checked for correct centering and if not, corrected and cropped so that the maximum value was at 0 degrees. This was done to correct for any misalignment caused by mounting and un-mounting of specimens which would cause the maximum intensity of light to not fall exactly at "point zero", i.e. where the detector is exactly opposite the laser source. This is not generally a problem in terms of the shape of the data, as that would remain the same, just shifted to the negative or positive side. However, when subtracting the background and averaging the two sides of the curve, it would introduce error. To prevent this, for each dataset the maximum value was detected and if it was not at 0 degrees, the edges of the dataset were cropped accordingly, to bring the maximum to the midline. As the data at the edges was not used for fitting, this did not affect the final results. The data was then averaged and the background noise removed. Data from goniometry of a pure water sample was used as the background. The data was then smoothed using a moving average filter and angle correction was applied to compensate for the effect of the convex lens used as a mount. The process is illustrated in Figure 2.12.
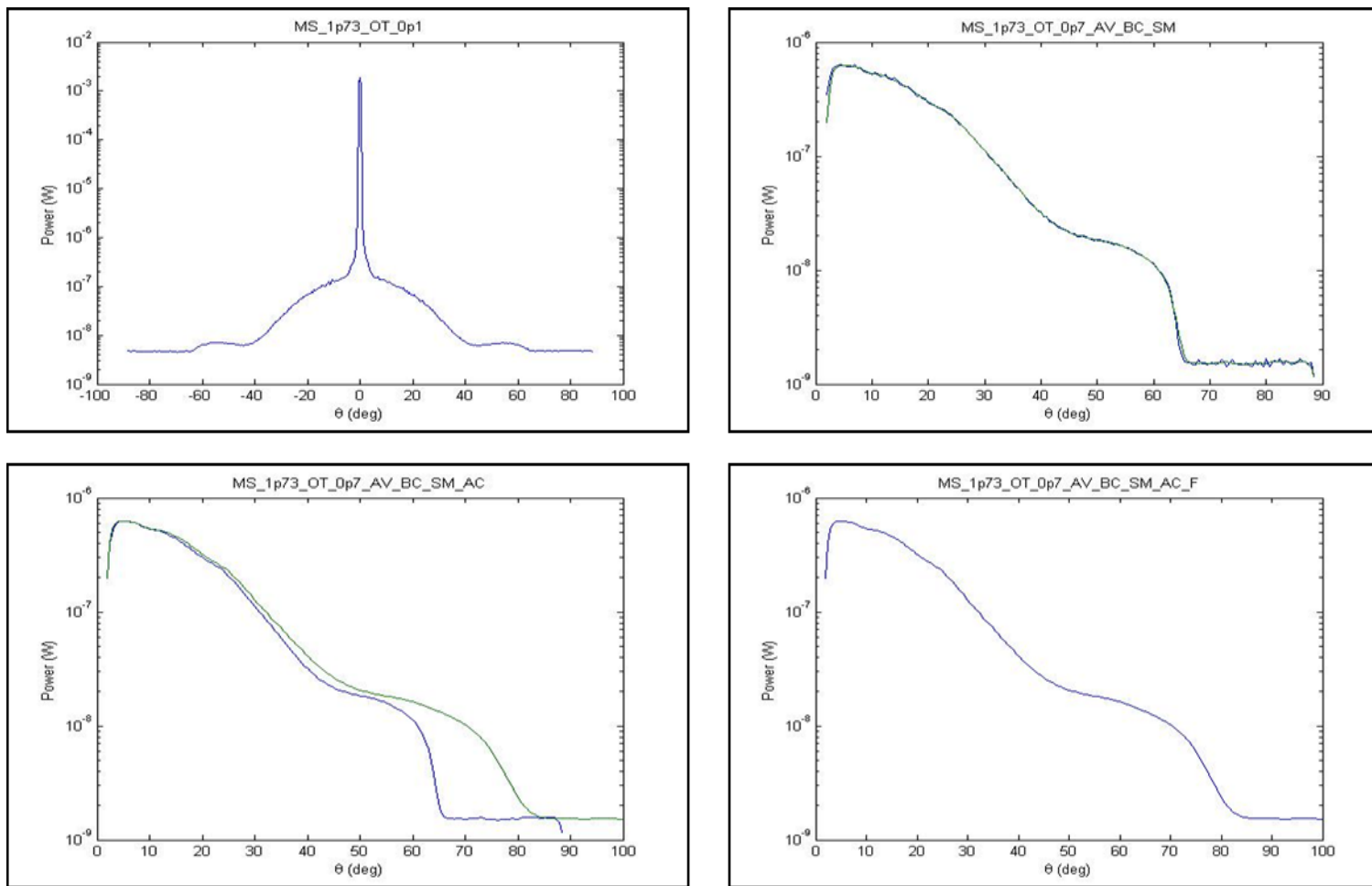
Figure 2.12: Processing of goniometry data. From top left to bottom right: original curve, averaged and smoothed curve (with background subtraction), angle correction and final curve.

Finally, the data was transformed into linear form and fitting of the Henyey-Greenstein function was carried out using a scaling factor K to account for differences in the normalisation of the curves.

Taking the Henyey-Greenstein function with the added scaling factor K as shown below:

$$p(\theta) = \frac{K(1-g^2)}{(1+g^2-2g\cos\theta)^{\frac{3}{2}}} \qquad (2)$$

If $x = \cos\theta$ and $q = p(\theta)^{-\frac{2}{3}}$ then plotting q against x gives a straight line, where the gradient (m) and intercept (c) are defined as follows:

$$m = -2g\left(K(1-g^2)\right)^{-\frac{2}{3}} \qquad (3)$$

$$c = \left(K(1-g^2)\right)^{-\frac{2}{3}}(1+g^2) \qquad (4)$$

By fitting the goniometry data in its linear form to the linear version of the HG function, the scaling factor K and g can be worked out:

$$K = \frac{\left(\sqrt{2}\sqrt{\frac{m\left(-c+\sqrt{c^2-m^2}\right)}{-c^2+m^2}}\right)}{m^{\frac{3}{2}}} \qquad (5)$$

$$g = \frac{\left(-c+\sqrt{c^2-m^2}\right)}{m} \qquad (6)$$

The process of linear transformation and fitting is shown in Figure 2.13.
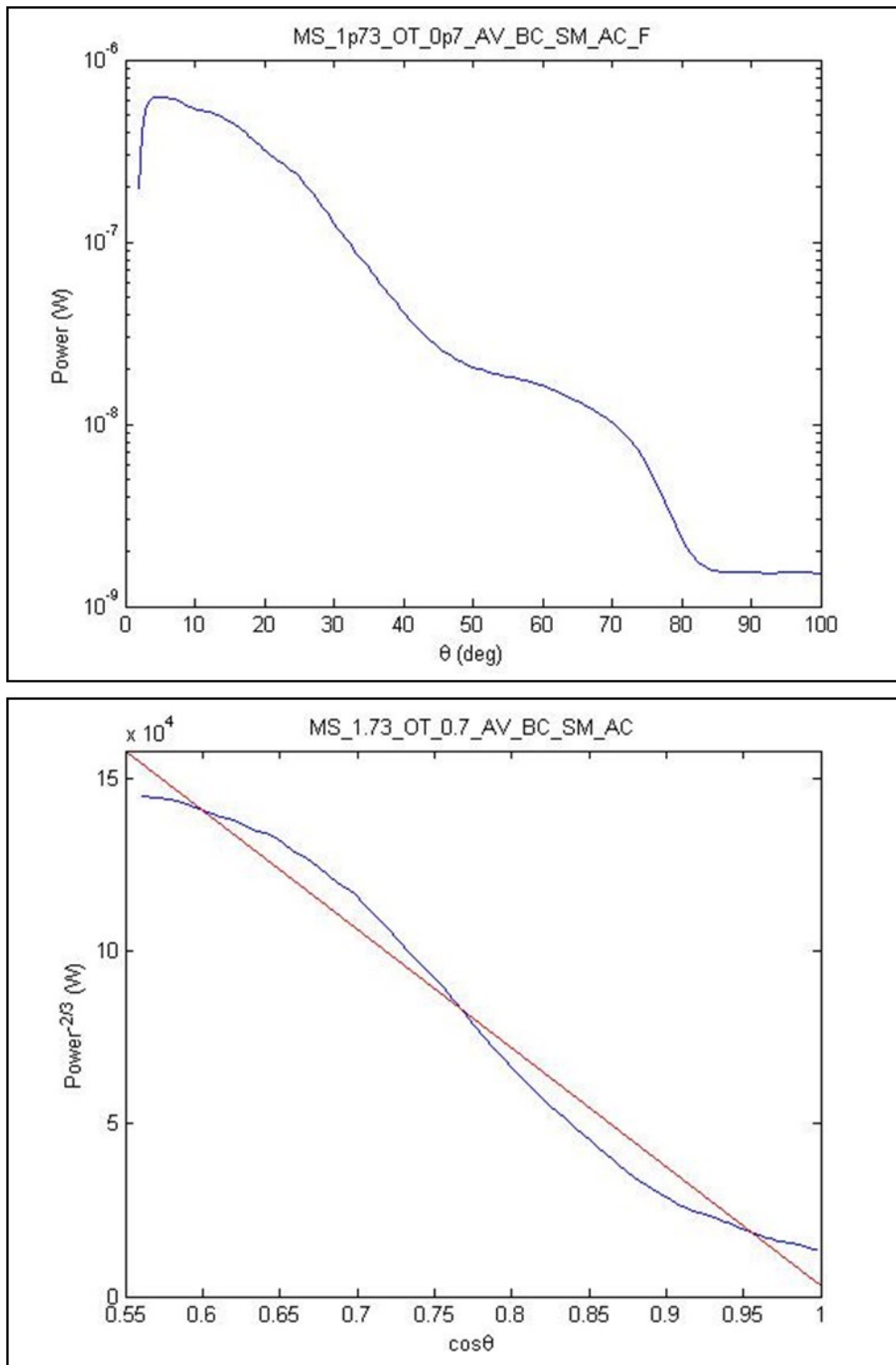
Figure 2.13: Linear transformation of goniometric measurement. The original curve (top) is linearly transformed and fitted with a linear Henyey-Greenstein function (bottom).

2.2.4 Results

Figure 2.14 depicts averaged measurement curves for the first set of phantoms at different concentrations (expressed as dilution ratios to the original stock solution). The change in measured values of *g* against microsphere concentration for the first set of phantoms is laid out in Table 2.2 and illustrated in Figure 2.15.
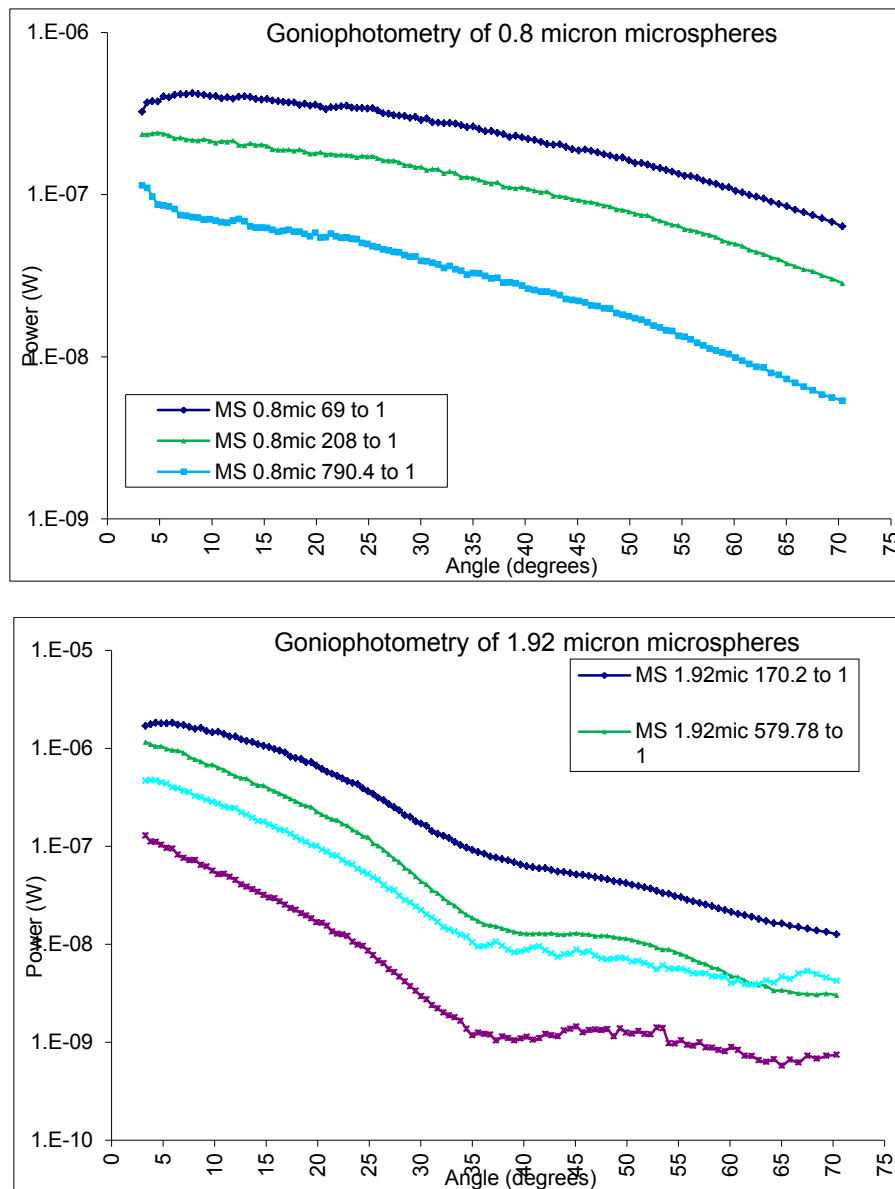
Figure 2.14: Goniophotometric measurements for selected concentrations of 0.8 μm and 1.92 μm microspheres are shown.

0.8 μm spheres		1.92 μm spheres	
Optical Thickness	g (pred: 0.734)	Optical Thickness	g (pred: 0.914)
0.098	0.554	0.024	0.822
0.297	0.505	0.033	0.830
0.375	0.453	0.049	0.823
1.131	0.429	0.099	0.792
3.004	0.383	0.292	0.796
3.706	0.375	0.995	0.747
5.657	0.358	1.127	0.737
7.413	0.320	3.322	0.654
11.14	0.313	7.882	0.423
19.51	n/a	11.315	0.378
22.28	n/a	16.612	0.335
		22.63	0.329

Table 2.2: Optical thickness and measured *g* values of 0.8 μm and 1.92 μm microsphere phantoms.
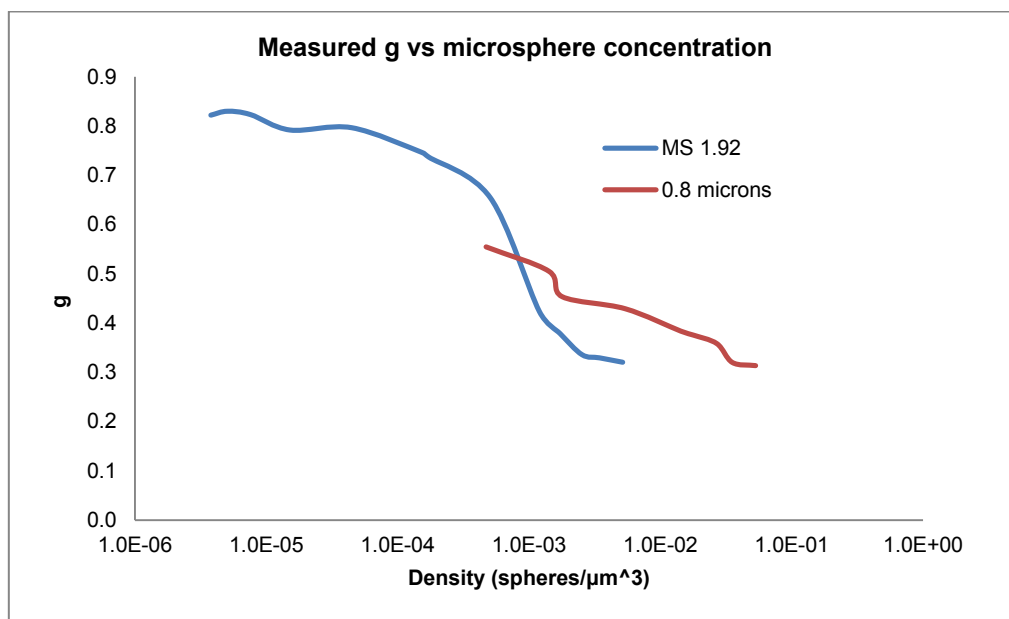
Figure 2.15: A graph of measured values of *g* vs. microsphere concentration for of 0.8 μm and 1.92 μm microspheres.

The measured values of *g* against microsphere concentration for the second set of phantoms are shown in Table 2.3 and illustrated in Figure 2.16 and Figure 2.17.

Optical thickness	Measured g			
	0.58 µm	0.69 µm	0.96 µm	1.73 µm
Predicted g	0.597	0.704	0.792	0.902
0.05	0.295	0.373	0.554	0.901
0.1	0.303	0.425	0.571	0.903
0.3	0.329	0.390	n/a	n/a
0.5	0.324	0.382	0.544	n/a
0.7	0.324	0.379	n/a	0.880
1	0.324	0.377	0.530	0.819
3	0.351	0.361	0.436	0.623
5	0.361	0.353	n/a	0.490
7	0.384	0.356	0.361	n/a
10	0.428	0.385	0.364	0.370
20	0.488	0.438	0.403	0.356
30	0.540	0.483	0.445	n/a
40	n/a	n/a	0.466	n/a

Table 2.3: Measured values of g vs. optical thickness for 0.58 µm, 0.69 µm, 0.96 µm and 1.73 µm microspheres.
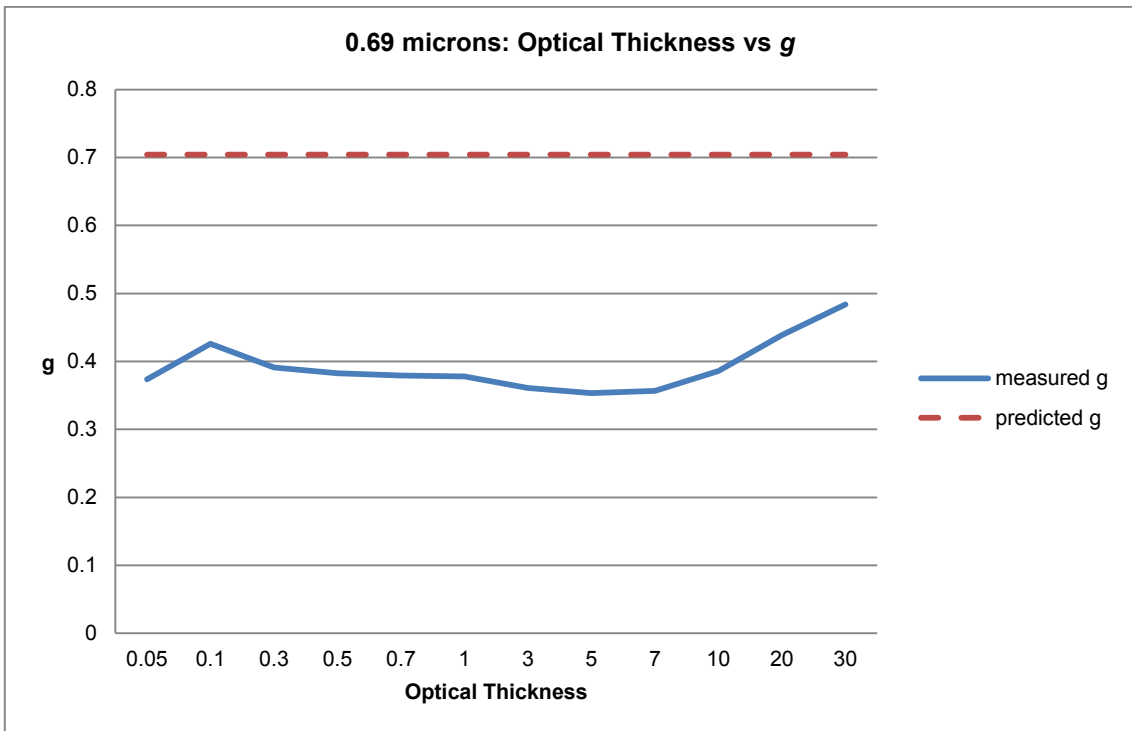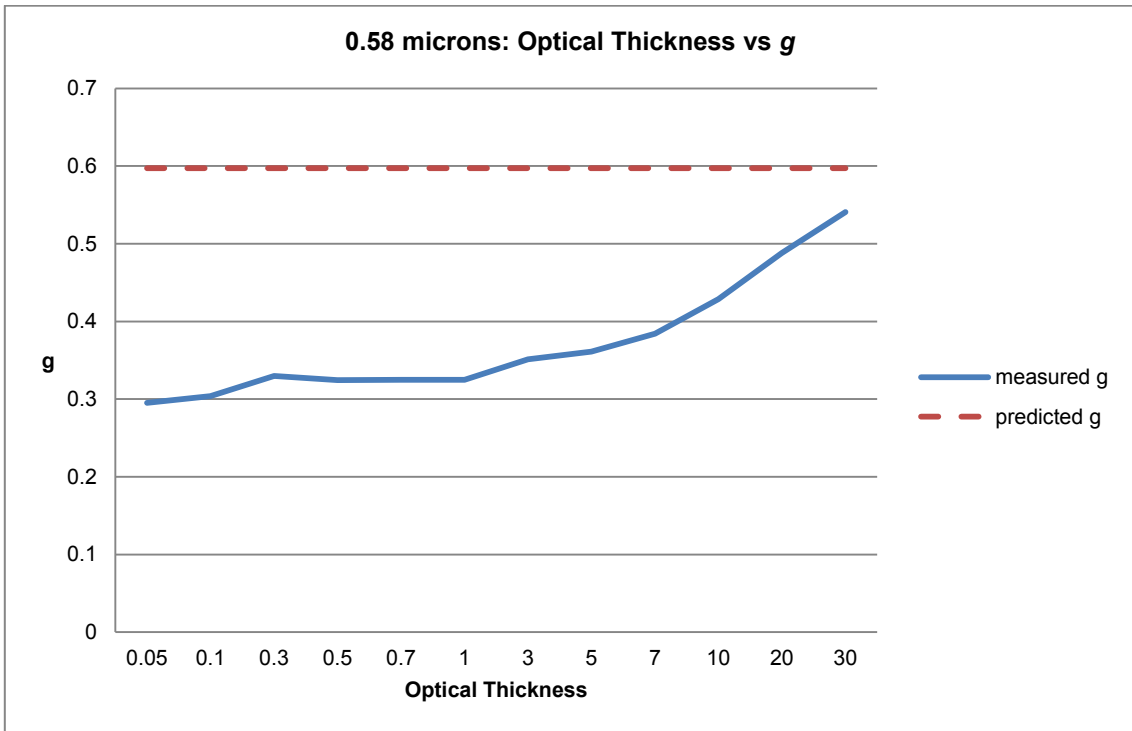
Figure 2.16: Measured *g* vs optical thickness for 0.58 µm and 0.69 µm microspheres.
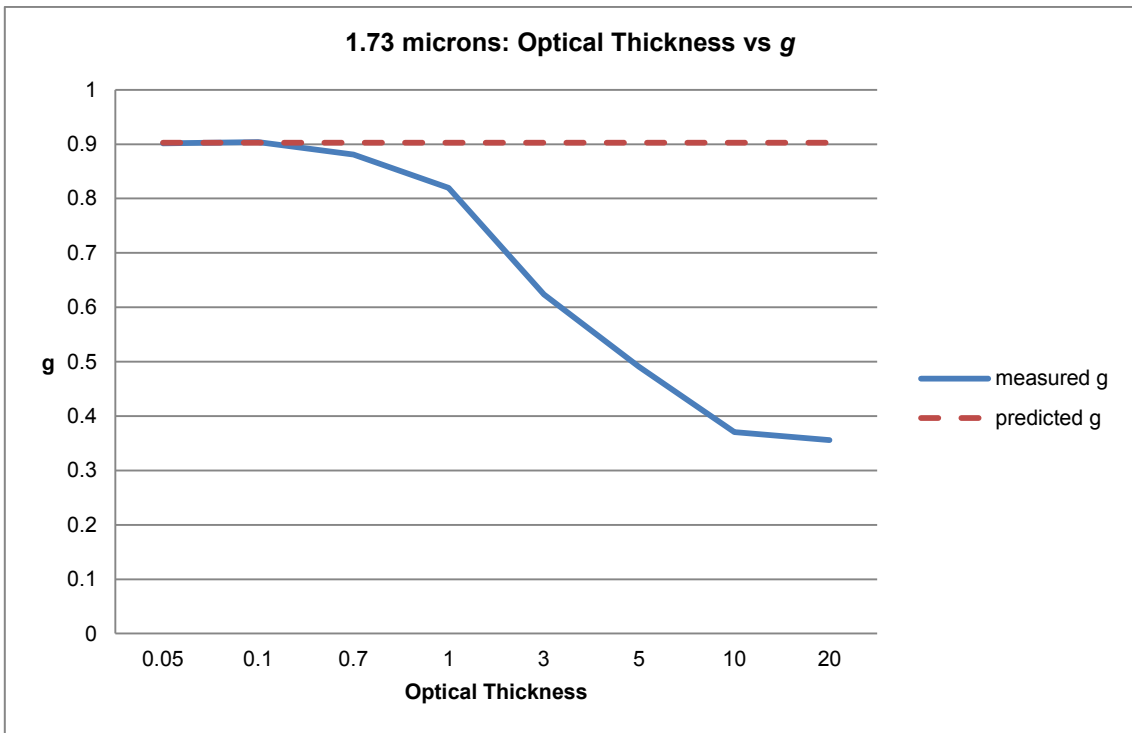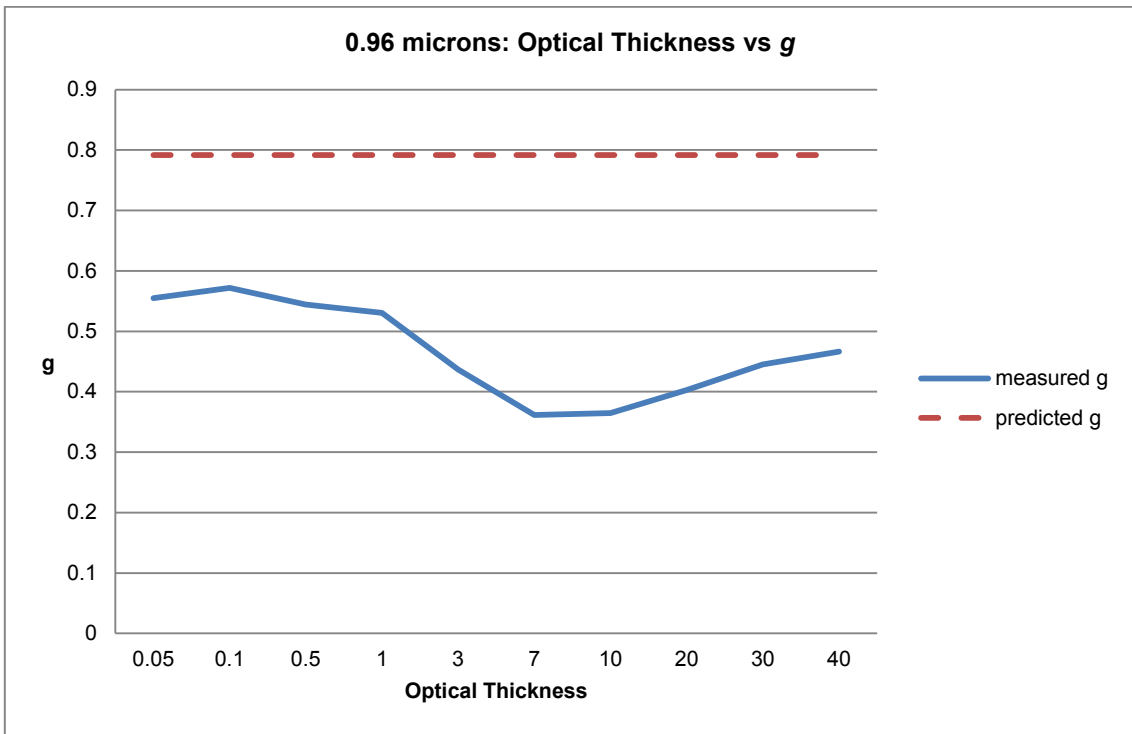
Figure 2.17: Measured *g* vs optical thickness for 0.96 μm and 1.73 μm microspheres.

2.2.5 Discussion

The results from microsphere phantoms are a great improvement on the measurements carried out on the epoxy resin phantoms with the previous goniophotometer setup. Starting with the first set of microspheres, the *g* values obtained for the two sizes (0.83 for 1.92 µm and 0.55 for 0.8 µm) are closer to those predicted by Mie theory (0.914 and 0.734 for 1.92µm and 0.8µm respectively) at low concentrations. Additionally, the trend of *g* with sample concentration (or optical thickness) follows the expected pattern, with measured values of *g* being higher and closer to the predicted ones as sample concentration decreases and single scattering dominates. A plateau is reached at both ends of the spectrum, where single or multiple scattering dominate.

This is also seen with the 1.73µm spheres in the second set, but not with the other sizes. In the latter not only is the trend not as expected with decreasing concentrations, but the values of *g* are also far from the predicted ones. Only the 0.96µm spheres are slightly better in terms of the trend, but the measured *g* (0.55) is not close to the predicted one (0.79).

Reasons for our results could include limitations in our goniometer design, data extraction methodology or phantom choice and preparation. When compared to the literature, our goniophotometer design has some similarities and some differences. The first one is the wavelength of choice. Since optical properties, including *g* are wavelength dependent [44], we chose to match the wavelength of the OCT system as closely as possible (1300nm), so that measurements from the two systems could be comparable. In the literature the wavelengths tend to be much shorter ranging from 450nm to 660nm [31, 44, 60, 61, 64, 65, 80, 82].

Our mounting system also differs from what described before: as we would be working with both very thin tissue slices and liquid-based phantoms we opted for a design that could hold both (with minor modifications) and that did not require major accuracy in terms of alignment. For the latter reason we chose to only use a front lens which was also justified by the fact that tissues are forward scattering [44, 59, 61], meaning we would not lose significant information by only measuring scattered light in the -90º to 90º range [31, 79]. We also avoided the use of a water tank, as it increases complexity but can also impede measurements due to the absorption peaks of water at infrared wavelengths [60]. In accordance with the literature [31, 60, 80], we used a cuvette for liquid samples and a glass slide for holding tissue samples (described in section 2.3).

Microsphere phantoms are commonly used in the literature for goniometer calibration [31, 36, 60, 64, 65]. As mentioned previously, g depends on the sphere sizes and for that reason a range of sphere sizes was chosen in order to test the goniometers capability to measure and distinguish between different values of g. This, however, was not seen in the literature, where the authors tended to pick one or two sizes, which were similar to or bigger than the wavelength of the light source used [31, 60]. After all, the angular scattering phase function and, therefore, the anisotropy coefficient g are strongly dependent on the size of scatterers relative to the wavelength of incident light [44]. It seems that, in order to obtain good results, the sphere size should be similar or greater than the wavelength, which could explain the disappointing results for our smaller microsphere sizes. In our experiments we agitated the phantoms before use, whereas some authors made them fresh each time [31]. We did not initially expect this to have a great effect on the final results, however, it could be that the smaller microspheres clumped together despite agitation before measurements, creating particles of larger sizes and leading to inaccuracies in our measurements. This would fit with the pattern seen in Table 2.3, where at low concentrations there is an increasing trend in the measured values of g with increasing microsphere diameter for the same optical thickness, as predicted by Mie theory, whereas at higher concentrations, the smaller microspheres actually yield higher values of g than the large ones.

Most studies in the literature use Mie theory to model light scattering from phantoms or tissues and then the Henyey-Greenstein function to calculate g from their scattering data [31, 59-61, 64]. They do not all clearly describe how they carry out the fitting [60] and some use variations, such as fitting the HG function separately to the forward and backward scattering components [61, 65]. In general though, they tend to focus more on the forward scattered light and ignore the very first 1-2 degrees which comprise mainly of unscattered light, which is passing straight through the sample [31, 79]. We have applied the same principles to our data processing and calculation of g from our phantoms. Polarization is also important and authors do mention that using polarized light yields more accurate results with Mie theory [31]. In the case of our system our light source –being a laser beam- is polarised, but we could experiment with different polarisation settings for the detector.

In general, experimental values are expected to be less accurate than predicted ones, as the latter do not take into account factors such as experimental error, noise, equipment sensitivity limitations and particle size distributions in the microsphere suspensions. It is

also difficult to directly compare our results with the literature or indeed compare those of other studies, as there is a disparity in wavelengths, phantoms and goniophotometer designs. In the end, we have shown in this study that the goniometer can produce consistent measurements and distinguish between phantoms with different properties and of different concentrations. The next stage is to carry out a pilot study with human tissues and work on further improvements to our design and methodology.

2.3 Pilot study in goniophotometry of human tissue samples

2.3.1 Background

The goniometry work described in the previous sections focused on developing a functional goniophotometer that can be used to investigate the anisotropy factor of human tissues and identify differences between healthy and diseased tissue, particularly cancer. In this section we describe early measurements carried out on human tissue, including the methodology used to procure specimens, section them and measure their anisotropy factor.

2.3.2 Aims

To develop a protocol for procurement, storage and sectioning of human gastrointestinal tissue samples.

To measure the anisotropy coefficient g of normal gastrointestinal tissue and tumours.

2.3.3 Methods

The same patients and surgical specimens were used for this study as described in Sections 3.1.3 and 3.2.3. Tissue blocks were taken from specimens with an abundance of either healthy tissue and tumour in agreement with the histopathologists, who ensured the integrity of the main specimen was not affected and the ability to provide a histological diagnosis for the patient was not compromised. These blocks were taken fresh from specimens arriving from theatres and frozen in a -20°C freezer until they were sectioned for goniophotometry. The sections were approximately 1cm x 1cm x 1cm in size, large enough to obtain meaningful measurements, yet not too large for microtoming. They were cut while frozen by a senior lab technician using a standard cryomicrotome, the same used for frozen sections from operations.

The size of the specimens was much larger than standard frozen sections, so there was an element of trial and error in the cutting process. Slices of varying thicknesses were taken to check the quality of the sections, which was done by staining and examination under the microscope by the histopathologist. With slices above 20 microns, and until the cutting technique was further refined, an obvious "Venetian blinds" effect was present (alternating

thick and thin tissue zones) and there were tears in the specimen, both caused by the frozen block being too hard to cut smoothly. It was decided to obtain slices at 8, 12 and 15 microns for all further specimens.

The tissue sections were mounted on glass slides. Glycerol was placed on both sides of each sections, to reduce the refractive index variation between the tissue, glass slide and coverslip. As this would hinder the fixing and staining process necessary for histological reporting on the sections, an adjacent section was obtained for each specimen, which was stained and used for reporting.

In total 4 specimens from 4 different patients were sectioned, of which 2 were healthy tissue and 2 tumour. As one of the tumour samples was larger, it was split in half and used as 2 separate specimens. The same methodology described in Section 2.2.3 was used to obtain goniometric measurements and values of g.

As mentioned above, 2 adjacent sections were obtained for each specimen. The first remained unstained and was used for goniophotometric measurements, whereas the second was stained and processed using the standard histopathological techniques described in Section 1.4.2. The stained slides were then examined by a histopathologist to confirm whether the measured sample was tumour or normal GI tissue.

