Cortical thickness estimation of the proximal femur from multi-view dual-energy X-ray absorptiometry

Nikolaos Tsaousis

Trinity College

Department of Engineering

University of Cambridge

A thesis submitted for the degree of

Doctor of Philosophy

February 2015

To my family...

Declaration

This dissertation is the result of my own work and includes nothing which is the outcome of work done in collaboration except where specifically indicated in the text. It is not substantially the same as any that I have submitted, or, is being concurrently submitted for a degree or diploma or other qualification at the University of Cambridge or any other University or similar institution. I further state that no substantial part of my dissertation has already been submitted, or, is being concurrently submitted for any such degree, diploma or other qualification at the University of Cambridge or any other University or similar institution.

It contains **71** figures, **13** tables, and approximately **43,000** words, which do not exceed the prescribed limits set by the relevant Degree Committee. Some of the figures are uninterpretable if not viewed in colour.

Nikolaos Tsaousis

Engineering Department, University of Cambridge,
Friday, 27th of February, 2015

Acknowledgements

I am extremely grateful to many people, without whom my endeavour to perform the work presented herein would have been impossible. First, and foremost, I express my sincere gratitude and admiration to my supervisor, Doctor Andrew Gee, for his patience, support and guidance. I have derived much motivation from his words of encouragement, and have learned how to tackle hard problems using proper scientific practices. I also thank my advisor, Doctor Graham Treece, for his continuous help and guidance. I really enjoyed the time we spent together at the conferences we attended. I would also like to thank my co-supervisor, Doctor Kenneth Poole and all my friends from the engineering department for their valuable input throughout these years. I owe a great amount to Miss Mary Armstrong, founder of the W. D. Armstrong Trust Fund, through which my PhD was funded. Finally, I would like to thank my family. To my mother Denny, my father George, and my brother Dimis I am forever indebted for their unparalleled love and support. I wish some day to be lucky enough and have an equally strong and blessed family of my own. Yet matching what they have provided for me is going to be nearly impossible.

Abstract

Hip fracture is the leading cause of acute orthopaedic hospital admission amongst the elderly, with around a third of patients not surviving one year post-fracture. Current risk assessment tools ignore cortical bone thinning, a focal structural defect characterizing hip fragility. Cortical thickness can be measured using computed tomography, but this is expensive and involves a significant radiation dose. Dual-energy X-ray absorptiometry (DXA) is the preferred imaging modality for assessing fracture risk, and is used routinely in clinical practice. This thesis proposes two novel methods which measure the cortical thickness of the proximal femur from multi-view DXA scans.

First, a data-driven algorithm is designed, implemented and evaluated. It relies on a femoral B-spline template which can be deformed to fit an individual's scans. In a series of experiments on the trochanteric regions of 120 proximal femurs, the algorithm's performance limits were established using twenty views in the range $0^\circ - 171^\circ$: estimation errors were 0.00 ± 0.50 mm. In a clinically viable protocol using four views in the range -20° to 40° , measurement errors were -0.05 ± 0.54 mm.

The second algorithm accomplishes the same task by deforming statistical shape and thickness models, both trained using Principal Component Analysis (PCA). Three training cohorts are used to investigate (a) the estimation efficacy as a function of the diversity in the training set and (b) the possibility of improving performance by building tailored models for different populations. In a series of cross-validation experiments involving 120 femurs, minimum estimation errors were 0.00 ± 0.59 mm and -0.01 ± 0.61 mm for the twenty- and four-view experiments respectively, when fitting the tailored models.

Statistical significance tests reveal that the template algorithm is more precise than the statistical, and that both are superior to a blind estimator which naively assumes the population mean, but only in regions of thicker cortex. It is concluded that cortical thickness measured from DXA is unlikely to assist fracture prediction in the femoral neck and trochanters, but might have applicability in the sub-trochanteric region.

Keywords: hip fracture, cortical thickness, DXA, CT, 3D reconstruction

Contents

Contents	v
List of Figures	viii
List of Tables	xi
Nomenclature	xiii
1 Introduction	1
1.1 Motivation	1
1.1.1 Hip Fracture Epidemiology	1
1.1.2 Fracture Predictors	2
1.1.3 The importance of the femoral cortex	4
1.2 Fracture Type Classification and Treatment	5
1.2.1 Intracapsular fractures	7
1.2.2 Extracapsular fractures	10
1.3 Imaging the femur	12
1.3.1 Bone quality screening	12
1.3.2 Diagnostic imaging	14
1.3.3 Summary	14
1.4 Objectives and overview of the thesis	14
2 3D Reconstructions from Projections	17
2.1 Introduction	17
2.2 Problem definition	18
2.3 Analytical Methods	20
2.3.1 Direct Fourier reconstruction	20
2.3.2 Backprojection and filtered backprojection	23
2.3.3 Theoretical limits to reconstruction resolution	23

2.3.4	Practical implications when measuring femoral cortical thickness	27
2.4	Iterative reconstruction methods	28
2.4.1	Algebraic reconstructions (ART, ILST and SIRT)	28
2.4.2	Statistical iterative reconstructions	29
2.5	Compressed sensing	30
2.6	Model-based methods	31
2.6.1	Template-based methods	32
2.6.2	Statistical model-based methods	36
2.6.3	Discussion of model-based approaches and limitations	40
3	Template-based cortical thickness estimation	44
3.1	Introduction	44
3.2	Template Definition	45
3.2.1	B-spline cross-sectional modelling	46
3.3	2D/3D Correspondence	47
3.3.1	Initialisation	47
3.3.2	Attenuation Projection Profiles (APP)	48
3.3.3	Synthetic Attenuation Profiles (SAP) and Sampling	48
3.4	Template Deformation	51
3.4.1	Optimisation parameters and cost function	51
3.4.2	Regularisation	53
3.5	Evaluation protocol	54
3.5.1	“CT method” and cortical thickness gold standard	54
3.5.2	Thickness orientation	56
3.5.3	Registration between CT and DXA estimates and thickness direction	58
3.5.4	Registration to a canonical femoral morphology	60
3.5.5	Experimental Data and Setup	60
3.6	BERN results (HRpQCT/QCT resolution)	63
3.6.1	Spline approximation	63
3.6.2	Initialisation (blind estimator)	66
3.6.3	Density and blur regularisation	67
3.6.4	Shape and thickness regularisation	68
3.6.5	Evaluating against the bronze standard	73
3.6.6	Evaluating performance in the region where estimation is expected to be best	78
3.7	FEMCO results (QCT resolution only)	84
3.8	Discussion	87

3.9	Summary	94
4	Model-based cortical thickness estimation	95
4.1	Introduction	95
4.2	Model definition	96
4.2.1	Statistical Shape Model	96
4.2.2	Statistical Thickness Model	99
4.3	2D/3D correspondence	99
4.3.1	Registration of DXA and CT estimates and thickness direction .	102
4.4	Model deformation	102
4.4.1	Optimisation	102
4.4.2	Regularisation	103
4.5	Experimental data, setup and evaluation protocol	104
4.5.1	Homogeneous and heterogeneous models	104
4.5.2	Cross-validation testing	108
4.5.3	Model size	108
4.5.4	Number of Principal Components and Horn’s Parallel Analysis .	111
4.6	FEMCO Results	114
4.6.1	Blind estimators	114
4.6.2	CT solution — validation and upper performance bound	116
4.6.3	20-view and 4-view experiments	124
4.7	Discussion	134
5	Conclusions and future work	141
5.1	Comparison of Template and Statistical methods	141
5.2	Conclusions	146
5.2.1	Contributions and main findings	146
5.2.2	Challenges and noteworthy observations	148
5.2.3	Clinical usefulness	150
5.3	Future work	150
	Appendices	152
A	Projection Theorem (aka. central-slice theorem)	153
B	Digitally Reconstructed Radiographs (DRRs)	155
	References	157

List of Figures

1.1	Colour map examples	6
1.2	Classification of fracture regions	7
1.3	Anterior-Posterior(AP) and lateral femoral views.	8
1.4	Subcapital fracture	8
1.5	Garden classification	9
1.6	Extracapsular fractures	11
1.7	CT scan	13
2.1	Parallel-beam geometry	19
2.2	Inverse Radon transform	21
2.3	Direct Fourier reconstruction	22
2.4	Filtered backprojection	24
2.5	Parallel and fan-beam geometries	25
2.6	Filtered backprojection resolution	26
3.1	Template definition	46
3.2	Template semi-automatic positional initialisation	47
3.3	2D/3D correspondence	49
3.4	Synthetic X-ray casting	50
3.5	APs and SAPs	52
3.6	Shape and thickness regularisation examples	54
3.7	Stradwin cortical thickness estimation from CT	55
3.8	Thickness orientation	57
3.9	Comparison of thickness definitions	57
3.10	Registration of cortical thickness estimates	59
3.11	Femoral registration using wxRegSurf	60
3.12	Spline approximation errors and thickness distribution	65
3.13	Blind estimator errors	68

LIST OF FIGURES

3.14 Blind estimator error colour maps	69
3.15 Density and blur regularisation	70
3.16 Cortical thickness estimation efficacy	71
3.17 ROE errors	71
3.18 Shape and thickness regularisation errors	72
3.19 Template method final results	73
3.20 Mean cortical thickness colour maps	74
3.21 Mean error colour maps	75
3.22 Standard deviation and standard deviation improvement colour maps .	76
3.23 Cortical thickness estimation efficacy	77
3.24 Template method error histograms	77
3.25 Restricted view experiments geometry	78
3.26 Vertex angle misalignment	79
3.27 Restricted view experimental errors	81
3.28 Restricted view results	82
3.29 Restricted view colour maps	83
3.30 Clinical resolution experimental results	86
3.31 Clinical resolution thickness estimation efficacy	87
3.32 Clinical resolution mean colour maps	88
3.33 Clinical resolution mean error colour maps	89
3.34 Clinical resolution standard deviation and standard deviation improve- ment colour maps	90
4.1 Statistical model creation steps	97
4.2 Statistical shape model modes	98
4.3 Statistical cortical thickness model modes	100
4.4 2D/3D Registration and placement of APs	101
4.5 FEMCO demographics	108
4.6 Homogeneity/heterogeneity of FEMCO data	109
4.7 Differences between models of different size	112
4.8 Colour maps of thickness principal components	113
4.9 Statistical model eigenvalues and cumulative variance	115
4.10 Mean statistical model errors and thickness distributions	117
4.11 Blind estimator colour maps	118
4.12 Blind estimator colour maps by case	119
4.13 Cortical thickness modelling efficacy when reducing the dimensionality of the model	121

LIST OF FIGURES

4.14	Errors as a function of principal components	122
4.15	Thickness estimation efficacy as a function of PCs	127
4.16	Thickness estimation efficacy for homogeneous and heterogeneous models	130
4.17	Mean statistical model errors and thickness distributions	131
4.18	Mean thickness colour maps	132
4.19	Mean error colour maps	133
4.20	Standard deviation colour maps	134
4.21	RMS improvement colour maps	135
5.1	Regions of thickness estimation per method	142
A.1	Projection theorem	154
B.1	Digitally Reconstructed Radiographs	155

List of Tables

3.1	Summary error statistics of the template method for the BERN experiments	64
3.2	Template method, restricted range error statistics	80
3.3	Summary error statistics of the template, FEMCO experiments	85
4.1	Training and testing cohorts, and three-way experimental setup	105
4.2	FEMCO demographics	107
4.3	FEMCO partitioning of data	107
4.4	Statistical method, cross-validation experimental scheme	110
4.5	Population variance captured and error statistics when reducing the model dimensionality	123
4.6	20-view summary error statistics for the statistical method	125
4.7	4-view summary error statistics for the statistical method	126
4.8	Experimental results when fitting the homogeneous and heterogeneous models to different test sets	128
4.9	Experimental results using the correct registration matrices	139
5.1	Comparison of template and statistical methods	145

Nomenclature

AAM Active Appearance Model

aBMD Areal Bone Mineral Density

AP Anterior-Posterior

APP Attenuation Projection Profile

ART Algebraic Reconstruction Technique

BMD Bone Mineral Density

BR Buckling Ratio

CSA Cross Sectional Area at the mid-neck region

CS Compressed Sensing

CT Computed Tomography

DRR Digitally Reconstructed Radiographs

DXA Dual X-ray Absorptiometry

FNAL Femoral Neck Axis Length

FWHM Full Width Half Maximum

GPA Generalised Procrustes Analysis

HA Hydroxyapatite

HAL Hip Axis Length

HAP Horn's Parallel Analysis

HRpQCT High Resolution peripheral Quantitative Computed Tomography

HSA TM	Hip Structure Analysis
HU	Hounsfield Units
ILST	Iterative Least-Squares Technique
LAD	Locally Affine Deformation
LOO	Leave-One-Out
MDCT	Multi-Detector Computed Tomography
MRI	Magnetic Resonance Imaging
NSA	Neck-Shaft Angle
PCA	Principal Component Analysis
PC	Principal Component
PSF	Point Spread Function
QCT	Quantitative Computed Tomography
RMS	Root Mean Squared
ROE	Residual Optimisation Error
SAPP	Synthetic Attenuation Projection Profile
SIRT	Simultaneous Iterative Reconstruction Technique
SM	Section Modulus
SSM	Statistical Shape Model
STM	Statistical Thickness Model
THR	Total Hip Replacement
TPS	Thin Plate Spline
WHO	World Health Organisation

Chapter 1

Introduction

1.1 Motivation

1.1.1 Hip Fracture Epidemiology

Hip fracture has been recognised as the leading cause of acute orthopaedic admission amongst senior citizens [120], accounting for 30% of all hospitalised patients in 2003 in the United States (US) [49], and the most prevalent cause of injury related death [75]. Extrapolating the results of a study performed in 58 countries reveals that approximately 2.7 million fractures occurred worldwide in 2010 [117]. Their incidence is projected to increase to 6.3 million by 2050 due to the rising life expectancy and changes in lifestyle throughout all five continents [40; 142; 156].

Especially saddening are published mortality rates amongst the elderly, attributed to complications directly related to their hip injury. One-year fatality estimates vary from 12% to 37% [2; 70; 95; 118; 168], although it is promising that this number might be declining [14]. Survivors face pain, reduced quality of life and disability: strikingly half of them are deprived of independent living [75; 112], a concern which older people often report as being worse than death [75].

Additionally, hip fractures are a serious economic burden for our society. They account for an estimated annual direct medical care cost of \$10.3 to \$15.2 billion just in the United States of America [33; 44; 163], with the surgical costs alone being substantial [39], since approximately one third of fracture patients proceed to receive a hip replacement [49].

The prevailing risk factors leading to fractures are falls and osteoporosis, the commonest skeletal condition affecting bone strength. Although this disorder is responsible for fractures at various skeletal sites, the most severe — from both a health hazard and a medical cost perspective — are hip fractures [2]. Since an estimated 30% to 60% of

community-dwelling older adults suffer a fall per annum [140], it is not surprising that a simple fall from standing position is responsible for approximately 90% of fractures amongst the elderly [66]. Moreover, women are on average three times more likely to sustain either an intracapsular or an intertrochanteric extracapsular fracture [16] due to being more prone to osteoporosis, having approximately an 18% lifetime fracture risk, which compares to 6% for men [70; 109]. Several large studies and reviews thereof in the US report that patients between the ages of 65 and 99 years are equally likely to sustain an intertrochanteric fracture [102], whereas subtrochanteric fractures appear as a bimodal distribution which peaks between 20 to 40 years and over 60 years [16]. Young, active individuals are more prone to isolated trochanteric and less severe avulsion fractures, with 85% of such cases attributed to patients under 20 years old [116; 162]. Finally, demographics suggest that the average person who undergoes a femoral neck fracture is 77 years old for women and 72 for men [66], with those originating from a low socio-economic status having the highest rates [62; 65; 110].

It is therefore apparent that a key research aim is the development of new tools to improve the current means of hip fracture risk assessment, prevention, treatment planning and monitoring.

1.1.2 Fracture Predictors

Currently, the state-of-the-art in hip fracture risk assessment is the World Health Organization (WHO) FRAXTM tool, which uses a combination of clinical risk factors and femoral neck bone mineral density (BMD) measured by dual-energy X-ray absorptiometry (DXA) to generate an absolute ten year hip fracture risk for an individual [79; 80]. DXA-measured BMD — a measure of the average mass thickness in g/cm², also known as areal BMD (aBMD)— has been found to be a strong predictor of hip fractures: a large study in 2005 involving 9,891 males and 29,082 females respectively from 12 cohorts and a follow-up period of 16.3 years examined the correlation between fracture risk and change in BMD, and reported similar predictive ability for both men and women [76]. More specifically, some of their findings include a relative risk increase of 2.94 (2.02–4.27, 95% confidence interval) for each standard deviation decrease in BMD for men at the age of 65. Women had a corresponding 2.88 (2.31–3.59, 95% confidence interval) increase. Other minor studies, performed in the laboratory, also support these findings [13; 26].

However, assessing hip fracture risk solely using the BMD value is not enough. Fritscher et al. [50] state that “BMD alone is not sufficient to predict bone failure” and that “additional parameters have to be determined for this purpose”. Faulkner et al. [47] indicate that “there is a significant overlap between osteoporotic and normal

individuals and BMD alone is not sufficient to predict bone failure load”, and Ahmad et al. [1] state that “more than half of women who suffer a hip fracture do not have a low aBMD as measured by DXA”. While the incorporation of clinical risk factors in the FRAXTM tool, such as age, sex, previous fragility fractures, parental history of fractures, smoking and/or alcohol intake, various diseases etc. [79] is a major advance in hip fracture prediction, patient selection remains a hard task since, despite its high specificity, its sensitivity is low [79]. Better analysis of bone structure has the capacity to further improve prediction of fractures in individuals. To this extent, many research groups have investigated the correlation between fracture risk and several biomechanical geometric parameters, the most important of which are the Cross Sectional Area at the mid-neck region (CSA), the Section Modulus (SM), the Buckling Ratio (BR), the Neck-Shaft Angle (NSA), the Femoral Neck Axis Length (FNAL) and the Hip Axis Length (HAL) [5; 7; 84]. The most widespread implementation of such measurements is provided by the HSATM (Hip Structure Analysis) software which is implemented in most modern Hologic DXA scanners (Hologic Inc, Bedford, MA, USA), and provides automated measures for the first three of the aforementioned parameters by analysing mineral profile lines (similar to the Attenuation Projection Profiles (APPs) discussed in Section 2.2) at three different sites: narrow-neck, intertrochanteric and femoral shaft [5]. Nevertheless, opinions in the literature differ on whether such geometrical measurements can be used as predictors of hip fracture. Some studies classify HAL and FNAL as possible predictors [9; 47; 57; 113] but others [58; 81] found no significant correlation. Furthermore, opinions differ for NSA, where some groups report high correlation [57; 58], whereas others found no significant relationship [47; 127]. Kolta et al. [84] suggest that such discrepancies may be due to each group studying different populations that suffered different kinds of fractures (e.g. cervical or intertrochanteric) or due to the sensitivity of such measurements to the patient positioning during the scan. Moreover, the latter research group suggest that an important measurement to be performed is the SM, since it encapsulates size information about the femoral neck diameter which they found to be a significant predictor in women. Finally Beck [5; 7; 8] presents a possible geometric interpretation of why individuals with lower density bones might be prone to fractures. Age and osteoporosis are continuously changing the shape of the femur, as well as the distribution of bone mineral. If the bone area did not change, a lowering of the BMD, i.e. loss of mineral mass, would directly imply loss of structural strength. But this is not the case, as the expansion of outer dimensions with ageing is partially responsible for the BMD decline. This expansion aids in the preservation of bending strength using less material, but an expanded bone may become unable to withstand unaccustomed loading forces, such as the ones exerted during a fall.

1.1.3 The importance of the femoral cortex

One limitation of current screening techniques is that they are incapable of detecting focal structural weaknesses. In particular, there is growing evidence that the distribution of cortical bone in the proximal femur is a key factor affecting its strength characteristics. It is well known that a substantial thinning of the cortex is observed with ageing, especially in the upper femoral neck [5; 38; 74; 107; 161], something that compromises the loading capacity of the femur independently of osteoporosis [107]. This can be explained by the progressive under-loading of the superior cortex, which eventually leads to atrophic thinning [107]. In addition, during sideways and backwards falls, maximum tensile and compressive stresses are observed in the inferior and the superior cortices respectively [5; 38; 107; 161]. These forces are opposite to the ones exerted during normal gait and everyday musculo-skeletal functions [38; 161], consequently the bone is unaccustomed to them. To this end, macro- and micro-finite element analysis models, as well as computational analysis using beam theory have predicted that compressive yielding, or compressive buckling of the superolateral cortex might initiate fracture, especially when the cortex becomes critically thin [5; 107; 161]. Indeed, this prediction has been experimentally confirmed by De Bakker et al. [38]: using a high speed camera and synchronously coupled load measurements, they were able to macroscopically identify the aforementioned compressive fracture initiation. Other scientific groups have accentuated the importance of the cortex by separating its contribution to strength from the contribution due to the trabecular bone. Pistoia et al. [128] used a micro finite element analysis model to investigate the relationship between the strength of the human radius and bone mass. They reported minor loss of strength when the trabecular bone mass was reduced, as opposed to a great loss when the cortical thickness was reduced. Moreover, Verhulp et al. [161] documented that most of the load in the femur is carried by the cortical shell, with the trabecular core hardly loaded in osteoporotic cases. Finally, Holzer et al. [68] evaluated the contribution of the cortical bone in vitro by completely removing the trabecular core from the femoral neck. Hollow bones had a less than 10% reduction in bone strength in all cases, irrespective of other structural parameters.

As far as studies involving the actual measurement of the cortex from Quantitative Computed Tomography (QCT) data are concerned, one needs to be cautious of the limitations imposed by the limited spatial resolution of the scanners. Since the cortical thickness in regions of the femur lies well below the Point Spread Function (PSF) of typical clinical apparatus, blurring leads to an overestimation of thickness and an underestimation of density [156]. Moreover, thin regions may be totally missed. The general agreement is that threshold-based approaches suffer from this problem for cor-

tices below approximately 2.5 mm [156]. Such errors exceeded 100% for sub-millimetre cortices found in the trochanteric and femoral neck regions [156]. Nevertheless, it is worthwhile to note that studies employing such methods report that cortical bone area was a significant predictor of bone strength [26; 31; 149].

Recently, Treece et al. [155; 156; 158] proposed a model-based solution capable of producing unbiased cortical thickness estimates (using clinical Multi-Detector Computed Tomography (MDCT)) down to 0.3 mm, which is well below the > 1 mm PSF of this technology. The technique, which initially assumed a constant cortical density across the whole proximal femur, but whose latest iteration relaxes this assumption by calculating localised cortical density estimates using a prior estimate of the imaging blur, results in thickness estimation errors of 0.12 ± 0.39 mm for thicknesses between 1.0 and 3.0 mm and -0.15 ± 0.23 for thicknesses in the range 0.3 – 1.0 mm. This opens up new research avenues, as even very thin regions of the cortex can now be studied in detail. In fact, the resulting thickness maps (see Figure 1.1 for an example) have been used in cohort studies to identify focal regions of interest that might predispose a hip to fracture [131], and to localize cortical thickening in response to pharmaceutical therapy [130; 132; 135; 166].

1.2 Fracture Type Classification and Treatment

The femur, the largest bone of the human body, forms a “ball and socket” joint with the pelvis. The femoral head (ball) fits in the acetabulum (socket) and is surrounded by a strong and flexible capsule filled with synovial fluid, which helps in joint movement by providing lubrication and stability [83]. In addition to the femoral head, the ligamentous hip joint capsule envelops the femoral neck, and is used to define the first of the two broad categories: **intracapsular fractures** (commonly referred to as femoral neck fractures). Fractures below the insertion of the capsule and down to 5 cm below the lesser trochanter [23] fall into the second category and are known as **extracapsular fractures** (see Figure 1.2). Further subdivision of these is possible according to the precise anatomical location, level of displacement and presence of comminution, as discussed in the following sections. The type of fracture serves as a guide to the form of treatment and/or surgery, and detailed examination of common examples may reveal structural regions of interest.

Figure 1.1: Example colour maps, which allow the intuitive visualisation of cortical thickness across the femoral surface. (a) Mean cortical thickness averaged across multiple proximal femurs after mapping them to a canonical morphology, and the errors of the model-based algorithm proposed by Treece et al. (obtained from [155]). (b) Cortical thickness change after 36 months of Denosumab treatment, as obtained by performing a statistical analysis across many individuals (obtained from [130]).

Figure 1.2: Classification of fracture regions

1.2.1 Intracapsular fractures

Femoral neck fractures account for 45% to 53% of all hip fractures, with a 1:3 male-to-female ratio [108]. Most cases are the result of elderly people falling, although they might also follow a high-energy trauma in younger patients [36]. They are sub-categorised as subcapital (proximally, below the femoral head), transcervical (mid-femoral neck) and basicervical (distally, above the greater and lesser trochanters) [16; 138]. When comparing the X-rays of a healthy femur (Figure 1.3) with a typical example of an intracapsular fraction (Figure 1.4), one can easily identify a) the increased apparent density of the neck, due to overlapping, b) the loss of integrity of cortical bone and c) the clear view of the lesser trochanter in the AP view, due to external rotation [138].

Treatment depends on the age and fitness of the patient and the severity of the fracture, where the Garden System, a commonly used classification method which separates stable (Garden I and II, i.e. non-displaced and without deformity) and unstable (Garden III and IV, i.e. displaced) fractures (see Figure 1.5), can be used as a guide [16; 138; 154].

Undisplaced fractures

The general consensus dictates that even stable fractures need to be treated surgically using operative pinning with three cannulated screws parallel to the femoral neck cortex [108; 146], instead of conservative, non-operative management. Although there is some evidence that the outcome of both approaches is similar [123; 137], conservative

Figure 1.3: a) A standard 0° Anterior-Posterior (AP) view of a healthy femur, and b) its lateral projection. X-ray images obtained from [138].

Figure 1.4: a) A standard 0° (AP) view of a subcapital femoral fracture, and b) its lateral projection. X-ray images obtained from [138].

Figure 1.5: Type I-IV fractures according to the Garden classification of intracapsular femoral fractures. Image synthesised using figures obtained from [138].

Garden I: Trabeculae angulated distally with intact inferior cortex and no displacement.

Garden II: Visible fracture line from inferior to superior cortex but without significant displacement.

Garden III: An easily detectable fracture line with partially displaced and/or rotated fragments.

Garden IV: Complete, rough displacement of femoral head.

treatment leads to a 40% secondary displacement rate, and lacks the early mobilisation following surgical treatment [103; 121]. Furthermore, the chances of succumbing into one of the commonest potential complications, such as avascular necrosis, malunion or nonunion, all of which require a hip replacement re-operation, are reduced [36; 108]. It should be noted that less fit elderly patients might be considered for arthroplasty even in the event of an undisplaced fracture [146].

Displaced fractures

Unstable fractures demand surgical treatment, as otherwise the hip becomes functionless and painful [120]. Hence, all patients undergo either operative reduction followed by internal fixation, or unipolar hemiarthroplasty (femoral head is replaced), or total hip replacement (THR) (both the femoral head and the acetabulum are replaced) [83].

It is preferable to treat young individuals with pinning, as they exhibit a high rate of healing in the absence of osteoporosis. Doing so eliminates the worry of re-operation later in life due to prosthetic wear and loosening, a common problem of arthroplasty, at the expense of higher rates of non-union, secondary displacement or avascular necrosis [108; 146]. In cases where arthroplasty is necessary, highly active young patients with a reasonable life expectancy, or patients with a preexisting joint condition should have THR considered as their primary treatment [122; 146]. Although the results of hemiarthroplasty are better at first and a lower incidence of early dislocation is reported¹ [52; 72; 87], after three to five years THR performs better [42; 96; 150].

It is discouraged to treat older, less fit patients using reduction followed by fixation, as this operation leads to non-union and/or avascular necrosis 30–50% of the time and has an increased re-operation rate [10; 106], although it has a marginally lower mortality rate [37; 124]. Hence, unipolar hemiarthroplasty is usually the treatment of choice for the elderly [108; 146].

1.2.2 Extracapsular fractures

Extracapsular fractures of the hip are sub-categorised into intertrochanteric and subtrochanteric, based on whether they occur proximally or distally to the lesser trochanter [23; 138] (see examples in Figure 1.6). Moreover, depending on the amount of displacement and comminution they are also characterised as stable or unstable [108].

All incidents should be surgically treated, unless other medical contradictions prevail, as conservative management leads to prolonged hospitalisation, high cost per quality ad-

¹It is expected that 10–20% of THR will lead to a dislocation [119].

Figure 1.6: Standard, 0° , (AP) views of a) an intertrochanteric, and b) a subtrochanteric femoral fracture. X-ray images obtained from [138].

justed life year, and high morbidity, especially amongst the elderly [123; 146]. Usually treatment involves reduction followed by internal fixation using various proprietary extramedullar (sliding screws and plates) or intramedullar (Gamma nail) implants [108; 120; 146], although in rare exceptions treatment may involve arthroplasty [154].

Intertrochanteric fractures

Epidemiology studies reveal that their incidence is similar to that of intracapsular fractures: they account for approximately 38–50% off all cases and are distributed in a 1:3 male to female ratio [16; 108]. Post-operative results of stable fractures using intra- or extra-medullar implants display no significant difference [108; 146]. Most patients are treated by screwing a side plate to the femoral shaft, and attaching it to a sliding screw that provides impaction of the proximal fragment to promote union [108]. On the other hand, unstable fractures are better treated using a wide variety of proprietary intramedullar implants: instead of using a side plate, better fixation is achieved by placing a screw or nail in the femoral shaft with a small incision [108].

Subtrochanteric fractures

The least common type of hip fracture, accounting only for 5–15% of all cases [16; 108; 146], occurs between the lesser trochanter and the femoral isthmus (i.e. the proximal end of the femoral shaft). As already mentioned in Section 1.1.1, incidences follow a bimodal distribution which peaks between 20 to 40 years and over 60 years [16]. Young

patients suffer them almost exclusively as a result of a high energy impact, whereas osteoporotic, older individuals might be prone even after a relatively minor trauma [154]. Subtrochanteric are generally the least stable of all hip fractures, and although both intra- and extra-medullar treatment is possible, the former approach has higher rates of successful fixation and reduced operation time [89; 146].

1.3 Imaging the femur

1.3.1 Bone quality screening

X-Ray and Dual X-ray Absorptiometry (DXA)

The WHO classifies DXA as the best densitometric method and recognizes it as the reference technique for measuring and assessing BMD due to its good precision and reproducibility and its acceptable accuracy [2; 45]. Fracture risk assessment and treatment monitoring can be accurately performed from DXA scans, but extra care should be taken to ensure the proper acquisition of such scans, since serious errors in diagnosis and therapy are possible [45]. Such errors frequently originate from improper positioning of the patient when undergoing the scan, or improper calibration of the equipment.

The main advantage of DXA scans is that they only depict mineral content by eliminating any soft tissue [5]. More specifically, by using two different X-ray spectra, and by exploiting the dependence of their attenuation coefficient on the photon energy and the atomic number of the tissue, two different types of material can be inferred by measuring the transmission factors of the two different energy beams [2; 115]. This enables the areal densities (i.e. mass per unit projected area) to be measured [115]. The tissues of interest are bone mineral (i.e. hydroxyapatite (HA)) and soft tissue. Hence each pixel of the scan depicts the sum of the mass of HA (in g/cm^2) that is found along the straight line connecting the source from the detector, which is finally converted to aBMD [5; 115].

In addition, DXA is preferred for the measurement of BMD in multiple skeletal sites due to the very low effective radiation dose delivered to the patient (typically 6.7–31 μS ieverts using a fan beam system for femoral scans — 1/10 of that of a standard chest X-Ray) [1; 99; 115]. When one compares these values with the 10 μS ieverts of natural background radiation received every day by a human, it is immediately apparent that the medical hazard is minor. Finally, the ease of use of the equipment, the short investigation time (typically 5–6 minutes) [11; 99] and the relatively low cost acquisition and operation, add to the reasons why this is the preferred method for everyday clinical practice.