2.3.4 Results

Figures 2.18 to 2.20 show plots of goniometric measurements of healthy and cancerous tissue, at different slice thicknesses and the measured values of g after fitting of the Henyey-Greenstein function are summarised in Table 2.4.
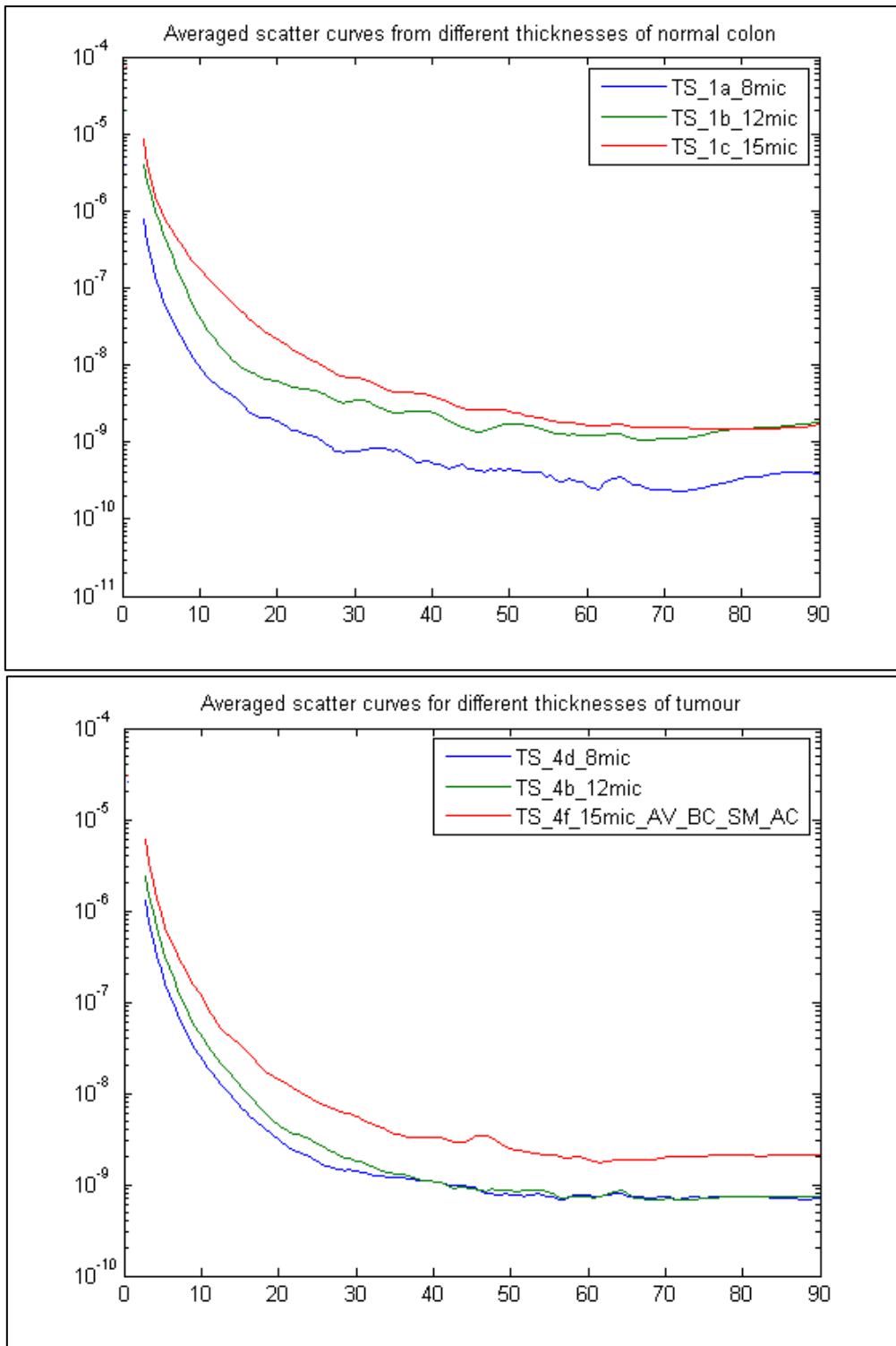
Figure 2.18: Averaged scatter curves from different thicknesses of normal colon and tumour.
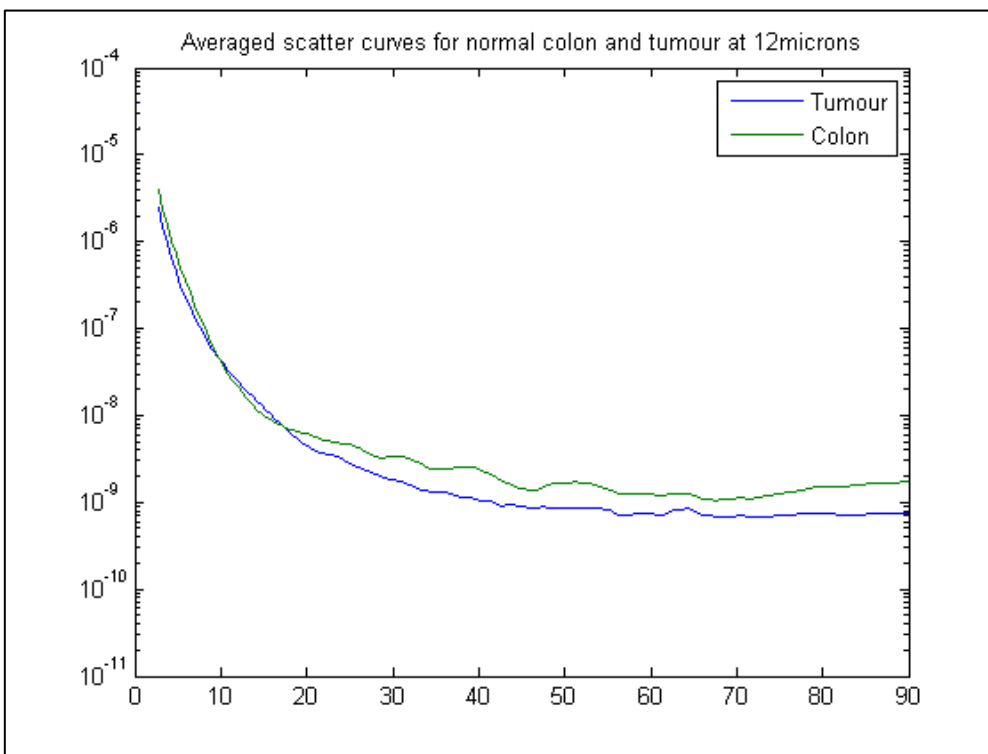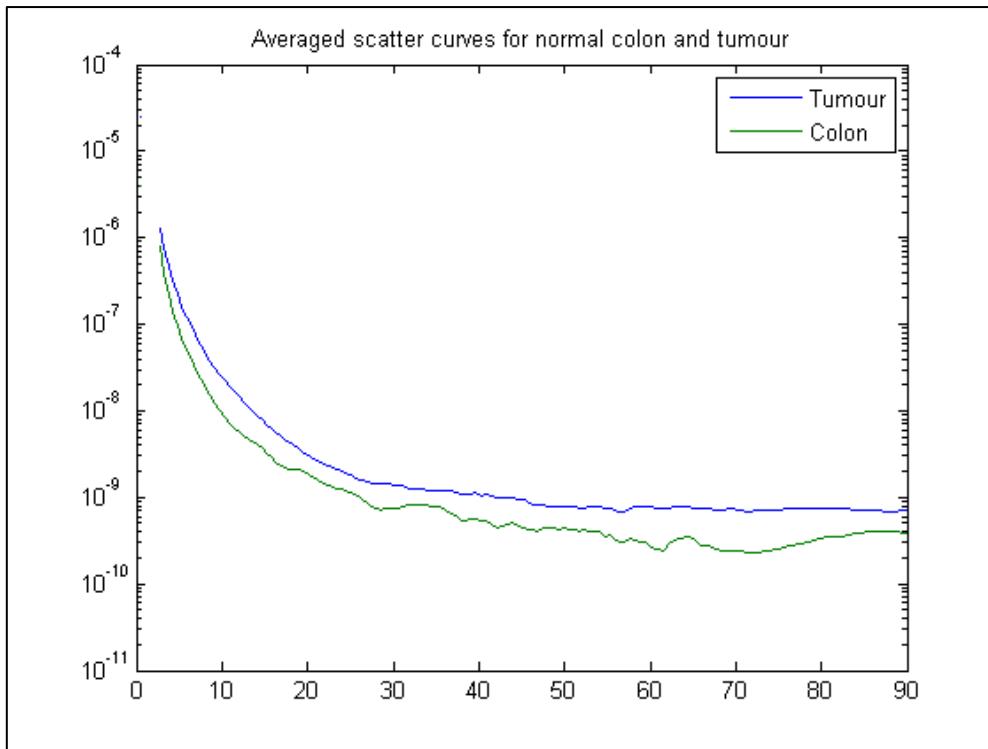
Figure 2.19: Averaged scatter curves for normal colon and tumour at 8 µm (top) and 12 µm (bottom).
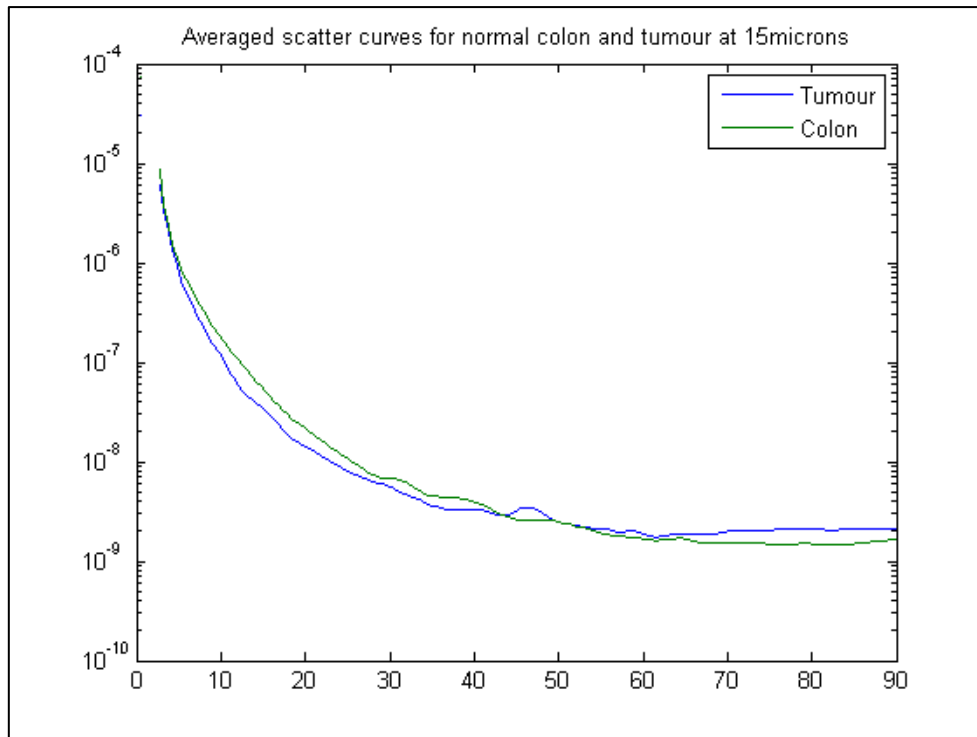
Figure 2.20: Averaged scatter curves for normal colon and tumour at 15 µm.

Thickness	Measured g				
	Normal Colon	**Normal Stomach**	**Tumour 1**	**Tumour 2a**	**Tumour 2b**
8 mic	0.664	0.708	0.769	0.659	0.655
12 mic	0.662	0.781	0.688	0.716	0.636
15 mic	0.797	0.750	0.714	0.610	0.714

Table 2.4: Measured g for normal colon, stomach and tumour at 8 µm, 12 µm and 15 µm.

2.3.5 Discussion

In general the results of this study are in keeping with our expectations and with what has been previously described in the literature. Our measurements show a strongly positive *g* value, in agreement with the range of *g* values reported in the literature (0.7 to 0.99) [44, 61, 65, 79]. No great difference in *g* between healthy and tumour tissue can be shown here, which can most probably be attributed to the low sample numbers used. Some authors have demonstrated differences in *g* between healthy and malignant tissues, such as Ghosh et al, who showed an increase in *g* from healthy to malignant breast tissue (0.88 and 0.96 respectively) [65], but others have shown no significant difference between healthy tissues and cancer (such as cervical tissue [35]) or different tissue types (such as uterus, lung and liver [61]).

A great strength of this study compared to the literature is the kind of tissues used and the way they were processed. Not many studies have carried out goniophotometric measurements on human tissues. Most have used animal tissues, such as pig, beef, chicken or rat [31, 44]. Some studies have used human tissues but used small sample numbers [59, 61, 65, 79] or processed them in a way that alters their structure and, consequently, their optical properties [80].

A variety of methods have been previously described for processing tissues before goniophotometric measurements. Generally authors tend to freeze tissues in order to preserve them until they are sectioned, some using liquid nitrogen [44] or carbon dioxide [79], but most using a standard laboratory freezer [31]. Not all studies mention the temperature at which the tissues were frozen and in some studies the tissues were not frozen straight away [65, 79], which could alter their properties. In this study the tissues were frozen immediately after excision. Some authors rehydrated their samples in saline before measurements [61, 79], but most did not. Furthermore, some studies use cadaveric tissue [31, 40, 80], which also has altered optical properties compared to fresh. In our study we used tissues that were frozen within 20 minutes of ischaemia.

We used a microtome to obtain tissue slices. In the literature authors have previously used microtomes [44, 59, 61] or dermatomes [79, 80], whereas some have simply used ground tissue [31]. Even though microtoming is probably the most accurate way to obtain thin tissue slices of predefined thickness, it is not without problems. Normally tissues are fixed in

paraffin (as described in section 1.4.2) which is relatively soft and easily cut by a microtome. In our case however, to preserve the optical properties we used unfixed tissues which were cut whilst frozen using the microtome reserved for frozen sections. Frozen tissue blocks are harder to cut and the ones we used are larger than the frozen sections normally sent for quick diagnosis from operating theatres. This meant that the wear and tear on microtome blades was increased and the quality of the slices produced was not as good, requiring a period of learning and adjustment. A common problem with early slices was the "Venetian blinds" effect, where multiple horizontal tears could be seen in the specimen, as a result of the blade slipping and tearing the tissue. This does not affect histological reporting, but would affect our measurements, so these slices were discarded. Furthermore, the tissue blocks do not naturally have smooth surfaces, so before slicing, they had to be "polished" by repeatedly running the microtome blade across the tissue block till an even surface was produced. Additionally, as the slices used for measurements were covered in glycerol, they could not be stained afterwards for histological diagnosis, so for each slice used for measurements an adjacent slice was also cut for histological diagnosis. As a result of the above –and the resulting increase in microtoming time and cost per tissue block- the number of slices produced for measurements per tissue block was further limited.

The size and thickness of our slices was in good agreement with the literature. Previous studies have also used slices of between 5 mm x 5 mm and 2 cm x 2 cm [44, 61, 65, 79]. Tissue thickness has been quite variable ranging from 20 μm to 120 μm [31, 44, 60, 61, 65, 79, 80]. At the larger end of the scale authors have acknowledged that multiple scattering has probably occurred [44], but generally technical difficulties prevent them from producing thinner slices. In general authors state that multiple scattering is not a problem if sample thickness is no greater than 3-7 times the MFP of the sample [31, 79].

Our study is also limited by a small sample number and it has highlighted the difficulties in obtaining large numbers of human tissue samples, particularly tumours. Samples for goniophotometry could only be taken from surgical specimens, as they could provide large enough slices. Biopsies are smaller than the diameter of the laser beam used in our setup. They could have been squeezed between glass slides to spread out, but this would have affected their structure and optical properties. To ensure the validity of measurements, the width of the samples needs to be significantly larger than the beam, so sections of around 1x1cm were aimed for.

The sections were mainly obtained from lower GI specimens for the following three reasons: firstly, the vast majority of upper GI specimens have small or even completely flat tumours, as it is standard practice in most cases to precede surgery with neoadjuvant chemotherapy, which reduces tumour size and spread, thus increasing the chances of a good surgical outcome. Therefore, in most cases it was impossible to obtain a reasonably sized sample without compromising the diagnostic quality of the specimen for histological reporting. Secondly, in tumours of the oesophagus and gastro-oesophageal junction (GOJ), it is important to preserve the structure of tissues around the level of the tumour, so that the pathologist can identify whether the circumferential margin is positive or not. Consequently, in some specimens the tumour could not be accessed. This was not a problem in lower GI specimens, other than rectum tumours, as there is no circumferential margin there. Finally, lower GI specimens tended to be much larger in general, providing more healthy tissue away from diagnostically important areas such as resection margins.

The number of sections in our study is comparable to the literature. Arnfield et al. used 9 tissue sections from 2 different rat tumours [44], whereas Jacques et al. used cadaveric human tissue samples from 4 subjects [80]. Gosh et al. used tissues from 10 patients undergoing radical mastectomies. They studied the differences between healthy tissues and ductal cell carcinoma using 10 paired healthy vs cancer tissue samples [65]. Finally, Treweek et al. used skin samples from 4 human subjects [79]. In fact, there are no goniometric measurements of human gastrointestinal tissue in the literature.

This pilot study has shown some promising results. Our measurements fall within the values described in the literature and we have developed a protocol for processing human specimens. Larger numbers and improvements to our setup are now needed to produce some statistically significant results.

3 OPTICAL COHERENCE TOMOGRAPHY

3.1 Development of a specimen collection protocol and preliminary analysis of Optical Coherence Tomography data

3.1.1 Background

As described previously, a complicated process is involved in preparing specimens for histological reporting. Detailed guidelines must be adhered to in order to carry out a comprehensive histological examination and provide a reliable diagnosis and prognosis for each case. Moreover, departmental procedures must be followed, to avoid adverse clinical events, such as destruction, misplacement or mislabelling of specimens and to ensure the safety of staff and patients.

In a large academic histopathology department in particular, there is an immense workload and a lot of time pressure. Multiple fixed and unfixed specimens are being processed and reported on at any one time, a lot of which are also used for research purposes. Specimens must be available for processing, cut-up rooms kept tidy and equipment kept clean. Lab resources and staff time must be used as economically as possible and reporting must be completed by specific deadlines, so that results can be discussed in multi-disciplinary meetings and made available to patients and their clinicians. It is therefore imperative that the research methodology does not deviate too much from the standard specimen pathway, research specimens are clearly labelled, histopathologists are fully aware of what is being done to specimens and what they need to do differently in terms of processing and reporting and that, ultimately, the research does not impact negatively on the quality of clinical reporting.

In this section we describe the development of a protocol that allowed us to collect data from fresh surgical specimens without compromising their diagnostic value. Furthermore, preliminary analysis of OCT data using gradient analysis software is described.

3.1.2 Aims

To develop a specimen and OCT data collection methodology that is compatible with histopathological processes.

To develop a preliminary OCT data processing algorithm and investigate differences between healthy and pathological gastrointestinal tissues using OCT signal gradient analysis

3.1.3 Methods

Data collection was carried out within the histopathology department at St Mary's Hospital (Paddington, London). This allowed for quick access to and processing of specimens and immediate placement within formalin for histological analysis. It also meant supervision and advice from senior histopathologists was readily available, as well as help from laboratory staff with the processing of specimens and preparation of slides for diagnosis.

All specimens were imaged using a commercial SS-OCT system (OCS1300SS, Thorlabs Incorporated, Newton, New Jersey), which incorporates a high-speed frequency swept external cavity laser (1325 nm central wavelength) having a 3 dB spectral bandwidth (> 100 nm) and an average output power of 10 mW. The frequency clock for the laser is provided by a built-in Mach-Zehnder Interferometer (MZI, Thorlabs INT-MZI-1300) and the main output of the laser is coupled into a fibre-based Michelson interferometer and split into the reference and sample arm using a 50/50 coupler (Thorlabs FC1310-70-50-APC).

In the reference arm of the interferometer, the light is reflected back into the fibre by a stationary mirror. In the sample arm, it is fibre coupled into the microscope head, collimated and then directed by the XY galvo-scanning mirrors towards the sample. The axial scans (A-scans) are performed at 16 kHz, which is the sweeping frequency of the laser. The transverse scan (B-scan) is controlled by the galvo-scanning mirrors and determines the frame rate of the OCT imaging. The sample is placed on a stage, providing XY and rotational translation. A pair of XY galvo-mirrors sequentially scans the probe beam across the sample surface area, and the 3D volume data set under this area is acquired (C-scan).

This OCT system produces high-resolution cross-sectional images of the gastrointestinal tissues with axial and transverse resolution of 9 and 15 mm, respectively. The interference signal is detected using a high-impedance gain balance photodetector which also applies noise correction. The fast Fourier transform (FFT) is used to convert the time to frequency of the interference signal.

Data for this study was collected at St Mary's Hospital, Paddington, London, from February to November 2010. Patients undergoing elective gastrointestinal resection surgery at St. Mary's Hospital, London, were included in the study. Patients who refused to be part of the study or who could not give informed written consent, as well as those under the age of 18 were excluded from the study. The purpose of the research and its implications were outlined to all patients during a pre-operative consultation before obtaining written consent. The suitability of both the patient and resected tissue was assessed in conjunction with the surgeons and histopathologists, to ensure that external handling and OCT imaging did not adversely affect tissue quality for histopathological investigation.

For anonymisation and data processing purposes, for every patient that was included in the study a specific patient code was generated consisting of the date and time of consent in reverse, hospital code and initials of the consenting researcher. For example, for a patient consented on 06.07.2011 at 19:30 at St Mary's Hospital (SMH) by the author (IA) the code would be 201107061930SMHIA. This ensured that each patient code was unique and provided for potential multi-centre recruitment too.

For the purposes of this study, specimens were collected from theatres in warm normal saline (0.9% sodium chloride) to maintain hydration. Formalin was not used till after imaging, as it is a fixative, causing cross-linking of proteins and effectively changing the structural properties of tissues and their optical properties too [83]. The specimens were collected as soon as they were excised (following a call from a member of theatre staff just before excision in each operation) and taken to the histology lab immediately, along with any additional specimens (e.g. resection margins, lymph nodes) and a histology form.

There, they were opened and gently rinsed according to the guidelines described in section 1.4.2, exposing the mucosa and any lesions and preserving the circumferential margin where necessary. This was initially done under supervision of a histopathologist. This became logistically difficult during the day, when histology staff was busy and it also meant that out of hours specimens, such as oesophagectomies, were routinely missed. Therefore, after a period of training, it was deemed safe to allow this specimen processing to be carried out independently, allowing the collection of late specimens and speeding up the whole process considerably.

In order to hold them open and mark sites of interest for imaging, research specimens were pinned onto corkboards after opening. Large pins were used to secure the specimen and smaller ones to mark areas from which OCT imaging was carried out, apart from tumour sites, where the thickness of the wall and tumour would not allow the pins to reach the corkboard (causing them to fall out when handling the specimen). The specimens were then carried to the OCT lab in a closed tray.

Optical coherence tomography imaging was carried out using the system described previously. As time for imaging was limited by specimen degradation once excised (histopathology advice was to keep the time between excision and placement in formalin below 45 minutes), imaging settings were adjusted to provide a balance between multiple high-resolution C-scans and minimal time consumption.

Image dimensions were set to 3mm x 3mm x 3mm. The OCT software always generates data to a fixed depth of 3 mm, regardless of the on-screen depth set by the user, which is for viewing purposes only, so the other two dimensions were set to match that, in order to obtain a cube-shaped c-scan. The lateral resolution was set to 512 pixels and the number of b-scans to 512. As the axial resolution of the system is fixed at 512 pixels, the resulting c-scans had 512 x 512 x 512 pixels.

Averaging was set to 8, contrast and brightness were set to "default". Before each specimen was imaged, noise removal was carried out, using the in-built noise removal function. To do this, the OCT probe was tilted away from the sample and pointed at a piece of black, non-reflective, light-shielding cardboard (same as the one used for construction of the goniometer, section 2.1) that was approximately 15 cm away and the noise removal button pressed on-screen.

In general, specimens were much thicker than the penetration depth of the OCT beam (maximum 3mm), so they did not require any special mounting and were imaged as they lay pinned on the corkboard. All specimens were kept hydrated by pouring some normal saline over the imaging sites in between data acquisition. As effectively all sites are rendered ischaemic and exposed at the same time, the order of imaging was randomised for each specimen, to standardise the average time before imaging for each diagnostic category (tumour, stomach, oesophagus etc.).

The OCT probe was lowered over the tissue whilst in 2D mode, till the surface of the tissue was approximately three quarters of the way up from the bottom of the frame. This was done to account for variations in height as the laser beam scanned along the tissue surface and to avoid, therefore, the surface of the tissue reaching the top of the frame, which would cause problems in the surface detection algorithms when processing the data. The system was then switched to 3D mode and a C-scan was recorded for each site.
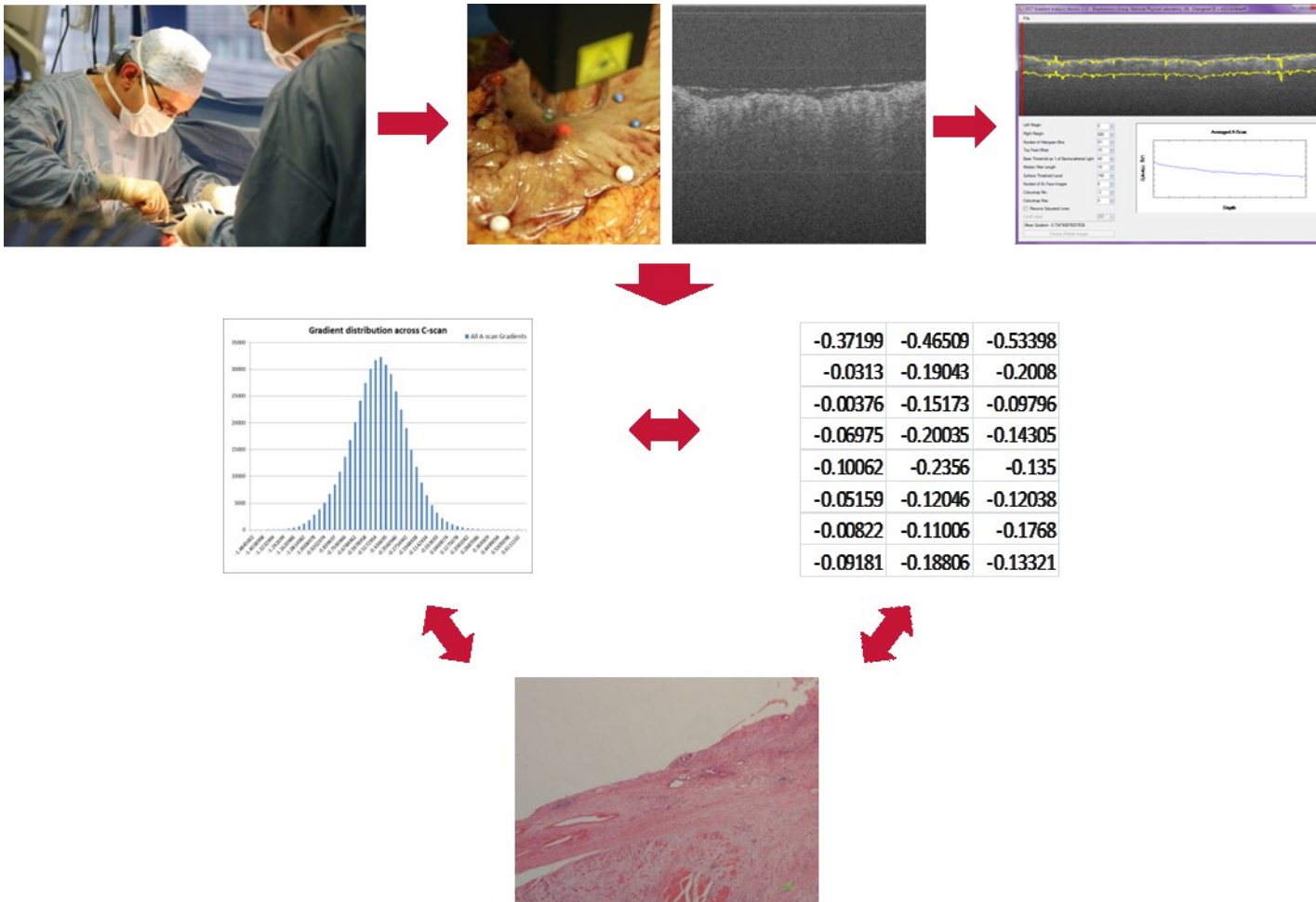
Figure 3.1: Summary of data collection methodology.

In the majority of C-scans, especially in surgical specimens, there were artefacts in the images, caused by reflection at the tissue surface and saturation of the OCT light detector. These manifested as vertical bright lines, which started from the top of the frame and extended all the way to the bottom and whose thickness ranged from a few pixels to significant portions of the image (even up to half the frame width). If there were only a few of these lines per C-scan, the data was kept, as the error correction algorithms described later were designed to filter them out. If, however, it was deemed that a significant part of the C-scan was affected (by watching the imaging process frame by frame), the data was discarded and the imaging repeated. Good hydration with normal saline, dabbing of excess mucus and fluid off the surface of the tissue and imaging with the probe at a slight angle to the tissue surface helped reduce artefacts.

For each site or biopsy that was imaged, a letter was appended to the unique patient code described previously, starting from "A" for the first site imaged and carrying on in alphabetical order. Thus, data for an individual patient would be saved as 201107061930SMHIA-A, 201107061930SMHIA-B, etc. The measurement sites (marked with small pins where possible, as mentioned above) were identified using their respective letters on a "measurement map" produced for each specimen, as shown in Figure 3.2.

After imaging was complete, tissues were fixed with 10% formalin and returned to pathology for routine histological processing.

Initial gradient analysis of OCT data was carried out using software originally written by Dr Pete Tomlins and modified by Dr Gavin Erry (both based at the National Physical Laboratory, Teddington, London). The software works on JPEG or BMP format images which have been created using the "export" function in the original Thorlabs imaging software. A single image or multiple images can be selected, one of which is shown on the main software screen. The user can set the region of interest (ROI) margins, which define which part of the image is used to generate the average a-scan. The lateral margins (marked in green and red in Figure 3.3) are useful for limiting the processing to areas of the tissue that might have a smoother surface or to avoid any blank areas in the image, such as the edge of a biopsy.
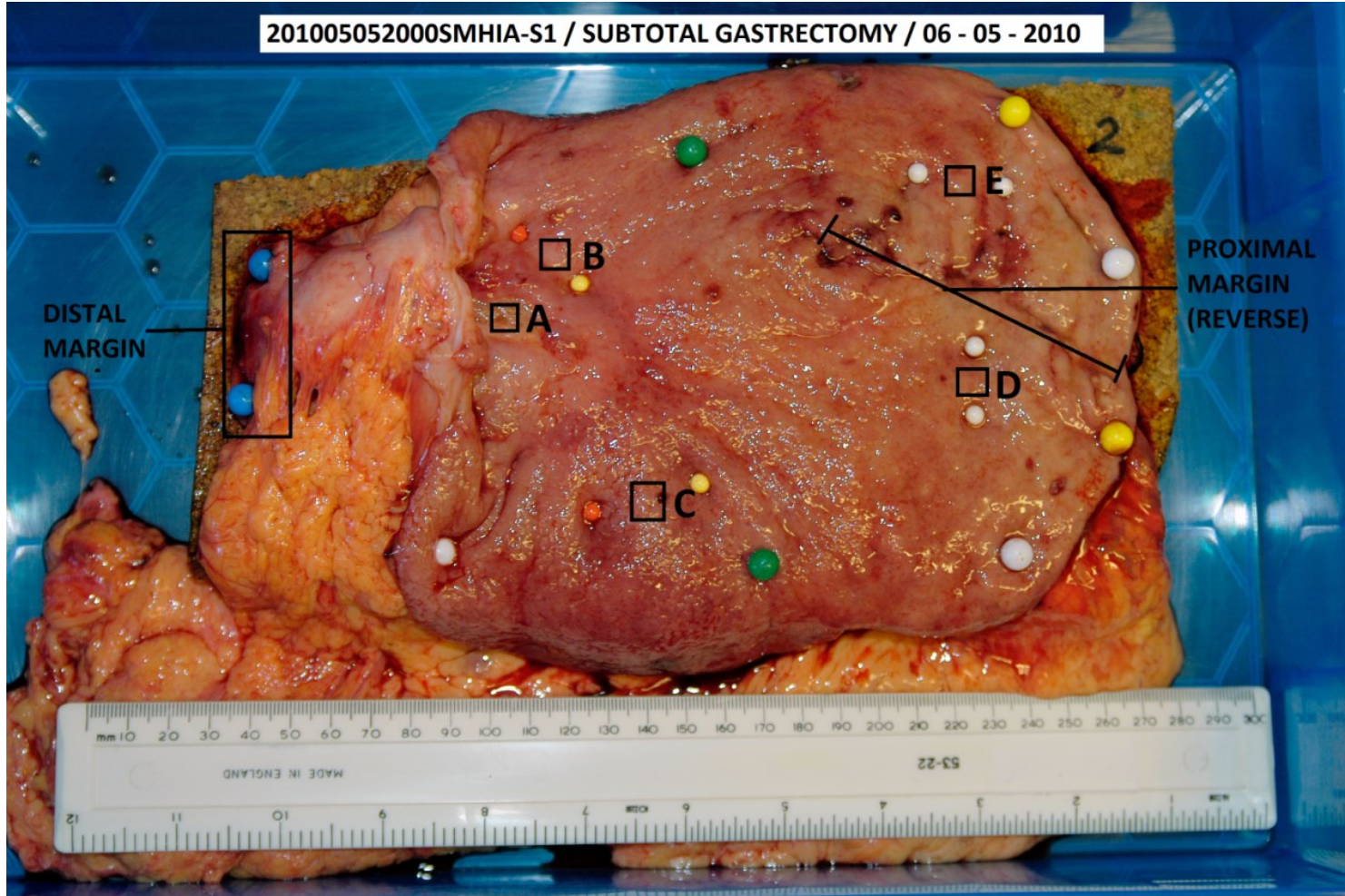
Figure 3.2: Typical measurement map for a surgical specimen

The surface of the tissue is detected using a median filter and a surface threshold set by the user. The intensity curve is smoothed using the median filter and the depth at which the intensity exceeds the set threshold is defined as the surface of the sample. A top pixel offset can also be set, so that the actual ROI top margin can be below the surface, where high intensities from specular reflection can affect the results. The bottom margin is defined using a base threshold, which is defined as the depth where the area under the curve from the surface to that point, divided by the total area under the whole curve, is equal to the percentage set by the user. The intensity is laterally averaged across the ROI and plotted against the average number of vertical pixels between the top and bottom margin, producing an average a-scan, which is displayed in the bottom right corner. A linear least squares fit is carried out and the gradient of that fit displayed at the bottom left corner.