Quantitative Computed Tomography (QCT)

Figure 1.7: A CT scan is comprised of many parallel cross-sectional slices through the body, four of which are shown in this example. Each slice is reconstructed by combining X-ray images from multiple orientations, as described in Chapter 2. Interpolation between them can fill in the gaps, resulting in a complete three dimensional voxel array.

When it comes to detailed geometrical structural analysis and whole bone strength, qCT is doubtlessly the preferred imaging technique. In contrast to X-ray scans, which are 2D projections of the tissue, lacking all depth information, this non-invasive volumetric imaging method incorporates information about the whole 3D structure within the body (see Figure 1.7). It is the only one with the ability to measure the true volumetric density (in mg/cm^3 , using a calibration phantom for conversion from Hounsfield Units) of the trabecular and cortical bone, which inherently provides a better assessment of osteoporosis, accurate calculation of the mechanical characteristics, estimation of bone quality at multiple sites of interest, and great potential for fracture risk assessment [27; 50; 70; 115]. Moreover, the possibility of examining in detail any slice through the data allows the measurement of, but is not limited to, the areas of the cortical and trabecular regions of the femur and the cross-sectional moment of inertia [2]. Finally, one of the major problems in the acquisition of DXA scans, patient positioning, is not an issue here [84].

However, all these advantages over DXA scans are obtained at a cost. QCTs are associated with a high effective radiation dose to the patient: typically 1.2–6 mSieverts, two orders of magnitude larger than a typical DXA scan, making the latter more desirable and the former prohibitive for osteoporosis follow-up [1; 70; 90; 91; 115]. Finally, important financial costs are associated with both acquiring the required infrastructure and the operation of the equipment [2; 70].

1.3.2 Diagnostic imaging

Undisplaced fractures, which account for approximately 15% of cases are not always clinically obvious. Infrequently, radiographic changes might be minimal [125; 138], and in 1% of cases they will not be visible at all on plain X-rays, the currently preferred screening method for diagnostic imaging [24]. Hence, further investigation will be required, with **Magnetic Resonance Imaging (MRI)** being the modality of choice as it can reveal soft tissue pathologies nearby the fracture site such as muscle strains or greater trochanteric bursitis [108]. A plethora of studies confirm the efficacy of MRI in identifying femoral fractures [61; 78; 85], with coronal T1-weighted hip MRI achieving even 100% accuracy [136].

Alternatively, if MRI is contra-indicated or is not available within 24 hours, **CT** or **nuclear imaging** is performed [23; 138]. In the latter case, the radio-pharmaceutical technetium polyphosphate (Tc 99m) has been shown to be up to 98% accurate in identifying occult fractures [54; 67], but it is recommended to allow two to three days of patient bed rest before the scan for best detection chances [82].

1.3.3 Summary

A brief description of various non-invasive imaging techniques of the femur was outlined above, both for fracture risk estimation, and for diagnostic purposes. In summary, DXA is the preferred method, and routinely used in clinical practice for femoral fracture risk assessment and treatment monitoring. On the other hand, MDCT is preferred amongst research groups, as the whole 3D data structure can be studied in detail. This leads to a better assessment of osteoporosis and accurate measurement of potential, novel fracture predictors, such as the biomechanical geometric parameters discussed in Section 1.1.2 or the precise distribution of bone mass [50; 70; 115]. However, this imaging technique is not yet viable in the clinical environment, due to the high radiation dose associated with it, its high infrastructure and operational costs, and the long investigation times needed to analyse the data [1; 2; 70; 91; 115].

1.4 Objectives and overview of the thesis

The present study is an effort to develop a clinically admissible tool which has the potential to improve femoral fracture risk prediction, prevention, treatment monitoring and planning.

As already mentioned in Section 1.1.2, while the FRAXTM tool [79; 80] is a major advance in hip fracture prediction, it ignores focal structural defects. Section 1.1.3 refers

to scientific evidence which suggests that cortical bone thinning is such a defect, and consequently a critical component in characterizing hip fragility. The investigation of different fracture sites in Section 1.2, highlights the importance of performing localised cortical thickness measurements at multiple locations; the development of a sensitive and specific tool, able to target those regions of bone that are at risk is likely to improve fracture risk assessment. Finally, Section 1.3 discusses why DXA scanning technology would be the ideal imaging technique for the measurement of cortical thickness, because of its wide availability, minimal radiation and low cost.

Objectives

1. Design and develop a template-based tool, capable of performing 3D cortical thickness measurements using a clinically viable number of multi-view DXA scans.
2. Design and develop a tool that deforms a statistical shape and cortical thickness model to fit a particular set of multi-view DXA scans, and assess the cortical thickness estimation efficacy.
3. Validate, benchmark, and compare the performance of these tools.
4. As discussed later in Section 2.6.3, significant structural and anatomical variations are observed amongst people of different age, gender and ethnic origin. To this end, the final objective of this work is to examine whether cortical thickness estimation can be improved by using tailored models for different populations.

Thesis structure

This thesis is structured as follows: Chapter 2 sets the foundations of 3D reconstructions from planar projections, and discusses various solutions to this problem which resemble the ones developed and investigated herein. Then, Chapter 3 presents a novel tool that deforms a template of the femoral shape and cortical thickness to fit an individual's set of multi-view DXA scans, and assesses its performance. Moreover, it outlines the tools and materials used throughout this study. Subsequently, Chapter 4 discusses an alternative tool that exploits prior knowledge of the anatomical femoral variations, in the form of a statistical model, to achieve personalised cortical 3D reconstructions. In addition, it compares the efficacy of the algorithm when using a “homogeneous” and a “heterogeneous” model. Finally, a comparison of the two methods is presented in Chapter 5, along with the conclusions and the suggestions for future work.

The terminologies “**DXA template method**” or “**DXA model method**” are used henceforth to refer to the tools presented in Chapters 3 and 4 respectively. If there is

no explicit referral to “template” or “model”, then any method able to measure cortical thickness from DXA scans is assumed. Similarly, the terminology “**CT method**” (which is explained in Section [3.5.1](#)) is used for the estimation of cortical properties from MDCT scans.

Chapter 2

3D Reconstructions from Projections

2.1 Introduction

The calculation of 3D cortical thickness maps is impossible if the mineral content distribution within a femur is unknown. However, if a spatial grid (aka. voxel array) filled with bone density samples is available, a robust technique exists to extract thousands of cortical thickness estimates by identifying the location of the endocortical and periosteal surfaces [158] (see Section 3.5.1). Obtaining such voxel arrays is possible using the QCT imaging technology, which, despite being blurry and of limited resolution, provides information-rich, regularly sampled 3D datasets of the femoral mineral distribution.

Section 1.3.1 discusses the respective merits and drawbacks of DXA and QCT, the two most popular imaging techniques of the femur. Unfortunately, since DXA is the primary, and arguably the sole clinically viable imaging tool for fracture risk assessment, such a voxel array is not usually available and needs to be reconstructed from a set of radiographs (planar projections). To capture the 3D geometry of the femur, these projections should be obtained from different angles (similarly to the principle of operation of CT scanners), but their number should be minimised to reduce radiation exposure.

This chapter first discusses the mathematical foundation of the 3D reconstruction problem. Then, by relating the number of projections to the reconstruction resolution, we explain why the use of an *a-priori* model of the femur is a necessary trade-off to achieve the desired reduction of radiation dose. To this end, many research groups have utilised model-based approaches to fully reconstruct the shape and/or density distribution of the femur from radiographs. However, to the best of the author's knowledge, none of

them were used to examine the femoral cortex. Nonetheless, Section 2.6 examines some of the different techniques, since the reconstruction aspect of their approaches shares many of the same challenges to the ones examined in this thesis.

2.2 Problem definition

The intensity across an X-ray, or DXA scan, represents the attenuation of X-rays as they travel on a straight line through different tissues, which, depending on their mineral content, are described by a different attenuation coefficient μ . Thus, X-rays or DXA scans are a 2D representation (projection) of the underlying 3D data, lacking all depth information. However, the combined knowledge of multiple projections of the same object from different orientations, restricts the possible 3D tissue distributions, since depth information can be inferred from, for example, two orthogonal views.

The 3D reconstruction problem is reduced to a 2D one by separately examining each slice — parallel to the imaging direction — of the 3D voxel array. Hence, the problem simplifies to calculating a 2D distribution of attenuation coefficients $\mu(x, y)$, from a set of 1D projection intensity profiles $I_\theta(r)$ (see Figure 2.1). The viewing direction θ , defines a new coordinate system (r, s) for each projection, which is related to the original (x, y) system as follows:

$$\begin{bmatrix} r \\ s \end{bmatrix} = \begin{bmatrix} \cos \theta & \sin \theta \\ -\sin \theta & \cos \theta \end{bmatrix} \begin{bmatrix} x \\ y \end{bmatrix} \quad (2.1)$$

$$\begin{bmatrix} x \\ y \end{bmatrix} = \begin{bmatrix} \cos \theta & -\sin \theta \\ \sin \theta & \cos \theta \end{bmatrix} \begin{bmatrix} r \\ s \end{bmatrix} \quad (2.2)$$

The 1D intensity projection profile $I_\theta(r)$ can be calculated using the linear attenuation coefficients $\mu(x, y)$ and Equation 2.2:

$$\begin{aligned} I_\theta(r) &= I_0 \cdot e^{-\int_{L_{r,\theta}} \mu(x,y) ds} \\ &= I_0 \cdot e^{-\int_{L_{r,\theta}} \mu(r \cdot \cos \theta - s \cdot \sin \theta, r \cdot \sin \theta + s \cdot \cos \theta) ds} \end{aligned} \quad (2.3)$$

where $L_{r,\theta}$ denotes a line which makes an angle θ with the y -axis and is at a distance r from the origin.

To mathematically simplify the problem, each intensity projection profile is then

2. 3D RECONSTRUCTIONS FROM PROJECTIONS

Figure 2.1: Parallel-beam geometry of X-rays through a 2D-slice of an object, in which $\mu(x, y)$ represents the distribution of the linear attenuation coefficient. The beams of initial intensity I_0 are attenuated and produce a 1D projection intensity profile, $I_\theta(r)$. Darker grey corresponds to a higher attenuation coefficient.

2. 3D RECONSTRUCTIONS FROM PROJECTIONS

converted into an Attenuation Projection Profile (APP), which is defined as

$$\begin{aligned} p_\theta(r) &= -\ln \frac{I_\theta(r)}{I_0} \\ &= \int_{L_{r,\theta}} \mu(r \cdot \cos \theta - s \cdot \sin \theta, r \cdot \sin \theta + s \cdot \cos \theta) ds \end{aligned} \quad (2.4)$$

Finally, if we allow θ to vary, and we stack all projections $p_\theta(r)$ into an (r, θ) coordinate system, for θ ranging from 0 to π , we obtain the *sinogram* of the function $\mu(x, y)$. The transformation which maps a function to its *sinogram* is called the *Radon transform*:

$$\begin{aligned} p(r, \theta) &= \mathcal{R}\{\mu(x, y)\} \\ &= \int_{-\infty}^{\infty} \mu(r \cdot \cos \theta - s \cdot \sin \theta, r \cdot \sin \theta + s \cdot \cos \theta) ds \end{aligned} \quad (2.5)$$

Hence, the problem of reconstructing the original 2D function $\mu(x, y)$ from a set of its projections (or sinogram), reduces to calculating the inverse radon transform:

$$\mu(x, y) = \mathcal{R}^{-1}\{p(r, \theta)\} \quad (2.6)$$

Solution of discrete inverse Radon transform

The inverse Radon transform is uniquely solvable, provided that $p(r, \theta)$ is known for all values of $\theta \in [0, \pi)$ (since $p(r, 0) = p(-r, \pi)$), and all values of $r \in \mathbb{R}$. In reality, only a limited number of $p(r, \theta)$ samples are available, as both r and θ are discrete variables. In this case the inverse Radon transform is an ill-posed problem and might have multiple plausible solutions, as different 2D distributions may result in the same set of limited projections (see simple example of Figure 2.2). The following sections discuss different methods of estimating the original X-ray attenuation coefficient distribution from a finite number of APPs.

2.3 Analytical Methods

2.3.1 Direct Fourier reconstruction

The direct Fourier reconstruction provides a mathematical way of calculating the inverse radon transform, using the *projection theorem* (aka. the *central slice theorem*, see Appendix A).

2. 3D RECONSTRUCTIONS FROM PROJECTIONS

Figure 2.2: Demonstration of non-uniqueness of the inverse Radon transform, when only a limited number of projections is available. In this example, where only the APPs at 0° and 90° are available, assuming *Case 1* (i.e. the shaded regions have a constant, and the blank ones have a zero attenuation coefficient) is equally correct to assuming *Case 2*.

2. 3D RECONSTRUCTIONS FROM PROJECTIONS

This is achieved by first calculating all the 1D Fourier transforms of all the projections $p_\theta(r)$:

$$P_\theta(k) = \mathcal{F}_{1D}\{p_\theta(r)\} \quad (2.7)$$

and placing them in a polar (θ, k) grid. Then, the 2D function $F(k_x, k_y)$ is determined by interpolating the data samples in a Cartesian grid, as shown in Figure 2.3.

Figure 2.3: A 2D function $P(\theta, k)$ is obtained when all 1D Fourier transforms $P_\theta(k)$, of all projections $p_\theta(r)$, are appropriately placed on a polar grid. Then, the function $F(k_x, k_y)$, defined in the Cartesian grid, can be calculated by interpolating the data samples.

Finally, according to the projection theorem, the original function $f(x, y)$ is simply obtained by calculating the inverse 2D Fourier transform of $F(k_x, k_y)$:

$$f(x, y) = \mathcal{F}_{2D}^{-1}\{F(k_x, k_y)\} \quad (2.8)$$

It is worthwhile to note that a poor interpolation of the data samples $P(k, \theta)$ can lead to significant artifacts. Thus, researchers tend to prefer reconstructions by filtered backprojection (see Section 2.3.2 below) which does not suffer from this problem, or an iterative reconstruction scheme variant (see Section 2.4).

2.3.2 Backprojection and filtered backprojection

The simplest and most intuitive solution to the 3D reconstruction problem is known as backprojection and can be achieved by “smearing”, or “back-projecting” each APP line along its line of projection, and then superimposing the resulting 2D images (see Figures 2.4(a) and (b)). This can be achieved using the formula

$$\mu'(x, y) = \int_0^\pi p(x \cos \theta + y \sin \theta, \theta) d\theta \quad (2.9)$$

However, this approach has no mathematical basis, and produces unsatisfactory results.

Instead, it follows from the *central slice theorem* (see Appendix A) that each APP needs to be filtered in the frequency domain by the *ramp filter* $|k|$ before back-projecting, or in mathematical notation

$$\begin{aligned} \mu''(x, y) &= \int_0^\pi p^*(x \cos \theta + y \sin \theta, \theta) d\theta \\ \text{where } p^*(r, \theta) &= \mathcal{F}^{-1}\{P^*(k, \theta)\} \\ \text{and } P^*(k, \theta) &= P(k, \theta) \cdot |k| \end{aligned} \quad (2.10)$$

Examples of filtered backprojection can be seen in Figures 2.4(c) and (d). The *ramp filter* is introduced by the Jacobian when converting from polar to Cartesian coordinates in the polar variant of the inverse 2D Fourier transform:

$$\begin{aligned} k_x &= k \cos \theta \\ k_y &= k \sin \theta \\ J &= \begin{vmatrix} \partial k_x / \partial k & \partial k_x / \partial \theta \\ \partial k_y / \partial k & \partial k_y / \partial \theta \end{vmatrix} = \begin{vmatrix} \cos \theta & -k \sin \theta \\ \sin \theta & k \cos \theta \end{vmatrix} = |k| \end{aligned} \quad (2.11)$$

In practice, the *ramp filter* $|k|$ is cut off at a frequency k_{max} , since the useful high frequency Fourier content is limited by the spatial sampling rate of the scanner (i.e. the spacing between neighbouring detectors), and is called the Ram-Lak filter.