There are also options for setting the number of *en face* images to be produced and the colourmap to be used. The *en face* images are representations of the tissue as viewed from the surface downwards (longitudinal plane) and superimpose the gradients for each a-scan (not the gradient of the mean a-scan) over the original OCT image, using the defined colourmap to set the colours for each pixel. Each image (which corresponds to a b-scan) contributes a row of pixels to the *en face* image, with the first image contributing the bottom row and the last image the top row. The first *en face* image corresponds to the tissue surface, with subsequent images depicting deeper levels at regular intervals up to the number of images set by the user. Finally, an error correction function can be activated and a cut off value defined for error detection. The function detects areas of detector saturation, based on the cut off value and sets the software to ignore the data in the columns affected by saturation when calculating the mean a-scan and gradient. When producing the *en face* images, these areas of saturation are shown as black regions. Originally the software piled all of these to one side, distorting the image, but later versions were corrected to leave the errors in their original places.

The above settings are used to process every exported OCT image selected by the user and the ROI, mean a-scan and mean gradient for each frame are displayed on the main screen while the software is working. The software outputs comma-separated value (.CSV) files which contain the gradient of every a-scan, the mean, mode and median gradient of each b-scan and a histogram of mean b-scan gradients. *En face* images are also produced in the JPEG format check. The main software screen is illustrated in Figure 3.3.

The compiled software was used to analyse all surgical data collected between February and November 2010 (tables 3.1, 3.2 and 3.3). The data was acquired as described in the previous section and the "export" function of the Thorlabs software was used to generate the JPEG images for analysis. These were visually assessed for any obvious problems, such as blank areas or significant saturation artefacts and any severely affected images were deleted. As a similar screening process had also been applied whilst imaging, the deleted images were only a small fraction of the dataset.

The images were then analysed using the OCT gradient analysis software. The base threshold was set to 40%, which visually corresponded to the top layer of each site imaged. The lateral margins were set to -5 pixels bilaterally and the top pixel offset was set to 10 pixels to avoid the highly-scattering areas at the surface.

The software produced a ".CSV" file for each C-Scan as described above. These were collated and matched to the histological reports for each site. For the purposes of this study, a mean attenuation gradient for each C-Scan was calculated and these were compared across different pathologies.

Tissue Type	Upper GI	Lower GI	Total
Patients	25	6	**31**
Measurements	87	18	**105**
Non Cancer	56	14	**70**
Cancer	31	4	**35**

Table 3.1: Surgical specimens analysed with the compiled OCT gradient analysis software.
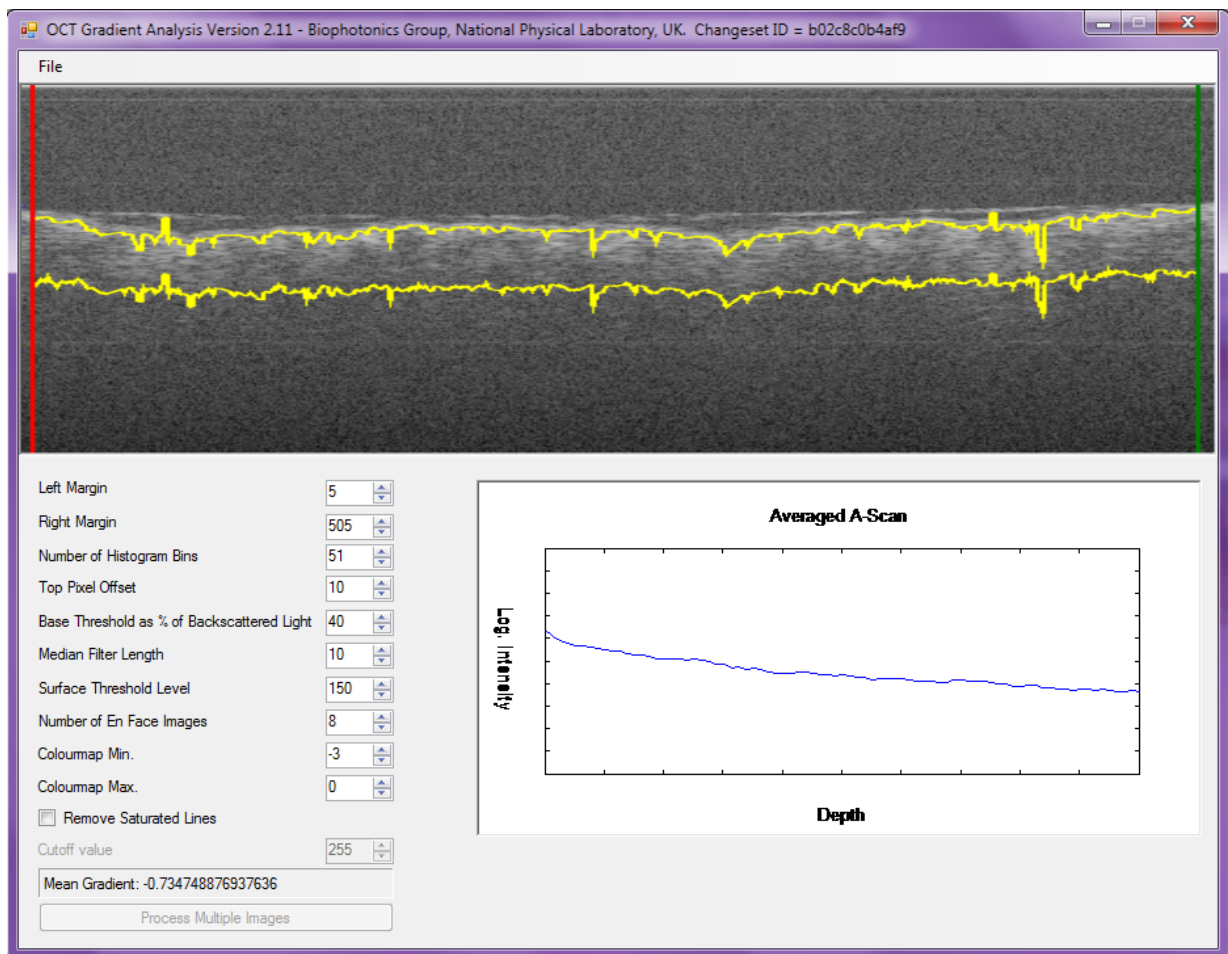
Figure 3.3: Screenshot of the OCT gradient analysis software. The OCT image, margins of the region of interest, settings, averaged a-scan and mean gradient for the region of interest can be seen.

Surgery Compiled Gradient Analysis: Upper GI Measurements Breakdown									
Histology		Mid Oesoph	Lower Oesoph	GOJ	Cardia	Fundus	Body	Antrum	Pylorus	Total
Normal		1	2				8	2		13
Inflammation	Mild	8			1		8	2		19
	Moderate					2	5			7
	Severe				1	1				2
Barrett's										0
Int Metaplasia					1	1	11			13
Dysplasia	Low Grade		1					1		2
	High Grade									0
Adenocarcinoma	Poor Diff		1	6			4			11
	Mod Diff		1	2		1	8	5		17
	Well Diff									0
Squamous Cell Ca	Poor Diff									0
	Mod Diff			3						3
	Well Diff									0
Total		9	5	11	3	5	44	10	0	87

Table 3.2: Breakdown of upper gastrointestinal tissue measurements used in the compiled software gradient analysis of surgical data.

Surgery Compiled Gradient Analysis: Lower GI Measurements Breakdown										
Histology		**Ileum**	**Caecum**	**Appendix**	**Asc Colon**	**Tran Colon**	**Desc Colon**	**Sigmoid**	**Rectum**	**Total**
Normal		3		1	6			1		**11**
Inflammation	Mild				1					**1**
	Moderate									**0**
	Severe									**0**
Crohn's						2				**2**
Dysplasia	Low Grade									**0**
	High Grade									**0**
Adenocarcinoma	Poor Diff				1					**1**
	Mod Diff		1		1			1		**3**
	Well Diff									**0**
Squamous Cell Ca	Poor Diff									**0**
	Mod Diff									**0**
	Well Diff									**0**
Total		**3**	**1**	**1**	**9**	**2**	**0**	**2**	**0**	**18**

Table 3.3: Breakdown of lower gastrointestinal tissue measurements used in the compiled software gradient analysis of surgical data.

3.1.4 Results
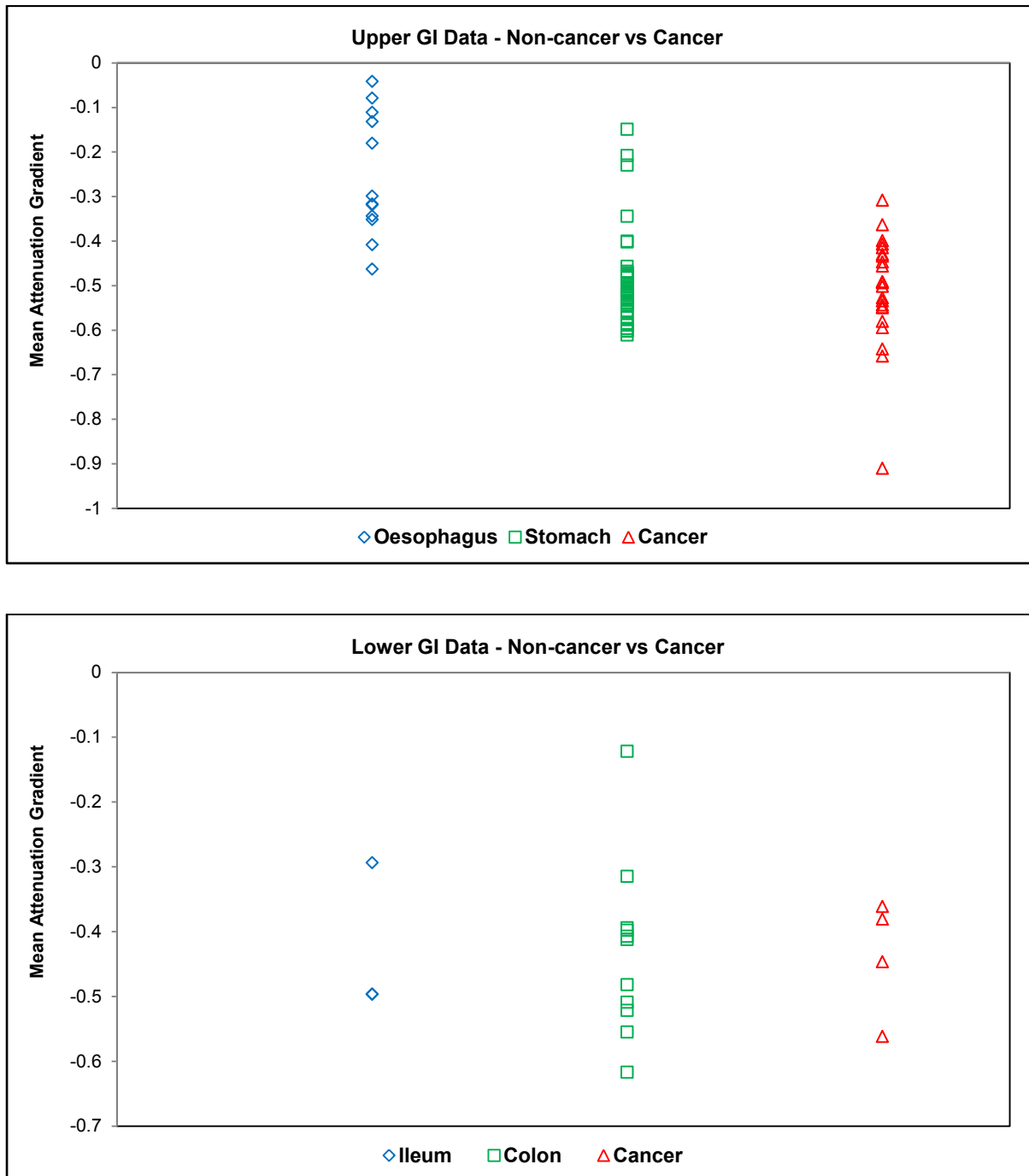
Figure 3.4: Mean attenuation gradients for surgical specimens after compiled software gradient analysis. Data is split into cancer and non-cancer (encompassing normal and inflammation). Top: upper gastrointestinal tissues. Bottom: lower gastrointestinal tissues.
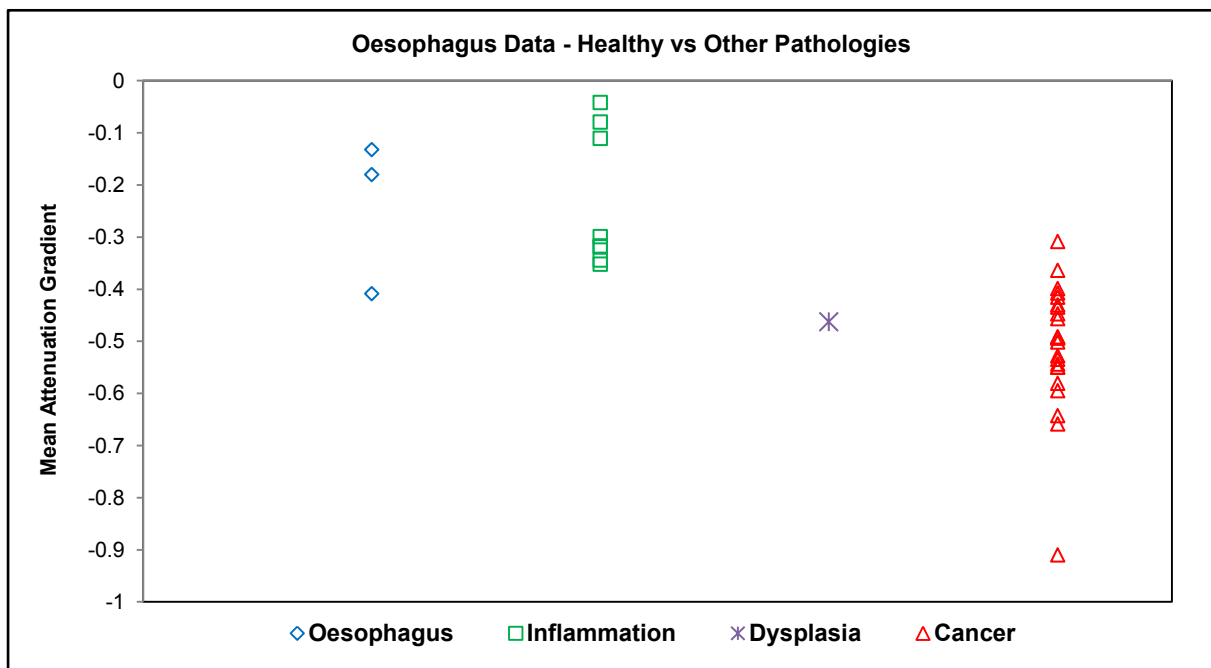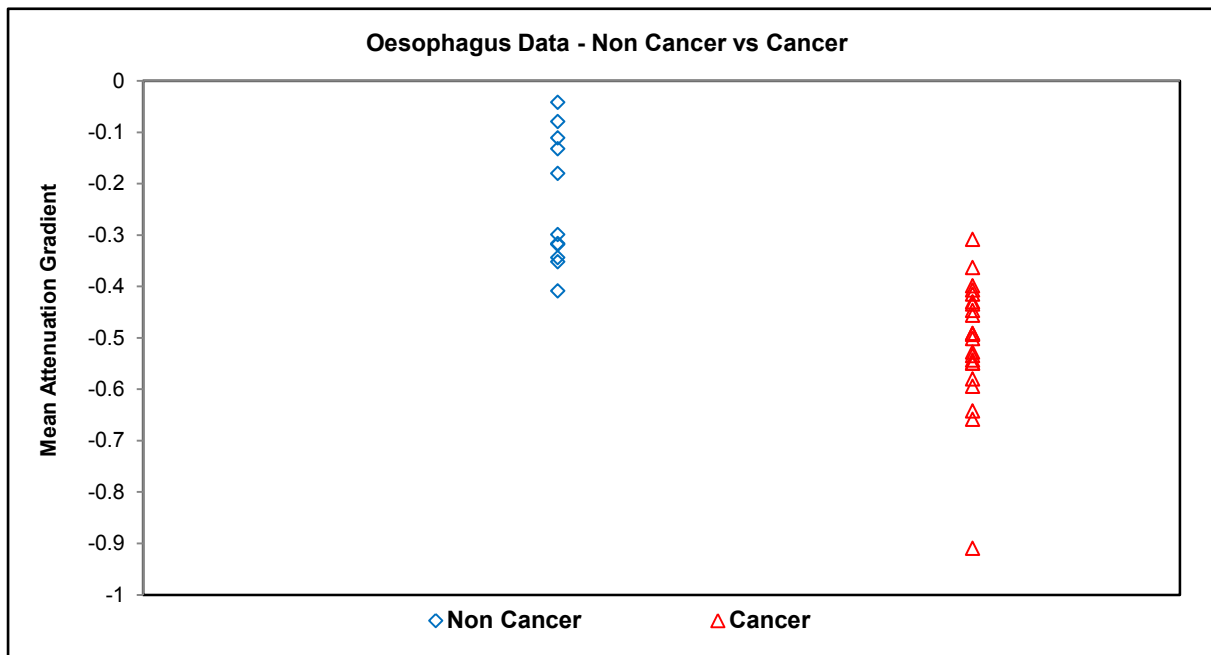
Figure 3.5: Mean attenuation gradients for surgical oesophageal specimens after compiled software gradient analysis. Top: cancer vs non-cancer. Bottom: all pathologies.
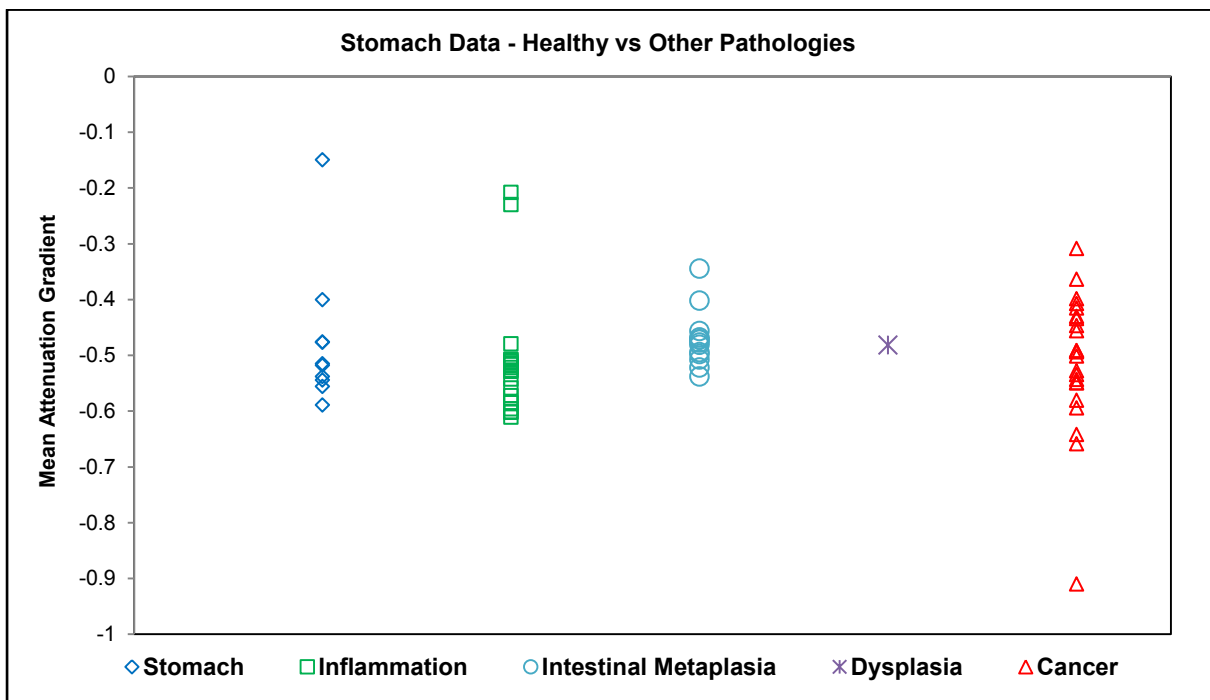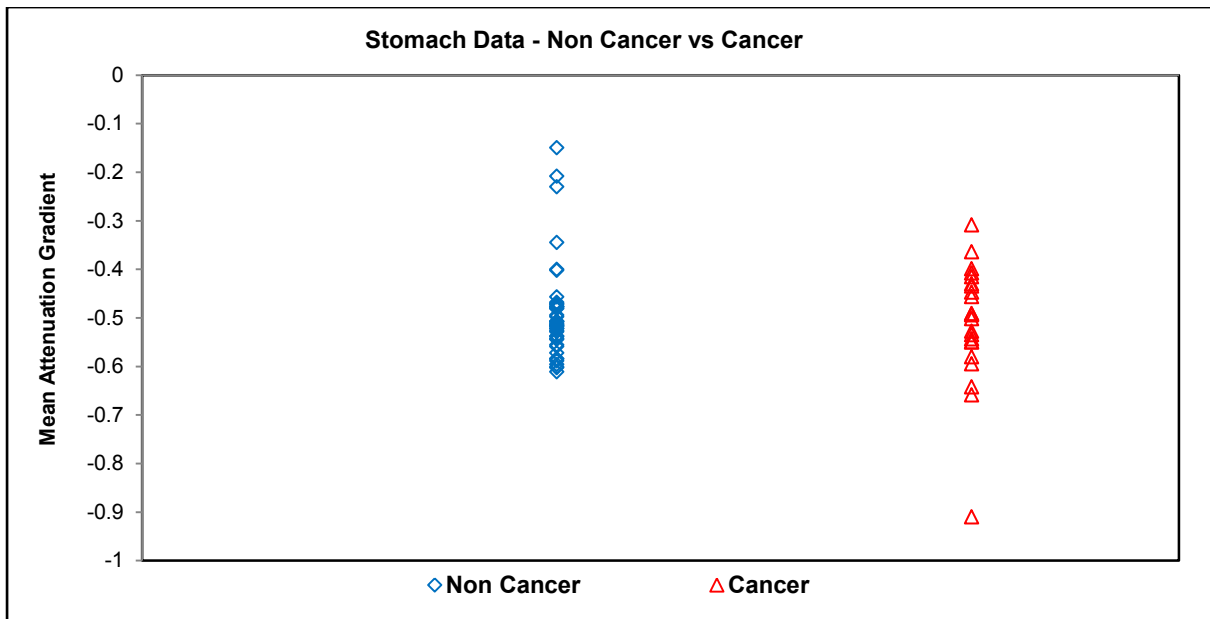
Figure 3.6: Mean attenuation gradients for surgical gastric specimens after compiled software gradient analysis. Top: cancer vs non-cancer. Bottom: healthy vs. all pathologies.
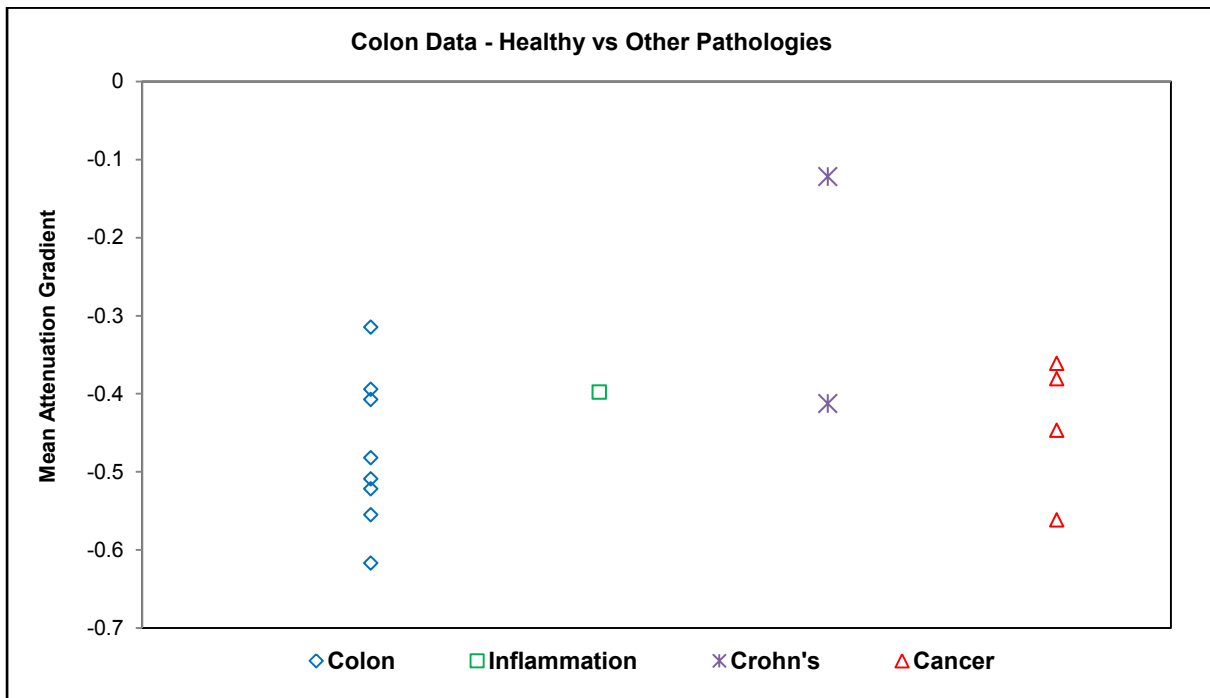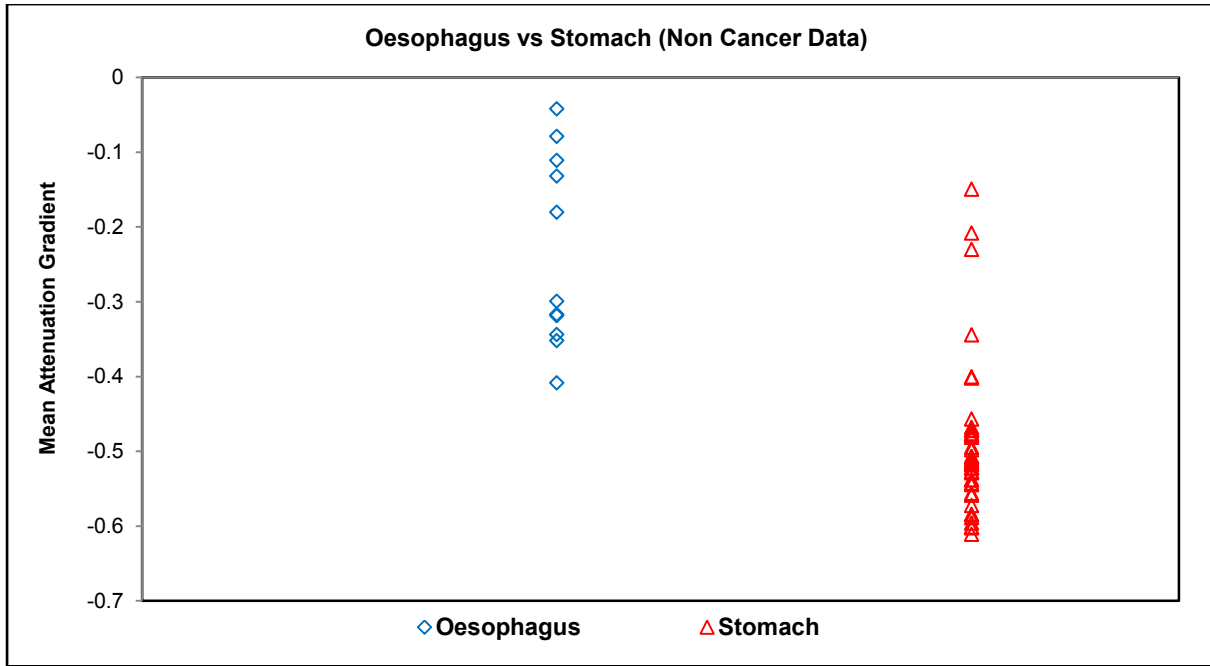
Figure 3.7: Mean attenuation gradients for surgical specimens. Top: oesophagus vs stomach (all non-cancer pathologies). Bottom: colon data.
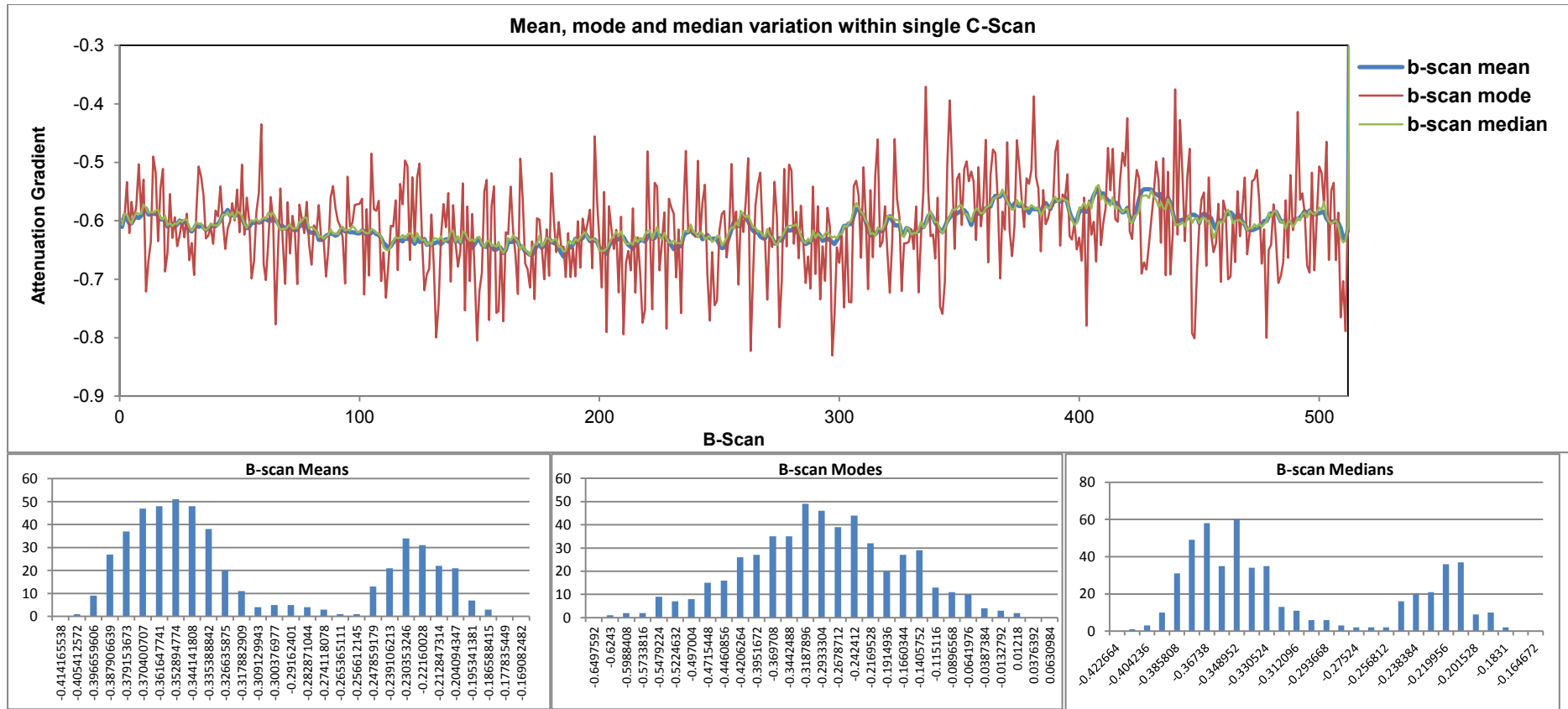
Figure 3.8: Graph and histograms showing the variation in mean, mode and median attenuation gradients distribution within a single C-Scan.
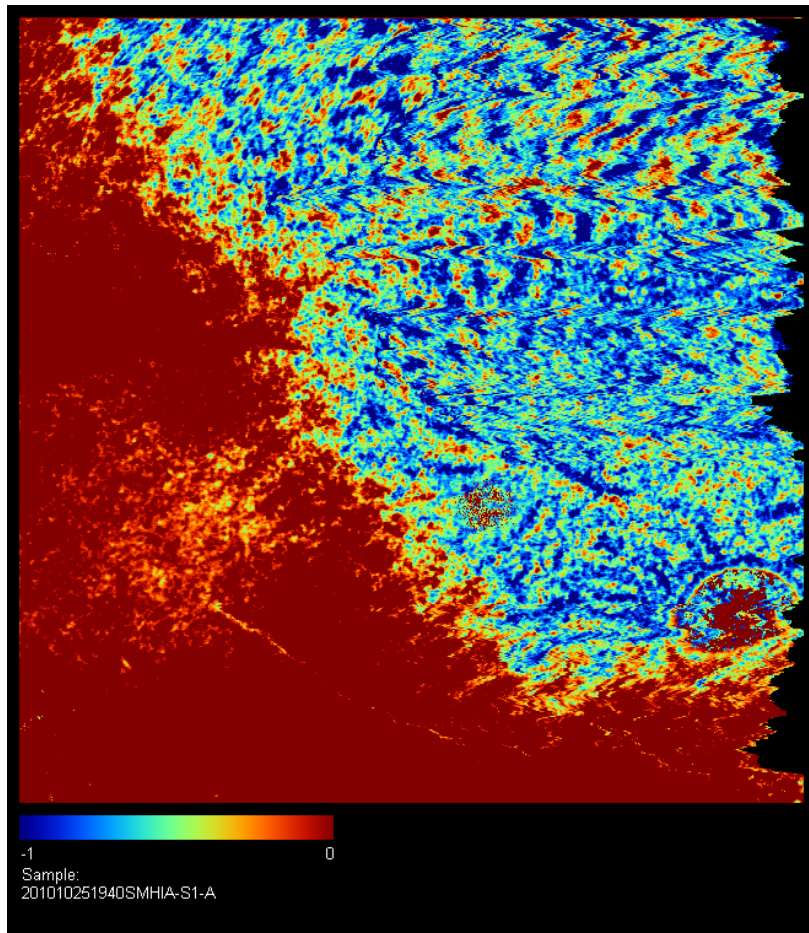
Figure 3.9: Colourmap of tissue sample after compiled gradient analysis. Two areas of significantly different attenuation can be seen, signifying different tissue types within the same measurement.

3.1.5 Discussion

In terms of the data collection methodology, this study fulfilled its aims. Through a process of step by step improvements and, at times, trial and error, an efficient and well-standardised specimen collection and imaging technique was developed.

Thanks to a multidisciplinary approach, suitable patients were prospectively identified and consented, while specimens were collected and processed in a timely manner and in accordance with histological protocols. Imaging was carried out in a standardised manner and with minimal delay after specimen excision. All relevant clinical information was collected, anonymised and saved in a database specifically developed to hold this

information, as well as key histological data for each specimen and measurement site and results of gradient analysis. As a result, a high quality dataset was produced for further processing.

Compared to the literature, there are positive aspects to our data gathering methodology. No previous study in this field has included as many patients or yielded as many different types of gastrointestinal tissues or pathologies. Moreover, other ex vivo studies have not been able to procure specimens as fresh as ours. In most cases specimens are cadaveric or frozen/refrigerated and are imaged with great delay from the time of excision. In our study, imaging of the vast majority of specimens was initiated within 15 minutes and completed within 30 minutes from the time of excision. The key to achieving this was the multidisciplinary approach (involving surgeons and histopathologists) used to identify patients and procure samples, as well as the fact that the study was carried out by a clinician who had direct access to operating theatres and the histology lab and who was trained through this process to independently collect and process specimens.

The gradient analysis method used in our study has been described in the literature and offers a quick and easy way to analyse OCT data, as well as a visual representation in the form of the colour map produced at the end of the process [77, 84]. The results in our case show a variation in the attenuation gradient between different pathologies and different tissues. With the Upper GI data there is a clear difference in attenuation gradients between oesophagus and cancer, showing an increased gradient (i.e. increased scattering) in cancer, as expected from the literature. There is also an increase in scattering in stomach tissues compared to oesophagus. The difference between stomach and cancer is not so clear and the same can be seen in Lower GI tissues.