2.3.3 Theoretical limits to reconstruction resolution

As already mentioned in Section 2.2, the reconstruction resolution is dependent on the density of the APP samples, $p(\theta, r)$, which in turn is dependent on the X-ray beam width (which determines the X-ray detector spacing) and the number of projections. It

2. 3D RECONSTRUCTIONS FROM PROJECTIONS

Figure 2.4: Reconstructions of the Shepp-Logan phantom using 4 (a, c) and 20 (b, d) views in the ranges 0° – 51° and 0° – 180° respectively, by simple (2nd row) and filtered backprojection (3rd row). The figure was created using MATLAB 2012b, The MathWorks, Inc., Natick, Massachusetts, United States.

2. 3D RECONSTRUCTIONS FROM PROJECTIONS

Figure 2.5: (a) Parallel and (b) fan beam geometries used to calculate the reconstruction resolution from a limited number of projections. For fan beam geometry, R represents the reconstruction radius, and R_s is the distance of the X-ray source from the origin, O .

is relatively straightforward to calculate how the former affects the sample spacing in the r direction using the Nyquist criterion [151]:

$$\Delta r \leq \frac{\Delta s}{2} \quad (2.12)$$

where Δr is the sampling distance and Δs is the beam width. However, as far as the reconstruction problem of this thesis is concerned, this limitation is *not* of paramount importance as it is a fixed variable dependent on the imaging equipment.

On the other hand, it is of much greater interest to relate the achievable resolution with the number and orientation of projections. This is a highly challenging problem and a rigorous mathematical proof is hard to determine [77; 151]. Joseph and Schulz [77] present a very intuitive mathematical analysis of the fan beam geometry shown in Figure 2.5b, which although cannot qualify as a proof, leads to the following equation:

$$N_{min} = \frac{2\pi R_s R v_m}{R_s - R} \quad (2.13)$$

where N_{min} is the minimum number of projections required for an **artifact-free** reconstruction of an object with maximum spatial frequency v_m within a region of radius R , and R_s is the distance of the fan-beam source(s) from the centre of the imaging region.

A similar formula can be derived for the parallel beam geometry of Figure 2.5a, by

2. 3D RECONSTRUCTIONS FROM PROJECTIONS

letting $R_s \rightarrow \infty$ in Equation 2.13:

$$\begin{aligned} \lim_{R_s \rightarrow \infty} \{N_{min}\} &= \frac{\overset{1}{R_s}}{R_s - R} 2\pi R v_m \\ &= 2\pi R v_m \end{aligned} \quad (2.14)$$

It is reassuring that this formula is identical to the one calculated by Snyder and Cox [148] after a long and thorough analysis for a parallel beam geometry of N uniformly distributed views in the range 0° – 180° . Moreover, Logan’s [100] analysis leads to the same result and goes on to prove that Equation 2.14 is valid even if the projections are arbitrarily distributed in the range 0° – 180° . However, as Suetens [151] notes, the latter analysis does not investigate the noise propagation effects of the unequally large gaps in the Fourier space.

Experimental validation

Figure 2.6: The original pictures (a) and (d) with maximum spatial resolution 6.4 and 25 line pairs per 100 pixels respectively, were reconstructed in (b) and (e) by filtered backprojection using 20 and 79 projections respectively. Finally, (c) and (f) are thresholded images of the reconstructions which exhibit an artifact free radius $R = 100$ pixels, as predicted theoretically.

2. 3D RECONSTRUCTIONS FROM PROJECTIONS

To this end, the following experiment was devised to relate the reconstruction resolution of a thin layer of femoral cortex, with the number of projections. The spatial frequency in this analysis is given in line pairs per mm (lp/mm), which, as the name suggests, reflects the number of distinguishable black-and-white line pairs in a millimetre.

We start by assuming that the pixel size of the original distribution is equal to the scanner’s detector spacing Δs : a typical value for clinical resolution CT is around 0.5 mm. Hence, a slightly bigger than average cross-section of the greater trochanteric region, measuring around 10 cm across its longest axis, would appear 200 pixels wide. Equation 2.14 then suggests that 20 equally distributed projections in the range 0° – 180° will guarantee an artifact free reconstruction region of radius 100 pixels, down to a resolution of 0.032 lines per pixel (6.4 lp/100 pixels). The simulation of Figure 2.6c verifies the above prediction, as *line pairs* are evidently no longer distinguishable outside the 100 pixel radius.

The above result is equivalent to a cortical thickness of

$$\frac{0.5 \text{ mm/pixel}}{0.032 \text{ lines/pixel}} = 15.6 \text{ mm/line.}$$

Undoubtedly more views are required for a model-free, accurate reconstruction, since the cortex is typically thinner by at least one order of magnitude.

Following the previous observation, Equation 2.14 can be also adopted to estimate the number of projections required to reconstruct a 1 mm wide cortical layer:

$$N_{min} = 2\pi Rv_m = 2\pi \cdot 100 \text{ pixels} \cdot \frac{1 \text{ lp}}{4 \text{ pixels}} \cong 79$$

Once more, this calculation is visually confirmed in Figure 2.6f. Considering that a) we are aiming to calculate thicknesses in the sub-millimetre range and that b) 79 projections already defeats the purpose of this research — since this would deliver a radiation dose comparable to a CT scan, albeit slightly less (Section 1.3.1) — an alternative method should be devised.

2.3.4 Practical implications when measuring femoral cortical thickness

The analysis of the previous section (Section 2.3.3) discusses the number of projections required for a **guaranteed** reconstruction resolution within a certain radius. However, this does not preclude the possibility of reconstructing high detail features, of higher spatial resolution than predicted, using a small number of views. The simplest example of such a case would be to image at 0° and 90° a high density, thin, horizontal line in

2. 3D RECONSTRUCTIONS FROM PROJECTIONS

a low density background: the line will be properly back-projected, and the only factor limiting the accuracy of its width would be the PSF of the scanner.

Nevertheless, imaging the femur imposes a practical limitation to the range of possible views: typically, a maximum range of 60° is possible before the acetabulum or the contra-lateral femur obscures the pertinent femur’s trochanteric region [1; 84]. The effect of this constraint is clearly visible in the left column of Figure 2.4, where only regions of the thin shell “parallel” to the X-rays are relatively distinguishable.

To conclude, relying solely on projection data when estimating the femoral cortical thickness from a small number of projections is expected to produce very poor results, thus a *prior*-model is encouraged to assist in filling the missing Fourier space gaps.

2.4 Iterative reconstruction methods

One of the key benefits motivating the development of the DXA method is the reduced radiation exposure when compared to MDCT. However, this comes at the cost of reduced information. In the previous section (Section 2.3.3), the reconstruction resolution is examined as a function of the number of DXA scans. It is demonstrated that, unfortunately, “back-projection” techniques are unable to cope with the sparseness of the data in a clinically realistic scenario. A potential trade-off would be to assist the reconstruction process using some *prior* anatomical information. To this end, the study of iterative reconstruction methods is motivated by the ease of incorporating a *prior* model, for example by initialising the solution based on the expected distribution [15; 126], or by constraining the possible outcomes. The following sections first lay the foundations by briefly discussing the commonly used algebraic iterative reconstructions (which are solely data-driven and do not require a prior model) and subsequently examine the more important statistical-based iterative methods.

2.4.1 Algebraic reconstructions (ART, ILST and SIRT)

In the case of algebraic iterative techniques, the reconstruction problem is treated as a set of linear equations. It is described by

$$\mathbf{b} = \mathbf{P}\mathbf{x} \tag{2.15}$$

where $\mathbf{b}$ is a vector holding all projections, $\mathbf{P}$ the set of linear equations that define the mapping from the object to its projections, and $\mathbf{x}$ a vector of all voxels to be determined [126]. Due to noise, the system in practice is not directly invertible — even if it was, inverting matrix $\mathbf{P}$ poses a significant computational challenge due to its size.

2. 3D RECONSTRUCTIONS FROM PROJECTIONS

Hence, the problem can be regarded as an optimisation problem with objective to minimise the difference

$$\mathcal{O}(\mathbf{x}) = |\mathbf{P}\mathbf{x} - \mathbf{b}| \quad (2.16)$$

On each step of the iteration, voxels along the direction of an X-ray are all simultaneously modified to fit the projections. The update formula of the voxels defines the correction sequence [15].

Other correction sequences include the Algebraic Reconstruction Technique (ART)¹, the Simultaneous Iterative Reconstruction technique (SIRT) and the Iterative Least-Squares Technique (ILST). ART is the least computationally expensive, but is susceptible to noise, since its solution is heavily dependent on the last projection examined. On the other hand, ILST and SIRT cope better with noisy data by converging to the least squared difference solution, and adapting to the fact that not all projections will be accurately predicted [15].

2.4.2 Statistical iterative reconstructions

Up until recently, the application of iterative reconstruction methods with a statistical basis was confined to nuclear imaging. CT datasets are usually larger than their nuclear counterparts and hence computationally very challenging to process. Moreover, nuclear imaging is susceptible to Poisson noise governing the radioactive process, which explains the necessity of a statistical approach.

On the other hand, although CT imaging does not suffer from this problem, the evolution of CT scanners towards a lower radiation dose and the need for 3D reconstructions from fewer projections led to datasets that are increasingly noisy. To this end, iterative reconstruction algorithms with a statistical basis have become more popular as they are particularly suited to cope with noise. Of course, this is only possible because of the significant recent advancements in computing infrastructure.

Statistical iterative reconstructions are very similar to ARTs, but the formula defining the correction sequence aims to satisfy an objective function which usually derives from the Bayesian theorem:

$$p(\Lambda|Q) = \frac{p(Q|\Lambda)p(\Lambda)}{p(Q)} \quad (2.17)$$

¹Variants include the Additive Algebraic Reconstruction Technique (AART) or the Multiplicative Algebraic Reconstruction Technique (MART).

2. 3D RECONSTRUCTIONS FROM PROJECTIONS

where Λ represents the reconstructed image and Q the observations (projection data). The probability of the observations, $P(Q)$, is constant and can be ignored.

Maximising the posterior probability $P(\Lambda|Q)$ is termed the *maximum-a-posteriori* (**MAP**) approach. Assuming that a mathematical expression for the *prior* probability $p(\Lambda)$ is available (for example by imposing constraints on the possible outcomes in the form of regularisation), the objective function is then given by

$$\arg \max_{\Lambda} p(\Lambda|Q) = \arg \max_{\Lambda} (\ln p(\Lambda|Q) + \ln p(\Lambda)) \quad (2.18)$$

Maximising the log-likelihood is possible, since the logarithm is a monotonically increasing function.

On the other hand, it is common practice to assume that all possible reconstructions have the same *a-priori* probability of being correct, $p(\Lambda)$ (mainly because it is very hard to mathematically express it), so this factor can also be ignored. In that case, the objective function boils down to maximising $p(Q|\Lambda)$, i.e. the likelihood that a particular reconstruction is correct, given the projections. This approach is called *maximum-likelihood* (**ML**). The system of linear equations defining the imaging process can then be solved with guaranteed convergence, using algorithms such as the *expectation-maximization* (**ML-EM**).

2.5 Compressed sensing

Compressed sensing (CS) is a relatively recently invented approach for acquiring and reconstructing sparsely sampled signals by harnessing optimisation methods. Approximately a decade ago, Candes, Tao, and Donoho [18; 19; 41] demonstrated for the first time that a signal may be perfectly recovered using fewer samples than previously thought. Until then, Nyquist-Shannon’s criterion determined the threshold for lossless reconstructions, which states that the sampling rate needs to be at least twice the highest frequency component observed in the signal. However, this groundbreaking idea applies only to datasets which are known to be sparse in some domain. Since then, this technique has been successfully applied to many signal processing areas; MRI is a prominent example in the medical imaging field, with CT being a noteworthy candidate.

The first step requires the decomposition of a signal to a basis whereby the coefficients of the components are sparse (i.e. their vast majority are equal to zero). Subsequently, reconstruction of a sampled signal can be performed in the usual way by recovering the coefficients. However, sparse sampling leads to a linear system with more unknowns than

2. 3D RECONSTRUCTIONS FROM PROJECTIONS

equations, which typically has an infinite number of solutions. Therefore, the problem needs to be regularised before it can be uniquely solved. The key to CS is imposing a non-linear constraint involving the $\mathbf{L}^1$ -norm, which is analogous to minimising the number of non-zero coefficients, as explained in the next paragraph.

Typical approaches used in medical imaging, such as the filtered-backprojection method (Section 2.3.2) aim to minimise the total “energy” in the image, which is equivalent to regularising based on the Euclidean distance $\mathbf{L}^2$ -norm. In other words, they assume that all unobserved Fourier frequency coefficients are zero, a technique that hardly produces acceptable results, with many streak artifacts when a small number of projections are used (consider the example in the bottom right corner of Figure 2.4) [18]. On the other hand, Candes et al. [19] showed that optimising based on the “Manhattan” distance $\mathbf{L}^1$ -norm (that is examining neighbouring pixels only in the horizontal and vertical directions, and not the diagonals) leads with “overwhelming probability” (sic) to the same solution that would be obtained if the $\mathbf{L}^0$ -norm was used instead. The $\mathbf{L}^0$ -norm refers to the number of non-zero coefficients, which is generally not used in practice as it results in very computationally expensive algorithms, in contrast to its $\mathbf{L}^1$ counterpart for which very efficient solving methods exist. When an optimisation function whose objective is the minimisation of the $\mathbf{L}^0$ -norm is used, impressive reconstructions can be achieved. For example Chartrand [25] presents an algorithm which manages to reconstruct an exact replica of the Shepp-Logan phantom using just nine equiangular radial lines, or just 3.5% of the MRI k-space. Trzasko et al. [160] achieved the same result using ten radial lines using a homotopic $\mathbf{L}^0$ -minimisation (which was replicated in different phantoms as well), but failed to do so when $\mathbf{L}^1$ -minimising.

Therefore, provided that a suitable domain that allows the representation of femoral scans in a sparse manner can be identified, CS techniques could be applied to reconstruct the femoral cortex using a small number of projections. Identifying the basis for such a decomposition seems like a challenging task, and would most likely require modeling the femur using a linear combination of components identified from statistical analysis.

2.6 Model-based methods

Model-based 3D reconstructions have been attempted using just one radiographic scan or multiple ones obtained from different imaging angles. They can be subdivided into “**template-based**”, which involve a femoral atlas that can be deformed to fit a particular set of radiographs, and “**statistical**”, which incorporate some mathematical *prior* knowledge of a particular anatomy of the femur from a population study [2].

2.6.1 Template-based methods

Template-based methods rely on *a-priori* knowledge of the subject’s anatomy, but ignore statistical variations during the reconstruction process. Although the template *per se* may be the outcome of a statistical study, its transformation is data-driven. The implementation of such approaches in the literature varies amongst research groups to suit their particular needs. Nevertheless, all of them share the following four main components: template definition, 2D/3D correspondence protocol, template deformation means and reconstruction evaluation. These are examined separately below, with examples of their application to femoral reconstructions from radiographs.

Template definition

The first step involves the formation of a mathematical representation of the template. As far as femoral reconstructions are concerned, a triangulated surface mesh is usually used. This generic model can be either obtained by segmenting a single bone surface from CT [51; 84; 93; 94], or by calculating the mean shape from a population study [91; 152]. A prerequisite for the latter approach is the alignment of all training subjects to a common reference axis before the mean calculation, which is usually performed by Generalised Procrustes Analysis (GPA) based on landmarks [59; 91; 143].

The choice of template primarily depends on the objective of the study, and its design should allow a straightforward measurement of the variables of interest. In the aforementioned examples, reconstructing the femoral shape was the sole intention: this could be very useful in preoperative assessment. More extensive models which, for example, are intended to allow femoral strength measurements could also incorporate material properties and separate different tissues. To this end, in Chapter 3 we design, implement and evaluate a template that is particularly suited to allow cortical thickness measurements.

2D/3D correspondence

Once the template has been created, the protocol for establishing correspondences between the 3D model and the 2D radiographic images must be established. Usually this involves the extraction of clearly identifiable geometrical entities, such as point sets (aka. landmarks), edges or contours, which are detectable in both the template (and/or their projections) and the 2D scans [20; 21; 51; 84; 93; 94]. However, this step can be avoided if an intensity-based registration is performed (see Section 2.6.2 for the discussion of such an approach using a statistical model).