The study does suffer from a number of limitations, which would be expected to have a strong effect on the results. Firstly, the sample numbers are low, particularly for Lower GI tissues, where only 18 measurements from 6 patients were taken. More data was available for Upper GI tissues (87 measurements from 26 patients), but this was also inadequate to demonstrate a clear difference in attenuation gradients for gastric tissues.

Secondly, the data collection process and matching to histology has room for improvement. As can be seen Figure 3.8, there was a lot of inhomogeneity in individual measurements.

Within a single C-scan there was a large variation in attenuation gradients, even when these were averaged across B-scans to reduce errors. This could simply be due to inhomogeneity within each tissue type, but at least in some cases this was more likely due to a mixture of tissues being imaged within each C-scan. This is clearly demonstrated in the colour map shown in Figure 3.9, where there is a distinct difference between two areas of attenuation gradients, likely reflecting a measurement taken at the border between two different tissue types, such as tumour and healthy tissue. This kind of sample would be labelled as "cancer" by the histopathologist, if it contained malignant tissue, but the large proportion of healthy (or other) tissue would clearly have a strong effect on the numbers obtained.

Finally, the data analysis carried out in this study could be improved in many ways. The data quality can be improved by working on raw OCT data, as described in the next section. This has also been mentioned in the literature [77]. Furthermore, a single number for each C-scan (such as the mean attenuation gradient used in this study) may not be the best way to represent data with such a high variability. A better way would be to use a descriptor of the range of values that correspond to each tissue type and try to define a cut-off that can be used to predict with a degree of accuracy which tissue type a measurement belongs to [77].

3.2 Gradient analysis of Optical Coherence Tomography data from surgical and endoscopic specimens

3.2.1 Background

In the previous section we described the development of a data collection protocol as well as preliminary gradient analysis of OCT data. We demonstrated some promising results and identified areas that needed improvement, which we have aimed to address in this section.

Firstly, the correlation of OCT data to diagnoses has been improved by better marking of specimens, quantification of pathological tissues in each measurement site and reporting by a single pathologist. We have also re-examined the tissues used in the preliminary study and reclassified them where necessary.

Secondly, more surgical specimens as well as biopsies from endoscopies have been collected, which should lend more significance to our results and has also provided us with a wider range of pathologies, such as Barrett's oesophagus, which is of particular interest as a precancerous lesion.

Finally, an improved data processing algorithm has been employed. The basic concept of signal gradient analysis remains, but in this section we work with raw OCT data and employ more sophisticated error correction processes. The technique is tested on microsphere phantoms and then used to distinguish between healthy and pathological tissues.

3.2.2 Aims

To demonstrate differences in OCT signal attenuation between healthy and pathological gastrointestinal tissues in surgical and endoscopic specimens through linear gradient analysis.

To demonstrate differences in the OCT signal gradient between pathological subsets (e.g. inflammation, dysplasia, cancer).

3.2.3 Methods

The data in this study was collected between February 2010 and June 2011. Patients undergoing elective GI surgery or endoscopy at St. Mary's Hospital, London, were included in the study. Patients who declined to be part of the study or who could not give informed written consent, as well as those under the age of 18 were excluded from the study. Patients were selected and consented as described previously.

Surgical specimens were collected as outlined in the previous section. Data collected from the previous study was also used for this one and new specimens were collected and imaged too. Additionally, biopsies from endoscopies were also collected. These are normally placed into labelled pots pre-filled with formalin and taken to the histology lab at a later stage. Multiple biopsies may be placed in the same pot together or in different pots in small groups (e.g. if from the same location) or individually.

For the purposes of this study, the biopsies were transported in normal saline instead and placed in formalin immediately after imaging. Biopsies were imaged whole, without any previous dissecting or pinning. Instead, imaged biopsies were placed individually in specially labelled pots, so that the histological diagnoses for each pot could be matched to the respective OCT data. As the patient turnover on any given endoscopy list was high and biopsies were produced in large numbers, a continued presence in the endoscopy department was necessary, to ensure biopsies were collected, imaged and placed in formalin as quickly as possible, to avoid dehydration or autolysis. As the endoscopy department is situated directly under Histopathology, it was possible to carry out this process in minutes.

Image dimensions were set to 3 x 3 x 3 mm for large specimens and 2 x 2 x 3 mm for biopsies. The same image settings were used as previously described. As before, large specimens did not require any special mounting and were imaged as they lay pinned on the corkboard. With biopsies, however, the beam penetrated through the full thickness of the tissue and was reflected off the surface they lay on, causing interference in the image.

To counteract this, biopsies were placed on top of a thick glass slide, with a thin layer of saline between the tissue and glass. This allowed the beam to penetrate through the glass and dissipate, without affecting the imaging of the tissue above. Alcohol wipes were used to

clean the surface of the glass slide for disinfection purposes and glass slides were periodically discarded.

All specimens were kept hydrated by pouring some normal saline over the imaging sites in between data acquisition. Biopsies were imaged in the order they were obtained by the endoscopist for each patient, to minimise the time spent before entering formalin. In large specimens, where effectively all sites are rendered ischaemic and exposed at the same time, the order of imaging was randomised for each specimen, to standardise the average time before imaging for each diagnostic category (tumour, stomach, oesophagus etc.).

Specimens were marked as before, with the addition of inking of the precise sites where imaging was carried out (after imaging was complete, so that the ink would not interfere in any way with the imaging). Furthermore, the author (I.A.) was present in the dissecting room during block-taking, to ensure blocks were taken accurately from the imaged areas.

Histological reporting was carried out by the same histopathologist (H.D.) for all surgical and endoscopic specimens. Reporting was modified for the purposes of this study, taking into account the amount and location of pathologies within each slide. For example, whereas a sample consisting mainly of normal tissue with a very small amount of cancer would be reported as "cancer" for clinical purposes, for our study it was classified as "normal tissue". The same was done for specimens where the location of pathological tissue was more than 2mm from the surface of the specimen, regardless of quantity, as our OCT system does not penetrate that deep into tissue. Slides from specimens used in the previous study were also re-examined and reclassified where necessary.

The technique used for quantitative analysis of the OCT data in this study was also different. Normally, a number of data processing steps are involved between the raw spectroscopic OCT data and the final images produced on screen in the Thorlabs software or those exported as BMP/JPEG files, which were used as the starting point in our previous study. These steps mainly involve smoothing and averaging of the data to remove noise, as well as tweaking of contrast, brightness and other settings to improve the final image. As the aim of our work is not to compare the final images qualitatively, but rather to quantitatively analyse the raw OCT data, code was written to directly access and process the raw OCT data in MATLAB (Mathworks, Natick, MA, USA). This not only avoided any quality loss

associated with the data post-processing, but also allowed faster processing and manipulation of multiple files using time-saving automations, choice of post-processing to be applied to raw data and analysis using different techniques, as described later.

In order to process the OCT data in MATLAB, the conversion process from fringe (.FRG files) data to images that is carried out by the Thorlabs system had to be replicated. As a first step, custom code was written in MATLAB to read the .FRG files, reshape the data, carry out Fourier transformation and produce the images comprising each c-scan. The data was processed frame by frame and the resulting images were saved as three-dimensional matrices (MAT files) that could be further processed in the MATLAB environment (see Appendix 6.1). This was used for initial testing and development of further steps in the process such as the error correction code. However, the discovery of certain artefacts in some images that were not present on-screen during specimen imaging, led to a discussion with the Thorlabs support team, who provided MATLAB code written by the company itself. This was modified and used for the conversion of all OCT data to MAT files, as described below.

The Thorlabs code was quite complicated and designed to offer various options for data processing and functions to deal with errors, corrupted data etc. It was also designed to produce high quality images, employing a variety of smoothing functions and contrast / brightness settings. The processing steps included a moving average smoothing filter, conversion to 8bit images, Decibel scaling (0-255) of intensity values and image apodization (another smoothing process). These were removed or set to "off" in the import code used for this work (Appendix 6.1.2 and 6.1.3).

The next step when importing OCT data into MATLAB was the removal of errors and preparation of the data for further processing. During imaging of specimens, areas of specular reflection saturate the detector. These cause bright areas to form in the OCT images above the surface of the specimen and dark columns under the reflection, that span the whole depth of each A-scan below it. The error correction algorithm detected these areas and replaced the intensity values with "NaN" values, that would later be ignored by further processing algorithms. This served to maintain the 512 pixels x 512 pixels size of images, which would be affected if areas of errors were just deleted and cause further problems downstream (Appendix 6.1.4).

Image alignment followed. Here, the tissue surface in each image was detected and the pixels above (corresponding to air) were deleted, bringing the surface to the top of each B-scan and padding the bottom with "NaN" values again to maintain original image dimensions. This facilitated further processing of images by ensuring the surface was always at the top and the rest of the image was properly aligned. As illustrated in Figure 3.10, the dark area normally present at the top of OCT images (corresponding to air) has now been removed, "dragging" the surface pixels to the top of the image (yellow and red areas) and filling the bottom with "NaN" values (blue area). This allowed all further processing to begin from the top of the image and work its way downwards, knowing that only tissue and no air was being analysed.
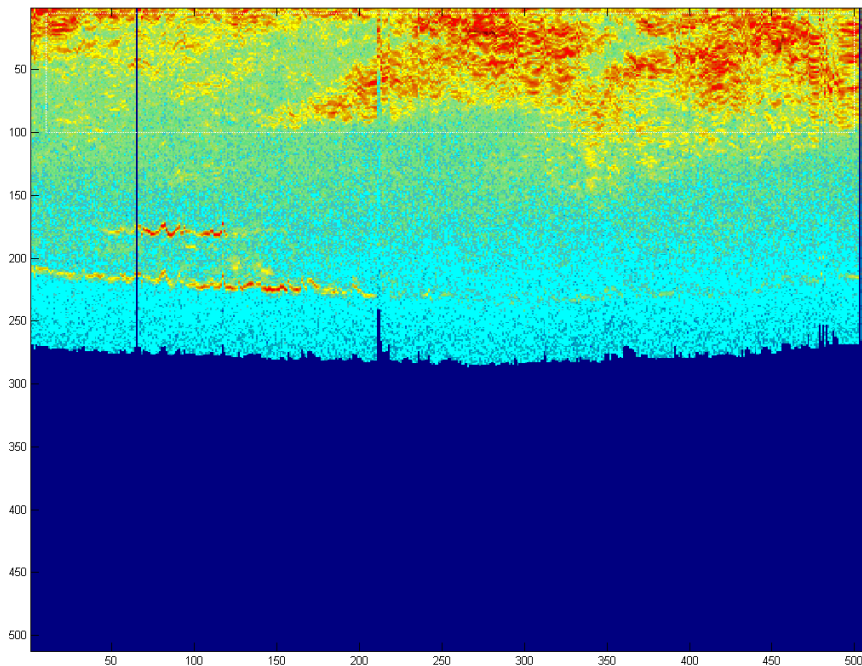
Figure 3.10: Optical coherence tomography image after error correction and surface alignment.

The gradient analysis code was designed to work on the error-corrected, aligned OCT images produced by the error correction code described previously. First, a region of interest was defined, which was used for all B-scans. Typically this was set to avoid the highly reflective surface of the sample and the deeper parts of the image which contain little information. The A-scans within that region of interest were averaged horizontally, to produce a single mean A-scan for each B-scan. As previously mentioned, areas of error,

which have been replaced by "NaN" values, were ignored and did not affect the final mean A-scan values. The mean A-scan was then smoothed using a 5-pixel moving average filter and the attenuation gradient was calculated using robust linear-fitting. Thus, for each 512-column B-scan, a single mean attenuation gradient was produced and for each C-scan there were 512 mean attenuation gradients. These were saved into a gradient matrix named after the original measurement.

Secondly, the mean gradient matrices for each C-scan were grouped together into larger matrices according to pathology (e.g. all oesophageal C-scans or all cancer C-scans etc.) and boxplots were generated, which allowed comparisons to be made between measurements of the same pathology from different parts of a specimen and different specimens altogether. This is shown in Figure 3.11, where measurements of normal oesophagus (from different specimens) have been grouped together and boxplots generated. As can be seen, there is variability between different measurements of the same type of tissue.

Thirdly, grouped gradient matrices from different pathologies were reshaped as columns, padded with "NaN" values in order to be of the same length (as the sample numbers for each pathology are not equal) and concatenated into a larger comparison matrix for further boxplot generation and derivation of simple descriptive statistics. The comparison matrices were named after the comparison being carried out (e.g. Oesophagus vs. Cancer) and were used for further analysis. An example is shown in Figure 3.12.
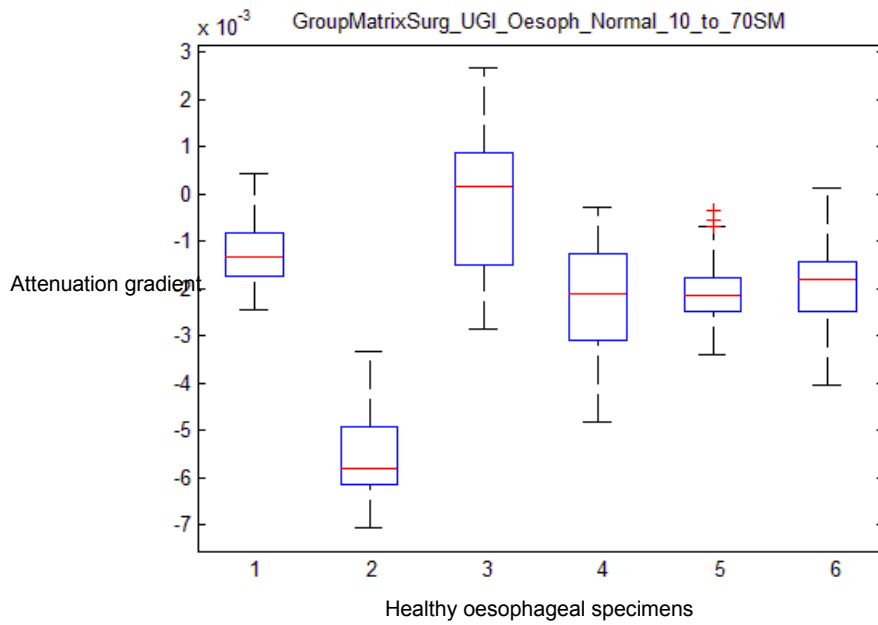
Figure 3.11: Boxplot of gradient matrices from same tissue type (normal oesophagus).
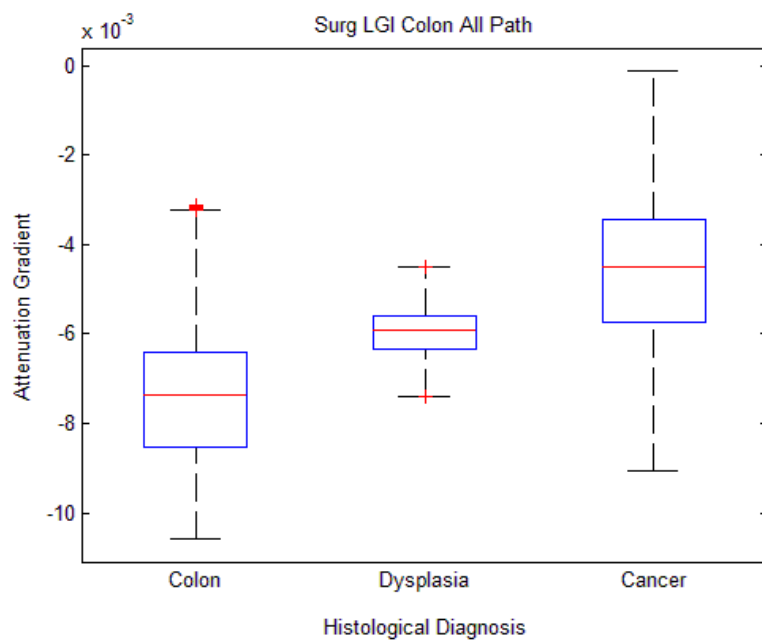
Figure 3.12. Comparison of mean attenuation gradients of different colon tissues.

As described in Section 2.2, a set of microsphere phantoms was created with different scattering properties to test the goniophotometer. The same phantoms were used to test the gradient analysis code discussed above. If the OCT data import, error correction and gradient analysis codes worked well, then the attenuation gradients obtained from different phantoms should be lower (less negative) and higher (more negative) for less and more scattering phantoms respectively. The results, in the form of boxplots, can be found in section 3.2.4.

The breakdown of patients and tissue types can be seen in Tables 3.4 to 3.9. Compared to the preliminary gradient analysis study described previously, there were additional surgical specimens and also endoscopic specimens in this study. The latter allowed for comparison of different pathologies to before, such as dysplasia and Barrett's oesophagus. As can be seen in Table 3.5, a small number of biopsies were autolysed after being left in saline too long. These were not included in the study and did not affect the diagnostic value of the endoscopies for those patients, but steps were taken to minimise the time biopsies spent in saline to ensure this did not happen again.

Tissue Type	Upper GI	Lower GI	Total
Patients	32	7	39
Measurements	111	21	132
Non Cancer	75	17	92
Cancer	36	4	40

Table 3.4: Surgical specimens analysed in MATLAB.

Tissue Type	Upper GI	Lower GI	Total
Patients	51	56	107
Measurements	129	173	302
Non Cancer	121	150	271
Cancer	6	15	21
Autolysed	2	8	10

Table 3.5: Endoscopic specimens analysed in MATLAB.

Before gradient analysis was carried out on the whole dataset, an analysis was carried out on data from 11 patients to determine the best region of interest for discrimination between

different pathologies using gradient analysis. To do this, a Savitzky-Golay smoothing filter was applied to the mean A-scan for each C-scan and the resulting plots were overlapped. This allowed comparison of the gradient variability across the OCT image depth and selection of the area with the greatest variation in gradient for the analysis. As can be seen in Figure 3.14, the area of greatest gradient variation between the selected pathologies was between 10 and 70 pixels from the surface. Thus, gradient analysis for all specimens in this study was carried out at that depth range.

The attenuation gradients across different pathologies were compared pair-wise using the Mann-Whitney U test (Wilcoxon rank sum test), a non-parametric test that compares the medians of independent samples. This test was chosen as the distributions of the data were not normal, as can be seen in the normal probability plot demonstrated in Figure 3.13.
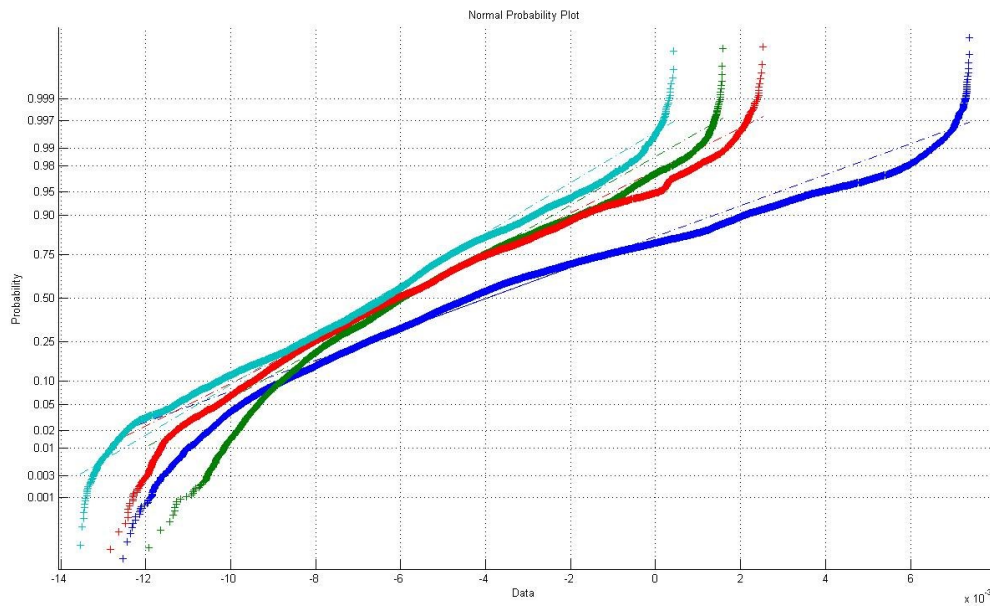
Figure 3.13: Normal probability plot for mean attenuation gradients of all surgical lower gastrointestinal pathologies. Different colours represent different tissue types.

Surgery Upper GI Measurements Breakdown - Second Study										
Histology		Mid Oesoph	Lower Oesoph	GOJ	Cardia	Fundus	Body	Antrum	Pylorus	Total
Normal		2	4				13	2		**21**
Inflammation	Mild	9	1		1	1	12	3		**27**
	Moderate					1	8			**9**
	Severe				1	1				**2**
Barrett's										**0**
Int Metaplasia					1	1	12			**14**
Dysplasia	Low Grade		1					1		**2**
	High Grade									**0**
Adenocarcinoma	Poor Diff		1	4			5			**10**
	Mod Diff		2	6		1	10	5		**24**
	Well Diff									**0**
Squamous Cell Ca	Poor Diff									**0**
	Mod Diff			2						**2**
	Well Diff									**0**
Total		**11**	**9**	**12**	**3**	**5**	**60**	**11**	**0**	**111**

Table 3.6: Breakdown of upper gastrointestinal surgical specimens analysed in MATLAB.

Surgery Lower GI Measurements Breakdown - Second Study										
Histology		Ileum	Caecum	Appendix	Asc Colon	Tran Colon	Desc Colon	Sigmoid	Rectum	Total
Normal		3		1	7	3		1		**15**
Inflammation	Mild									**0**
	Moderate									0
	Severe									0
Crohn's										0
Dysplasia	Low Grade				2					**2**
	High Grade									0
Adenocarcinoma	Poor Diff				1					**1**
	Mod Diff		1		1			1		3
	Well Diff									0
Squamous Cell Ca	Poor Diff									0
	Mod Diff									0
	Well Diff									0
Total		**3**	**1**	**1**	**11**	**3**	**0**	**2**	**0**	**21**

Table 3.7: Breakdown of lower gastrointestinal surgical specimens analysed in MATLAB.

Endoscopy Upper GI Measurements Breakdown										
Histology		Mid Oesoph	Lower Oesoph	GOJ	Cardia	Fundus	Body	Antrum	Pylorus	Duodenum	Total
Normal		7						4		31	**42**
Inflammation	Mild	6	4	2			13	15		2	**42**
	Moderate		2					3		1	**6**
	Severe		3				2				**5**
Barrett's			20	1							**21**
Int Metaplasia				1				1	2		**4**
Dysplasia	Low Grade										**0**
	High Grade		1								**1**
Adenocarcinoma	Poor Diff										**0**
	Mod Diff		3								**3**
	Well Diff										**0**
Squamous Cell Ca	Poor Diff										**0**
	Mod Diff	3									**3**
	Well Diff										
Autolysed										2	**2**
Total		**16**	**33**	**4**	**0**	**0**	**15**	**23**	**2**	**36**	**129**

Table 3.8: Breakdown of upper gastrointestinal endoscopic specimens analysed in MATLAB.

Endoscopy Lower GI Measurements Breakdown											
Histology		**Ileum**	**Caecum**	**Asc Colon**	**Hep Flexure**	**Tran Colon**	**Spl Flexure**	**Desc Colon**	**Sigmoid**	**Rectum**	**Total**
Normal		11	12	12	1	9	2	7	19	14	**87**
Inflammation	Mild		5	4		3	3	1	5	7	**28**
	Moderate		1	1		1		1	4	5	**13**
	Severe									1	**1**
Crohn's											**0**
Dysplasia	Low Grade		1	2			2	1	11	3	**20**
	High Grade		1								**1**
Adenocarcinoma	Poor Diff					3	3		5		**11**
	Mod Diff									4	**4**
	Well Diff										**0**
Squamous Cell Ca	Poor Diff										**0**
	Mod Diff										**0**
	Well Diff										
Autolysed			2	2					3	1	**8**
Total		**11**	**22**	**21**	**1**	**16**	**10**	**10**	**47**	**35**	**173**

Table 3.9: Breakdown of lower gastrointestinal endoscopic specimens analysed in MATLAB.
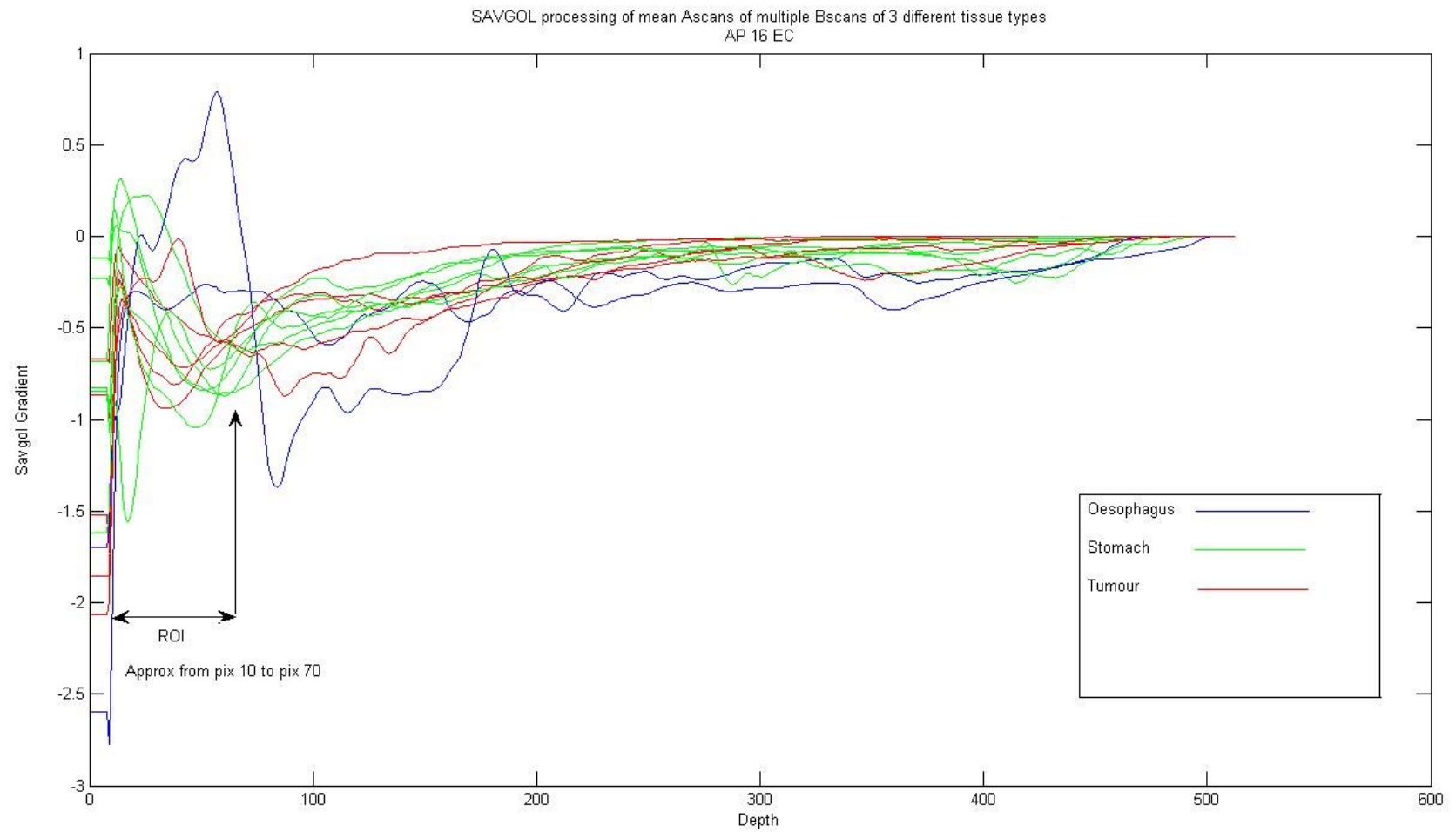
Figure 3.14. Savitzky-Golay analysis of selected attenuation gradients.

3.2.4 Results

The measurements carried out on the microsphere phantoms showed a good correlation between phantom concentrations and attenuation gradient. Measurements were carried out on all four microsphere sizes and the results were in agreement, showing an increased attenuation gradient with increasing concentration for all four sizes. A representative example is shown in Figure 3.15. This is followed by the statistics and boxplots for different tissue types from surgical and endoscopic specimens. In total, 132 measurements were taken from 39 surgery patients and 302 measurements from 107 endoscopy patients. Details of different comparison groups (e.g. healthy stomach vs stomach cancer) are shown in respective tables and graphs.
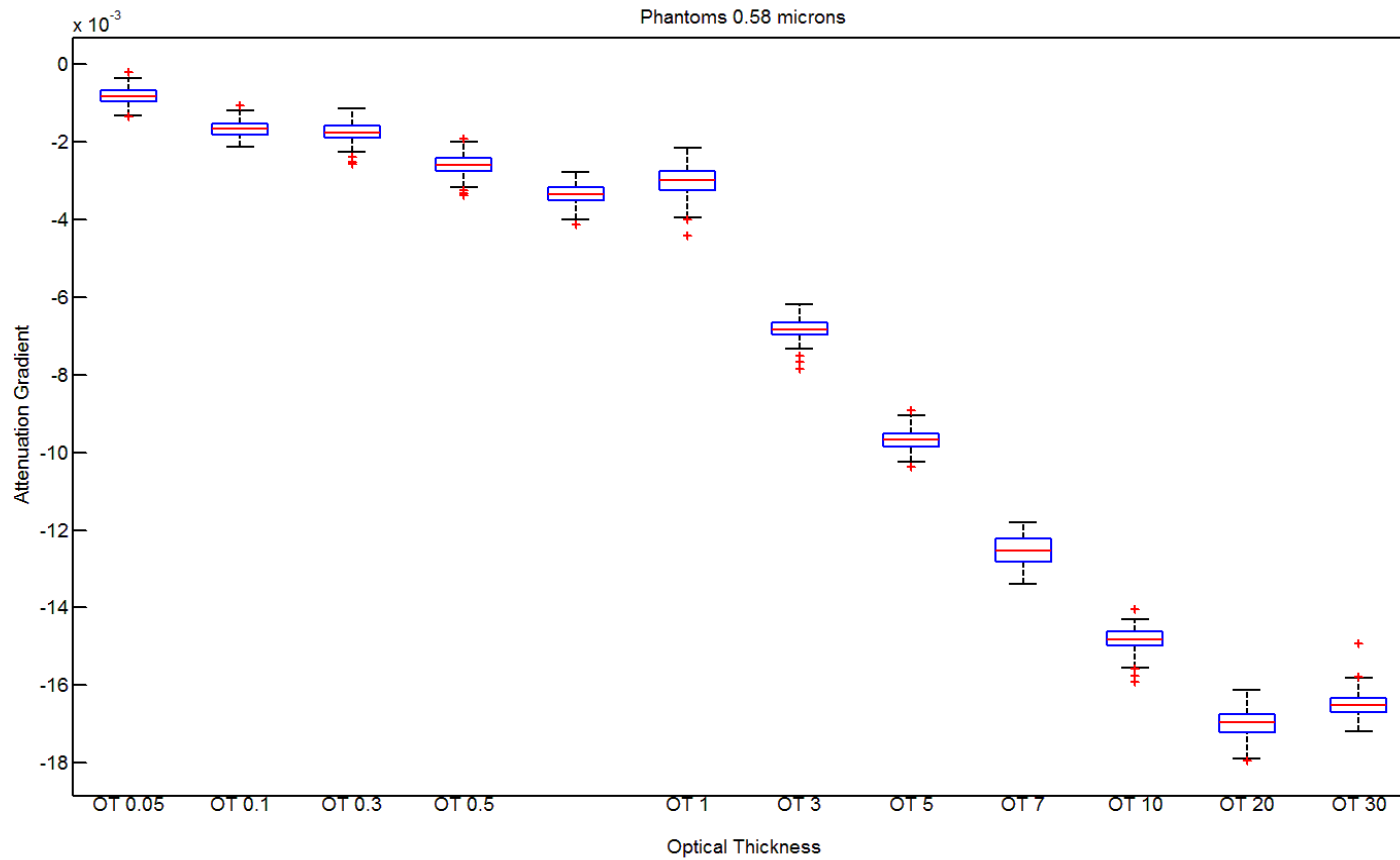
Figure 3.15: Boxplots of the attenuation gradients of different concentrations of 0.58 micron microsphere phantoms. Red crosses represent outliers, which were filtered out and not used in analysis.

Histology	Oesophagus	Oesoph Inflam	Stomach	Stomach Inflam	Stomach Int Met	Dysplasia	Cancer Regressed	Cancer
Median	-0.00180	-0.00180	-0.00650	-0.00740	-0.00600	-0.00580	-0.00560	-0.00640
Range	0.00740	0.01400	0.01200	0.01000	0.00730	0.00450	0.01000	0.01200
Min	-0.00580	-0.00720	-0.01200	-0.01200	-0.00950	-0.00810	-0.01000	-0.01200
Max	0.00150	0.00680	-0.00007	-0.00220	-0.00210	-0.00360	-0.00002	0.00002
1st Quantile	-0.00260	-0.00360	-0.00830	-0.00850	-0.00690	-0.00640	-0.00670	-0.00800
3rd Quantile	-0.00120	0.00120	-0.00510	-0.00610	-0.00510	-0.00530	-0.00330	-0.00480
Interquartile Range	0.00140	0.00480	0.00320	0.00240	0.00180	0.00110	0.00340	0.00310
No Of B-scans	2574	4736	7430	13685	6368	991	3650	11246

Table 3.10: Statistics of attenuation gradients obtained from gradient analysis of surgical upper gastrointestinal data in MATLAB.

Histology	Ileum	Colon	Crohn's	Dysplasia	Cancer
Median	-0.00589	-0.00735	-0.00521	-0.00592	-0.00449
Range	0.01025	0.00744	0.00789	0.00292	0.00894
Min	-0.00914	-0.01057	-0.00786	-0.00741	-0.00907
Max	0.00112	-0.00313	0.00002	-0.00449	-0.00013
1st Quantile	-0.00706	-0.00853	-0.00587	-0.00632	-0.00573
3rd Quantile	-0.00182	-0.00641	-0.00389	-0.00559	-0.00344
Interquartile Range	0.00523	0.00212	0.00197	0.00073	0.00229
No Of B-scans	1152	3409	874	1004	1151

Table 3.11: Statistics of attenuation gradients obtained from gradient analysis of surgical lower gastrointestinal data in MATLAB.
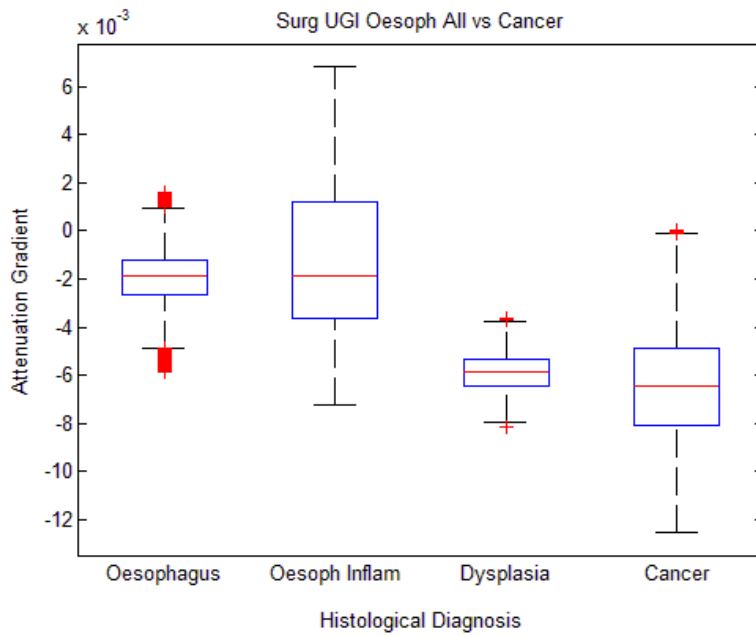
Figure 3.16: Boxplots of attenuation gradients for normal surgical oesophageal tissue and pathologies.
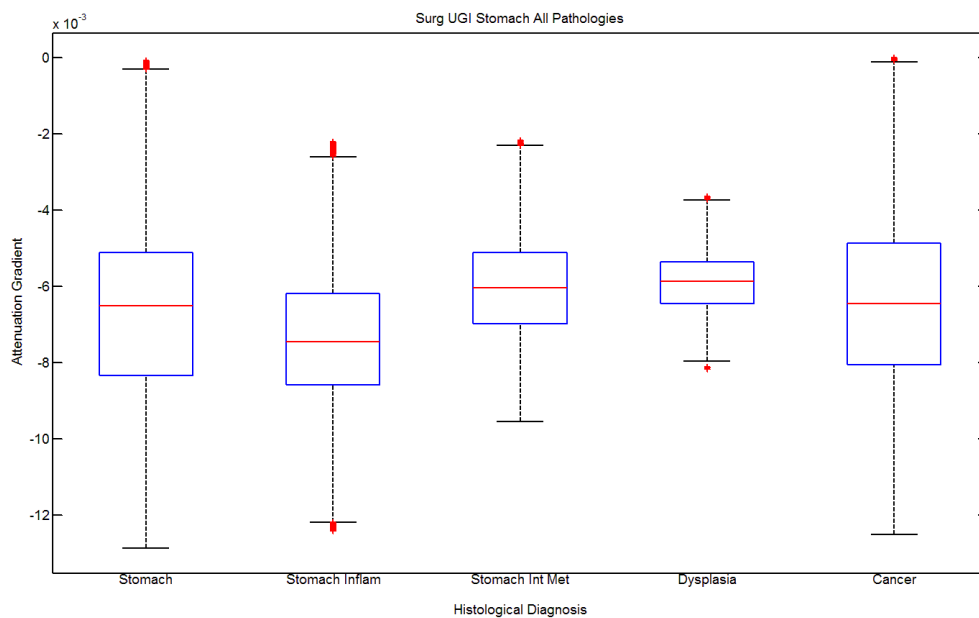
Figure 3.17: Boxplots of attenuation gradients for normal surgical gastric tissue and pathologies.
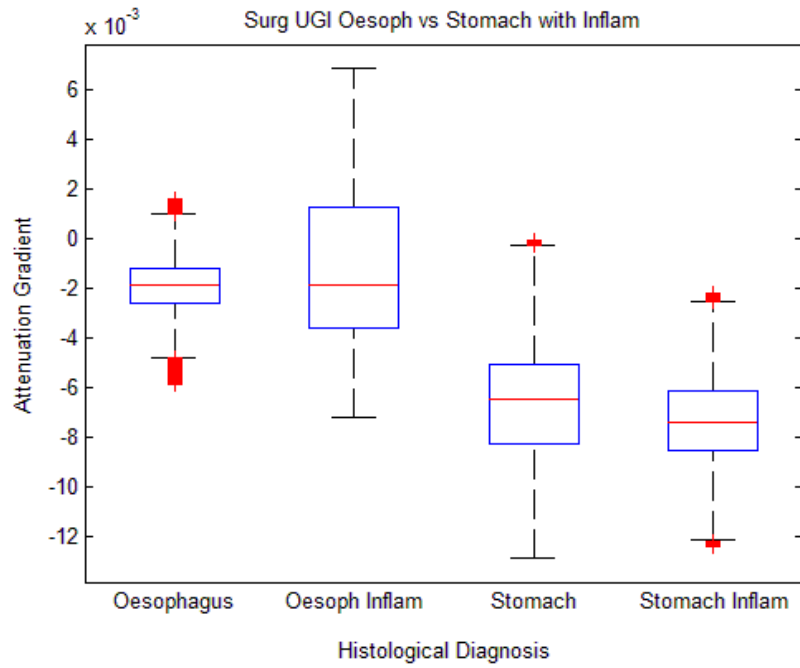
Figure 3.18: Boxplots of attenuation gradients for surgical oesophageal and gastric tissue.
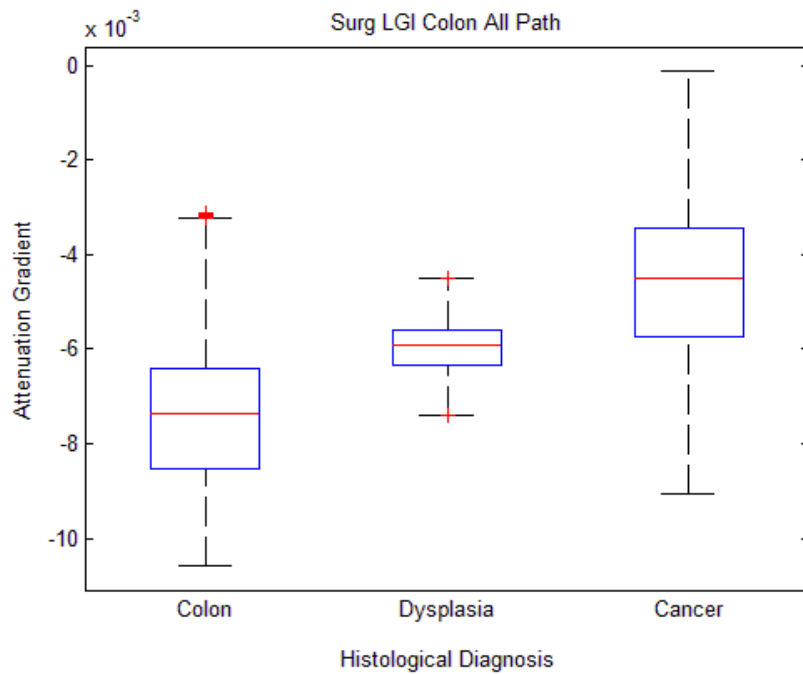
Figure 3.19: Boxplots of attenuation gradients for normal surgical colonic tissue and pathologies.
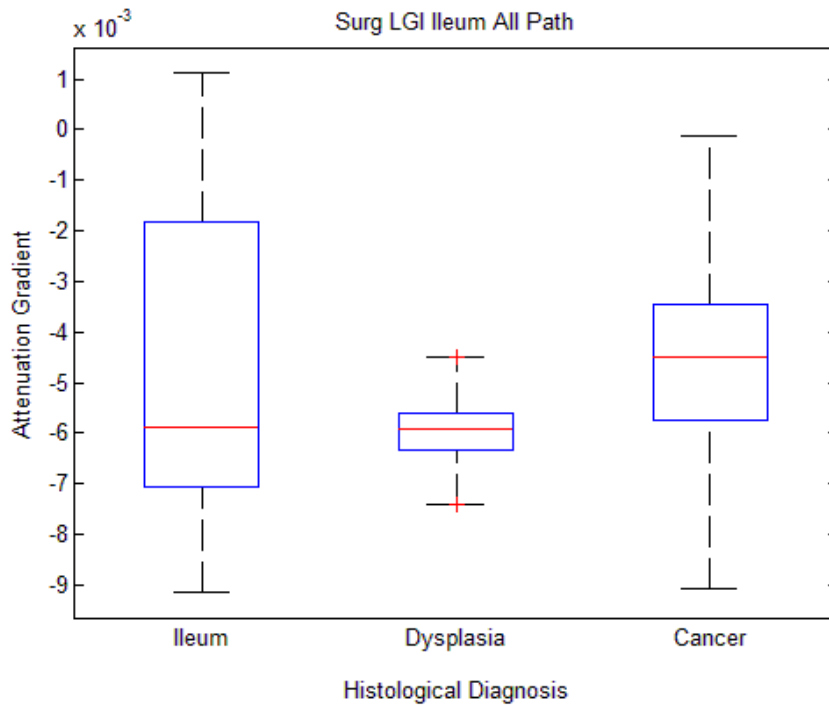
Figure 3.20: Boxplots of attenuation gradients for normal surgical ileal tissue and pathologies.

Comparison Groups	p
Oesophagus_vs_Oesoph Inflam	4.48E-08 (<0.001)
Oesophagus_vs_Dysplasia	0.00E+00 (<0.001)
Oesophagus_vs_Stomach	0.00E+00 (<0.001)
Oesophagus_vs_Cancer	0.00E+00 (<0.001)
Oesophagus_vs_Cancer Regressed	0.00E+00 (<0.001)
Stomach_vs_Stomach Inflam	8.52E-132 (<0.001)
Stomach_vs_Stomach Int Met	7.19E-71 (<0.001)
Stomach_vs_Dysplasia	9.03E-28 (<0.001)
Stomach_vs_Cancer	4.75E-07 (<0.001)
Stomach_vs_Cancer Regressed	5.02E-136 (<0.001)
Stomach Inflam_vs_Dysplasia	1.74E-168 (<0.001)
Stomach Inflam_vs_Cancer	3.91E-271 (<0.001)
Stomach Inflam_vs_Cancer Regressed	0.00E+00 (<0.001)
Dysplasia_vs_Cancer	4.62E-16 (<0.001)
Cancer Regressed_vs_Cancer	1.90E-131 (<0.001)

Table 3.12: Results of Mann-Whitney testing of attenuation gradient comparison groups for surgical upper gastrointestinal data.

Histology	Oesophagus	Oesoph Inflam	Stomach	Stomach Inflam	Duodenum	Duod Inflam	Barretts	Stomach Int Met	Cancer
Median	-0.0034	-0.0055	-0.0034	-0.0043	-0.0047	-0.0052	-0.0044	-0.0038	-0.0060
Range	0.0103	0.0186	0.0114	0.0202	0.0127	0.0101	0.0120	0.0143	0.0116
Min	-0.0087	-0.0148	-0.0090	-0.0144	-0.0111	-0.0096	-0.0102	-0.0100	-0.0120
Max	0.0016	0.0038	0.0024	0.0058	0.0016	0.0005	0.0018	0.0044	-0.0004
1st Quantile	-0.0046	-0.0078	-0.0048	-0.0067	-0.0063	-0.0067	-0.0057	-0.0060	-0.0074
3rd Quantile	-0.0019	-0.0033	-0.0020	-0.0016	-0.0032	-0.0038	-0.0028	-0.0014	-0.0047
Interquartile Range	0.0027	0.0045	0.0028	0.0051	0.0031	0.0028	0.0029	0.0046	0.0027
No Of B-scans	3584	7446	2007	16837	15209	1528	10463	2047	2962

Table 3.13: Statistics of attenuation gradients obtained from gradient analysis of endoscopic upper gastrointestinal data in MATLAB.

Histology	Ileum	Colon	Colon Inflam	Rectum	Rectum Inflam	Dysplasia	Cancer
Median	-0.0049	-0.0052	-0.0062	-0.0039	-0.0056	-0.0061	-0.0064
Range	0.0119	0.0202	0.0148	0.0202	0.0204	0.0153	0.0140
Min	-0.0107	-0.0145	-0.0129	-0.0125	-0.0148	-0.0128	-0.0135
Max	0.0012	0.0058	0.0020	0.0077	0.0056	0.0025	0.0004
1st Quantile	-0.0063	-0.0076	-0.0081	-0.0065	-0.0079	-0.0081	-0.0083
3rd Quantile	-0.0033	-0.0025	-0.0042	-0.0002	-0.0024	-0.0040	-0.0048
Interquartile Range	0.0030	0.0051	0.0039	0.0062	0.0055	0.0041	0.0035
No Of B-scans	5600	30553	14504	7168	6653	10064	7585

Table 3.14: Statistics of attenuation gradients obtained from gradient analysis of endoscopic lower gastrointestinal data in MATLAB.
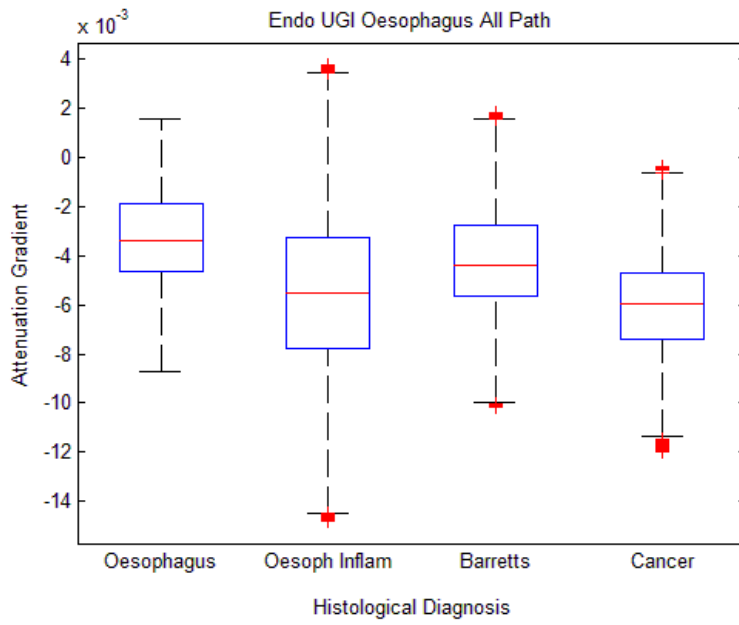
Figure 3.21: Boxplots of attenuation gradients for endoscopic oesophageal tissue specimens.
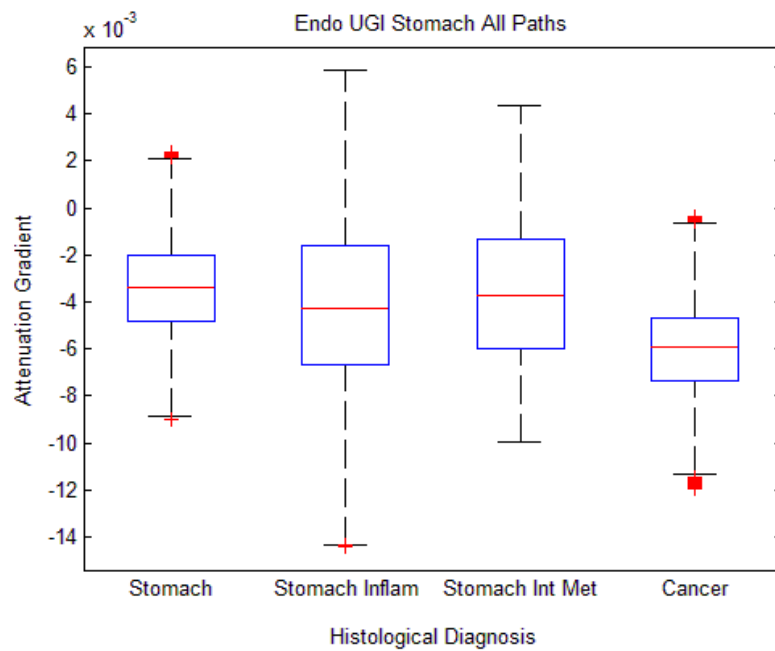
Figure 3.22: Boxplots of attenuation gradients for endoscopic gastric tissue specimens.
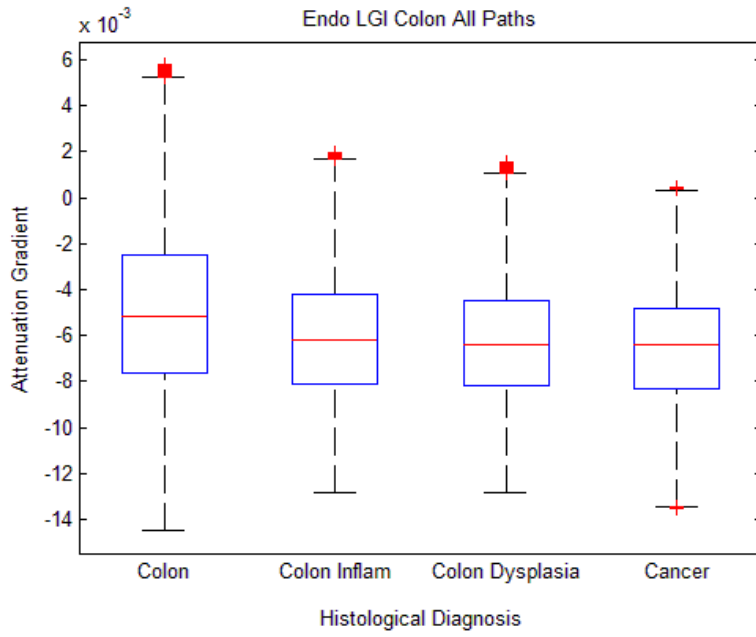
Figure 3.23: Boxplots of attenuation gradients for endoscopic colonic tissue specimens.
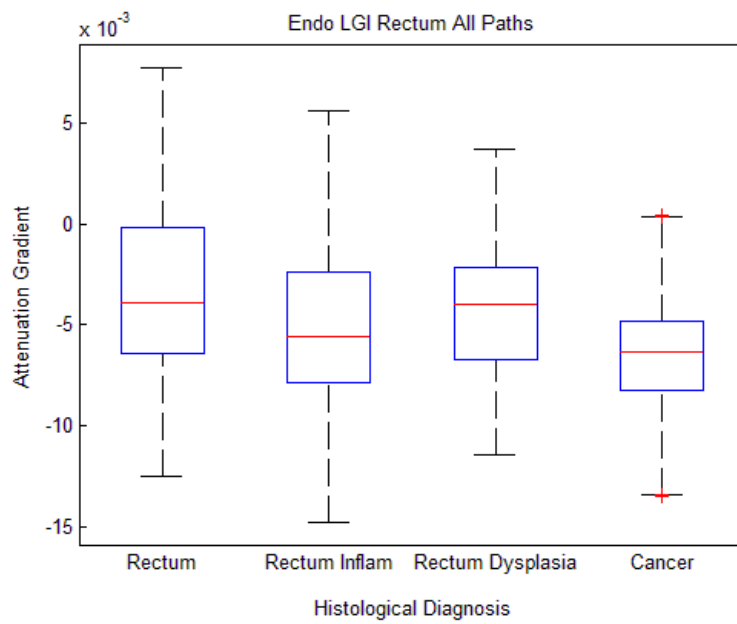
Figure 3.24: Boxplots of attenuation gradients for endoscopic rectal tissue specimens.

Comp Group	p
Oesophagus_vs_Oesoph Inflam	2.47E-303 (<0.001)
Oesophagus_vs_Barretts	2.59E-106 (<0.001)
Oesophagus_vs_Cancer	0.00E+00 (<0.001)
Oesoph Inflam_vs_Barretts	1.61E-196 (<0.001)
Oesoph Inflam_vs_Cancer	1.11E-24 (<0.001)
Stomach_vs_Stomach Inflam	6.27E-28 (<0.001)
Stomach_vs_Stomach Int Met	2.92E-04 (<0.001)
Stomach_vs_Cancer	2.37E-307 (<0.001)
Stomach Inflam_vs_Cancer	2.86E-194 (<0.001)
Colon_vs_Colon Inflam	2.020E-158 (<0.001)
Colon_vs_Dysplasia	3.3042E-98 (<0.001)
Colon_vs_Cancer	2.383E-231 (<0.001)
Colon Inflam_vs_Cancer	2.559E-27 (<0.001)
Rectum_vs_Dysplasia	1.542E-289 (<0.001)
Rectum_vs_Cancer	0 (<0.001)

Table 3.15: Results of Mann-Whitney testing of attenuation gradient comparison groups for endoscopic data.

Comparisons were also carried out between histologically normal and abnormal tissues for oesophagus, stomach, colon and rectum, with "abnormal" encompassing all pathologies (cancer, dysplasia, metaplasia, inflammation etc.). The aim was to see if gradient analysis could identify tissues where a biopsy would yield a positive result. The results of these comparisons are presented below. The median attenuation gradients and their ranges for normal vs abnormal tissues were: for oesophagus -0.00331 (-00874 to 0.0103) vs -0.00491 (-0.01174 to 0.001952), for stomach -0.00336 (-0.00904 to 0.002383) vs -0.00458 (-0.01399 to 0.004926), for colon -0.00488 (-0.01446 to 0.005777) vs -0.00615 (-0.01383 to 0.001511) and for rectum -0.00359 (-0.01253 to 0.00771) vs -0.00584 (-0.01478 to 0.003291) respectively.

Comparison Group	p
Colon vs AbnormalColon	0 (<0.001)
Rectum vs AbnormalRectum	0 (<0.001)
Oesoph vs AbnormalOesophagus	3.90E-302 (<0.001)
Stomach vs AbnormalStomach	8.48E-70 (<0.001)

Table 3.16: Results of Mann-Whitney testing of attenuation gradients for paired normal and abnormal endoscopic data.
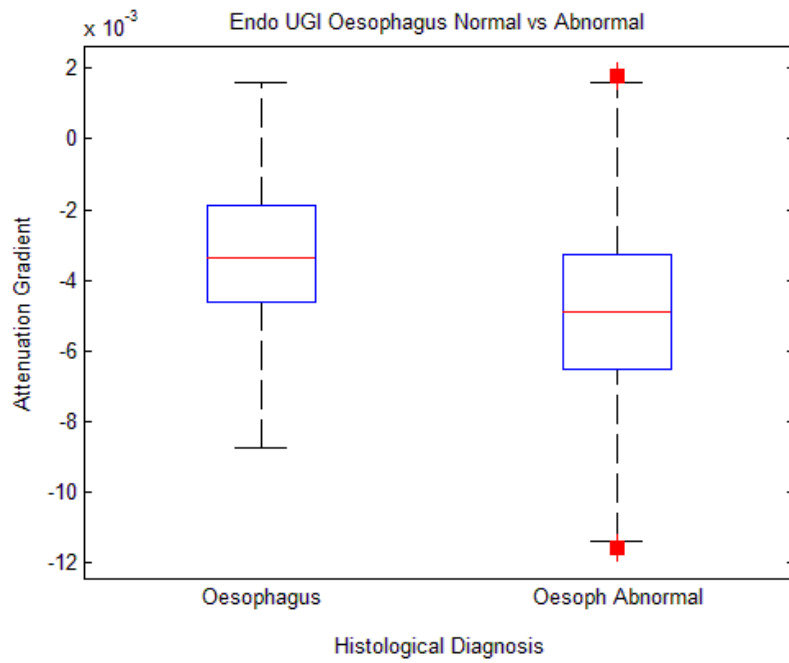
Figure 3.25: Boxplots of attenuation gradients of normal vs. abnormal oesophageal tissues.
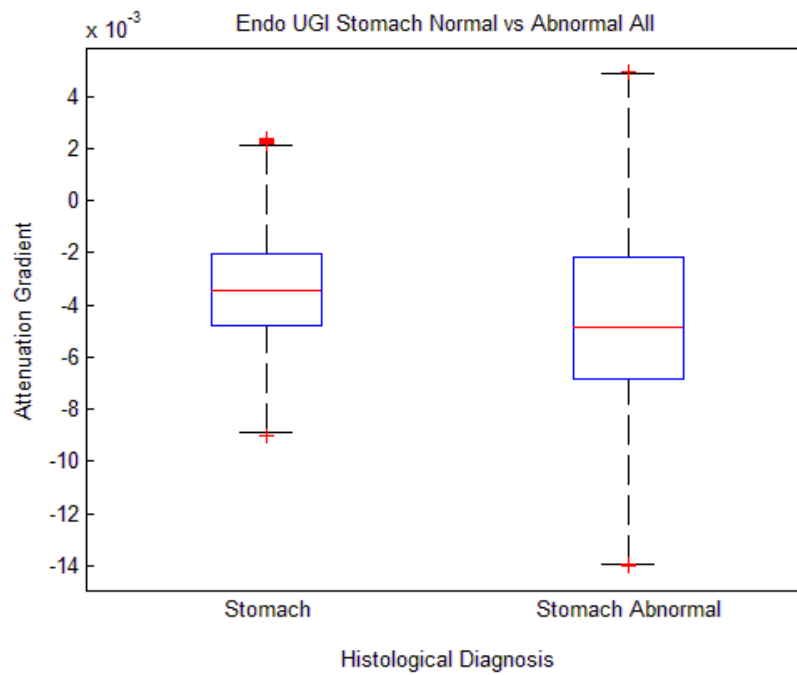
Figure 3.26: Boxplots of attenuation gradients of normal vs. abnormal gastric tissues.
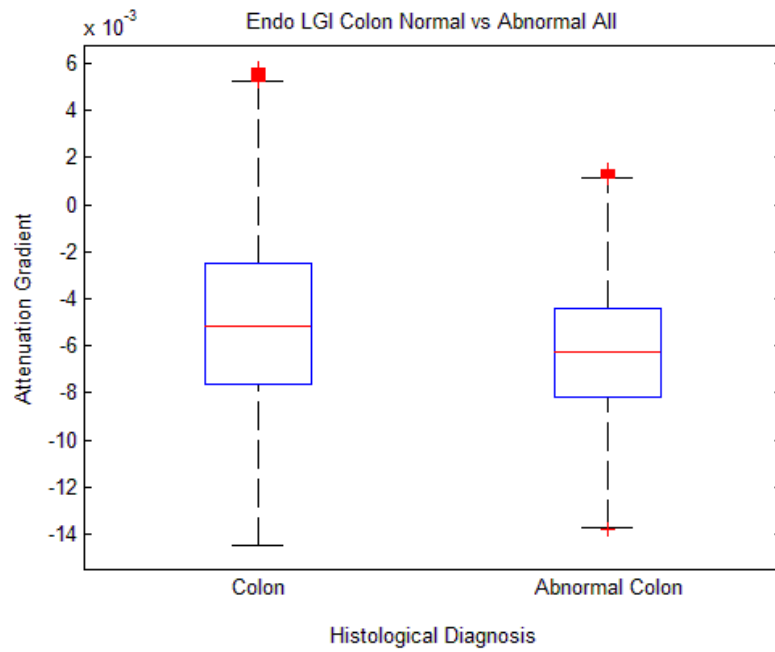
Figure 3.27: Boxplots of attenuation gradients of normal vs. abnormal colonic tissues.
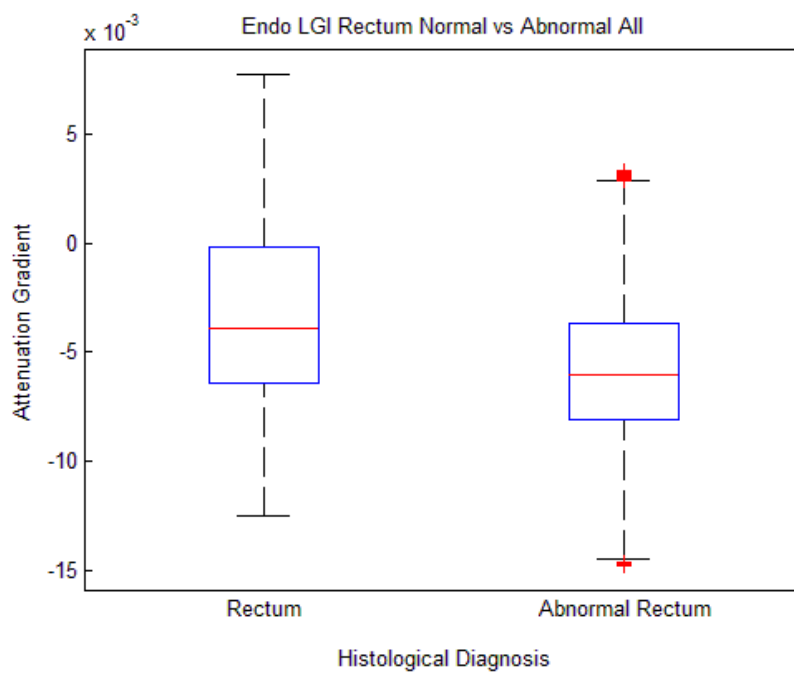
Figure 3.28: Boxplots of attenuation gradients of normal vs. abnormal rectal tissues.

3.2.5 Discussion

Testing on phantoms has yielded the expected results, with attenuation gradients increasing with more concentrated microsphere suspensions of all sizes. This shows that our system is able to consistently distinguish between less and more scattering media.

Concerning the data from surgical specimens, oesophageal data once again display the most convincing differences between healthy oesophagus, dysplasia and cancer, as seen in Figure 3.16. Inflammation is not so easy to distinguish from healthy, but this can be justified by the fact that there are many grades of inflammation and there are no major structural changes occurring by inflammation alone. In the stomach data there seems to be more overlap in the distributions of healthy and diseased gastric tissues, with some discernible differences between different tissue categories, albeit not necessarily following the expected pattern. Again, there is a clear difference between oesophagus and stomach (i.e. squamous and glandular epithelium). In the lower GI data there is a difference between healthy, dysplasia and cancer, but the reverse of what is expected is shown, i.e. the scattering seems to decrease with cancer compared to healthy tissues. This can be seen in Figure 3.19 and Figure 3.20.

Endoscopic oesophageal data is again very good, showing clear differences between healthy oesophagus, inflammation, Barrett's and cancer (Figure 3.21). The gastric data is also better here, with a clear difference between healthy stomach and cancer, and less difference between healthy, inflammation and int. metaplasia. The pattern is also as expected, with increased scattering from healthy to diseased tissues. Colon and rectum data also shows a difference between healthy, dysplasia and cancer and this time the trend is also as expected (Figure 3.23 and Figure 3.24). In the comparison of normal vs abnormal, oesophageal tissues demonstrated the greatest difference again in terms of the boxplots, though a difference could be seen for all tissues, which is promising for the future potential of this technique in guiding biopsy-taking during endoscopy.

Mann Whitney statistical analysis shows significant p values across all comparisons. This shows that there is a difference in attenuation gradients, even if not clearly visible from the boxplots alone, but a classifier is still needed in order to guide decision making in real-time.

Compared to our previous study, this work has definitely improved on many aspects, including specimen collection, histological reporting and data processing. The results are

statistically significant and more in line with what is expected, partly due to improvements in our technique and partly due to increased sample numbers. A large number of both surgical and endoscopic specimens were collected and a variety of pathologies were studied.

An improvement on the preliminary study, which stemmed directly from the data processing and particularly the matching of histological diagnoses to OCT measurements, was the use of ink to mark measurement sites in surgical specimens. In the preliminary study only the small pins and "measurement maps" were used to guide block-taking. Additionally, the author was present during block-taking in this study. This ensured that imaged areas were correctly identified by the histopathologist, who could actually see the ink on the slides and report on the tissues directly under the surface. However, even ink is not the perfect solution, as it can also get washed away or smear during fixing in formalin. Taking smaller tissue blocks that only included the area imaged was also tried, but this resulted in tissues getting twisted and sheared off. In any case, it worked better than in our previous study and further improvements can be carried out in the future.

Histological reporting was carried out by the same histopathologist for all specimens (surgical and endoscopic) and for specimens collected previously too. Instead of the standard technique used for clinical purposes, where the presence of any pathology in the slide would lead to classification under that pathology, this time attention was paid to the quantity and location of pathological tissues. If, in a section consisting of healthy tissue for example, there was a very small amount of cancer tissue in the area imaged by the OCT system, or the cancer tissue was not directly under the area imaged (i.e. under the inked part of the slide) or if it was deeper than 2mm -the maximum depth of penetration of the OCT light beam- then the slide would be classified as healthy tissue, rather than cancer. This led to the reclassification of a small amount of data collected in the first study, the slides of which were re-examined before pooling all available data together.

As described previously, this time processing was carried out on raw OCT data, rather than the JPEG images exported by the system's software. This allowed us access to uncompressed data of higher quality, as well as greater flexibility in applying correction algorithms and fitting our data without unwanted features such as apodization or Decibel scaling. The software also worked much faster and could be automated to handle large volumes of data. We also opted to carry out averaging only across each b-scan for noise reduction, but pooled

all averaged b-scans from each c-scan of the same diagnostic category together, as seen in the literature [77]. This was more representative of the true variation of attenuation gradients in each sample and allowed for filtering of outliers across a complete range of values for each diagnostic category. For example, values that may have been considered non-outliers for each individual c-scan, became outliers when all c-scans were combined.

No en-face colourmaps of tissues were produced this time, as that still relies on subjective interpretation. Instead, once a suitable classifier has been applied in future work, a simple traffic light system could be used to highlight "normal tissue" or "pathology".

Nevertheless, there are still areas which could be improved in the future. In terms of specimen processing, marking of imaged areas and histological reporting can still be improved. Marking with ink is not perfect solution, as it tends to smear or even come off completely and is not always visible on slides, as layers can get mixed up in the cutting process. In general, it can be really difficult to get a precise slice from the area of imaging, as there is a large volume of clinical specimens being processed on a daily basis and lab staff are busy. Ideally, dedicated histopathology staff would carry out block-taking and reporting of research specimens, but in practice that is difficult to apply.

Our data processing algorithm can also be improved. We have used a limited model that assumes single scattering and uses a simple fit to model the signal attenuation. In future we can try with different fits, such as exponentials or work on a model that simulates multiple scattering and attempt to extract other parameters, such as the anisotropy factor g.

All in all, our study has shown strongly positive results and a lot of originality compared to the literature. To our knowledge, this is the first study to carry out gradient analysis of OCT data from gastroenterological tissues and to investigate such a variety of GI tissues and pathologies. It is also unique in the large sample numbers and freshness of specimens and biopsies, something that we could achieve thanks to our multidisciplinary approach and the fact the study was driven by a clinician.

The large sample numbers also lend credibility to our results. Studies so far have been carried out on small sample numbers, a lot of them mainly as proof of principle. Here, particularly for our oesophageal data, we have shown differences between healthy and

diseased tissues, which can form the basis of future in vivo diagnostics. In stomach the diagnostic ability of OCT is more limited, as the thick and highly scattering gastric mucosa restricts imaging depth and contrast [68, 73].

The greater difference in oesophagus can probably be explained by the fact that it is squamous, so when it becomes dysplastic or (adeno-) cancerous, there is a big difference as it turns columnar. In the glandular epithelium of gastric and colonic tissues the difference is less marked, particularly for early, well-differentiated tumours which do not disrupt the anatomical layers so much.

3.3 Morphological analysis and K- Nearest Neighbours classification of Optical Coherence Tomography data from surgical and endoscopic specimens

3.3.1 Background

In our previous studies we used linear gradient analysis to model OCT signal attenuation in different tissues and established that there is a significant difference in attenuation gradients between healthy and diseased tissues, which could be used for diagnostic purposes in vivo. We concluded that a classifier was needed that would be powerful enough to diagnose gastrointestinal disease using OCT with good sensitivity and specificity values. Here we look at a different method of analyzing OCT data, called morphological analysis and use k-Nearest-Neighbours and Principal Components Analysis to classify data into different pathologies.

Morphological analysis is a form of image segmentation, which separates the image into different clusters according to parameters such as grey-level intensity of individual pixels [85]. It has been previously demonstrated in images such as mammograms [85] or spectroscopic scatter images of healthy and cancerous tissue [86].

Generally, an algorithm is used to detect differences in pixel intensity and to separate regions of the image accordingly. Algorithms previously used include fuzzy C-means and κ - means. The former allows each pattern to be associated with multiple intensity clusters, whereas the latter strictly associates each pattern with one cluster [85, 87]. The number of clusters is defined by the user, but the initial seed points (intensity values around which the clusters are defined) are usually set randomly [85, 87]. The κ – means is an example of a squared error algorithm and is one of the simplest most commonly used, as it has low processing time requirements [87]. It starts with a random initial partition for separating values into clusters and keeps reassigning them until a pre-defined endpoint is reached. This can be when no further reassignment is possible or when the squared error no longer decreases significantly after a number of iterations [87].