2. 3D RECONSTRUCTIONS FROM PROJECTIONS

Although landmark positioning needs to be performed manually, as in Langton et al.'s [91] work, the advantage of contour based approaches is that they can be implemented to run without user intervention [104]. For example, Gamage et al. [51] proposed a method which automatically extracts the outer contours in the projections of the 3D template using a ray tracing technique [104], whereas inner ones are identified using the saddle points of the surface curvature. Successively, an edge extraction algorithm is used to identify the same contours in the patient specific DXA scans, whose main steps involve adaptive image filtering (to enhance the edges), thresholding with hysteresis (to define transitional regions), edge relaxation (to remove pixels erroneously identified as edges) and a final filtering to skeletonise the output.

However, contour based algorithms that require an operator to “draw” contours on radiographs also exist, such as the Non-Stereo-Corresponding Contour (NSCC) algorithm [84; 92; 93; 94]. Specifically, nine different anatomical areas are defined on the generic shape (femoral head, inferior and superior femoral neck, greater and lesser trochanters, and medial, lateral, anterior and posterior parts of the diaphysis), whose bounding volumes can be used to automatically define contours on the template projections. For a personalised reconstruction, the operator needs to manually identify the same contours on the radiographs, a step prone to user error [104].

Template deformation and 2D/3D registration

Once correspondences between the 3D template and the 2D radiographic projections have been established, an algorithm is employed to minimise the distance between them. If the personalised reconstruction involves fitting the model to just one scan, as in Langton et al. [91], it is possible to simply warp the template to the test scan by aligning the landmarks (using GPA or affine/similarity transformations for example) and interpolating in between using Thin Plate Spline (TPS) [12] deformations. An unavoidable drawback of this approach is the necessary assumption that the bone depth scales proportionally with bone length and width, as it is impossible to infer the full 3D geometry from a single projection; Langton et al. acknowledged that, but stated that this assumption proved reasonable for the majority of cases.

However, predominantly, 3D reconstructions involve the registration of the template to multiple radiographs obtained at different orientations [20; 21; 51; 84; 93; 94]. Hence, the alignment of the correspondence features to all of the scans simultaneously is no longer feasible before convergence, as this would mean that the template perfectly fits all the projections and hence is the solution of the personalised reconstruction. Consequently, the algorithm responsible for performing the 2D/3D registration is a two-step iterative procedure, which alternates between the correspondence and transformation

2. 3D RECONSTRUCTIONS FROM PROJECTIONS

steps [104]: after each incremental deformation a new set of optimal correspondences should be calculated.

For example, Le Bras et al. [93; 94] proposed a semi-automatic full 3D reconstruction of the proximal femur, using information from two X-ray radiographs (also known as stereo-radiography). The radiographs were obtained at the AP and lateral views in vitro. The Non-Stereo-Corresponding Contour (NSCC) method used, first calculates a global rigid transformation to minimise the distance between the features in a least squares sense using an optimisation procedure, and then iteratively elastically deforms the template by means of a Kriging algorithm [159] and retro-projects it in the AP and lateral views. The deformation is such so that the contours of the anatomical regions of interest in the 2D radiographs match those of the retro-projections. The same algorithm was used by Kolta et al. [84], but instead of relying on standard X-rays they used a pair of DXA scans per specimen.

Gamage et al. [51] also based their personalised reconstruction on two DXA scans but deformed their model using a different approach. Two 2D translational fields were defined in the lateral and anterior views, defined by the point correspondence between the 2D projected contours from the model and the 2D contours from the DXA images. These fields were used to define a full 3D, non-rigid translational field. This was achieved by means of a TPS algorithm which interpolated the sparse translations. The template was then deformed according to the 3D translational field defined above, and the procedure was repeated until convergence.

Reconstruction evaluation

Le Bras et al. [93; 94] performed 23 successful, and 2 unsuccessful reconstructions of excised proximal femurs (both male and female), and evaluation was performed by comparing them to their corresponding segmentations from CT. Each subject was first registered to the segmented surface using translations and rotations according to a least squares matching scheme. The accuracy of the reconstruction algorithm was described by the mean point to surface distance, Root Mean Squared (RMS) and maximum errors, which were 0.7 mm, 1.0 mm and 6.7 mm respectively. Unfortunately they do not quantify the shape variance amongst the test specimens (although this is hard for a template matching method like this one), nor the errors obtained when best fitting the undeformed template to the test femurs.

Kolta et al. [84], reconstructed 25 proximal femurs using the same algorithm (NSCC), but used DXA scan pairs instead of plain X-rays (single X-ray absorptiometry). Moreover, a more extensive evaluation of the reconstruction accuracy was performed, as they also presented measurement errors of the following 3D geometric parameters:

2. 3D RECONSTRUCTIONS FROM PROJECTIONS

femoral head diameter, Femoral Neck Axis Length (FNAL), mid neck cross-sectional area and neck-shaft angle (NSA). Reconstruction time per specimen was 10 minutes, and the gold standard comprised segmented surfaces from CT as before. They reported a mean shape error of 0.06 ± 1.02 mm, 95% of errors ($= 2 \times \text{RMS}$) less than 2.1 mm and a maximum error of 7.8 mm across all subjects. However, they do not state the amount of error obtained when best-fitting their generic model — a separate CT femoral segmentation of an 89 year old female — to the test specimens (prior to any deformation). Thus there is no baseline against which their algorithm’s deformation effectiveness can be compared. As far as the recovery of the 3D parameters is concerned, a coefficient of variation of less than 5% was reported.

Both studies mentioned above report maximal errors at the trochanters. They attribute this observation to the high variability between subjects, as it is known that the strong muscles attaching to these regions create strong local deformations [84].

In the work of **Gamage et al.** [51], testing involved an extremely small and patient-specific dataset, and hence the results presented are unlikely to be representative. Furthermore, they do not state whether the template was generated using a separate dataset. Just six cadavers (three left/right pairs) were reconstructed and the maximum/minimum Euclidean errors were 1.26 mm/−1.16 mm. The errors of the best-fitted undeformed template are not specified, nor the intra-test set shape variation.

Finally, **Langton et al.** [91] used two template models to assist the personalised reconstructions from a single Digitally Reconstructed Radiograph (DRR). The first one was trained using nine low resolution CT scans, 1.08 mm CT pixel size, and the second using thirteen high resolution scans, 0.674 mm CT pixel size. Twenty separate femurs were used as a test set to assess the accuracy of their algorithm (nine low, and eleven high resolution). Their results are presented as “depth” and “offset” errors. These correspond to the distance between the lateral and medial femoral surfaces from a reference plane parallel to the DRR orientation and were obtained using a ray-casting technique. They reported mean absolute depth errors of 3.40 mm and mean absolute offset errors of 2.97 mm for the low resolution CT test scans, and 1.73 mm/1.33 mm for the high resolution respectively. These results are comparable to the previous studies, although the others employed more projections in their experiments. It is worth noting that their low resolution test set comprised very similar femurs: the SDs of the depth and offset were just 1.74 mm and 1.47 mm respectively. However, they do not state the errors obtained when best fitting the undeformed generic model, neither for the low, nor for the high resolution experiments, so the effectiveness of their reconstruction algorithm is not clear.

2.6.2 Statistical model-based methods

The principal modes of shape and/or density variations of bones within a training cohort from their mean values can be expressed as a SSM and/or a Statistical Appearance Model (SAM) respectively. Search functions can be then used to fit these to observed data and achieve personalised reconstructions. This procedure is known as Active Shape Modelling (ASM) or Active Appearance Modelling (AAM) [143]. ASMs and AAMs have been implemented in various different ways, and a discussion of some of the approaches follows, emphasising both the model construction protocol and the choice of fitting algorithm.

Whitmarsh et al.

Whitmarsh, Humbert et al. [70; 164; 165] presented a way to fully and automatically reconstruct both the shape and BMD distribution using both single- and multi-view (1–4 views) DXA images and a statistical atlas. Each of their three studies utilised an AAM [30] built from different training populations, consisting of a) 60, b) 85 (both male and female, single view experiments) and c) 44 (all female, multi-view experiments) QCT femoral scans. The main steps involved in the creation of the atlases are summarized below.

All segmented surfaces are registered to a reference femur by means of affine transformations (translation, rotation and scaling), followed by a multi-scale B-spline registration. Care was taken to constrain the movement of the control points in order to guarantee a diffeomorphic deformation field [141]. Next, Principal Component Analysis (PCA) was applied on the transformed vertices of the meshes to create a statistical model of these surface transformations.

A similar approach led to creation of a PCA model for the BMD distribution. In order to decouple BMD variations from shape variations, the segmented specimen surfaces were deformed using TPS interpolation [12] to the mean reference shape, before creating the model by applying PCA on each voxel. Apart from examining the performance of separate shape and density models, they also investigated the accuracy of a combined model which encoded both statistics.

For studies (a) and (c), no DXA scans paired with corresponding QCT scans were available. Hence, they resorted to independent testing using artificially produced DXA images from thirty (study a) and twenty (study c) separate femoral QCT scans. These were synthesised by means of a ray-casting technique (DRRs). The DRRs, along with the statistical model, were used to perform a full volumetric reconstruction. This was achieved by iteratively maximising the similarity between the DRR of the test specimen

2. 3D RECONSTRUCTIONS FROM PROJECTIONS

and the DRR of the deformed model. Study (b) made use of thirty independent real DXA scans, hence no DRR synthesis was necessary. The number of PCs to retain was determined by Horn’s Parallel Analysis in (a) (11 shape modes and 14 density modes), Cattells scree test in (b) (11 combined modes) and not specified in (c).

Evaluation was performed by comparing the reconstructed femur to the original QCT scan. When using a single DRR view and separate thickness and shape models, they reported a mean shape reconstruction error of 1.2 mm (1.6 mm RMS) and a BMD accuracy of 4.6% across the whole range of density values. The corresponding values obtained when using real DXA scans and the combined shape and thickness model were 1.1 mm (1.45 mm RMS) and 4.9%. For two views (frontal-sagittal), reported errors decreased to 0.9 mm (1.65 mm RMS) and 3.2%. Further views added small improvements, with errors reaching 0.7 mm (1.0 mm RMS) and 2.9% when using 4 views. Unfortunately they do not mention the training and test set population variances, nor the errors obtained after fitting the mean models to each individual. Thus it is not possible to quantify the extent to which the model deformations improve the “blind guess” estimation.

Ahmad et al.

Ahmad et al. [1] used a similar procedure to create a statistical atlas of both the shape and the BMD of the femur. When compared to Whitmarsh’s et al. [164; 165] approach described above, the following discrepancies are observed. The statistical atlas was trained using a larger dataset of 99 in-vivo QCT femoral scans of females, and there was a larger test set comprising 48 femurs (all obtained in-vivo from elderly women). Moreover, the latter dataset compromised real DXA scans instead of DRRs, obtained at $-21, 0, 20$ and 30 degrees relative to the AP view.

To create the statistical atlas, first they fitted a tetrahedral model to the segmented surface of each of the 99 QCT scans, which incorporated both density and shape information. This allowed them to create an average tetrahedral model femur. Then, deformation fields were applied to this average shape in order to make it fit to each subject’s segmented surface. Finally, PCA was performed on these deformation fields to determine the principal modes of variation. Eight PCs were retained without justification.

Reconstruction of each individual femur was performed using the same iterative technique as Whitmarsh et al. [164; 165]. In this study however, an additional step was performed after the convergence of the iterative optimization process. To further improve the estimation of aBMD, the AP DRR projection was compared to the AP DXA projection, and the BMD values of all voxels contributing to the projected pixels

2. 3D RECONSTRUCTIONS FROM PROJECTIONS

were linearly scaled so that the intensities of the two images matched exactly.

However, the most significant difference between the two approaches was the performance assessment protocol. In this work, performance was gauged based on the ability to predict correctly several geometric parameters, such as, but not limited to, the FNAL and the CSA across the narrow-neck (NN) and intertrochanteric regions. Furthermore, the extent to which the volumetric BMD was correctly predicted was used to evaluate the reconstruction accuracy (measured at the same regions of interest: NN and intertrochanteric — 10 mm thick slab slices).

All evaluation parameters were highly correlated when compared to their equivalents measured from the ground truth QCT scans, with the linear coefficients varying from $r = 0.81$ – 0.98 . Unfortunately, once more, the authors omitted to disclose the test and training population diversities, although they state that their test set was obtained from females aged 82 ± 2.4 years old, which might hint at a relatively low test set anatomical variance.

Thevenot et al.

An alternative to this technique was proposed by Thevenot et al. [152]. They associated eight geometric parameters from seven AP training radiographs to the surface morphologies of the femurs obtained from CT scans. To do so, the top of the femoral head was modelled as a half-sphere, the femoral neck was split into ten non-circular cross-sections, whose shapes were defined by measuring the radii every 45° degrees. Unfortunately, they do not specify any details on how they defined the skeleton for the trochanteric regions, nor on the method they used to associate the parameters with variations in shape. Moreover, since their objective was to build a finite element 3D model of the femur, their model also separated the trabecular from the cortical compartments, and set material properties for these two tissues and a transitional layer in-between them. However, a limiting factor of their approach is the assumption of a 1 mm minimum cortical thickness, which is an overestimate [156; 158].

Independent testing was performed by reconstructing 21 separate femurs, using measurements of the same eight geometric parameters as input. No details on the anatomical diversity of their test set is communicated. They report separate shape reconstruction errors for the femoral head ($1.03 \text{ mm} \pm 1.17 \text{ mm}$), neck ($1.27 \text{ mm} \pm 0.60 \text{ mm}$) and trochanteric region ($2.24 \text{ mm} \pm 1.33 \text{ mm}$). However, they do not mention the average variation of the test femurs from the mean shape, so these numbers alone cannot tell how well the algorithm performs.

2. 3D RECONSTRUCTIONS FROM PROJECTIONS

Hurvitz et al.

Hurvitz et al. [71] use a CT-like intensity atlas to perform personalised 3D reconstructions of the proximal femur from four DRRs (simulating fluoroscopic X-ray images and not DXAs). The model was a combination of an AAM, built using Procrustes Analysis followed by PCA, and an average template intensity image representing the mineral distribution within the bone.

Personalised reconstructions were performed using the following steps. The statistical shape model was fitted to each individual's X-rays by means of an affine transformation, followed by a deformation of the mean shape using the first K Principal Components (PCs), which accounted for 95% of the training population variance. Then the template intensity image was backwards-wrapped to the deformed model, and DRRs were generated from the atlas. Subsequently, a 2D/2D intensity-based deformable registration between these DRRs and the test X-rays allowed the re-estimation of the transformations required to best fit the two sets of images. This procedure was repeated until convergence.

To evaluate the algorithm, a Leave-One-Out (LOO) cross validation scheme with 17 femurs was used, for which no demographics are provided. To capture 95% of the variance of the training population 11 PCs were retained. The reconstruction error was 1.40 ± 0.55 mm. Once again, the diversity of the dataset was not specified, neither the error when fitting the mean statistical model.

Kurazume et al.

Kurazume et al. [88], constructed a Statistical Shape Model (SSM) of the proximal femur from 56 femoral CT scans using PCA. For personalised reconstruction, their search function was driven by a silhouette-based 2D/3D registration between the model and two X-ray images of each test specimen using distance maps constructed by the Level Set Method.

The efficacy of the algorithm was tested using both DRRs and in-vivo data, although both test sets were very small. The former comprised ten femurs, five of which were included in the construction of the model (although results are presented separately for the independent testing). The latter involved real in-vivo fluoroscopic X-rays from four patients. In both cases, two views per individual assisted in the reconstructions, although the projection angles of the final setup are not disclosed. Reconstructions were performed by retaining up to 10 PCs, and it was deduced that performance saturated when five were used. Shape accuracy was measured by comparing against CT segmentations of the test femurs. Errors for the simulated experiments were 0.81 ± 0.07 mm for the five

2. 3D RECONSTRUCTIONS FROM PROJECTIONS

specimens which were included in the model training dataset and 0.91 ± 0.15 mm for the five which were unseen. On the other hand, corresponding errors for the in-vivo experiments ranged from 0.8 ± 0.7 mm to 1.1 ± 1.0 mm. Notably, this was the only study which includes the blind estimator’s errors, i.e. the estimation efficacy using none of the PCs. For the simulated experiments, efficacy was 1.69 ± 0.54 mm, whereas for the in-vivo it was approximately 1.3 ± 0.9 – 1.0 mm. Therefore, it is apparent that the performance increase obtained using this reconstruction algorithm is relatively small. Finally, it would be interesting to confirm these error statistics using larger datasets. When doing so, the authors should disclose the demographics of the training and test sets (they neglected to do so in this work). In that way, readers will be able to evaluate whether this algorithm generalises well to diverse data, or whether tailored models are required for different populations.