Various features can be then extracted from the values comprising the different clusters in an image. These can be descriptive statistics, such as mean and standard deviation, degrees of entropy or homogeneity, or more abstract concepts, such as distances on three-

dimensional axes [85]. These features are then generally analysed using classifiers, which aim to group them according to parameters set by the user, such as diagnostic/histological labels, and then predict to which group new values should be assigned to. Classifiers employed previously include k-Nearest-Neighbours (kNN) and Principal Components Analysis (PCA) [85], both of which are used in this work.

The kNN classifier assigns an unclassified value or vector using the closest κ vectors in a training set [67, 85, 86]. The number of points from which the distance is measured (neighbours) can be set by the user [67], as can the type of distance measured, such as Euclidean, which measures proximity in two or three-dimensional space [67, 85, 86] or Mahalanobis, which assigns weights to each feature to avoid distortions caused by certain features being more frequent than others [67, 87]. To prevent some features being more heavily weighted than others, the dataset needs to be normalised before applying kNN classification [85].

To the author's knowledge, morphological analysis has not been previously applied to OCT images. In the following sections, morphological analysis of OCT images from surgical and endoscopic specimens and classification using PCA and kNN according to histological diagnosis is described.

3.3.2 Aims

The aim of this study is to classify gastrointestinal specimens into diagnostic categories through morphological analysis and kNN classification of OCT images.

3.3.3 Methods

The same surgical and endoscopic dataset described in Section 3.2 was used for this study. As described previously, raw OCT data was imported into the MATLAB environment for processing, where error correction and image alignment was carried out.

Morphological analysis was then carried out on the resulting images. This involved a statistical analysis of the light intensity distribution in each B-scan belonging to each measurement. The variation in brightness in the vertical direction was computed and the

image was split into vertical segments, corresponding to areas with similar intensity distributions. The κ – means technique was used for this and it involved three steps: firstly, a set of initial κ – clusters was defined; secondly each intensity value –corresponding to a pixel in the image- was assigned to the cluster whose mean is closest to the value; thirdly, the process is repeated, with clusters evolving and changing means as new observations are added, resulting in the reassignments of previous observations till a final equilibrium is reached, with no more new observations and no reassignments.
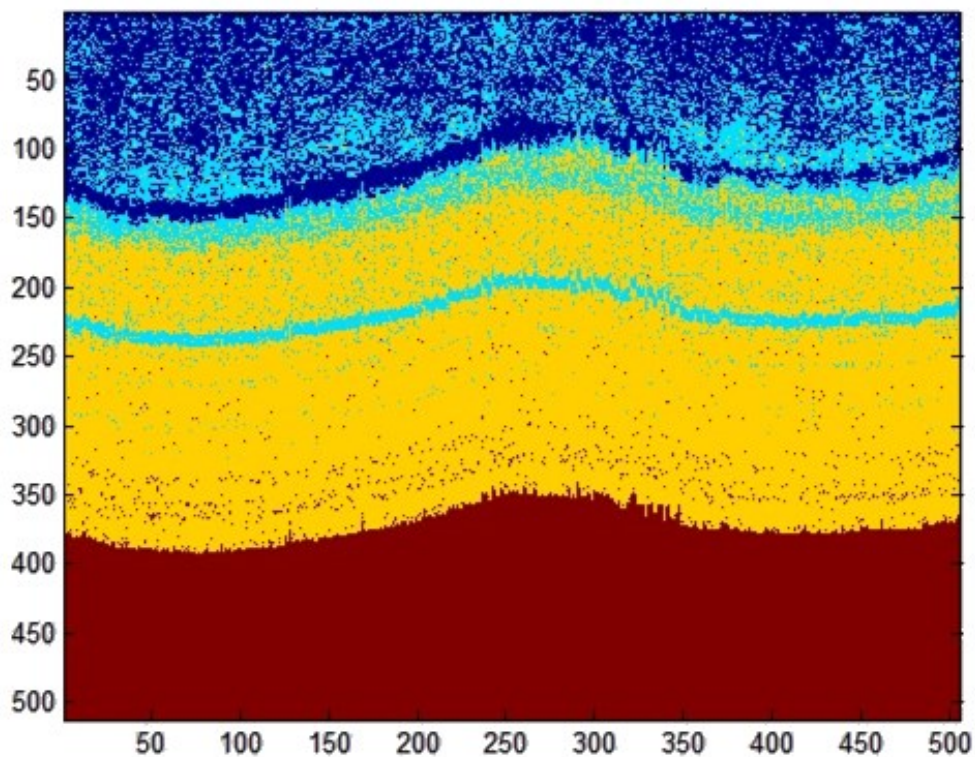
Figure 3.29: Example of B-Scan after morphological analysis: the image has been split into different layers according to the morphological features.

To improve the performance of the algorithm, the vertical intensity distribution of each B-scan was calculated and the initial clusters were centred around evenly distributed κ percentiles, where κ is the number of regions the image was split in and was defined as 3 for this study. The number can be set to match the number of tissue layers, so that the algorithm may attempt to map intensities to tissue layers, but in our case we optimised it for processing efficiency, choosing the smallest number that was capable of distinguishing

between tissue pathologies with good accuracy in order to reduce computing requirements [67]. A typical segmented B-scan is shown in Figure 3.29.

For each segment four statistical parameters of the intensity distribution were calculated: the mean, the standard deviation, skewness and kurtosis. Additionally, the ratio of the area of the segment compared to the total area of the B-scan was calculated for each segment. Therefore, 5 parameters were calculated for each segment, giving a total of 15 parameters for each B-scan.

The final step involved kNN analysis of our data. The number of neighbours was set to 1 and the Euclidean distance was used. The classification algorithm was applied on the morphological data, using the 15 parameters per B-scan described above to classify C-scans from a testing dataset into different diagnostic categories according to a training dataset. The cross-validation function in Matlab was used for this purpose. Firstly, the morphological data from surgical specimens was split into 3 parts. One part formed the test set, whereas the other two formed the training set. The accuracy, sensitivity, specificity, negative predictive value (NPV) and positive predictive value (PPV) per diagnostic category compared to the others was computed using standard Matlab algorithms and the process was repeated for each dataset. The performance of the classifier was then calculated by averaging the resulting errors from the three analyses , thus eliminating the dependency of classification results on the training or test sets. The process was repeated for endoscopic data. The code was configured in a way that allows the user to pick the data subsets to be compared (e.g. oesophageal data only, or oesophagus vs stomach, or all normal specimens vs all cancer specimens).

Two pathways were implemented. In the first one, the morphological features were first analysed using PCA and then the kNN classifier was run on the PCA scores. Graphs were also produced showing the separation of datasets in 3 PCA axes. In the second pathway, kNN classification was applied directly to the morphological features of the OCT data. This allowed us to run two classification techniques (PCA and kNN) on the datasets and also investigate whether the combination of PCA and kNN improved our results compared to kNN alone.

Surgical data was processed using both pathways. As no real advantage was noticed with the application of PCA as a first step, and because of processing power demands, the step was skipped for endoscopic data.

3.3.4 Results
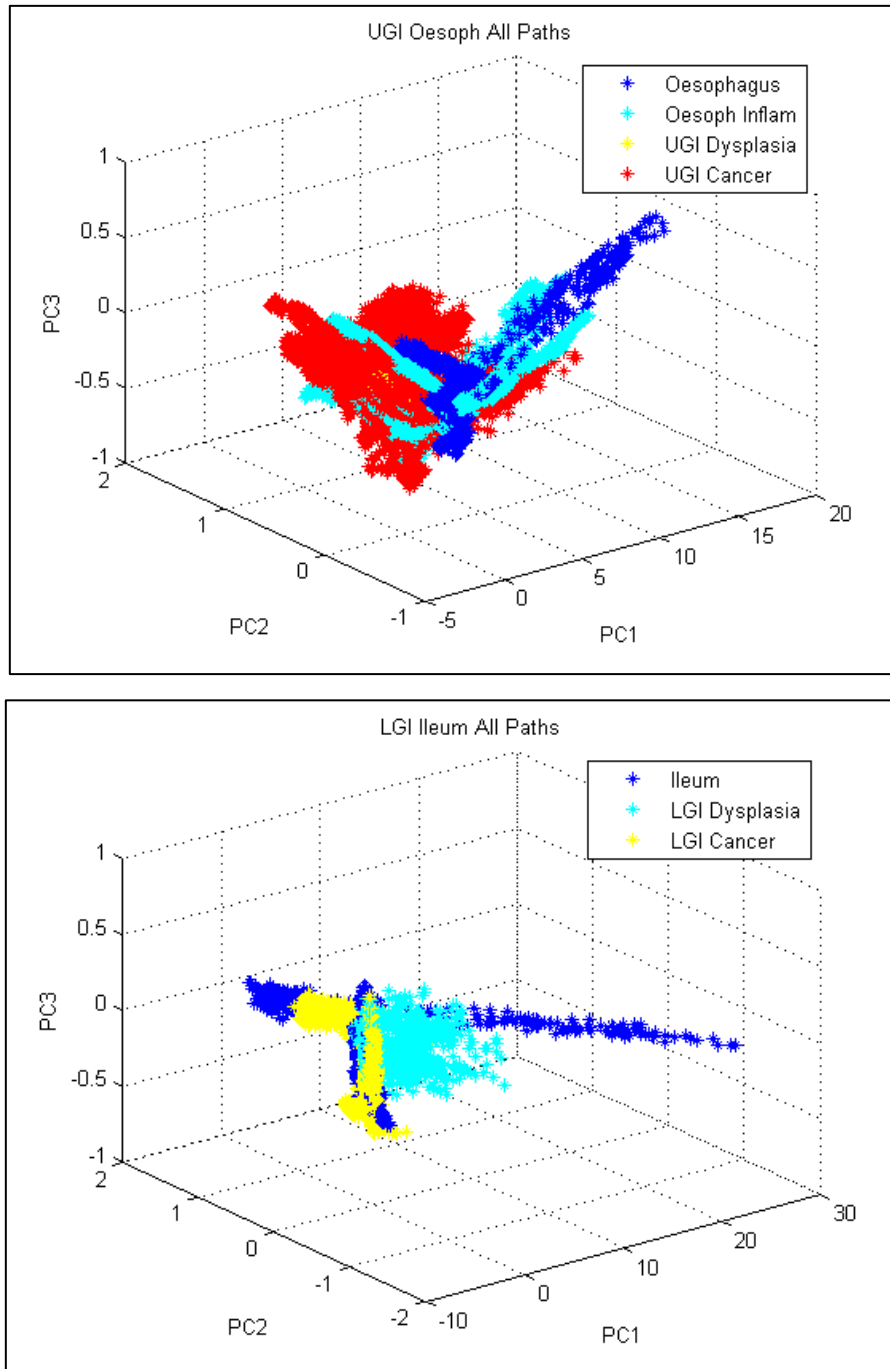
Figure 3.30: Results of Principal Components Analysis of morphological features of oesophageal (top) and ileal (bottom) surgical data.

Surgery Oesophagus					
Histology	Sensitivity	Specificity	PPV	NPV	Accuracy
Without PCA					
Oesophagus	1	0.978	0.887	1	0.981
Oesoph Inflam	1	0.996	0.989	1	0.997
UGI Dysplasia	1	0.591	0.116	1	0.612
UGI Cancer	0.999	0.999	0.999	0.998	0.999
With PCA					
Oesophagus	1	0.951	0.774	1	0.958
Oesoph Inflam	0.999	0.984	0.952	0.999	0.988
UGI Dysplasia	1	0.304	0.0718	1	0.339
UGI Cancer	0.992	0.997	0.997	0.990	0.994

Table 3.17: kNN analysis on data from surgical specimens of oesophagus with and without PCA.

Surgery Stomach					
Histology	Sensitivity	Specificity	PPV	NPV	Accuracy
Without PCA					
Stomach	1	0.872	0.640	1	0.895
Stomach Inflam	0.998	0.982	0.967	0.999	0.987
Stomach Int Met	1	0.840	0.544	1	0.865
UGI Dysplasia	1	0.372	0.0394	1	0.388
UGI Cancer	0.997	0.987	0.968	0.999	0.990
With PCA					
Stomach	1	0.875	0.646	1	0.898
Stomach Inflam	1	0.990	0.983	1	0.994
Stomach Int Met	1	0.856	0.570	1	0.879
UGI Dysplasia	1	0.157	0.0296	1	0.178
UGI Cancer	0.999	0.995	0.988	0.999	0.996

Table 3.18: kNN analysis on data from surgical specimens of stomach with and without PCA.

Surgery Ileum					
Histology	**Sensitivity**	**Specificity**	**PPV**	**NPV**	**Accuracy**
Without PCA					
Ileum	0.995	0.994	0.986	0.998	0.994
LGI Dysplasia	1	0.938	0.855	1	0.954
LGI Cancer	0.998	0.965	0.957	0.999	0.980
With PCA					
Ileum	1	0.943	0.884	1	0.960
LGI Dysplasia	0.999	0.971	0.927	0.999	0.978
LGI Cancer	0.998	0.971	0.964	0.999	0.983

Table 3.19: kNN analysis on data from surgical specimens of ileum with and without PCA.

Surgery Colon					
Histology	**Sensitivity**	**Specificity**	**PPV**	**NPV**	**Accuracy**
Without PCA					
Colon	1	0.998	0.998	1	0.999
Crohn's	1	0.957	0.785	1	0.962
LGI Dysplasia	1	0.866	0.557	1	0.885
LGI Cancer	1	0.960	0.885	1	0.969
With PCA					
Colon	0.986	0.985	0.984	0.987	0.985
Crohn's	1	0.929	0.688	1	0.938
LGI Dysplasia	1	0.907	0.644	1	0.920
LGI Cancer	0.999	0.768	0.569	0.999	0.822

Table 3.20: kNN analysis on data from surgical specimens of colon with and without PCA.

Endoscopy Oesophagus					
Histology	**Sensitivity**	**Specificity**	**PPV**	**NPV**	**Accuracy**
Oesophagus	0.999	0.978	0.889	0.999	0.981
Oesoph Inflam	0.999	0.980	0.959	0.999	0.986
Barrett's	0.998	0.990	0.988	0.998	0.993
UGI Cancer	1	0.796	0.313	1	0.813

Table 3.21: kNN analysis on data from endoscopic specimens of oesophagus.

Endoscopy Stomach					
Histology	Sensitivity	Specificity	PPV	NPV	Accuracy
Stomach	1	0.667	0.231	1	0.697
Stomach Inflam	0.960	0.998	0.999	0.893	0.969
Stomach Int Met	1	0.761	0.234	1	0.777
UGI Cancer	1	0.842	0.389	1	0.857

Table 3.22: kNN analysis on data from endoscopic specimens of stomach.

Endoscopy Colon					
Histology	Sensitivity	Specificity	PPV	NPV	Accuracy
Colon	0.973	0.990	0.989	0.975	0.982
Colon Inflam	0.999	0.831	0.634	0.999	0.869
LGI Dysplasia	0.999	0.814	0.508	0.999	0.844
LGI Cancer	1	0.621	0.266	1	0.667

Table 3.23: kNN analysis on data from endoscopic specimens of colon.

Endoscopy Rectum					
Histology	Sensitivity	Specificity	PPV	NPV	Accuracy
Rectum	0.999	0.962	0.886	0.999	0.970
Rectum Inflam	0.999	0.832	0.612	0.999	0.867
LGI Dysplasia	0.995	0.960	0.922	0.997	0.971
LGI Cancer	0.999	0.835	0.659	0.999	0.875

Table 3.24: kNN analysis on data from endoscopic specimens of rectum.

Endoscopy normal tissues vs. all of their corresponding pathologies					
Histology	Sensitivity	Specificity	PPV	NPV	Accuracy
Oesophagus	0.999	0.980	0.890	0.999	0.982
Stomach	1	0.632	0.198	1	0.663
Colon	0.974	0.989	0.989	0.975	0.982
Rectum	0.998	0.963	0.887	0.999	0.971

Table 3.25: kNN analysis of paired endoscopic data (normal vs. all corresponding pathologies. The corresponding abnormal groups have been left out of this table.

3.3.5 Discussion

In this study we have demonstrated a novel methodology for analysis and classification of OCT data. In the first step a morphological analysis is carried out, splitting each B-scan into 3 regions and producing 5 parameters for each one (mean, standard-deviation, skewness, kurtosis and relative area of each segmented region to total B-scan area), resulting in a total of 15 parameters per B-scan. A kNN classifier was then used to predict the diagnostic group each dataset belongs to after splitting data into training and testing datasets. In contrast to other segmentation techniques in the literature, this technique does not require any previous information about tissue anatomy as the segmentation and ROI selection is carried out automatically based on intensity [67]. A similar technique for morphological analysis has been previously used for classification of tissue density in mammographic data [85], but this is the first time it has been used on OCT data.

The basic principle of morphological analysis is the same as that of gradient analysis, namely that the changes in tissue structure affect the light scattering within the tissue and thus the light intensity at different depths. This approach, however, is more sophisticated and takes into account the whole image, not just the surface pixels which we defined earlier as our ROI. This way more information is included in the analysis, which is particularly important in tissue sections where the pathology does not lie directly below the surface.

For demonstration purposes we have shown some results from PCA analysis of surgical data. These show a good clustering and separation between certain pathologies in the three-dimensional space. In oesophagus there is a distinction between cancer and healthy tissue, with data from inflammation clustering in-between. In ileal tissue there are distinct clusters belonging to healthy, dysplasia and cancer groups. As can be seen in Tables 3.17-3.20, PCA analysis did not lead to significant improvements in kNN classification, so it was not carried out on the endoscopic dataset.

In the surgical data, the best results from kNN classification were obtained from oesophagus whereas the worst were from stomach. Sensitivity and specificity values for healthy oesophagus and cancer reached 100% / 97.87% and 99.99% / 99.91% respectively, whereas corresponding values for stomach were 100% / 87.22% and 99.78% / 98.73%. From the pathology subsets, dysplasia was the one with the lowest accuracy in each group, with a

specificity ranging from 37.24% in stomach to 86.62% in colon. This would be in accordance with the fact that dysplasia is a stage in between healthy tissue and cancer in the process of carcinogenesis and would thus contain elements of both, making it hard to distinguish. However, sample numbers are probably a contributing factor here, as can be seen when comparing dysplasia and cancer across the surgery and endoscopy groups. As can be seen in Tables 3.6 to 3.9, there were only 4 measurements of dysplasia and 40 of cancer in the surgical data, whereas in the endoscopic group the numbers were 22 and 20 respectively. Looking at the kNN results from endoscopic data, dysplasia has a much higher accuracy, reaching a sensitivity and specificity of 99.55% / 96.01% in rectal tissue. On the other hand, the accuracy of detecting cancer has dropped across all tissues in comparison to the surgical data, where the sample numbers were higher. Nevertheless, oesophagus results are among the best along with rectum, whereas stomach tissues provide the worst results again, supporting the observations made in the gradient analysis study. Particularly noteworthy is the high accuracy in detecting Barrett's oesophagus, with sensitivity and specificity values of 99.80% and 99.02%.

Limitations of this study again include the sample numbers, particularly for certain pathology subsets, such as dysplasia in surgical specimens, as mentioned above. A specific limitation of the new technique employed here which dictated to a large extent the methodology of this study is the high processing power requirement. This severely limited our ability to test different configurations. For example, after carrying out a preliminary analysis, we settled for 3 segments per B-scan in the morphological analysis, which was a compromise between accuracy and processing power [67]. In the kNN analysis, we chose 1 neighbour for the study. As stated, increased numbers of neighbours make boundaries between classes less distinct, but also improve accuracy [67]. With a less demanding technique we could have run the analysis with different number of neighbours and compared the results. The same is true for the cross-validation parameter, which was set to 3 (splitting the data into 3 groups, 2 for training and 1 for testing). By changing the parameter and, therefore, the number of groups we could have compared results with different settings. Finally, to properly test the algorithm a blind study would be necessary, where the testing dataset would be independently collected from the training set, rather than be selected from it as in the current cross validation technique.

All in all, the method described in this study has produced good results and has been shown to be better that previously applied textural techniques [67]. Compared to the gradient analysis shown in the previous section it is cumbersome, but much more sophisticated and accurate. The next stage would be to test it on more tissues and ideally in vivo using an endoscopic system.

4 SUMMARY

This thesis has principally focused on quantitative analysis of OCT data to distinguish between healthy and pathological gastrointestinal tissues. A secondary area of work was the construction of a goniophotometer for the measurement of the anisotropy coefficient of the same tissues, with the aim of providing useful information that could improve our OCT signal analysis and validate anisotropy measurements from OCT data.

The goniophotometer used in this work was constructed based on designs in the literature, integrating different concepts for specimen mounting, light shielding etc. The light source was chosen to closely match the wavelength of our OCT system, whereas the ones in the literature were of shorter wavelengths. For initial calibrations we used titanium oxide phantoms, which were produced at varying concentrations and sliced to varying thicknesses to simulate variations in scattering coefficients of tissues. For further studies microsphere phantoms were used, whose properties were calculated using Mie theory. To produce measurements of g, the Henyey-Greenstein function was fitted to the data. Results with phantoms were mixed, approaching the predicted values and showing the expected trends with larger microsphere sizes. Our choice of phantoms -particularly microsphere sizes-, mounting system or data processing could be the reasons behind the discrepancies in the results. Nevertheless, we proceeded with measurements on human tissues, which produced scattering curves akin to those in the literature, as well as values of g denoting forward-scattering, as expected.

Because of the difficulties in procuring suitable specimens for goniometry and the added time and staff needs for microtoming and examining of tissue slides, the number of goniophotometric measurements on tissues was too limited to produce any statistically meaningful results. However, as a pilot study in human tissue goniophotometry it was successful and further work can be carried out in this area to collect more data and improve our technique for extraction of g.

For our optical coherence tomography work we used a bench top OCT system, although the ultimate aim would be to carry out this work in vivo with an endoscopic probe. However, certain factors cannot be controlled during endoscopy, such as movement of the probe relative to the mucosa (because of peristalsis and the inherent difficulty of holding the tip of an endoscope completely motionless against the mucosa), mucus secretions and pressure on the tissue surface which can affect the images and optical properties obtained [72, 74].

Therefore, using an ex vivo approach as a first step is a good strategy for developing the methodology and validating the data processing and classification techniques under controlled conditions. It is also much easier to precisely mark the imaged areas for histological examination and does not add time to or complicate the logistics of operations and endoscopies.

In our study we used both surgical specimens and endoscopic biopsies. This allowed us to study a larger variety of pathologies than if we had focused on just one type, as there was more cancer in the surgical specimens for example and more dysplasia in the endoscopic ones. Barrett's oesophagus was only seen in the endoscopic specimens, though there were probably also areas with Barrett's which we did not image. Surgical specimens were particularly useful when developing our data gathering methodology, as in general the diseased and healthy areas were clearly separated and it was possible to carry multiple measurements from different areas of each specimen and clearly mark the measurement sites for the histologists. Surgical specimens were also large enough to provide tissue samples for goniophotometry, though in many it was not possible to sample tumour tissue due to the location or size of the tumour, particularly post-chemotherapy. In terms of sample numbers, endoscopy provided us with far more than surgery.

Regardless of the type of specimen used, proper handling of tissues was of paramount importance for this study, both for the quality of our measurements and in order to ensure that the histopathological diagnosis for clinical purposes was not compromised. This was achieved in our work, where both types of specimens were collected fresh, imaged within minutes of excision and placed in formalin without sacrificing any quality in the histological examination. A key factor was the fact that the study was run by a clinician, which meant good access to operating theatres and endoscopy suites, the ability to independently process specimens (for example dissection of surgical specimens) out of hours and a good understanding of the histological examination process. A large number of both types of specimens was collected from multiple patients, encompassing a variety of pathologies, in contrast to the majority of studies in the literature which have worked with small numbers of frozen tissues, which have sometimes been homogenised too. In terms of the number and quality of gastrointestinal tissues used, our study is unique.

We employed two techniques for the quantitative analysis of our OCT data. The first one was gradient analysis of the OCT signal, which looks at the attenuation in the signal caused by scattering in tissues. The linear model used is relatively quick and simple, but is limited to the surface of the samples, where single scattering is assumed to occur. Our pilot study with the compiled software produced some promising results with oesophageal specimens, but mainly served to test our methodology. Indeed, we found some problems, such as the fact that some of the areas assumed to contain tumour in fact contained only fibrous tissue due to regression after chemotherapy or contained tumour at a deeper level than what was imaged by our OCT system. This led to a change in the way research slides were reported on as described in the relevant section. Improvements were also carried out on the gradient analysis software. We developed code that allowed us to analyse the raw OCT data and carry out statistical corrections before applying our gradient analysis. The results from the second study where statistically significant and showed particularly good separation of healthy and pathological values in oesophageal and rectal tissue.

The second technique involved a novel form of image segmentation called morphological analysis followed by classification using k-Nearest-Neighbours. Morphological analysis has been used before, but this is the first time it has been applied to OCT images. The technique split the OCT images into segments based on the light intensity at different depths and calculated statistical parameters for each segment, resulting in 15 parameters per image (or B-scan). These parameters were then fed into our classifier, which split the data into training and testing sets and used kNN to classify individual measurements into diagnostic groups (healthy, dysplasia, cancer etc.). Results were very good across all categories, with high sensitivity and specificity values. The best results were again obtained from oesophageal measurements with sensitivity and specificity values for healthy oesophagus and cancer in surgical specimens reaching 100% / 97.87% and 99.99% / 99.91% respectively and high accuracy in detecting Barrett's oesophagus in endoscopic specimens, with sensitivity and specificity values of 99.80% and 99.02%. Results in rectal tissue where also noteworthy, with detection of dysplasia reaching a sensitivity and specificity of 99.55% / 96.01%.

Despite not meeting our aims regarding goniophotometry and anisotropy measurements, we have produced good results with our quantitative analysis of OCT data. We have employed original techniques such as morphological analysis and predictive classification of our data using kNN. Our detection accuracy is high enough to potentially be used as a

screening test for oesophageal and rectal pathologies, but our results need to be validated with a blind study. In its current state our work has set the foundations for further research *in vivo*. By modifying our techniques to work with an endoscopic OCT probe, we would be able to address the area in which we see the most potential benefit, namely the missed diagnoses and non-selective biopsies in endoscopy described in the introduction to this thesis.

There is certainly a lot of potential for further work in this area, regardless of whether this is carried out *ex vivo* or *in vivo*. One point which has already been made is that natural next step is to work with an endoscopic OCT system [88-91]. This would allow us to work with perfused tissues, detect and image areas of interest in real time and carry out clinical studies without necessarily relying on biopsies. We can also compare the ability of endoscopists to detect suspicious areas to the results of OCT imaging of those same areas. The effect of polarised OCT on the diagnostic ability would be another interesting area of research, possibly offering increased accuracy and better signal-to-noise ratio [90, 92].

In terms of software, a great leap would be to make the software more efficient and use hardware which would allow real-time processing. Certain groups have looked into using graphics-processing units for example [93]. An in vivo system with real-time diagnostics would be a very effective research tool and could lead directly to a clinical diagnostic tool which could help guide biopsy-taking.

Continuing data collection would allow us to build up even more statistically significant sample numbers, which would then make the investigation of even smaller subgroups possible. For example, with sufficient tumour specimens it may be possible to distinguish between differentiated and non-differentiated tumours or between squamous and adenocarcinoma. The tumour response after chemotherapy could potentially also be assessed and conditions such as dysplasia could be classified into high-grade or low-grade.

To conclude, the detection of gastrointestinal pathologies using quantitative analysis of OCT data is a promising technique with good *ex vivo* results. Transferring the methodology to the *in vivo* domain holds a lot of potential as a future quick and reliable diagnostic test for endoscopic screening or guiding biopsy-taking.

5 REFERENCES

1.   Standring, S., ed. *Gray's Anatomy - The anatomical basis of clinical practice*. 39th ed. 2005, Elsevier.

2.   Kumar, V., et al., *Robbins and Cotran - Pathologic basis of disease*. 8th ed. 2010: Saunders, Elsevier.

3.   Odze, R.D., *Pathology of the gastroesophageal junction.* Seminars in diagnostic pathology, 2005. 22(4): p. 256-265.

4.   Sarbia, M., K.F. Becker, and H. Hofler, *Pathology of upper gastrointestinal malignancies.* Seminars in oncology, 2004. 31(4): p. 465-475.

5.   Dent, J., *Barrett's esophagus: A historical perspective, an update on core practicalities and predictions on future evolutions of management.* Journal of Gastroenterology and Hepatology, 2011. 26: p. 11-30.

6.   Shaheen, N.J. and J.E. Richter, *Barrett's oesophagus.* Lancet, 2009. 373(9666): p. 850-861.

7.   Feakins, R., et al., *Tissue pathways for gastrointestinal and pancreatobiliary pathology.* 2009, The Royal College of Pathologists

8.   Pinto-Santini, D. and N.R. Salama, *The biology of Helicobacter pylori infection, a major risk factor for gastric adenocarcinoma.* Cancer Epidemiol Biomarkers Prev, 2005. 14(8): p. 1853-1858.

9.   Nagini, S., *Carcinoma of the stomach: A review of epidemiology, pathogenesis, molecular genetics and chemoprevention.* World J Gastrointest Oncol, 2012. 4(7): p. 156-169.

10.   Lanza, G., et al., *Colorectal tumors: the histology report.* Digestive and liver disease : official journal of the Italian Society of Gastroenterology and the Italian Association for the Study of the Liver, 2011. 43 Suppl 4: p. S344-355.

11.   Treanor, D. and P. Quirke, *Pathology of colorectal cancer.* Clinical Oncology, 2007. 19(10): p. 769-776.

12.   Hall, N., *Colorectal cancer: features and investigation.* Medicine, 2007. 35(6): p. 302-305.

13. Dorundi, S. and A. Bannerjea, *Colorectal cancer: early diagnosis and predisposing causes.* Surgery (Oxford), 2006. 24(4): p. 131-136.

14. *The Global Burden of Disease 2004 Update*. 2008: World Health Organisation.

15. Cunningham, D., et al., *Colorectal cancer.* Lancet, 2010. 375(9719): p. 1030-1047.

16. Suvakovic, Z., et al., *Improving the detection rate of early gastric cancer requires more than open access gastroscopy: a five year study.* Gut, 1997. 41(3): p. 308-313.

17. Todd, J.A., et al., *Gastric ulcer and malignancy--is there a need for follow-up endoscopy?* Aliment Pharmacol Ther, 2004. 19(9): p. 989-991.

18. Raftopoulos, S.C., et al., *A cohort study of missed and new cancers after esophagogastroduodenoscopy.* Am J Gastroenterol, 2010. 105(6): p. 1292-1297.

19. *Cancer Research UK: CancerStats Incidence 2008 - UK* 2008 06 June 2011]; Available from: http://info.cancerresearchuk.org/cancerstats/incidence/.

20. Khushalani, N., *Cancer of the esophagus and stomach.* Mayo Clinic proceedings. Mayo Clinic, 2008. 83(6): p. 712-722.

21. Gallo, A. and C. Cha, *Updates on esophageal and gastric cancers.* World Journal of Gastroenterology: WJG, 2006. 12(20): p. 3237-3242.

22. Amin, A., et al., *Gastric adenocarcinoma missed at endoscopy.* Journal of the Royal College of Surgeons of Edinburgh, 2002. 47(5): p. 681-684.

23. Kida, M., *Endoscopic tumor diagnosis and treatment.* Endoscopy, 2002. 34(11): p. 860-870.

24. Araujo, S.E., P.R. Alves, and A. Habr-Gama, *Role of colonoscopy in colorectal cancer.* Revista do Hospital das Clinicas, 2001. 56(1): p. 25-35.

25. Mapstone, N.P., *Dataset for the histopathological reporting of oesophageal carcinoma*. 2007, The Royal College of Pathologists.

26. Novelli, M.R., *Dataset for the histopathological reporting of gastric carcinoma*. 2007, The Royal College of Pathologists.

27. Williams, G.T., P. Quirke, and N.A. Shepherd, *Dataset for colorectal cancer*. 2007, The Royal College of Pathologists.

28.   Voutilainen, M.E. and M.T. Juhola, *Evaluation of the diagnostic accuracy of gastroscopy to detect gastric tumours: clinicopathological features and prognosis of patients with gastric cancer missed on endoscopy.* Eur J Gastroenterol Hepatol, 2005. 17(12): p. 1345-1349.

29.   Hosokawa, O., et al., *Detection of gastric cancer by repeat endoscopy within a short time after negative examination.* Endoscopy, 2001. 33(4): p. 301-305.

30.   Lee, H.L., et al., *When do we miss synchronous gastric neoplasms with endoscopy?* Gastrointest Endosc, 2010. 71(7): p. 1159-1165.

31.   Flock, S.T., B.C. Wilson, and M.S. Patterson, *Total Attenuation Coefficients and Scattering Phase Functions of Tissues and Phantom Materials at 633nm.* Medical Physics, 1987. 14(5): p. 835-841.

32.   Bargo, P.R., et al., *In vivo determination of optical properties of normal and tumor tissue with white light reflectance and an empirical light transport model during endoscopy.* Journal of Biomedical Optics, 2005. 10(3):034018.

33.   Bashkatov, A.N., et al. *Optical properties of human stomach mucosa in the spectral range from 400 to 2000 nm - art. no. 653513.* in *2006 Saratov Fall Meeting.* 2006. Saratov, RUSSIA: Spie-Int Soc Optical Engineering.

34.   Wei, H.J., et al., *Determination of optical properties of normal and adenomatous human colon tissues in vitro using integrating sphere techniques.* World J Gastroenterol, 2005. 11(16): p. 2413-2419.

35.   Drezek, R., et al., *Light scattering from cervical cells throughout neoplastic progression: influence of nuclear morphology, DNA content, and chromatin texture.* Journal of Biomedical Optics, 2003. 8(1): p. 7-16.

36.   Drezek, R., A. Dunn, and R. Richards-Kortum, *Light scattering from cells: finite-difference time-domain simulations and goniometric measurements.* Applied Optics, 1999. 38(16): p. 3651-3661.

37.   Cheong, W.F., S.A. Prahl, and A.J. Welch, *A Review of the Optical-Properties of Biological Tissues.* IEEE Journal of Quantum Electronics, 1990. 26(12): p. 2166-2185.

38.   Salomatina, E., et al., *Optical properties of normal and cancerous human skin in the visible and near-infrared spectral range.* Journal of Biomedical Optics, 2006. 11(6):064026-064026-9.

39.   Thueler, P., et al., *In vivo endoscopic tissue diagnostics based on spectroscopic absorption, scattering, and phase function properties.* Journal of Biomedical Optics, 2003. 8(3): p. 495-503.

40.   Wilson, B.C., M.S. Patterson, and S.T. Flock, *Indirect Versus Direct Techniques for the Measurement of the Optical-Properties of Tissues.* Photochemistry and Photobiology, 1987. 46(5): p. 601-608.

41.   Maier, A., et al., *In vivo determination of tumor optical parameters in esophageal carcinoma.* Lasers in Surgery and Medicine, 2000. 27(4): p. 350-357.

42.   Holmer, C., et al., *Optical properties of adenocarcinoma and squamous cell carcinoma of the gastroesophageal junction.* Journal of Biomedical Optics, 2007. 12(1): 014025-014025-8.

43.   Marchesini, R., et al., *Ex-Vivo Optical-Properties of Human Colon Tissue.* Lasers in Surgery and Medicine, 1994. 15(4): p. 351-357.

44.   Arnfield, M.R., J. Tulip, and M.S. McPhee, *Optical Propagation in Tissue with Anisotropic Scattering.* IEEE Transactions on Biomedical Engineering, 1988. 35(5): p. 372-381.

45.   Tomlins, P.H. and R.K. Wang, *Theory, developments and applications of optical coherence tomography.* Journal of Physics D-Applied Physics, 2005. 38(15): p. 2519-2535.

46.   Wei, H.J., et al., *Differences in optical properties between healthy and pathological human colon tissues using a Ti : sapphire laser: an in vitro study using the Monte Carlo inversion technique.* Journal of Biomedical Optics, 2005. 10(4): 44022.

47.   Charvet, I., et al., *Diagnosis and grading of gastritis by non-invasive optical analysis.* European Journal of Gastroenterology & Hepatology, 2004. 16(11): p. 1189-1198.

48.   Pitris, C., et al., *Feasibility of optical coherence tomography for high-resolution imaging of human gastrointestinal tract malignancies.* Journal of Gastroenterology, 2000. 35(2): p. 87-92.

49.   Zuccaro, G., et al., *Optical coherence tomography of the esophagus and proximal stomach in health and disease.* American Journal of Gastroenterology, 2001. 96(9): p. 2633-2639.

50.   Dacosta, R.S., B.C. Wilson, and N.E. Marcon, *New optical technologies for earlier endoscopic diagnosis of premalignant gastrointestinal lesions.* Journal of Gastroenterology and Hepatology, 2002. 17: p. S85-S104.

51.   Familiari, L., et al., *Optical coherence tomography evaluation of ulcerative colitis: The patterns and the comparison with histology.* American Journal of Gastroenterology, 2006. 101(12): p. 2833-2840.

52.   Thrane, L., H.T. Yura, and P.E. Andersen, *Analysis of optical coherence tomography systems based on the extended Huygens-Fresnel principle.* Journal of the Optical Society of America a-Optics Image Science and Vision, 2000. 17(3): p. 484-490.

53.   Levitz, D., et al., *Determination of optical scattering properties of highly-scattering media in optical coherence tomography images.* Optics Express, 2004. 12(2): p. 249-259.

54.   Turchin, I.V., et al., *Novel algorithm of processing optical coherence tomography images for differentiation of biological tissue pathologies.* Journal of Biomedical Optics, 2005. 10(6): 064024-064024-11.

55.   van der Meer, F.J., et al., *Localized measurement of optical attenuation coefficients of atherosclerotic plaque constituents by quantitative optical coherence tomography.* IEEE Transactions on Medical Imaging, 2005. 24(10): p. 1369-1376.

56.   van der Meer, F.J., et al., *Temperature-dependent optical properties of individual vascular wall components measured by optical coherence tomography.* Journal of Biomedical Optics, 2006. 11(4): 041120.

57.   Xu, C.Y., et al., *Characterization of atherosclerosis plaques by measuring both backscattering and attenuation coefficients in optical coherence tomography.* Journal of Biomedical Optics, 2008. 13(3): 034003-034003-8.

58.   Hoffmann, M., O. Schewtschenko, and O. Minet. *Measuring of optical properties of biological samples by low cost gonlometrical equipment - art. no. 653501.* in *2006 Saratov Fall Meeting*. 2006. Saratov, RUSSIA: Spie-Int Soc Optical Engineering.

59.   Prahl, S.A., *Light transport In tissue - PhD Thesis.* The University of Texas at Austin, 1988.

60.   Bolt, R.A. and F.F.M. de Mul, *Goniometric instrument for light scattering measurement of biological tissues and phantoms (vol 73, pg 2211, 2002).* Review of Scientific Instruments, 2002. 73(8): p. 3133-3133.

61.   Marchesini, R., et al., *Extinction and Absorption-Coefficients and Scattering Phase Functions of Human-Tissues Invitro.* Applied Optics, 1989. 28(12): p. 2318-2324.

62.   Pogue, B.W. and M.S. Patterson, *Review of tissue simulating phantoms for optical spectroscopy, imaging and dosimetry.* Journal of Biomedical Optics, 2006. 11(4): 041102.

63.   Mourant, J.R., et al., *Mechanisms of light scattering from biological cells relevant to noninvasive optical-tissue diagnostics.* Applied Optics, 1998. 37(16): p. 3586-3593.

64.   Hammer, M., et al., *Single scattering by red blood cells.* Applied Optics, 1998. 37(31): p. 7410-7418.

65.   Ghosh, N., et al., *Measurement of optical transport properties of normal and malignant human breast tissue.* Applied Optics, 2001. 40(1): p. 176-184.

66.   Fercher, A.F., *Optical coherence tomography - development, principles, applications.* Zeitschrift Fur Medizinische Physik, 2010. 20(4): p. 251-276.

67.   Garcia-Allende, P.B., et al., *Morphological analysis of optical coherence tomography images for automated classification of gastrointestinal tissues.* Biomed Opt Express, 2011. 2(10): p. 2821-2836.

68.   Bouma, B.E., et al., *High-resolution imaging of the human esophagus and stomach in vivo using optical coherence tomography.* Gastrointestinal Endoscopy, 2000. 51(4): p. 467-474.

69.   Poneros, J.M., et al., *Diagnosis of specialized intestinal metaplasia by optical coherence tomography.* Gastroenterology, 2001. 120(1): p. 7-12.

70.   Adler, D.C., et al., *Three-dimensional endomicroscopy of the human colon using optical coherence tomography.* Optics Express, 2009. 17(2): p. 784-796.

71.   Pfau, P.R., et al., *Criteria for the diagnosis of dysplasia by endoscopic optical coherence tomography.* Gastrointestinal Endoscopy, 2003. 58(2): p. 196-202.

72.   Westphal, V., et al., *Correlation of endoscopic optical coherence tomography with histology in the lower-GI tract.* Gastrointestinal Endoscopy, 2005. 61(4): p. 537-546.

73.   Wong Kee Song, L.M. and B.C. Wilson, *Endoscopic detection of early upper GI cancers.* Best Pract Res Clin Gastroenterol, 2005. 19(6): p. 833-856.

74.   Isenberg, G., et al., *Accuracy of endoscopic optical coherence tomography in the detection of dysplasia in Barrett's esophagus: a prospective, double-blinded study.* Gastrointestinal Endoscopy, 2005. 62(6): p. 825-831.

75.   Thrane, L., et al., *Extraction of optical scattering parameters and attenuation compensation in optical coherence tomography images of multilayered tissue structures.* Optics Letters, 2004. 29(14): p. 1641-1643.

76.   Andersen, P.E., et al., *Advanced modelling of optical coherence tomography systems.* Physics in Medicine and Biology, 2004. 49(7): p. 1307-1327.

77.   Adegun, O.K., et al., *Quantitative analysis of optical coherence tomography and histopathology images of normal and dysplastic oral mucosal tissues.* Lasers in Medical Science, 2012. 27(4): p. 795-804.

78.   Kirillin, M.Y., A.V. Priezzhev, and R. Myllyla, *Role of multiple scattering in formation of OCT skin images.* Quantum Electronics, 2008. 38(6): p. 570-575.

79.   Treweek, S.P. and J.C. Barbenel, *Direct measurement of the optical properties of human breast skin.* Medical & Biological Engineering & Computing, 1996. 34(4): p. 285-289.

80.   Jacques, S.L., C.A. Alter, and S.A. Prahl, *Angular Dependence of HeNe Laser Light Scattering by Human Dermis.* Lasers in the Life Sciences, 1987. 1(4): p. 309-334.

81.   Hale, G.M. and M.R. Querry, *Optical Constants of Water in the 200-nm to 200-µm Wavelength Region.* Appl. Opt., 1973. 12(3): p. 555-563.

82.   Hammer, M., A.N. Yaroslavsky, and D. Schweitzer, *A scattering phase function for blood with physiological haematocrit.* Physics in Medicine and Biology, 2001. 46(3): p. N65-N69.

83.   Hsiung, P.L., P.R. Nambiar, and J.G. Fujimoto, *Effect of tissue preservation on imaging using ultrahigh resolution optical coherence tomography.* Journal of Biomedical Optics, 2005. 10(6):064033.

84.   Adegun, O.K., et al., *Quantitative optical coherence tomography of fluid-filled oral mucosal lesions.* Lasers Med Sci, 2013. 28(5): p. 1249-1255.

85.   Oliver, A., et al., *A novel breast tissue density classification methodology.* IEEE Trans Inf Technol Biomed, 2008. 12(1): p. 55-65.

86.   Garcia-Allende, P.B., et al., *Automated identification of tumor microscopic morphology based on macroscopically measured scatter signatures.* Journal of Biomedical Optics, 2009. 14(3):034034.

87.   Jain, A.K., M.N. Murty, and P.J. Flynn, *Data clustering: a review.* ACM Comput. Surv., 1999. 31(3): p. 264-323.

88.   Kang, W., et al., *Endoscopically guided spectral-domain OCT with double-balloon catheters.* Optics Express, 2010. 18(16): p. 17364-17372.

89.   McLaughlin, R.A. and D.D. Sampson, *Clinical applications of fiber-optic probes in optical coherence tomography.* Optical Fiber Technology, 2010. 16(6): p. 467-475.

90.   Shakhova, N.M., et al., *Clinical aspects of the endoscopic optical coherence tomography and the ways for improving its diagnostic value.* Laser Physics, 2002. 12(4): p. 617-626.

91.   Tearney, G.J., et al., *In vivo endoscopic optical biopsy with optical coherence tomography.* Science, 1997. 276(5321): p. 2037-2039.

92.   Kuranov, R.V., et al., *Complementary use of cross-polarization and standard OCT for differential diagnosis of pathological tissues.* Optics Express, 2002. 10(15): p. 707-713.

93.   Rasakanthan, J., K. Sugden, and P.H. Tomlins, *Processing and rendering of Fourier domain optical coherence tomography images at a line rate over 524 kHz using a graphics processing unit.* Journal of Biomedical Optics, 2011. 16(2):020505.

6 APPENDIX

6.1 OCT Experiments

6.1.1 Preliminary code for importing FRG files

```
%Code for reading FRG files
%Dr Gavin Erry, Dr Iakovos Amygdalos
%create an index of filenames that we want processed; beware, if multiple
%files selected, a cell array is created, with each cell containing a
%filename; if one file selected, not a cell array, so {} in filename
%function below do not work

fileselect = uigetfile('.FRG','MultiSelect','on');
for fileindex = 1 : size(fileselect,2);
filename = fileselect{1,fileindex};
file = fopen(filename);

% File information according to Thorlabs Software manual
id =            fread ( file, 16, '*char' );
numimages =     fread ( file, 1, 'int32' );
width =         fread ( file, 1, 'int32' );
depth =         fread ( file, 1, 'int32' );
numframes =     fread ( file, 1, 'int32' );
num3dVols =     fread ( file, 1, 'int32' );
fftlength =     fread ( file, 1, 'int32' );
framesize =     fread ( file, 1, 'int32' );
recordlength =  fread ( file, 1, 'int32' );
fringemode =    fread ( file, 1, 'int16' );
average =       fread ( file, 1, 'int16' );
imagemode =     fread ( file, 1, 'int16' );
doppler =       fread ( file, 1, 'int16' );
redshift =      fread ( file, 1, 'int32' );
blueshift =     fread ( file, 1, 'int32' );
reserved =      fread ( file, 448, 'int8' );

%create suitable filename for saving
filename1 = strrep(filename,'.FRG','');
filename2 = strrep(filename1,'-','_');
filename3 = genvarname(filename2);

%create correct image width according to averaging
avwidth = width/average;
%create width to use for cropping
cropwidth = width-1;
%create a 3D cscan matrix, where the frames will be put into. depth is
%always 512, width is defined by the width/average and number of frames
%defined by numimages
cscan = zeros([512,avwidth,numimages]);

% Iterate all the frames
for index = 1 : numimages

    % This will now read the first frame...
    timeelapsed =   fread ( file, 1, 'int32' );
    systemtime =    fread ( file, 1, 'int32' );
    reserved2 =     fread ( file, 32, 'int8' );
    frame =         fread ( file, width * fftlength, 'int16' );
```

```matlab
    %reshape, crop and resize b-scan:
    frame2 = reshape ( frame, fftlength, width );
    depth = fft ( frame2, 2048 );
    amplitude = depth ( 1 : 1024, : );
    I0 = log10 ( abs ( amplitude ) );
    J0 = imcrop (I0,[1 1 cropwidth 511]);
    K0 = imresize (J0, [512 avwidth]);


    %place current bscan into cscan 3D matrix, according to index number
    cscan (:,:,index) = K0;


    clc
    fprintf('processing ')
    fprintf('cscan ')
    fprintf('%d ', fileindex);
    fprintf('of ')
    fprintf('%d ', size(fileselect,2));
    fprintf('bscan ')
    fprintf('%d ', index);
    fprintf('of ')
    fprintf('%d ', numimages);
end

eval([sprintf(filename3) '= cscan;']);
save([filename3,'.mat'], filename3);
%movefile ([filename3,'.mat'],'C:\Users\Iakovos\Desktop\Outputs\');

fclose ( file );
end
clear;
beep;
```

6.1.2 Secondary code for reading FRG files

```
% Main program for processing Thorlabs .FRG or .IMG file
% recorded from Thorlabs Swept source OCT system
% Copyright 2009-2010 Thorlabs Inc.
% Modified by Dr Iakovos Amygdalos March 2011


global fileindex
global fileselect
global filename
global average

octmap.contrast = 1.0;
octmap.brightness = 0;
octmap.db_range = 50;
octmap.method = 1;              %  Mapping method for conversion OCT image from
linear scale to log scale,
%  1: method 1;  2: method 2
apodiz_method = 'Gaussian';  % 'Cosine', 'Gaussian',
'super_Gaussian','Hann', 'Kaiser'
% 'Hamming','Blackman', 'Nuttall', 'Blackman_Harris'*,'none'

sliding_average = 'N';    % switch for sliding average, 'Y': sliding
average, 'N': none
window.shift_depth = 1;   % shift step along depth for sliding average
window
window.shift_width = 1;   % shfit step along width for sliding average
window
window.ave_depth = 2;     % window length along depth for sliding average
window
window.ave_width = 2;      % window length along depth for sliding average
window
colormap_show = 'gray';   % colormap for show images

fileselect = uigetfile('.FRG','MultiSelect','on');
numcscans = size(fileselect,2);

for fileindex = 1 : numcscans;
    [file_head, frame_head, code] = Open_File_Multi1;
    filename1 = strrep(filename,'.FRG','');
    filename2 = strrep(filename1,'-','_');
    filename3 = genvarname(filename2);
    filename4 = strcat(filename3, 'AP16noDB');

    depth = frame_head.depth;
    width = frame_head.width;
    avwidth = width/average;
    numframes = file_head.nFr;

    fprintf(['...Type of the file:', file_head.type])
    fprintf('...Number of frames in the file is: %d', file_head.nFr)

    cscan = zeros([depth,avwidth,numframes]);
    frame_start = 1;
    frame_stop = numframes;
    Frame_Locator(file_head.fid, frame_start, frame_head.length_bytes);

    for index_frame = frame_start : frame_stop;
```

```matlab
        clc;
        fprintf('processing file ');
        fprintf(filename1);
        fprintf(' (frame %d of ', index_frame);
        fprintf('%d, ', numframes);
        fprintf('cscan %d ', fileindex);
        fprintf('of %d)', size(fileselect,2));

    [frame_2D, frame_info, elapsed_time] = Load_Frame(file_head,
    index_frame, frame_head);
    file_head.type ='FRG';

    OCT_complex = FRG_Proc(index_frame, frame_2D, frame_head,
    apodiz_method);

    OCT_IMG_AS = OCT_Image_Calculation(OCT_complex, frame_head,
    sliding_average, window);

    OCT_Display = log10(OCT_IMG_AS);

    cscan (:,:,index_frame) = OCT_Display;

    end

    fprintf('...Number of frames processed: %d', frame_stop - frame_start +
      1);

    eval([sprintf(filename4) '= cscan;']);
    save([filename4,'.mat'], filename4);

    Close_File(file_head.fid);
    clear x*;
end

clear;
clc;
fprintf('processing finished');
beep;
```

6.1.3 Concatenation code

```
%Code for concatenating parts of C-Scans.
%Dr Iakovos Amygdalos

clc;
numcscans = input('...Please set number of concatenated cscans to be
created: ');

for I = 1:numcscans;

    tempgroup = uigetfile('.MAT','MultiSelect','on');
    tempgroup1 = sort(tempgroup);
    eval(sprintf('group%d = tempgroup1;', I));
    clear temp*;
end

for I = 1:numcscans;

    cscan=[];
    eval(sprintf('fileselect = group%d;', I));

    for fileindex = 1 : size(fileselect,2);
        filename = fileselect{1,fileindex};
        clc;

        fprintf('group ');
        fprintf('%d ', I);
        fprintf('loading file ');
        fprintf('%d ', fileindex);
        fprintf('of ');
        fprintf('%d ', size(fileselect,2));
        fprintf('cscan ');
        fprintf('%d ', I);
        fprintf('of ');
        fprintf('%d ', numcscans);
        load(filename);
        filename1 = strrep(filename,'.mat','');
        S=eval(sprintf('%s',filename1));
        clear x*;
        cscan = cat(3,cscan,S);
        clear S;
        %movefile(filename, 'G:\OCT DATA\Optical Properties of
GASTROINTESTINAL tissues OCT data\SURGERY RAW DATA\Non-Concatenated');

    end

    filenameend = (size(filename1,2)- 15);
    filename2 = filename1(1,1:filenameend);
    filename3 = strcat(filename2, '_AP16noDB');
    eval([sprintf(filename3) '= cscan;']);
    clear cscan;
    clc;
    fprintf('saving cscan ');
    fprintf('%d ', I);
    fprintf('of ');
    fprintf('%d ', numcscans);
    save([filename3,'.mat'], filename3);
```

```
    filename4 = strcat(filename3, '.mat');
  %movefile(filename4, 'G:\OCT DATA\Optical Properties of GASTROINTESTINAL
tissues OCT data\SURGERY RAW DATA\Concatenated');
    clear x*;

end

clear

beep
clc
fprintf('processing finished');
```

6.1.4 Error Correction Code

```
%Author: Beatriz Garcia Allende, Dr Iakovos Amygdalos
%Start date: 23/03/2011
%Aim: Reading either apodized or non apodized MAT files and save them
%(adding EC to the file name) for subsequent analysis. This preparation
%includes:
%1.- Infinitive values detection and removal (it substitutes them by the
first
%non infinitive value of the depth profile)
%2.- Outliers removal (so far we are using alignment in all our analysis
%so surface touching the surface = outliers, we can come back to this later
on if we are not happy with this)
%3.- Check if there is a bright line at the top and if so it removes it
%4.- Surface detection and alignment


clear
clc
close all

[file_name,file_path,filter]=uigetfile('*.MAT','Multiselect','on');
clear file_path filter
Nfiles=size(file_name,2);
OutTH=1.5; %extreme oultier removal, change for bigger values for mild
outlier removal (less strict) or smaller if you think too many outliers
have been left behind

for indFr=1:Nfiles %size(file_name,2)
    ActualFRname=char(file_name(indFr));

    %loop counter - which file is being loaded?
    clc;
    fprintf('Loading File ');
    fprintf(ActualFRname);
    fprintf(' (%d of ', indFr);
    fprintf('%d)', Nfiles);

    TissueSection = importdata (ActualFRname);
    clearvars -except file_name Nfiles indFr OutTH ActualFRname
      TissueSection
    %infinitive value removal
    TissueSection=InfValRemoval(TissueSection);
    %outlier removal
    TissueSection=OutlierRemoval(TissueSection,OutTH);
    TissueSection=BrightLinRemoval(TissueSection);
    [Nfil,Ncol,Nimag]=size(TissueSection);
    TissueSectionCO=zeros(Nfil,Ncol,Nimag);

    for indiceImag=1:Nimag
        ActualTom=TissueSection(:,:,indiceImag);

        posOut=find(sum(ActualTom,1)==0);
        ActualTom(:,posOut)=[];

[CompensatedCancerSection,CancerMasks1]=TomogramNoisRem(ActualTom,30,0.7);

[CompensatedCancerSection,CancerMasks2]=ImageMaskQU(CompensatedCancerSectio
n,75);
```

```matlab
            BinaryMask=CancerMasks2;

            posSurf=zeros(1,size(ActualTom,2));
            for indice=1:size(BinaryMask,2)
                if isempty(find(BinaryMask(:,indice),1,'first'))
                    posSurf(1,indice)=0;
                else
                    posSurf(1,indice)=find(BinaryMask(:,indice),1,'first');
                end
            end
            posOut2=find(posSurf==1|posSurf==2);
            ActualTom(:,posOut2)=[];
            posSurf(:,posOut2)=[];
            BinaryMask(:,posOut2)=[];
            posOut=[posOut posOut2];
            posOut=sort(posOut,'ascend');
            posNulS=find(posSurf==0);
            if not(isempty(posNulS))
                posSurfAv=posSurf;
                posSurfAv(posNulS)=[];
                posSurfAv=median(posSurfAv);
                posSurf(posNulS)=posSurfAv;
                posSurf=medfilt1(posSurf,10);
                posSurf=floor(posSurf);
            end
            posSurf(find(posSurf < 1)) = 1;
            %profMaxMax=max(posSurf);
            depth=size(ActualTom,1);

            for indicePixHor=1:size(BinaryMask,2)
                TissueSectionCO(1:depth-
posSurf(1,indicePixHor)+1,indicePixHor,indiceImag)=ActualTom(posSurf(1,indi
cePixHor):depth,indicePixHor);
            end
            if not(isempty(posOut))

TissueSectionCO(:,:,indiceImag)=OutlierPositionRecovering(TissueSectionCO(:
,:,indiceImag),posOut);
            end
            figure ( 1 );
            imagesc(TissueSectionCO(:,:,indiceImag));
            title(ActualFRname)
            pause(0.025);

            clc;
            fprintf('Processing File ');
            fprintf(ActualFRname);
            fprintf(' / Cscan ');
            fprintf('%d ', indFr);
            fprintf('of ');
            fprintf('%d ', Nfiles);
            fprintf(' / Bscan %d of ', indiceImag);
            fprintf('%d ', Nimag);

    end
    addstr='_EC';
    extensionF='.mat';
    codePA=ActualFRname(:,1:end-4);
    nameVar=strcat(codePA,addstr);
    nameFil=strcat(codePA,addstr,extensionF);
```

```
        assignin('base',nameVar,TissueSectionCO)

        clc;
        fprintf('Saving File ');
        fprintf(ActualFRname);

        save(nameFil,nameVar)
        clearvars -except file_name Nfiles indFr OutTH
end

clear;
clc;
close;
beep;
pause(0.25);
beep;
```

6.1.5 Code for SavGol Analysis

```
% SavGol plotting
% Author Dr Iakovos Amygdalos
%run this on SVG_M files to obtain SAVGOL plots. this
%will show variations in the gradient along the depth of the ascan and
%indicate areas of interest for gradient processing

close;
clear;
clc;

fileselect = uigetfile('.MAT','MultiSelect','on');
numSVG = size(fileselect,2);
MeanSavGolMatrix = zeros(numSVG,512);

for fileIndex = 1 : numSVG;
    filename = fileselect{1,fileIndex};
    filename1 = strrep(filename,'.mat','');

    clc;
    fprintf('Loading ');
    fprintf(filename1);
    fprintf(' MeanSavGol %d of ', fileIndex);
    fprintf(num2str(numSVG));

    MeanSavGolMatrix(fileIndex,:) = importdata (filename);

end

clc;
savename = input('Enter save name for MeanSavGolMatrix: ','s');
eval([sprintf(savename) '= MeanSavGolMatrix;']);
save([savename,'.mat'], savename);
savename2 = strcat(savename, '_Labels');
eval([sprintf(savename2) '= fileselect;']);
save([savename2,'.mat'], savename2);
savename3 = strcat(savename, '_FIG');

clc
fprintf ('Processing Finished');

fig1 = figure('units','normalized','outerposition',[0 0 1 1]);
transpMatrix = MeanSavGolMatrix';
transpLabels = fileselect';
plot(transpMatrix);
hold on;
leg1 = legend(transpLabels);
set(leg1,'Location','Best')
set(leg1,'Interpreter','none')
title1 = title(savename);
set(title1,'Interpreter','none')
saveas(fig1,savename3,'fig');
```

6.1.6 Code for gradient analysis of OCT data

```
% OCT Gradient analysis Author Dr Iakovos Amygdalos
%this code reads an EC Cscan MAT file, defines the ROI and averages the
%Ascans in the ROI (along the 2nd dimension), producing a MeanAscan for
%each Bscan, which is then smoothed and fitted with robustfit


fileselect = uigetfile('.MAT','MultiSelect','on');
numCscans = size(fileselect,2);

for fileIndex = 1 : numCscans;
    filename = fileselect{1,fileIndex};
    filename1 = strrep(filename,'.mat','');
    filename2 = strcat(filename1, '_GA7_');

    clc;
    fprintf('Loading ');
    fprintf(filename1);
    fprintf(' Cscan %d of ', fileIndex);
    fprintf(num2str(numCscans));

    Cscan = importdata (filename);
    noBscans = size(Cscan,3);
    noAscans = size(Cscan,2);

    %Define ROI margins
    RM = noAscans;        %Right Margin
    LM = 1;               %Left Margin
    TM = 10;              %Top Margin
    BM = 70;              %Bottom Margin
    FRM = 1;              %Front Margin
    BKM = noBscans;       %Back Margin

    filename3 = strcat(filename2, num2str(TM),'_to_',num2str(BM),'SM');

    FitCscan = Cscan(TM:BM,LM:RM,FRM:BKM);
    FitCscan(find(FitCscan==0)) = NaN;
    noFitBscans = size(FitCscan,3);
    noFitAscans = size(FitCscan,2);
    gradMatrix = zeros(noFitBscans,1,2);

    for bscanIndex = 1 : noFitBscans;

        clc;
        fprintf('processing ');
        fprintf(filename1);
        fprintf(' /Cscan %d of ', fileIndex);
        fprintf(num2str(numCscans));
        fprintf(' /bscan %d of ', bscanIndex);
        fprintf('%d ', noBscans);

        MeanAscan = nanmean (FitCscan(:,:,bscanIndex),2);
        FitAscan = smooth(MeanAscan,5);
        x_1 = (1:numel(FitAscan))';
        coeff = robustfit(x_1,FitAscan);
```

```
        gradMatrix(noFitBscans-bscanIndex+1,1,1) = coeff(2,1);
        gradMatrix(noFitBscans-bscanIndex+1,1,2) = coeff(1,1);

    end

    eval([sprintf(filename3) '= gradMatrix;']);
    save([filename3,'.mat'], filename3);

end

beep;
pause(0.25);
beep;

clear
clc
close

fprintf ('Processing Finished');
```

6.1.7 Code for grouping of OCT gradient matrices

```
%OCT Gradient Matrix Grouping
%This code allows the user to select a group of gradient matrices and put
%them all together as columns inside a larger matrix. This facilitates
%further processing eg. boxplot generation
%Author Dr Iakovos Amygdalos

clear;
clc;
close;

GroupName = input('Please enter the name of the cscan group (e.g.
Cancer):','s');
fileselect = uigetfile('.MAT','MultiSelect','on');
numMatrices = size(fileselect,2);
GroupMatrix = NaN(262144,numMatrices);

for fileIndex = 1 : numMatrices;
    filename = fileselect{1,fileIndex};
    filename1 = strrep(filename,'.mat','');
    clc;
    fprintf('Loading ');
    fprintf(filename1);
    fprintf(' GradMatrix %d of ', fileIndex);
    fprintf(num2str(numMatrices));

    GradMatrix = importdata (filename);
    gradients = GradMatrix(:,:,1);
    reshaped = reshape(gradients,[],1);
    padding = 262144 - (size(reshaped,1));
    reshaped2 = [reshaped;NaN(padding,1)];
    GroupMatrix(:,fileIndex) = reshaped2;

end

namelength = size(filename1,2);
Ending = filename1(37:namelength);
GroupName2 = strcat('GroupMatrix',GroupName,Ending);
GroupMatrix(find(GroupMatrix==0)) = NaN;
%GroupMatrix(find(~GroupMatrix))=NaN;
eval([sprintf(GroupName2) '= GroupMatrix;']);
save([GroupName2,'.mat'], GroupName2);

clc;
fprintf('Plotting');
fig1 = figure ( 1 );
boxplot(GroupMatrix);
hold on;
title(GroupName2,'Interpreter','none');
saveas(fig1,GroupName2,'fig');

fig2 = figure ( 2 )
boxplot(GroupMatrix,'plotstyle','compact');
hold on;
title(GroupName2,'Interpreter','none');
GroupName3 = strcat(GroupName2,'_C');
saveas(fig2,GroupName3,'fig');
```

6.1.8 Code for OCT group matrix processing

```
%Group Matrix Processing
%Author Dr Iakovos Amygdalos
%this code will load grouped gradient data, transform into columns and
%concatenate into a single matrix for boxplot plotting and statistical
%testing,

clc;
clear;
close;

NumGroups = input('Enter number of groups to be compared:');

for groupindex = 1 : NumGroups;
    fileselect{1,groupindex} = uigetfile('.MAT');
end

numMatrices = size(fileselect,2);

for matindex = 1 : numMatrices
    filename = fileselect{1,matindex};
    filename1 = strrep(filename,'.mat','');

    GroupMatrix = importdata (filename);
    GroupColumn = reshape(GroupMatrix,[],1);
    GroupColumn = GroupColumn(~isnan(GroupColumn));

    %now to remove outliers
    reshaped = reshape(GroupColumn,1,[]);
    psi=1.5;
    MaxQ=prctile(reshaped, [25 75]);
    MaxIQR=iqr(reshaped);
    A1=reshaped<MaxQ(1,1)-psi*MaxIQR;
    A2=reshaped>MaxQ(1,2)+psi*MaxIQR;
    outliers=A1|A2;
    outliersPOS=find(outliers);
    reshaped(:,outliersPOS)=NaN;
    GroupColumn = reshape(reshaped,[],1);

    filename2 = strrep(filename1,'Matrix','Column');
    filename3 = strcat(filename2,'.mat');
    fileselect{2,matindex} = filename3;
    fileselect{3,matindex} = size(GroupColumn,1);
    eval([sprintf(filename2) '= GroupColumn;']);
    save([filename2,'.mat'], filename2);
    clearvars Group* filename*
end

Sizes = cell2mat(fileselect(3,:));
Maxsize = max(Sizes);
CompareMatrix = NaN(Maxsize,numMatrices);

for matindex = 1 : numMatrices

    filename =  fileselect{2,matindex};
```

```
    filename1 = strrep(filename,'.mat','');
    column = importdata (filename);
    padding = Maxsize - size(column,1);
    column = [column;NaN(padding,1)];
    CompareMatrix(:,matindex) = column;
end

briefname = input('Enter a brief name for the comparison matrix, eg
"LGICancVsHealth":','s');
startchar = size(filename1,2)-16;
%nameend = filename1(startchar:size(filename1,2));
savename = strcat('Compare',briefname);
eval([sprintf(savename) '= CompareMatrix;']);
save([savename,'.mat'], savename);
labelname = strcat(savename,'_Lab');
labelnameS = strcat(labelname,'S');
Labels = fileselect(2,:);
ShortLabels = {};

for labindex = 1 : size(Labels,2);
    inputtext = strcat('Please enter a short name for the following group:
',Labels{1,labindex},'...');
    shortlab1 = input(inputtext,'s');
    ShortLabels{1,labindex} = shortlab1;
end

eval([sprintf(labelname) '= Labels;']);
save([labelname,'.mat'], labelname);
eval([sprintf(labelnameS) '= ShortLabels;']);
save([labelnameS,'.mat'], labelnameS);

fig1 = figure ( 1 )
boxplot(CompareMatrix,'labels',ShortLabels);
hold on;
titlename = strrep(briefname, '_', ' ');
title(titlename,'Interpreter','none');
xlabel({'';'Histological Diagnosis'});
ylabel('Attenuation Gradient');
saveas(fig1,savename,'fig');
saveas(fig1,savename,'bmp');

StatsMatrix = cell(15,(NumGroups+1));
StatsMatrix(1:15,1) = {'Histology','Mean','Standard
Deviation','Median','Mode','Range','Min','Max','1st Quantile','3rd
Quantile','Interquartile Range','Skewness','Kurtosis','Variance','No Of A-
scans'};

StatsMatrix(1,2:end) = ShortLabels;
StatsMatrix(2,2:end) = num2cell(nanmean(CompareMatrix));
StatsMatrix(3,2:end) = num2cell(nanstd(CompareMatrix));
StatsMatrix(4,2:end) = num2cell(nanmedian(CompareMatrix));
StatsMatrix(5,2:end) = num2cell(mode(CompareMatrix));
StatsMatrix(6,2:end) = num2cell(range(CompareMatrix));
StatsMatrix(7,2:end) = num2cell(nanmin(CompareMatrix));
StatsMatrix(8,2:end) = num2cell(nanmax(CompareMatrix));
StatsMatrix(9,2:end) = num2cell(quantile(CompareMatrix,.25));
StatsMatrix(10,2:end) = num2cell(quantile(CompareMatrix,.75));
StatsMatrix(11,2:end) = num2cell(iqr(CompareMatrix));
StatsMatrix(12,2:end) = num2cell(skewness(CompareMatrix));
```

```
StatsMatrix(13,2:end) = num2cell(kurtosis(CompareMatrix));
StatsMatrix(14,2:end) = num2cell(nanvar(CompareMatrix));
x=-5:5:5;

count = sum(histc(CompareMatrix,x));
StatsMatrix(15,2:end) = num2cell(count(1,:));
statsname = strcat(savename,'_Stats');
eval([sprintf(statsname) '= StatsMatrix;']);
save([statsname,'.mat'], statsname);
xlswrite(statsname,StatsMatrix);

clc;
fprintf('Processing Finished');
clear;
```

6.1.9 Code for Mann-Whitney statistical analysis

```
%Mann-Whitney U-test / Ranksum test
%Author: Dr Iakovos Amygdalos
%works with Compare files - choose number of comparisons to make, then
%assign a group number to each column you load and it will compare each
%group against the rest

clc
clear
CompMatrixName = uigetfile('.MAT');
CompMatrix = importdata(CompMatrixName);
ShortLabelMatrix = uigetfile('.MAT');
ShortLabels = importdata(ShortLabelMatrix);
NumColumns = size(CompMatrix,2)

CompName1 = input('What part of the GI tract are you comparing? eg "Colon"
or "OesophAndStomach" etc ...','s');

for Column1 = 1 : NumColumns-1;
    Var1 = CompMatrix(:,Column1);
    Var1 = Var1(~isnan(Var1));

    for Column2 = Column1+1 : NumColumns;

    Var2 = CompMatrix(:,Column2);
    Var2 = Var2(~isnan(Var2));
    [p,h,stats] = ranksum(Var1, Var2)
    Column1Name = ShortLabels{1,Column1};
    Column2Name = ShortLabels{1,Column2};
    CompName2 = strcat(Column1Name,'_vs_',Column2Name);
    CompName3 = strcat('MannWhit',CompName1, CompName2);
    save(CompName3,'p','h','stats');

    end
end

clear
```