2.6.3 Discussion of model-based approaches and limitations

Limitations and sources of error

The main reconstruction limitations and sources of error mentioned in the literature are the following:

- The use of Digital Reconstructed Radiographs (DRRs) instead of real DXA scans [91; 164; 165], as it is hard to acquire QCT scans with associated DXA scans. Although these resemble DXAs, discrepancies exist and further testing should be performed with real scans to assess the reconstruction accuracy, especially in-vivo.
- The large datasets required for the formation of statistical atlases to properly capture the various modes of variation, and the possible requirement of separate models for different cohort studies, since significant structural discrepancies (for example geometry/size [105; 114; 147], BMD [73; 76; 105; 114; 147; 153] and cortical thickness [73; 105; 114; 147; 153]) arise between people of different gender [27; 73; 139; 147], race [4; 105; 113; 114] and age [32; 73; 74; 107; 139; 147; 153]. The relatively small size of training and testing datasets is a limitation almost always mentioned by the authors of the studies reviewed in the previous sections.
- Anatomical model based approaches are inherently susceptible to bias, and care should be taken to avoid over-fitting, or constructing models from biased populations. Studies that involve the identification of potential defects or structural abnormalities in the femur are particularly susceptible, since these variations are by definition not captured well in the dominant modes of the models.

2. 3D RECONSTRUCTIONS FROM PROJECTIONS

- The practical limitation on the range of imaging angles in-vivo, as beyond about 60 degrees from the AP view (e.g. lateral view) the acetabulum bone, the pelvis or the contra-lateral femur may overlap with the femur in the DXA projection. Nevertheless, studies have claimed that this range of angles is adequately large for accurate femoral reconstructions [2; 84]. However, as seen in Section 2.3.3 where the theoretical limits to reconstruction resolution was discussed, these claims seem unlikely.

Overall findings

The objective of the studies discussed in the previous section was to reconstruct the femur in 3D from one or more 2D radiographs. They claim that doing so allows a) the pre-operative preparation of 3D anatomical models for image guided orthopaedic surgery, b) the accurate measurement of the geometry and the calculation of 3D structural measurements irrespective of proper patient scanning positioning, and c) the improvement of hip fracture risk assessment and the better diagnosis of osteoporosis.

Although CT/MRI scanners have long been capable of achieving all the above, developing methods that operate on 2D scans has many advantages: CT/MRI scanners are expensive and not available on many medical centers, imaging/labor costs and radiation exposure (compared to CT, not MRI) are greatly reduced, and require no change to the existing infrastructure as they make use of the imaging modality routinely used in diagnostic clinical practice.

The findings of these studies suggest that reconstructions of the proximal femur from four or less 2D scans are possible. Reported shape reconstruction errors range from 1.0 mm to 1.6 mm RMS, but none of the studies comments on whether this is acceptable performance for any particular task. Finally, Ahmad et. al. [1; 2] conclude that it is possible to obtain accurate 3D structural measurements from DXA scans, and Le Bras [93; 94] et. al. managed to improve failure load prediction by combining aBMD measurements with 3D geometric parameters measured from reconstructed femurs.

Training and test set anatomical diversity and model bias

One question which is not addressed by any of the aforementioned researchers is how well the training population fits the test set. On one hand, models might be particularly biased if the training and test cohorts are anatomically very similar. Thus, generalisation to unseen data might lead to significantly different results. On the other hand, if the training dataset does not resemble the test femurs, building tailored models might lead to better estimation performance. Proper interpretation of reconstruction results requires

2. 3D RECONSTRUCTIONS FROM PROJECTIONS

the authors to assess whether they are using a representative training population.

Furthermore, it is of particular importance to also reveal the diversity observed in both the test sets and the training data. Some of the examined studies disclose the *demographics* of the test and training data, which, although indicative, does not provide the complete picture. The reader should be given the chance to assess whether the models were built using a diverse enough population to qualify them suitable for a significant target audience. Furthermore, they should be provided with the test set anatomical variance, to determine the extent to which the proposed reconstruction algorithm generalises. Finally, it is worrying that some studies do not even reveal whether they are using separate femurs for training the models and evaluating performance.

Comparison with the naive blind estimator

Following one of the observations of the previous section, it is particularly troubling that researchers fail to realise that they should establish a comparative baseline, instead of only reporting absolute errors. Studies involving very limited test sets, especially when these are relatively homogeneous (as is the case with most of these studies), should also report the errors obtained after best fitting the mean template/statistical model to each individual — or at least the variance of the test set. Comparing against a naive estimator, such as a mean model, is the simplest approach to providing a reference against which the efficacy of each algorithm can be compared. If the testing population variance is low, authors may not realise that the errors they measure might be marginally improving, or even deteriorating, upon the mean estimator.

Surprisingly, all of the authors who examined template based methods failed to establish a comparative reference, with the exception of Gamage et al. who at least indicated the intra test set variance. As far as the statistical methods are concerned, only Kurazume et al. provided the relevant statistics.

Reconstructing the femoral cortex

As far as the femoral cortex is concerned, to the best of our knowledge there is no prior work involving the 3D reconstruction of cortical thickness from a single or multiple radiographic projections. The most relevant work identified comes from Beck et al. [6; 7; 8]. Their HSATM algorithm is able to produce estimates of the mean cortical thickness around three cross sections at the shaft, intertrochanteric and neck regions, using a single DXA image and by modelling the bone as circular annuli with the inner space filled with trabecular bone. With the introduction of many C-arm equipped DXA scanners into the market (such as the Hologic Discovery QDR Series, Hologic Inc, Bedford, MA, USA),

2. 3D RECONSTRUCTIONS FROM PROJECTIONS

that allow acquisition of scans from multiple imaging angles, novel avenues of research are opened up such as the one proposed in this thesis.

Chapter 3

Template-based cortical thickness estimation

3.1 Introduction

This chapter examines a template based method for measuring cortical thickness across the proximal femur from multi-view DXA scans (the **Template Method**). To the best of the author’s knowledge it is the first algorithm of its kind. Section 2.6.1 discusses the application of various template based methods to 3D femoral reconstructions, but they are all limited to shape and density. The problem discussed herein is harder, as the additional step of measuring cortical thickness should be accomplished along with the shape reconstruction.

Since cortical thickness cannot be measured from multi-view DXA scans yet, we opted to evaluate the Template Method against the current state-of-the art algorithm, which operates on MDCT data (the **CT Method**) and whose principles of operation are explained in Section 3.5.1. Hence, the ideal experimental dataset for the evaluation of the Template Method would consist of femurs scanned using both multi-view DXA and QCT. Unfortunately, during the course of this study such a dataset was not available. Therefore, we resorted to DRRs from MDCT as a surrogate for DXA scans. DRRs can be easily generated at any orientation, in contrast with DXAs which require careful patient positioning. On the other hand, as explained in Section 1.3.1, DXA scans only depict mineral content by eliminating any surrounding soft tissue. This can be simulated in DRRs, but it is not a trivial task. Appendix B explains in detail how the DRRs were constructed, and how the background was removed. Finally, the amount of blur present in DXAs and DRRs depends on the Point Spread Function (PSF) of the imaging device used. However, converting the in-plane and out-of-plane blur present in CT scans to a

3. TEMPLATE-BASED CORTICAL THICKNESS ESTIMATION

DRR blur value is non-trivial and depends on the DRR orientation.

This chapter is structured as follows. First, a template particularly suited for cortical thickness estimation is defined. Section 3.3 explains the 2D/3D correspondence protocol between the template and the DRRs. Then, an algorithm that iteratively deforms the template to best fit the DRRs is presented in Section 3.4. Section 3.5 describes the foundations of the evaluation protocol. It discusses how the “CT method” leads to gold and bronze standard thickness estimates, using High Resolution peripheral QCT (HRpQCT) and clinical resolution QCT femoral scans respectively. Then, it defines the registration protocol between DXA and CT thickness measurements. The presentation of experimental results is split into Sections 3.6 and 3.7: in the former, all femurs were scanned using both high and clinical resolution QCT, whereas the latter involves only femurs scanned at clinical resolution. Finally, the last section discusses the findings and outlines some conclusions.

3.2 Template Definition

As explained in Section 2.6.1, the template should be designed according to the reconstruction task. For the problem discussed herein, the ideal template would model the endocortical and periosteal surfaces to the required resolution, along with all the parameters which affect the appearance of the cortex in a DXA scan. Moreover, it would allow the flexible manipulation of the cortex boundaries in a low dimensionality framework, and ease the reconstruction process by minimising the unknown variables.

As always, the model complexity is limited by the amount of observable information. Allowing too many degrees of freedom can render the model underconstrained, which eventually may lead to over-fitting. On the contrary, an overconstrained template may result in suboptimal performance, or even misleading reconstructions. As the number of DXA views is reduced, and hence the available 3D information, we decide to trade (or simplify) potential variables for more prior assumptions, to allow for the best possible representation of the observed DRR data.

To this end, the following design decisions were made. Each proximal femur was modelled using 20 cross-sections, which in turn are defined by two B-spline contours representing the periosteal and endocortical surfaces respectively. Their shape and separation (i.e. cortical thickness) are initialised by sampling a canonical shape and thickness model across twenty planes (see Figure 3.1). In addition, three independent density values per contour were used for the cortical, trabecular and background tissues, all of which are modelled as uniform-density, homogeneous materials. Finally, it is assumed that the PSF of the imaging device, which describes the resolution charac-

3. TEMPLATE-BASED CORTICAL THICKNESS ESTIMATION

teristics, corresponds to that of blurring with a Gaussian filter with standard deviation σ .

Figure 3.1: **Template definition:** The canonical shape and thickness model is sampled across twenty planes to obtain the cross-sections, $c_{1..20}$. Each cross-section is modelled by two radial B-splines, which represent the endocortical and periosteal surfaces respectively, and three density values: background tissue (d_b), cortical (d_c) and trabecular (d_t). The center of the B-splines is defined at the centroid of each cross section.

3.2.1 B-spline cross-sectional modelling

The benefits of parametrising the endocortical and periosteal surfaces as B-spline contours are threefold:

1. As discussed shortly in Section 3.5.2, this allows for a regular, uniformly distributed thickness sampling around the cross-sections, and conveniently provides a straightforward thickness direction definition.
2. Provides a low-dimensionality framework with variable flexibility, as the number of control points can be adjusted. The positions of the control points were constrained to lie on equiangular lines radiating from a central point, as shown in Figure 3.1b, to even further reduce the degrees of freedom.
3. Finally, it significantly eases the shape regularisation of the contours, by inherently ensuring smooth curves.

3. TEMPLATE-BASED CORTICAL THICKNESS ESTIMATION

In detail, each B-spline is defined by n_{cp} periosteal (p^{out}) and endocortical (p^{in}) control points, which define an equal number of spline segments (successive segments share three of their control points to ensure second order parametric and geometric continuity). Their locations are calculated using (see also Figure 3.1b)

$$p_i^{out} = \begin{bmatrix} r_i \cos(\theta_i) \\ r_i \sin(\theta_i) \end{bmatrix}, \quad p_i^{in} = \begin{bmatrix} (r_i - t_i) \cos(\theta_i) \\ (r_i - t_i) \sin(\theta_i) \end{bmatrix} \quad (3.1)$$

$$\begin{aligned} \text{where,} \quad \theta_i &= \frac{2\pi}{n_{cp}}(i-1), \\ i &= 1..n_{cp} \end{aligned}$$

and r_i is the radius relative to the central point and t_i the radial spacing between each of the endocortical and periosteal control points.

Each spline is sampled at 100 locations, to produce an equal number of thickness measurements. Their location is calculated using the B-spline definition:

$$s_i(t) = \frac{1}{6} \begin{bmatrix} t^3 & t^2 & t & t \end{bmatrix} \begin{bmatrix} -1 & 3 & -3 & 1 \\ 3 & -6 & 3 & 0 \\ -3 & 0 & 3 & 0 \\ 1 & 4 & 1 & 0 \end{bmatrix} \begin{bmatrix} p_i \\ p_{i \oplus 1} \\ p_{i \oplus 2} \\ p_{i \oplus 3} \end{bmatrix} \quad (3.2)$$

3.3 2D/3D Correspondence

3.3.1 Initialisation

Figure 3.2: Registration of the template to a particular femoral specimen. The silhouette of the model is shown in cyan, and the points from the thresholding process are shown in yellow for three different DRR poses of the same bone, at angles 9° , 45° and 90° .

3. TEMPLATE-BASED CORTICAL THICKNESS ESTIMATION

We initialized the position of the template using a semi-automatic procedure, assisted by an iterative closest point algorithm. Specifically, after a rough manual alignment of the template’s silhouette to the projection of the femur on the DRRs, the Levenberg-Marquardt optimization method [111] was used to calculate the affine transformation that minimises the distance between the silhouettes of the model and the subject, as shown in Figure 3.2. The silhouette of the subject was obtained using a simple threshold approach, whereas that of the template was defined using ray-casting.

The mean and root mean squared registration errors¹ were -0.04 mm and 1.85 mm respectively. The initial radial separation between the periosteal and endocortical B-splines was set to the mean cortical thickness of a large number of femurs (calculated using the “gold-standard” algorithm of Section 3.5.1 and MDCT data).

3.3.2 Attenuation Projection Profiles (APP)

Section 2.2 explains how to reduce the dimensionality of a 3D reconstruction problem to 2D by examining the APPs of slices through the data. This observation inspired the modelling of the femur as a collection of B-spline contours, since an appropriately positioned set of APPs on multi-view scans can be used to infer the underlying femoral cross-sectional structure, using the principles of filtered backprojection.

Such analysis is only meaningful if all cross-sections are perpendicular to all DRRs, as shown in Figure 3.3. Otherwise, the projection of a cross-section onto an arbitrary DRR is not guaranteed to be contained within a single APP line, but rather spread across a 2D image patch. Hence, the normals of the cross-sectional planes should coincide with the DRR axis of rotation (which can be well approximated by the femoral shaft axis).

Moreover, their positioning can extend only up to the most distal point of the femoral head, as illustrated in Figure 3.3. This is because it is impossible to correctly interpret any APPs above this point without modelling the acetabulum and the pelvis: their projections encroach on top of the femur in most DRRs, leaving very few unobstructed views suitable for analysis.

3.3.3 Synthetic Attenuation Profiles (SAP) and Sampling

Similarly, by simulating X-ray attenuation through each cross-section in each view, as shown in Figure 3.4, the template model provides an alternative set of Synthetic

¹These errors were calculated considering only the shaft and trochanteric regions (i.e. only the regions where cortical thickness was to be estimated), using the 35 femurs of the first experimental dataset (see Section 3.5.5).

3. TEMPLATE-BASED CORTICAL THICKNESS ESTIMATION

Figure 3.3: 2D/3D correspondence between the cross-sections and their APPs on each DRR. The APPs are obtained by interpolating the original DRR pixel data using the Mitchell-Netravali cubic spline and re-sampling at 200 points. The region above the most distal point of the femoral head (where the APPs are shaded in red) cannot be evaluated without taking into account the acetabulum and the pelvis, as their projections overlap that of the femur most of the time.

3. TEMPLATE-BASED CORTICAL THICKNESS ESTIMATION

Figure 3.4: X-ray casting geometry and sampling of SAPs. When multiple views are examined simultaneously, the direction of the X-rays (dotted lines) uniquely identifies the position of the contours in 3D space. The areas of the cortical and trabecular compartments are calculated using the intersection of the sampling lines, x_i and x_{i+1} , with the periosteal (p_i) and endocortical (e_i) contours. The area of the background is found using the sampling width, $x_{i+1} - x_i$, and the DRR compounding thickness, t_1 . Finally, the sample s_i is calculated by multiplying these areas with their respective densities (d_c, d_t and d_b), summing them together, and blurring the result using the values of the neighbouring samples (notice how the APPs' non-zero values extend beyond the actual projection of the cross-section).

3. TEMPLATE-BASED CORTICAL THICKNESS ESTIMATION

Attenuation Profiles (SAPs). In an optimal model of the femur, the SAPs and the APPs would be identical, although in practice this is unlikely as illustrated in the example of Figure 3.5.

Blurring

Proper analysis of DXA data requires the modelling of imaging blur. This is particularly important when measuring thin structures, such as the cortex, whose appearance is heavily influenced under its presence. Herein, this is achieved by convolving all discrete SAPs with a sampled, truncated Gaussian kernel:

$$s'_i(\sigma^2) = \sum_{n=-M}^M s_{i-n} G(n, \sigma^2) \quad (3.3)$$

$$\begin{aligned} \text{where} \quad G(n, \sigma^2) &= \frac{1}{\Sigma} e^{-\frac{n^2}{2\sigma^2}} \\ \text{and} \quad M &= \lfloor 3.7\sigma \rfloor \end{aligned} \quad (3.4)$$

where s_i is a SAP sample, Σ is the normalisation constant and σ the standard deviation. The size, M , of the filter is chosen so that the tails are cut off when the values are less than 1/1000 of the peak value for efficiency purposes.

3.4 Template Deformation

3.4.1 Optimisation parameters and cost function

After initialising the position of the template relative to the DRRs, the Levenberg-Marquardt optimization method [111] was used to minimize the squared intensity difference between the SAPs and the APPs in all available DRRs simultaneously. This was accomplished by optimizing the following parameters for every cross-section: $\{r_{1..n_{cp}}, t_{1..n_{cp}}, d_c, d_t, \sigma\}$. The first two represent the radii and separation of the periosteal and endocortical spline control points respectively, d_c and d_t are the densities of the cortical and trabecular bone respectively, and σ is the variance of the assumed Gaussian imaging blur. The background density needs no optimisation, since all background voxels are replaced with the mean background value, as explained in Appendix B.

3. TEMPLATE-BASED CORTICAL THICKNESS ESTIMATION

Figure 3.5: The top graph plots the APP of the cross-section whose position is shown in green. The remaining plots compare this APP with the SAP of various model cross-sections. The last plot represents the SAP of the cross section whose thickness is obtained from the CT method, i.e. the ground truth estimate (see Section 3.5.1). Note that the error is non-zero, since the SAP is calculated assuming constant background, cortical and trabecular densities, and because the segmentation’s position does not necessarily coincide with the periosteal surface the CT method deduces (the segmentation of the bones, although manually post-refined, was primarily based on thresholding, which tends to overestimate the volume of the femur when the blur conceals the true boundary of the cortex). In addition, some of the error can be attributed to the imprecision of the CT method.

3.4.2 Regularisation

Density and blur

The optimization is ill-posed, in that more than one set of parameters can explain the APPs to the same degree. This is obviously the case with a small number of views — 3D anatomy cannot be uniquely determined from one or two planar projections — but even with many views there is a trade-off between cortical density and thickness in the presence of imaging blur. It was deduced that best results can be obtained (the relevant experimental results are discussed in Section 3.6.3) if the problem is regularized by assuming a constant cortical density throughout the proximal femur, which is also the approach taken in Treece et al. [156]. This fixed density was estimated in regions low down the femoral shaft, where the cortex is sufficiently thick that its true density is unambiguous in the DRRs despite the imaging blur. Hence, the algorithm first optimizes a number of cross-sections low down the femoral shaft, averages the resulting d_c solutions, and then optimizes the whole femur assuming this fixed value of d_c . In reality, the problem could be even further regularised by feeding the optimizer with the true value of the Gaussian imaging blur, which would probably be known since it depends on the scanner. However, this was not attempted herein, since converting the in-plane and out-of-plane blur present in the CT scans to a DXA blur value is non-trivial and dependent on the DRR orientation.

Shape and thickness

For a clinically viable protocol, where there are a small number of views covering a narrow range of angles, further regularization may be necessary. Otherwise, the optimizer may over-fit the APs to SAPs by converging to misshaped contours, like the ones depicted in the left column of Figure 3.6. To prevent this, a penalty function may be introduced to prevent the control points deviating too far from the initialization. However, satisfactory results were obtained when constraining the periosteal control points to move by less than 1.5 mm, and confining the thickness to vary by less than 1.5 mm (the relevant experimental results are discussed in Section 3.6.4). This value was empirically deduced as a good compromise between flexibility and regularisation, as higher values frequently led to deteriorated cortical thickness estimation, primarily in the four-view experiments.

Figure 3.6: Optimization examples, showing the initialization (green) and outcome (red), for the lesser trochanter (top) and the greater trochanter (bottom). (a) four DRR views without shape and thickness regularization, (b) four views with shape and thickness regularization, (c) 20 views without shape and thickness regularization.

3.5 Evaluation protocol

3.5.1 “CT method” and cortical thickness gold standard

Figure 3.7 explains how cortical thickness can be measured across the proximal femur from QCT data, using the the medical imaging software **Stradwin** [157]. The result is expressed as a thickness colour map: each vertex of the segmented surface is assigned a thickness estimate which is defined along the vertex normal. The algorithm used to identify the cortical boundaries depends on the resolution (the following two algorithms are explained in detail in Treece et al. [156; 158]).

Gold standard from HRpQCT scans

For the high resolution datasets (Section 3.5.5), a **full-width half-maximum** [133; 134; 156] technique was employed to extract the thickness estimates. It is known that, if the imaging PSF is small compared to the thinnest cortices, this is an unbiased and accurate estimator [133].

3. TEMPLATE-BASED CORTICAL THICKNESS ESTIMATION

Figure 3.7: Stradwin user interface for cortical thickness estimation from MDCT data. The femoral segmentation, seen as a green contour on the top left, defines the position of many lines perpendicular to the surface (an example is shown in cyan). Then, an optimiser estimates the thickness model's parameters that best fit the CT data samples along these lines. The bottom panel compares the actual data with the optimised model. The resulting thickness estimations can be collectively depicted using a colour map, as shown in the top right panel.

3. TEMPLATE-BASED CORTICAL THICKNESS ESTIMATION

Bronze standard from QCT scans

On the other hand, simple techniques, such as thresholding variants [17; 63] or full-width half-maximum [133; 134], become increasingly biased and inaccurate as the imaging resolution is reduced. When clinical resolution scans are examined, thickness estimation errors for sub-millimetre cortices can exceed 100% — the general consensus suggests that estimation is unreliable below 2.5 mm [43; 64; 156].

The current state of the art was recently proposed by Treece et al. [155; 156; 158], and involves fitting a restrictive thickness model to each point estimate. As such, cortical thickness over the proximal femur can be measured accurately down to 0.3 mm using clinical MDCT (accuracy 0.12 ± 0.39 mm for cortices in the range 1–3 mm), despite the > 1 mm PSF of this technology. To do so, the LevenbergMarquardt optimisation algorithm [111] is used to estimate the model parameters that best fit the data. Namely, these are the surrounding tissue density, the trabecular density, the cortical density, the cortical thickness and the imaging blur.

Deblurring thin laminar structures constitutes an ill-posed problem. This is because blurring a thin, high intensity step function might produce results indistinguishable from a blurred, thick, low intensity one. Therefore, the problem ought to be regularised using some prior information. The first iteration of Treece’s algorithm achieved that by fixing the cortical density, and not the blur level, to a constant value, as this approach produced the best results. The fixed density is obtained from regions where the cortex is sufficiently thick, and consequently the true density is unambiguous despite the imaging blur (typically from regions distal to the lesser trochanter). However, this assumption inevitably resulted in slightly wrong thickness estimates in regions where the local cortical density differed from the global assumed one. The second iteration of Treece’s algorithm improves upon this problem by adjusting the cortical density per point estimate using a prior estimate of the imaging blur [155].

3.5.2 Thickness orientation

The thickness of a thin laminar shell can be uniquely defined only if the orientation of measurement is also specified. Figure 3.8 presents six different valid ways of measuring thickness, all of which would produce different estimates (assuming a rotationally asymmetric thickness distribution, such as the one depicted in the figure). The expected amount of discrepancy between them is illustrated in Figure 3.9, as obtained from 10000 randomly generated cross-sections, each sampled at 100 points.

Each method is associated with its own advantages and disadvantages, hence the

3. TEMPLATE-BASED CORTICAL THICKNESS ESTIMATION

Figure 3.8: Comparison of six different thickness definitions, based on the direction of measurement. The shape and thickness variation is intentionally exaggerated, so that extreme cases can be visualised. In reality, femoral cortical cross-sections are smoother, and the discrepancies between algorithms are expected to be smaller.

Figure 3.9: Comparison of different thickness definitions on simulated data. The algorithm that generates thickness estimates between successive B-spline samples (Figure 3.8a) was chosen as the baseline ($y = 0$), since it produces the most evenly distributed results. This graph plots the mean thickness difference of each algorithm compared to the baseline, as a function of thickness.

3. TEMPLATE-BASED CORTICAL THICKNESS ESTIMATION

choice of protocol depends on the task at hand. For example, if the objective is to identify the minimum thickness at every possible location, then either of the two minimum distance algorithms (Figures 3.8c and 3.8d) could be used. On the other hand, if a more regular sampling of the thickness is required, matching B-spline samples or using lines radiating from the centre of gravity (Figures 3.8a and 3.8b) are expected to produce better results.

The approach that defines the thickness orientation based on surface normals (Figures 3.8e and 3.8f) is particularly useful when the exact position of both laminar boundaries is not known. As explained in the previous section, this approach can be used to define the direction of cortical thickness measurements from MDCT data, starting from an approximate segmentation of the femur.

3.5.3 Registration between CT and DXA estimates and thickness direction

The CT thickness estimates are defined in QCT space along the vertex normals of the segmented surface. On the other hand, DXA estimates are obtained by sampling the B-spline cross-sections in DRR space (as defined in Section 3.3.2). Conversion between DRR and CT coordinates is straightforward, since the position (and orientation) of the AP lines on multiple DRRs define the cross-sections uniquely in space. However, even when both measurement sets are defined on the same coordinate system, registration is not trivial. A nearest-neighbour based algorithm would not account for a) thickness direction normalisation — which may significantly skew the results as seen in the previous section — and b) spatial normalisation and interpolation.

The registration approach taken here tackles both issues by expressing the CT thickness estimates along the DXA measurement directions. In detail, this is achieved in the following steps:

1. The CT segmentation represents the periosteal femoral surface. The endocortical surface can be simply obtained by translating each vertex point along the surface normal, using the thickness values obtained from the CT method.
2. The planes of the template's B-spline contours were expressed in CT spatial coordinates, according to the positions of the AP lines. Then, an algorithm which calculates the intersection contour between a surface mesh and a plane was used to extract the endocortical and periosteal contours from CT. The location of the intersection planes was such that the DXA and CT thickness estimates were coplanar (see Figure 3.10).

3. TEMPLATE-BASED CORTICAL THICKNESS ESTIMATION

3. The periosteal position of each DXA thickness estimate was used to calculate its angle relative to the cross-section's centre (i.e. the intersection point of the radial lines which define the positions of the B-spline control points) and the 0° AP view.
4. Finally, the periosteal contours obtained in step 2 were sampled at the same angles¹, and the CT gold/bronze standard was obtained by measuring thickness **along the direction of the DXA estimates**.

Section 3.6.1 confirms that this choice outperforms a nearest neighbour based registration approach, even when the thickness direction is normalised to a common vector. Finally, measuring along the DXA thickness direction was preferred, as otherwise cross-sections would have to be interpolated (to find the cortex boundaries along the CT thickness direction), imposing further assumptions on the model.

Figure 3.10: Registration between DXA and CT cortical thickness estimates. The endocortical and periosteal contours (red) are calculated from the intersection of the CT segmentation and raw thickness (grey and red surface meshes respectively) with the plane of the template's cross-section. Then, the periosteal contour is sampled at angles θ_i , according to the location of the DXA thickness estimates (samples on the blue contours). Finally, the gold/bronze standard CT estimates are obtained along each of the DXA thickness directions.

¹Although unlikely, the radial sampling of the CT intersection contours might result in multiple matches per angle, in contrast with the template's contours which by definition will always have one. If this is the case, the nearest neighbour from the matching CT estimates is used.

3. TEMPLATE-BASED CORTICAL THICKNESS ESTIMATION

3.5.4 Registration to a canonical femoral morphology

Figure 3.11: The graphic user interface of wxRegSurf, which can align two different femoral morphologies using various algorithms. Here, the red surface is registered to the green using (a) a similarity and (b) a similarity followed by a free form B-spline deformation.

One of the objectives of this study is to identify the femoral regions where cortical thickness estimation is best, and determine whether they coincide with the fracture prevalent areas discussed in Section 1.2. Error colour maps are particularly suited to this task, as they depict the estimation efficacy as a function of location. However, the calculation of *average* thickness (and error) maps of a set of femurs is only possible if they are all first registered to a single canonical morphology. As such, proper averaging per *anatomical location* can be performed. The medical imaging software **wxRegSurf** (Figure 3.11), developed at Cambridge University Engineering Department, was used to accomplish this task. Following the author’s recommendation, we registered each femur to the canonical surface using a similarity transformation followed by a Locally Affine Deformation (LAD). Then, each vertex of the canonical model was assigned the thickness value of its nearest neighbour.

3.5.5 Experimental Data and Setup

The experiments used to validate and assess the performance of the template method are divided into the following two sections.

3. TEMPLATE-BASED CORTICAL THICKNESS ESTIMATION

HRpQCT & QCT experiments (BERN)

First, in Section 3.6, we examine 18 left and 17 right cadaveric proximal femurs, all from separate individuals (18 females and 17 males). They were obtained from a study ethically approved by the Medical University of Vienna, and their age ranged from 59 to 96 years old (mean was 77 years old). Each femur was prepared by stripping any soft tissue, submerging in a saline solution and scanned using both HRpQCT and QCT, after placing them in vacuum to remove any air bubbles. The QCT data was scanned using a Brilliance64 scanner at 120 kV, with voxel size $0.33 \times 0.33 \times 1.0 \text{ mm}^3$, converted from Hounsfield Units to density using a BDC calibration phantom. HRpQCT data was obtained using an XTremeCT scanner, with voxel size $0.082 \times 0.082 \times 0.082 \text{ mm}^3$, converted to density by the manufacturer-provided phantom. More details of this data can be found in [34; 35]. Finally, each QCT scan was used to generate a set of multi-view DRRs for each femur, using the algorithm explained in Appendix B.

For this dataset, we set as our gold standard the cortical thickness estimates obtained using the Full Width Half Maximum (FWHM) algorithm (Section 3.5.1) from the HRpQCT scans, and we arrive at a final design by assessing the performance under different template configurations. Then, we also compare the DXA thickness estimates against the bronze-standard, obtained using the CT “model-based” algorithm (Section 3.5.1) from the corresponding clinical resolution CT data. The primary objective of this set of experiments is to quantify how much of the DXA method error can be attributed to the “model-based” method uncertainty, before evaluating (in Section 3.7) the performance on the remaining 120 femurs, whose description follows.

QCT experiments (FEMCO)

Subsequently, in Section 3.7, we evaluate the performance against a larger set of 120 proximal femurs scanned only at clinical resolution. These were obtained from the FEMCO study [129] which started recruiting participants in Cambridge in 2007. However, several other UK centres contributed after the study was adopted onto the UK National Institute for Health Research (NIHR) portfolio. Females with healthy (controls) and fractured (cases) femurs were scanned using a Siemens Somatom Sensation 16, 64 or GE Lightspeed 64 scanner at 120 kV. The QCT data was converted to Hounsfield units using a Mindways 5-compartment solid phantom or using ClinicQCT asynchronous calibration (if the scanner was already calibrated by a phantom). The resolution of the Siemens scanner was $0.58 \times 0.58 \times 1 \text{ mm}^3$, whereas that of the GE was $0.58 \times 0.58 \times 1.25 \text{ mm}^3$. Detailed demographics for this dataset are tabulated in the next chapter where they are more relevant to the experiments.