6.1.10 Code for morphological analysis of OCT data

```
%Author: Dr Beatriz Garcia Allende
%Modifications by Dr Iakovos Amygdalos
%Start date: 11/02/2011
%Aim: Reading either corrected apodized or non apodized MAT files and
%performs and saves the morphological image features dividing each tomogram
%in the number of clusters indicted by the user

clear
clc
close all

[file_name,file_path,filter]=uigetfile('*.MAT','Multiselect','on');
clear file_path filter
Nfiles=size(file_name,2);
nClusters=input('Please input the number of regions for tomogram
segmentation: ');

for indFr=1:Nfiles

    ActualFRname=char(file_name(indFr));
    clc;
    fprintf('Loading ');
    fprintf(ActualFRname);
    fprintf ( ' cscan ' );
    fprintf(num2str(indFr));
    fprintf( ' of ')
    fprintf(num2str(Nfiles));
    load(ActualFRname);
    CancerSection=eval(ActualFRname(:,1:end-4));

    Nimag=size(CancerSection,3);
    FeatCan=zeros(5,nClusters,Nimag);
    % 5 morphological features:
    for indiceImag=1:Nimag
        clc
        fprintf('processing ');
        fprintf (ActualFRname);
        fprintf ( ' cscan ' );
        fprintf(num2str(indFr));
        fprintf( ' of ')
        fprintf(num2str(Nfiles));
        fprintf(' frame ');
        fprintf(num2str(indiceImag))
        fprintf(' of ')
        fprintf (num2str(Nimag));
        ActualTom=CancerSection(:,:,indiceImag);
        posOut=find(sum(ActualTom,1)==0);
        ActualTom(:,posOut)=[];
        ActualTomAlin=ActualTom;
        [NfilAlin,NcolAlin]=size(ActualTomAlin);
        porPer=100/(2*nClusters):100/nClusters:100;

    centroIni=(prctile(reshape(ActualTomAlin,1,NfilAlin*NcolAlin),porPer)
    )';
    [TomClusterizado,Centroids]=kmeans(reshape(ActualTomAlin,1,NfilAlin*N
    colAlin),nClusters,'start',centroIni, 'EmptyAction','singleton');
```

```matlab
        [CentroidsOrd,indeces]=sort(Centroids,'descend');

        for indClus=1:nClusters
            posiciones=find(TomClusterizado==indeces(indClus,1));
            TomClusterizado(posiciones)=nClusters+indClus;

        end

    TomClusterizado = TomClusterizado-
    repmat(nClusters,length(TomClusterizado),1);
      TomClusterizado=reshape(TomClusterizado,NfilAlin,NcolAlin);


    FeatCan(:,:,indiceImag)=ComputMorphFeat(ActualTomAlin,TomClusterizado
    );
    end

    %removing the morphological features of the zero cluster
    FeatCan=FeatCan(:,1:nClusters-1,:);

    codePA=ActualFRname(:,1:end-4);
    extensionF='.mat';
    addstr1='_MF';
    addstr2=sprintf('_%d',nClusters-1);
    nameVar=strcat(codePA,addstr1,addstr2);
    nameFil=strcat(codePA,addstr1,addstr2,extensionF);
    assignin('base',nameVar,FeatCan)
    clc;
    fprintf('Saving ');
    fprintf(nameFil);
    fprintf ( ' cscan ' );
    fprintf(num2str(indFr));
    fprintf( ' of ')
    fprintf(num2str(Nfiles));
    save(nameFil,nameVar)

    clearvars -except file_name indFr Nfiles nClusters

end
```

6.1.11 Code for categorization of morphological OCT data

```
%Data Categorization Code This code allows the user to select a group of
%MF_3 files and label them as belonging to a specific histological
%diagnosis Authors: Dr Iakovos Amygdalos, Dr Beatriz Garcia Allende

clear;
clc;
close;

fileselect = uigetfile('.MAT','MultiSelect','on');
Nfiles = size(fileselect,2);

UpperLower= input ('Is this group UGI or LGI?\nPlease enter "UGI" or
"LGI"\n','s');
GroupName = input('Please enter a name for the group (e.g. COLON,
COL_INFLAM)\nIN CAPITALS PLEASE:','s');
if strfind('UGI',UpperLower);
    clc;
    fprintf '\n\nUGI Diagnostic Group Classification Codes:\n';
    fprintf '\n 1 = Oesophagus\n 2 = Oesophagus Inflammation\n 3 =
Barretts\n 4 = Stomach\n 5 = Stomach Inflammation\n 6 = Stomach Int Met\n 7
= UGI Dysplasia\n 8 = UGI Cancer\n 9 = Duodenum\n 10 = Duodenum
Inflammation\n 22 = Oesophagus All Abnormal\n 23 = Stomach All Abnormal\n';
elseif strfind('LGI',UpperLower);
    clc;
    fprintf '\n\nLGI Diagnostic Group Classification Codes:\n';
    fprintf '\n 11 = Ileum\n 12 = Ileum Inflammation\n 13 = Colon\n 14 =
Colon Inflammation\n 15 = Crohns\n 16 = Rectum\n 17 = Rectum Inflammation\n
18 = LGI Dysplasia\n 19 = LGI Cancer\n 20 = Colon All Abnormal\n 21 =
Rectum All Abnormal\n';
end
ClassCode = input ('\nPlease enter one of the above codes for this group
"\n','s');
ClassCodeNum = str2num(ClassCode);
RenamingKey = {};
for indFr=1:Nfiles;
    ActualFRname=char(fileselect(indFr));
    load(ActualFRname)
    FeatCan=eval(ActualFRname(:,1:end-4));
    if strfind('UGI',UpperLower);
        FeatCan=cat(1,FeatCan,ones(1,size(FeatCan,2),size(FeatCan,3)));

FeatCan=cat(1,FeatCan,ClassCodeNum*ones(1,size(FeatCan,2),size(FeatCan,3)))
;
    elseif strfind('LGI',UpperLower);
        FeatCan=cat(1,FeatCan,zeros(1,size(FeatCan,2),size(FeatCan,3)));

FeatCan=cat(1,FeatCan,ClassCodeNum*ones(1,size(FeatCan,2),size(FeatCan,3)))
;
    end

    MeasNo = num2str(sprintf('%02d',indFr));
    ClassCode1 = sprintf('%02d',ClassCodeNum);
    NameEnd = ActualFRname(:,end-11:end-4);

nameVar=strcat('x',ClassCode1,'_',UpperLower,'_',GroupName,'_',MeasNo,NameE
nd,'_C');
```

```
    nameFil=strcat(nameVar,'.mat');
    assignin('base',nameVar,FeatCan);
    save(nameFil,nameVar);
    RenamingKey{indFr,1} = fileselect(indFr);
    RenamingKey{indFr,2} = nameVar;

    clearvars -except Nfiles indFr fileselect ClassCode ClassCodeNum
GroupName UpperLower RenamingKey

end
KeySaveName = strcat(UpperLower,'_',GroupName,'_RenamingKey');
save([KeySaveName,'.mat'], 'RenamingKey');
beep;
pause(0.25);
beep;

clc;
clear;
fprintf ('Processing Finished');
```

6.1.12 Code for PCA classification of morphological OCT data

```
%Author: Dr Beatriz Garcia Allende, Dr Iakovos Amygdalos
%Aim: Carrying out Principal Component Analysis on selected diagnostic
%groups

clear;
clc;
close all;

[file_name,]=uigetfile('*.MAT','Multiselect','on');
CompName = input('Please enter a name for this dataset\n...','s');
Nfiles=size(file_name,2);
FeaturesU=[];
WideLU=[];
PatU=[];
load('DiagnosisKey.mat');

for indFr=1:Nfiles;
    ActualFRname=char(file_name(indFr));
    load(ActualFRname);
    FeatCan=eval(ActualFRname(:,1:end-4));
    FeaturesU=cat(3,FeaturesU,FeatCan(1:end-2,:,:));

WideLU=cat(1,WideLU,FeatCan(size(FeatCan,1),1,1)*ones(size(FeatCan,3),1));
    PatU=cat(1,PatU,repmat(ActualFRname(:,2:25),size(FeatCan,3),1));
end

%Assign diagnostic labels according to Diagnostic Key
GroupingMatrix= num2cell(WideLU);
for Diagnosis = 1:size(DiagnosisKey,1);
    GroupingMatrix(WideLU ==
DiagnosisKey{Diagnosis,1})={DiagnosisKey{Diagnosis,2}};
end

% Features={'Relative Area','Mean','Standard
Deviation','Skewness','Kurtosis'};
Nfeatures=size(FeaturesU,1);
Nclusters=size(FeaturesU,2);
NumClust=sprintf('_%d',Nclusters);
Nfig=1;
for indFeat=1:Nfeatures;
    for indClus=1:Nclusters;
        BP = figure(Nfig);

boxplot(reshape(FeaturesU(indFeat,indClus,:),size(FeaturesU,3),1),GroupingM
atrix);
        BPname = sprintf('Fig%d',Nfig);
        BPname1 = strcat('PCA_MF',NumClust,'_',CompName,'_',BPname);
        saveas(BP,BPname1,'fig');
        saveas(BP,BPname1,'bmp');
        Nfig=Nfig+1;
    end
end

[NfeatU,NclustU,NtomU]=size(FeaturesU);
FeaturesU=reshape(FeaturesU,NfeatU*NclustU,NtomU);
```

```
[componentes,scores,latentv]=princomp(FeaturesU');

Key1=strrep(file_name,'x','');
Key2=char(Key1);
Key3=Key2(1:end,1:2);
Key4=str2num(Key3);
DiagIndex=unique(Key4);

Fig = figure(Nfig);
Labels = cell(1,size(DiagIndex,1));
cmap=jet(size(DiagIndex,1));
for Diagnosis = 1:size(DiagIndex,1);
    DiagnosisNum = DiagIndex(Diagnosis,1);
    DiagLoc = find([DiagnosisKey{:,1}]==DiagnosisNum);
    DiagName = DiagnosisKey(DiagLoc,2);
    Labels(1,Diagnosis)=DiagName;
    eval([sprintf('H%d',Diagnosis) '=
plot3(scores((find(WideLU==DiagnosisNum)),1),scores((find(WideLU==Diagnosis
Num)),2),scores((find(WideLU==DiagnosisNum)),3),''*'',''Color'',cmap(Diagno
sis,:));']);
    hold on;
end
grid on;
xlabel('PC1');
ylabel('PC2');
zlabel('PC3');
leg = legend(Labels,'Location','Northeast');
titlename = strrep(CompName, '_', ' ');
title(titlename,'Interpreter','none');
codePA='PCAmorphologicalF_Upper';
nameFil=strcat('PCA_MF',NumClust,'_',CompName,'.mat');
nameFig=strrep(nameFil,'.mat','');
saveas(Fig,nameFig,'fig');
saveas(Fig,nameFig,'bmp');
save(nameFil,'FeaturesU','componentes','scores','latentv','WideLU',
'GroupingMatrix', 'DiagIndex') ;

clear;
clc;
beep;
pause(0.25);
beep;
fprintf ('Processing Finished');

CloseFigs = input('\nClose figures y/n?...','s');
if CloseFigs == 'y';
    close all;
    fprintf ('Figures Closed');
else
end
```

6.1.13 Code for KNN classification of morphological OCT data

```
%Author: Dr Beatriz Garcia Allende, Dr Iakovos Amygdalos
%Start date: 01/06/2011
%Aim: First test on the classification with KNN (Mahalanobis distance) of
%the morphological features with/without PCA

% CVparameter=3; %3 and 10 are the most typical values for cross validation
% PCAyn=input('Would you like to apply PCA to the features (1 for yes / 0
for no):\n  ');
% Nneighbors=input('Number of neighbors for the KNN validation
(1/3/5/7/11...):  '); %k in the KNN classifier

clear;
clc;
close all;

FileSelect=uigetfile('*.MAT','Multiselect','on');
TF = ischar(FileSelect);
if TF==1;
    FileSelect = cellstr(FileSelect);
end
CVValues = input('Cross-validation values, e.g. 3,10 or [3,10] to run
both\n');
clc;
PCAValues = input('Would you like to apply PCA to the features (1 for yes /
0 for no / [0,1] for both):\n  ');
clc;
KNNValues = input('Number of neighbors for the KNN validation
(1/3/5/7/11... or \n[1,3,5 etc] for multiple comparisons)\n:  ');
clc;
load('DiagnosisKey.mat');
StartTime=datestr(now);

for FileIndex = 1:size(FileSelect,2);
    file_name = FileSelect(1,FileIndex);
    ActualFRname=char(file_name);
    load(ActualFRname);

    WideLU=WideLU';
    DiagIndex = unique(WideLU);
    NbinarySimulations=size(DiagIndex,2);
    ClassificationChar=cell(NbinarySimulations+1,6);

    for CVIndex = 1:size(CVValues,2);
        CVparameter = CVValues(1,CVIndex);

        for PCAIndex = 1:size(PCAValues,2);
            PCAyn= PCAValues(1,PCAIndex);
            %removal of principal components whose associated eigenvalues
are
            %smaller than 1e-4*maxium eigenvalue (noise removal, redundancy
elimination)
            if PCAyn==1;
                umbral=latentv(1,1)*1e-4;
                firstDisc=find(latentv<umbral,1,'first');
                scores=scores';
```

```matlab
        scores=scores(1:firstDisc-1,:);
        FeaturesU=scores;
        disp('FeaturesU has been substituted by PCA scores');
    end

    for KNNIndex = 1:size(KNNValues,2);
    Nneighbors = KNNValues(1,KNNIndex);

        %feature normalization
        FeatMeanValues=mean(FeaturesU,2);
        FeatMeanValues=repmat(FeatMeanValues,1,size(FeaturesU,2));
        FeatStdValues=std(FeaturesU,0,2);
        FeatStdValues=repmat(FeatStdValues,1,size(FeaturesU,2));
        FeaturesU=(FeaturesU-FeatMeanValues)./FeatStdValues;
        %score normalization
        ScoMeanValues=mean(scores,2);
        ScoMeanValues=repmat(ScoMeanValues,1,size(scores,2));
        ScoStdValues=std(scores,0,2);
        ScoStdValues=repmat(ScoStdValues,1,size(scores,2));
        scores=(scores-ScoMeanValues)./ScoStdValues;

        clear FeatMeanValues FeatStdValues ScoMeanValues
ScoStdValues;

        % classification accuracy, sensitivity, specificity,
negative
        % predictive value and prositive predictive value (5
numbers) are obtained per
        % diagnostic category from all others


        for indS=1:NbinarySimulations;
            CurrClass=[WideLU==DiagIndex(1,indS)];
            CurrClassLoc=
find([DiagnosisKey{:,1}]==DiagIndex(1,indS));
            CurrClassName = DiagnosisKey(CurrClassLoc,2);
            CVO = cvpartition(CurrClass,'k',CVparameter);
            ClassificationMeas=zeros(CVparameter,5);

            for indCV = 1:CVO.NumTestSets;

                trIdx = CVO.training(indCV);
                teIdx = CVO.test(indCV);
                Data4Training=FeaturesU(:,trIdx');
                TrainingClass=CurrClass(:,trIdx');
                Data4Test=FeaturesU(:,teIdx');
                TestClass=CurrClass(:,teIdx');

MeanPerDiagnosticCathegory=[mean(Data4Training(:,find(TrainingClass==0)),2)
mean(Data4Training(:,find(TrainingClass==1)),2)];

CovarianzeMatrices=zeros(size(Data4Training,1),size(Data4Training,1),2);
                for indClass=1:size(CovarianzeMatrices,3);

ClassSet=Data4Training(:,find(TrainingClass==(indClass-1)));
                        CovarianzeMatrices(:,:,indClass)=(ClassSet-
repmat(MeanPerDiagnosticCathegory(:,indClass),1,size(ClassSet,2)))*(ClassSe
t-repmat(MeanPerDiagnosticCathegory(:,indClass),1,size(ClassSet,2)))';
                    clc;
```

```matlab
                        fprintf('\nStart date & time: %s', StartTime);
                        fprintf('\nProcessing file: %s', ActualFRname);
                        fprintf(' (%d ', FileIndex);
                        fprintf('of %d) ', size(FileSelect,2));
                        fprintf('\nCV Parameter: %d', CVparameter);
                        fprintf(' (%d ', CVIndex);
                        fprintf('of %d) ', size(CVValues,2));
                        fprintf('\nPCA Status: %d', PCAyn);
                        fprintf(' (%d ', PCAIndex);
                        fprintf('of %d) ', size(PCAValues,2));
                        fprintf('\nKNN NoN: %d', Nneighbors);
                        fprintf(' (%d ', KNNIndex);
                        fprintf('of %d) ', size(KNNValues,2));
                        fprintf('\nNbinarySimulation %d of ', indS);
                        fprintf('%d ', NbinarySimulations);
                        fprintf('\nTestSet %d ', indCV);
                        fprintf('of %d ', CVO.NumTestSets);
                        fprintf('\nClassSet %d ', indClass);
                        fprintf('of %d ', size(CovarianzeMatrices,3));
                    end

TrainingCovariances=zeros(size(CovarianzeMatrices,1),size(CovarianzeMatrice
s,2),length(TrainingClass));
                        for indMcov=1:length(TrainingClass);
                            if TrainingClass(indMcov)==0;

TrainingCovariances(:,:,indMcov)=CovarianzeMatrices(:,:,1);
                            elseif TrainingClass(indMcov)==1;

TrainingCovariances(:,:,indMcov)=CovarianzeMatrices(:,:,2);
                            end
                            clc;
                            fprintf('\nStart date & time: %s', StartTime);
                            fprintf('\nProcessing file: %s', ActualFRname);
                            fprintf(' (%d ', FileIndex);
                            fprintf('of %d) ', size(FileSelect,2));
                            fprintf('\nCV Parameter: %d', CVparameter);
                            fprintf(' (%d ', CVIndex);
                            fprintf('of %d) ', size(CVValues,2));
                            fprintf('\nPCA Status: %d', PCAyn);
                            fprintf(' (%d ', PCAIndex);
                            fprintf('of %d) ', size(PCAValues,2));
                            fprintf('\nKNN NoN: %d', Nneighbors);
                            fprintf(' (%d ', KNNIndex);
                            fprintf('of %d) ', size(KNNValues,2));
                            fprintf('\nNbinarySimulation %d of ', indS);
                            fprintf('%d ', NbinarySimulations);
                            fprintf('\nTestSet %d ', indCV);
                            fprintf('of %d ', CVO.NumTestSets);
                            fprintf('\nMcov %d ', indMcov);
                            fprintf('of %d ', length(TrainingClass));
                        end
                        TestOutputs=zeros(1,length(TestClass));
                        for indQpoint=1:length(TestClass);

M=repmat(Data4Test(:,indQpoint),1,size(Data4Training,2));
                            M=M-Data4Training;
                            Distances=zeros(1,size(Data4Training,2));
                            for indTp=1:size(Data4Training,2);
```

```matlab
Distances(1,indTp)=M(:,indTp)'*TrainingCovariances(:,:,indTp)*M(:,indTp);
                          end
                          [sorted,inds] = sort(Distances);
                          neighbors_ind = inds(1:Nneighbors); % K nearest

TestOutputs(1,indQpoint)=[sum(TrainingClass(:,neighbors_ind))>(Nneighbors/2
)];
                          clc;
                          fprintf('\nStart date & time: %s', StartTime);
                          fprintf('\nProcessing file: %s', ActualFRname);
                          fprintf(' (%d ', FileIndex);
                          fprintf('of %d) ', size(FileSelect,2));
                          fprintf('\nCV Parameter: %d', CVparameter);
                          fprintf(' (%d ', CVIndex);
                          fprintf('of %d) ', size(CVValues,2));
                          fprintf('\nPCA Status: %d', PCAyn);
                          fprintf(' (%d ', PCAIndex);
                          fprintf('of %d) ', size(PCAValues,2));
                          fprintf('\nKNN NoN: %d', Nneighbors);
                          fprintf(' (%d ', KNNIndex);
                          fprintf('of %d) ', size(KNNValues,2));
                          fprintf('\nNbinarySimulation %d of ', indS);
                          fprintf('%d ', NbinarySimulations);
                          fprintf('\nTestSet %d ', indCV);
                          fprintf('of %d ', CVO.NumTestSets);
                          fprintf('\nQpoint %d ', indQpoint);
                          fprintf('of %d ', length(TestClass));
                      end
                      TP=sum([(TestClass==1)&(TestOutputs==1)]);
                      FN=sum([(TestClass==1)&(TestOutputs==0)]);
                      FP=sum([(TestClass==0)&(TestOutputs==1)]);
                      TN=sum([(TestClass==0)&(TestOutputs==0)]);
                      p=sum([TestClass==1]);
                      n=sum([TestClass==0]);
                      ClassificationMeas(indCV,1)=TP/(TP+FN);%sensitivity
                      ClassificationMeas(indCV,2)=TN/(FP+TN);%specificity
                      ClassificationMeas(indCV,3)=TP/(TP+FP);%PPV
                      ClassificationMeas(indCV,4)=TN/(TN+FN);%NPV
                      ClassificationMeas(indCV,5)=(TP+TN)/(p+n);%accuracy
                  end
                  ClassificationChar(1,1:6)={'Histology', 'Sensitivity',
'Specificity', 'PPV', 'NPV', 'Accuracy'};
                  ClassificationChar(indS+1,1)=CurrClassName;

ClassificationChar(indS+1,2:6)=num2cell(mean(ClassificationMeas,1));
              end
              ActualFRname2 = strrep(ActualFRname, '.mat','');
              ActualFRname3 = strrep(ActualFRname2, 'PCA_','');
              SaveName = strcat('KNN_', ActualFRname3,
'_PCA_',num2str(PCAyn),'_NoN_', num2str(Nneighbors));
              save(SaveName, 'CVparameter', 'Nneighbors', 'PCAyn',
'ClassificationChar');
              xlswrite(SaveName,ClassificationChar);
              end %(KNN NoN loop)
        end %(PCA loop)
    end %(CV loop)
    clearvars -except CVValues FileSelect KNNValues PCAValues StartTime
FileIndex DiagnosisKey
end
```

6.1.14 Diagnostic key for processing of OCT data

Code	Histology
1	Oesophagus
2	Oesoph Inflam
3	Barretts
4	Stomach
5	Stomach Inflam
6	Stomach Int Met
7	UGI Dysplasia
8	UGI Cancer
9	Duodenum
10	Duod Inflam
11	Ileum
12	Ileum Inflam
13	Colon
14	Colon Inflam
15	Crohns
16	Rectum
17	Rectum Inflam
18	LGI Dysplasia
19	LGI Cancer
20	Colon All Abnormal
21	Rectum All Abnormal
22	Oesoph All Abnormal
23	Stomach All Abnormal

6.2 Goniophotometry experiments

6.2.1 Code for correction of goniometry data in MATLAB

```
%Goniometry Raw Data Correction Code Author Dr Iakovos Amygdalos This code
%works on csv files obtained from goniometry and carries out averaging,
%background and angle correction

clear
clc
close
backgroundSwitch = input('Carry out background correction? y/n ','s');
anglecorrectSwitch = input('Carry out angle correction? y/n ','s');


if  backgroundSwitch == 'y'
    backgroundname = uigetfile('.CSV','Select the file to be used as
background');
    background = csvread(backgroundname);
end
fileselect = uigetfile('.CSV', 'Select files for processing','Multiselect',
'On');


for fileindex = 1 : size(fileselect,2);

    filename = fileselect{1,fileindex};
    filename2 = strrep(filename,'.csv','');
    filename3 = strcat('TS_',filename2);

    Var1 = csvread(filename);
    Var2(:,1) = Var1(:,1);
    Var2(:,2) = mean(Var1(:,2:end),2);

    %centering correction
    OT = str2num(filename2(1,12:end));
if OT < 20;
   [Max,Loc] = max(Var2(:,2));
   if Loc>181
       cut = 2*Loc-361;
    Var2 = Var2(cut:361,:);
    Var2(:,1)=-(Loc-cut)/2 : 0.5 : (Loc-cut)/2;
   elseif Loc<181
   lim = 2*Loc-1;
   Var2 = Var2(1:lim,:);
   Var2(:,1)=-(Loc-1)/2 : 0.5 : (Loc-1)/2;

   end
end


    fig1 = figure ( 1 );
    semilogy(Var2(:,1),Var2(:,2));
    title(filename3,'Interpreter','none');
    xlabel('\theta (deg)');
    ylabel('Power (W)');
    saveas(fig1, filename3,'fig');
```

```matlab
        eval([sprintf(filename3) '= Var2;']);
        save([filename3,'.mat'], filename3);

        %Data averaging to give half-curve with positive angles

        posAngleIndex = Var2(:,1)>=0;
        PosAngData = Var2(posAngleIndex,:);
        negAngleIndex = Var2(:,1)<=0;
        NegAngData = Var2(negAngleIndex,:);
        NegDataSort = sortrows(NegAngData,-1);
        AvgData(:,1) = PosAngData(:,1);
        AvgData(:,2) = (PosAngData(:,2) + NegDataSort(:,2))/2;

        fig2 = figure ( 2 );
        semilogy(AvgData(:,1),AvgData(:,2));
        filenameAvg = strcat(filename3,'_AV');
        title(filenameAvg,'Interpreter','none');
        xlabel('\theta (deg)');
        ylabel('Power (W)');
        saveas(fig2,filenameAvg,'fig');

        if  backgroundSwitch == 'y'

            BGPosIndex = background(:,1)>=0;
        BGPosData = background(BGPosIndex,:);
        BGNegIndex = background(:,1)<=0;
        BGNegData = background(BGNegIndex,:);
        BGNegSort = sortrows(BGNegData,-1);
        BGAvg(:,1) = BGPosData(:,1);
        BGAvg(:,2) = (BGPosData(:,2) + BGNegSort(:,2))/2;

        AvgDataBC(:,1) = AvgData(:,1);
        AvgDataBC(:,2) = AvgData(:,2) - BGAvg(1:size(AvgData,1),2);
         filenameAvgBC= strcat(filenameAvg,'_BC');


            fig3 = figure ( 3 );

semilogy(AvgData(:,1),AvgData(:,2),BGAvg(:,1),BGAvg(:,2),AvgDataBC(:,1),Avg
DataBC(:,2));
        title(filenameAvgBC,'Interpreter','none');
        xlabel('\theta (deg)');
        ylabel('Power (W)');
        saveas(fig3,filenameAvgBC,'fig');

        eval([sprintf(filenameAvgBC) '= AvgDataBC;']);
        save([filenameAvgBC,'.mat'],filenameAvgBC);

        %          fig3 = semilogy(Var2Cor(:,1),Var2Cor(:,2));
        %          title(filenameCor,'Interpreter','none'); xlabel('\theta
        %          (deg)'); ylabel('Power (W)');
        %          saveas(fig2,filenameCor,'fig');
        else
            AvgDataBC = AvgData;
            filenameAvgBC = filenameAvg;
        end

        %Smoothing
```

```matlab
    AvgDataBCSM(:,1) = AvgDataBC(:,1);
    AvgDataBCSM(:,2) = smooth(AvgDataBC(:,2),5);
    filenameAvgBCSM = strcat(filenameAvgBC,'_SM');

        fig4 = figure ( 4 );

semilogy(AvgDataBC(:,1),AvgDataBC(:,2),AvgDataBCSM(:,1),AvgDataBCSM(:,2));
        title(filenameAvgBCSM,'Interpreter','none');
        xlabel('\theta (deg)');
        ylabel('Power (W)');
        saveas(fig4,filenameAvgBCSM,'fig');

    if anglecorrectSwitch == 'y'
        AngCorData(:,1) = 0.00000001459*(AvgDataBCSM(:,1)).^6 -
0.000002641*(AvgDataBCSM(:,1)).^5 + 0.0001807*(AvgDataBCSM(:,1)).^4 -
0.005727*(AvgDataBCSM(:,1)).^3 + 0.08402*(AvgDataBCSM(:,1)).^2 +
0.5583*(AvgDataBCSM(:,1)) + 0.5468;
        AngCorData(:,2) = AvgDataBCSM(:,2);

        fig5 = figure ( 5 );

semilogy(AvgDataBCSM(:,1),AvgDataBCSM(:,2),AngCorData(:,1),AngCorData(:,2))
;
        filenameCorAng = strcat(filenameAvgBCSM,'_AC');
        title(filenameCorAng,'Interpreter','none');
        xlabel('\theta (deg)');
        ylabel('Power (W)');
        xlim([0 100]);
        saveas(fig5,filenameCorAng,'fig');

    else AngCorData = AvgDataBCSM;
        filenameCorAng = filenameAvgBCSM;
    end

    fig6 = figure ( 6 );
        semilogy(AngCorData(:,1),AngCorData(:,2));
        filenameCorAngFin = strcat(filenameCorAng,'_F');
        title(filenameCorAngFin,'Interpreter','none');
        xlabel('\theta (deg)');
        ylabel('Power (W)');
        xlim([0 100]);
        saveas(fig6,filenameCorAngFin,'fig');
    eval([sprintf(filenameCorAng) '= AngCorData;']);
    save([filenameCorAng,'.mat'], filenameCorAng);
    clearvars MS* Var* Avg* Ang*
end
beep;
close all
```

6.2.2 Code for fitting the Henyey-Greenstein function

```
%Linear fitting of Henyey-Greenstein function to goniometry data
%Dr Iakovos Amygdalos, Dr Gavin Erry

fileselect  = uigetfile('.MAT','Multiselect','on');
numfiles = size(fileselect,2);
g_matrix = {};
FitStartAngle = input('Please enter the start angle for fitting: ');
FitEndAngle = input('Please enter the end angle for fitting: ');

for fileindex = 1 : numfiles;
 filename = fileselect{1,fileindex};
 filename2 = strrep(filename,'.mat','');
Data = importdata(filename);
FitRange = find(Data(:,1)>FitStartAngle & Data(:,1)<FitEndAngle);
Xdata1 = cos(degtorad(Data(FitRange,1)));
Ydata1 = Data(FitRange,2).^(-2/3);
R = robustfit(Xdata1,Ydata1);
c = R(1,:);
m = R(2,:);
g = (-c+sqrt(c^2-m^2))/m;

graphname = strrep(filename2,'p','.');
fig1 = figure ( 1 );
plot(Xdata1,Ydata1);
hold on;
rline = refline(m,c);
set(rline,'Color','r');
title(graphname ,'Interpreter','none');
xlabel('cos\theta');
ylabel('{Power}^{-2/3} (W)','Interpreter','tex');
savename = strcat(filename2,'_RF');
saveas(fig1, savename,'fig');
pause (1);
g_matrix{fileindex,1} = graphname;
g_matrix{fileindex,2} = g;
close all

end

g_matrix = sortrows(g_matrix,1);
matrixname = input('Please enter a name for the g matrix: ','s');
eval([sprintf(matrixname) '= g_matrix;']);
save([matrixname,'.mat'],matrixname);
```

6.2.3 Dilution plan for first set of microsphere phantoms

Orig. density (number/ml)	New density (number/ml)	Dilution factor	New density (number / µm³)	Optical Thickness
		0.8 micron spheres		
3.72E+11	3.72E+11	none	0.37190	78.04
3.72E+11	9.30E+10	4.000	0.09298	22.28
3.72E+11	1.06E+11	3.503	0.10618	19.51
3.72E+11	3.53E+10	10.526	0.03533	11.14
3.72E+11	2.70E+10	13.793	0.02696	7.4138
3.72E+11	1.43E+10	25.974	0.01432	5.6579
1.06E+11	5.31E+10	7.005	0.05309	3.7069
3.53E+10	1.77E+10	21.053	0.01767	3.0045
2.70E+10	5.39E+09	68.966	0.00539	1.1316
1.43E+10	1.79E+09	207.792	0.00179	0.37557
5.39E+09	1.42E+09	262.226	0.00142	0.29761
1.79E+09	4.71E+08	790.084	0.00047	0.09877

Orig. density (number/ml)	New density (number/ml)	Dilution factor	New density (number / µm³)	Optical Thickness
		1.92 micron spheres		
2.58E+10	2.58E+10	none	0.0258	169.5
2.58E+10	5.16E+09	5.000	0.00516	33.9
2.58E+10	3.45E+09	7.491	0.00344	22.63
2.58E+10	2.53E+09	10.204	0.00253	16.612
2.58E+10	1.20E+09	21.505	0.00120	11.315
3.45E+09	1.72E+09	2.000	0.00172	7.882
2.53E+09	5.06E+08	5.000	0.000506	3.322
1.20E+09	1.72E+08	6.993	0.000171	1.127
1.72E+09	1.52E+08	11.364	0.000151	0.995
5.06E+08	4.45E+07	11.364	0.0000445	0.292
1.72E+08	1.51E+07	11.364	0.0000151	0.0991
1.51E+07	7.55E+06	2.000	0.00000755	0.0495
7.55E+06	5.04E+06	1.500	0.00000504	0.0330
7.55E+06	3.78E+06	2.000	0.00000378	0.0247

6.2.4 Dilution plan for second set of microsphere phantoms

0.58 microns			
Orig. density (number/ml)	Dilution factor	New density (number / μm³	Optical Thickness
0.904	none	none	40.44
0.904	1.348	0.670	30.00
0.904	2.022	0.446	20.00
0.904	4.044	0.223	10.00
0.904	5.777	0.156	7.00
0.904	8.088	0.111	5.00
0.111716	13.481	0.0670	3.00
0.111716	40.442	0.0223	1.00
0.111716	57.774	0.0156	0.70
0.111716	80.884	0.0111	0.50
0.011172	134.807	0.00670	0.30
0.011172	404.420	0.00223	0.10
0.011172	808.840	0.00111	0.05

0.69 microns			
Orig. density (number/ml)	Dilution factor	New density (number / μm³	Optical Thickness
0.541	none	none	54.10
0.541	1.353	0.400200	40.00
0.541	1.803	0.300150	30.00
0.541	2.705	0.200100	20.00
0.541	5.410	0.100050	10.00
0.541	7.729	0.070035	7.00
0.070035	10.821	0.050025	5.00
0.070035	18.034	0.030015	3.00
0.070035	54.103	0.010005	1.00
0.070035	77.290	0.007003	0.70
0.007003	108.206	0.005002	0.50
0.007003	180.343	0.003001	0.30
0.007003	541.030	0.001000	0.10
0.001000	1082.060	0.000500	0.05

0.96 microns			
Orig. density (number/ml)	Dilution factor	New density (number / µm³)	Optical Thickness
0.209	none	none	89.24
0.209	2.231	0.093770	40.00
0.209	2.975	0.070327	30.00
0.209	4.462	0.046885	20.00
0.209	8.924	0.023442	10.00
0.023442	12.749	0.016410	7.00
0.023442	17.848	0.011721	5.00
0.023442	29.747	0.007033	3.00
0.023442	89.240	0.002344	1.00
0.002344	127.486	0.001641	0.70
0.002344	178.480	0.001172	0.50
0.002344	297.467	0.000703	0.30
0.002344	892.400	0.000234	0.10
0.000234	1784.800	0.000117	0.05

1.73 microns			
Orig. density (number/ml)	Dilution factor	New density (number / µm³)	Optical Thickness
0.035	none	none	153.52
0.035	3.838	0.009197	40.00
0.035	5.117	0.006898	30.00
0.035	7.676	0.004599	20.00
0.004599	15.352	0.002299	10.00
0.004599	21.931	0.001610	7.00
0.004599	30.704	0.001150	5.00
0.004599	51.173	0.000690	3.00
0.000690	153.520	0.000230	1.00
0.000690	219.314	0.000161	0.70
0.000690	307.040	0.000115	0.50
0.000690	511.733	0.000069	0.30
0.000069	1535.200	0.000023	0.10
0.000069	3070.400	0.000011	0.05