3. TEMPLATE-BASED CORTICAL THICKNESS ESTIMATION

For these experiments, the bronze-standard was obtained using the “model-based” CT method described in Section 3.5.1. They led to the comparison of 240,000 point cortical thickness estimates, thus improving the statistics of the results which can be interpreted as error probability distributions. Furthermore, they verify that the algorithm performs similarly on datasets obtained *in-vivo* using different scanners, and allow the direct comparison of the Template Method with the Statistical Method presented in the next chapter.

Furthermore, for both experimental datasets described above, results are categorised according to the number of DXA views used. First, to estimate the limiting performance of the DXA method, we decided to use 20 DRRs, at angles of 0° to 171° relative to the standard DXA view, as a good compromise between data sparseness and computational load. For these “**validation experiments**”, 20 control points per spline contour were used to take advantage of the rich data. We then restricted the viewing angle range to -20° to 40° , mimicking the capabilities of commercial C-arm DXA scanners¹ [1]. For these “**clinical experiments**”, we used 4 DRR views at -20° , 0° , 20° and 40° , and it was experimentally deduced that more than 12 spline control points per contour led to deteriorated results, most probably due to over-fitting to the sparse, noisy information. This corresponds to a 30° angular separation between them.

We calculated the mean, standard deviation and root mean square (RMS) of the error $t_{\text{DXA}} - t_{\text{CT}}$, where t is the estimated cortical thickness. As an indication of the fit between APPs and SAPs, we also report the Residual Optimization Error (ROE), which is expressed as a percentage of the Residual Initialisation Error (RIE). The RIE and the ROE are calculated by estimating the area between the APPs and the SAPs (shaded regions in Figure 3.5). Moreover, various colour maps of error statistics are presented, which conveniently depict the performance of the algorithm in different parts of the femur. These were produced by first registering all femurs to a canonical morphology, as explained in Section 3.5.4, and then analysing the results per vertex location.

¹As mentioned in Section 2.6.3, typically, a maximum range of 60° is possible before the acetabulum or the contra-lateral femur obscures the pertinent femur’s trochanteric region. Unobstructed views of the proximal part of the femoral neck and the femoral head are even harder to obtain in typical DXA poses (see Figure 3.3), which is why they are not studied here.

3.6 BERN results (HRpQCT/QCT resolution)

The following sections present the outcome of many different experimental configurations, together with a discussion of their contribution to the final algorithm design and regularisation. For this purpose, only the femurs scanned in both high and low resolution CT are examined, so that the DXA estimates can be primarily compared against the CT gold-standard measurements.

We first examine the limits of performance by approximating the CT solution with the template’s B-splines. Then, we investigate the error of the initialisation, which is equivalent to that of the blind estimator. The following two sections outline our findings regarding different regularisation approaches, by presenting the effect of progressively adjusting different parameters. Finally, the amount of error that can be attributed to the uncertainty of the “model-based” CT bronze standard is investigated, along with the presentation of the final validation and clinical experimental results. All of this section’s error statistics are collated in Table 3.1.

3.6.1 Spline approximation

Validating the registration between CT and DXA cortical thickness estimates can be achieved by optimising the position of the template’s control points to best approximate the “CT solution”. In that case, the error observed should be minimal, and proportional to the number of control points used per spline; the more points used, the better the approximation. Most importantly, doing so allows us to calculate the limits of performance of the “validation experiments” (20 views & 20 control points) and the “clinical experiments” (4 views & 12 control points).

Figure 3.12 plots the error statistics across the whole range of thicknesses observed in the data, the distribution of which can be seen in the histogram of Figure 3.12c.

Table 3.1: Summary of error statistics for the BERN experiments (Sections 3.6.1-3.6.5). The top section presents the errors of the CT solution spline approximation when using different numbers of control points for the two CT methods. The template initialisation/blind estimator’s errors follow, before presenting the outcome of constraining the imaging blur variance, the cortical density, and the maximum deviation from the blind estimator. At the bottom, the final design of the DXA template method is compared to the FWHM (gold-standard) and “model-based” (bronze-standard) CT thickness estimates.

		CT method	mean error (mm)	std. deviation (mm)	RMS error (mm)	percentage error (%)	ROE (%)
CT solution approximation							
12 control point spline apr.		FWHM	-0.03	± 0.41	0.41	11.2	77.0
20 control point spline apr.		FWHM	-0.02	± 0.26	0.26	6.0	74.0
20 control point spline apr.		model	-0.02	± 0.28	0.28	6.9	77.1
template initialisation (blind estimator)							
2D/3D registration							
initialisation	affine	FWHM	-0.15	± 0.89	0.91	28.1	100.0*
initialisation	B-spline	FWHM	-0.13	± 0.88	0.89	26.6	71.1
density/blur regularisation							
		constrain					
20 views	variance of blur	FWHM	0.44	± 1.05	1.14	39.0	41.3
20 views	cortical density	FWHM	0.13	± 0.60	0.62	21.1	43.1
4 views	cortical density	FWHM	0.16	± 1.10	1.11	41.6	27.3
shape/thick. regularisation (final design)							
		constrain					
20 views	shape & th. 1.5 mm	FWHM	0.03	± 0.60	0.60	21.6	48.1
20 views	shape & th. 1.5 mm	model	0.00	± 0.71	0.71	28.7	52.7
4 views	shape & th. 1.5 mm	FWHM	-0.10	± 0.72	0.73	24.9	42.9
4 views	shape & th. 1.5 mm	model	-0.16	± 0.80	0.82	29.9	44.0

*baseline ROE

3. TEMPLATE-BASED CORTICAL THICKNESS ESTIMATION

Figure 3.12: Errors when approximating the CT solution with the template's B-splines as a function of thickness, using 12 (a) and 20 (b) control points per endocortical/periosteal contour. (c) presents the cortical thickness distribution across all 35 femurs for all examined locations.

3. TEMPLATE-BASED CORTICAL THICKNESS ESTIMATION

Smoothing of CT estimates

Different surface morphologies are used to define the DXA, gold and bronze standard thickness estimates. The DXA method uses the B-spline template, whereas the CT methods use slightly different — albeit representing the same bone — semi-manual registrations from the HRpQCT and QCT data respectively. In addition, the spatial localisation accuracy varies between them. Section 3.5.3 explains how this spatial registration and normalisation problem was solved when registering DXA to CT estimates.

However, to promote a fair comparison against the two different CT methods, we also opted to match the FWHM and model-based numbers of point measurements. After mapping both to a common segmentation using a rigid deformation with separate scaling factors for the three coordinate axis, a smoothing operator was used to calculate the thickness at 11,000 vertices per femur. To do so, every estimate was replaced by a weighted average of its neighbours, where the weights were inversely proportional to the ROE. This led to a consistent localisation accuracy across the femur, whilst preserving the bronze and gold standard resolutions through the cortex. In an effort to eliminate only outliers and noisy measurements, the amount of smoothing was adjusted so that the spline approximation error was almost identical in both cases (see Table 3.1), and approximately equal to the estimation accuracy of the gold-standard, which for this dataset is 0.3 mm [155].

3.6.2 Initialisation (blind estimator)

One of the primary motivational reasons for carrying out the present research is that currently no method exists capable of producing detailed cortical thickness maps from DXA scans. Thus, if a femur is imaged using only this technology — i.e. the current clinical routine practice for bone quality screening — one would have to resort to a blind guess. Statistically speaking, the guess with the least expected error for a population is the average cortical thickness of all of its femurs. Therefore, improving upon the efficacy of the blind estimator, which from now on will form our comparative baseline, is the least that needs to be done. It is worth noting that the blind estimator’s error is equivalent to that of the template’s initialisation, since by definition we start from the average shape and thickness of a large number of femurs.

Two ways of assessing the blind estimator’s error were considered, both of which resulted in almost identical error statistics. The first resorted to the semi-manual, silhouette-based, affine registration (Section 3.3.1) of the template to the 20 DRRs of the validation experiments. On the other hand, the second allowed a non-rigid deformation of the template shape, but *not* the thickness, so that the error between the APs and

3. TEMPLATE-BASED CORTICAL THICKNESS ESTIMATION

the SAPs was minimised. This was achieved by optimising the position of the periosteal control points, and simultaneously adjusting the endocortical control points to preserve the same thickness. The observation that both led to comparable results (see Table 3.1) suggests that the chosen initialisation method approximates well any newly observed data.

Close examination of the results of Table 3.1 (which are depicted in Figure 3.13) also reveals that the blind estimator’s performance is better than one might expect, due to the low variance between the examined femurs, which is best visualised in the bottom row of Figure 3.14. Unfortunately, this renders the task of this thesis extremely difficult, as the margin of improvement is relatively low, and directly comparable with the imaging resolution of the QCT scanners.

Note that the deviation from the diagonal line at low and high thicknesses should not be perceived as bias, and does not conflict with the definition of the blind estimator. This is the expected behaviour when regions of unusually thin or thick cortex (compared to the average at that location) are encountered: thicker outliers are going to be underestimated, and vice-versa in the opposite case. In addition, please note that the blind estimator was built using a large number of CT femoral scans of elderly females, and not the Bern population. This was an intentional decision to make sure that these results reflect what would be expected in an unseen population. This is apparent in Figure 3.14c, which would be completely green (zero error) if the estimator was constructed using only the Bern population.

3.6.3 Density and blur regularisation

As explained in Section 3.4.2, due to the density-thickness trade-off there exists more than one solution that explains the data to the same degree. Even worse, the interpretation becomes harder under the presence of imaging blur, which can render a thin, high density tissue layer indistinguishable from a thicker, but less dense one. These experiments investigate the effect of different regularisation solutions to this problem.

Figures 3.15 and 3.16 suggest that assuming a constant value of cortical density throughout the femur leads to the best possible performance. Otherwise the optimiser frequently interprets thin, dense regions as thick, less dense ones, or vice versa. It is also possible that over-fitting occurs, as the estimation error increases despite the ROE being minimised. Furthermore, constraining the value of the blur variance seems to have little effect; either way, in reality it would be a constant variable dependent on the quality of the scanner. Unsurprisingly, constraining the value of the trabecular density

3. TEMPLATE-BASED CORTICAL THICKNESS ESTIMATION

Figure 3.13: Template initialisation/blind estimator errors as a function of thickness for the two registration methods.

had devastating results (not presented herein), since this is far from reality. To that end, our attempt to improve the trabecular modelling using a mean density voxel grid (again the outcome of averaging many CT scans), and a single scaling factor s_t (so that $d'_{\text{trab}} = s_t \times d_{\text{trab. model}}$) per cross-section failed, as the performance deteriorated.

Finally, although the results of the validation experiments notably improve upon the performance of the blind estimator, the poor performance of the clinical experiments suggests that further regularisation might be necessary. A possible solution is investigated in the next section.

3.6.4 Shape and thickness regularisation

As seen in Section 3.4.2, when only four DRR views are used (clinical experiments), the optimizer may converge to misshaped contours. The investigation of Figure 3.17, which depicts the ROE per DRR for the clinical and validation experiments, may hint whether this occurs as a result of the SAPs being over-fit to the APs, or whether the optimiser is trapped in a local minimum. In fact, the former seems to be the case since the four-view solution has by far the lowest ROE compared to the other experiments. Thus, it might be possible to improve the results by further regularising the parameters.

3. TEMPLATE-BASED CORTICAL THICKNESS ESTIMATION

Figure 3.14: a) Mean FWHM CT thickness, b) mean template initialisation thickness, c) mean initialisation error and d) initialisation error standard deviation. The standard deviation can also be thought as the thickness variance per location. To properly interpret these figures, please consider that the thickness direction is always perpendicular to the DRR axis of rotation, and thus they might differ from maps which plot the thickness along the surface normals.

3. TEMPLATE-BASED CORTICAL THICKNESS ESTIMATION

(a) 20 views, variance of blur constrained

(b) 20 views, cortical density constrained

(c) 4 views, cortical density constrained

Figure 3.15: Thickness estimation errors of the template method for the validation (a, b) and clinical (c) experiments, when constraining the variance of the Gaussian blur (a), or the cortical density (b, c).

3. TEMPLATE-BASED CORTICAL THICKNESS ESTIMATION

Figure 3.16: Cortical thickness estimation efficacy of the template method for the validation and clinical experiments when constraining either the variance of the blur, or the cortical density. The x-axis corresponds to the measurement bias, whereas the y-axis to the precision. The size of the markers is proportional to the ROE; the blind estimator’s (initialisation) ROE is plotted for reference.

Figure 3.17: Polar plot of the ROE (radius) against the angle of each DRR view for the CT solution (purple), the validation experiments (red), the clinical experiments without shape and thickness regularisation (green), and when constrained not to deviate by more than 1.5 mm from the initialisation (blue). The initialisation ROE is shown as well for completeness (cyan), although it is by definition always 100%. The light purple dots correspond to the ROE of the CT solution after optimising its shape by fitting the SAPs to the APs. The reduction can be primarily attributed to the semi-manual CT surface segmentation error.

3. TEMPLATE-BASED CORTICAL THICKNESS ESTIMATION

In this section, the effect of constraining the deviation amount from the initialisation is investigated. The validation experiments are used to demonstrate that this approach does not affect the limiting performance of the algorithm — in fact, even the estimation efficacy of the validation experiments is slightly improved, as seen in Table 3.1. Similarly, for the clinical experiments, regularising the algorithm leads to a lower estimation error, despite the higher ROE (which was expected) which is also plotted in Figure 3.17. The value of the latter is now comparable to what is observed when using 20 views, or using the “CT solution” approximation. Figure 3.18 compares the estimation error before and after the shape and thickness regularisation, whereas Figure 3.19 displays the regularised results in detail using percentile shading. Finally, the colour maps of Figures 3.20 — 3.22 depict the performance of the template method per vertex location.

Figure 3.18: Thickness estimation errors of the template method for (a) the validation and (b) clinical experiments, when constraining the shape and thickness to deviate by no more than 1.5 mm from the initialisation. The error is plotted as a function of thickness; the solid line represents the mean estimation, whereas the dotted ones are plotted at ± 1 standard deviation. The diagonal line represents the perfect estimator. Please note that the regularised results (blue lines) are also plotted in detail using percentile shading in the next figure.

3. TEMPLATE-BASED CORTICAL THICKNESS ESTIMATION

Figure 3.19: Template final design thickness estimation errors for the a) 20-view and b) 4-view experiments, using the gold standard, high resolution CT estimates.

3.6.5 Evaluating against the bronze standard

This section summarises the results of the validation (20-views) and clinical (4-views) experiments according to the design decisions presented in the previous sections. The template method’s results are compared against the gold (FWHM) and bronze (model-based) ground-truth thickness estimates separately. The error statistics plotted in Figure 3.23 are collated in Table 3.1. In addition, Figure 3.24 directly compares the error distributions of the gold and bronze standards to highlight their discrepancies. It is deduced that approximately 0.1 mm of the standard deviation may be attributed to the bronze standard uncertainty. The rest can be reasonably associated with the assumptions of the DXA approach.

3. TEMPLATE-BASED CORTICAL THICKNESS ESTIMATION

Figure 3.20: a) Mean FWHM CT thickness, b) mean DXA thickness from the validation experiments (20 views) and c) mean DXA thickness from the clinical experiments (4 views).

3. TEMPLATE-BASED CORTICAL THICKNESS ESTIMATION

Figure 3.21: a) Mean DXA thickness estimation error from a) the validation experiments (20 views) and b) the clinical experiments (4 views). Errors are calculated using the gold standard.

3. TEMPLATE-BASED CORTICAL THICKNESS ESTIMATION

(a) 20 views

(b) 4 views

(c) 20 views

(d) 4 views

Figure 3.22: Standard deviation of error, and RMS change (negative represents improvement) compared to the blind estimator for the validation and clinical experiments. Errors are calculated using the gold standard.

3. TEMPLATE-BASED CORTICAL THICKNESS ESTIMATION

Figure 3.23: Cortical thickness estimation efficacy of the template method for the validation and clinical experiments. The x-axis corresponds to the measurement bias, whereas the y-axis to the precision. The size of the markers is proportional to the ROE; the blind estimator’s (initialisation) ROE is plotted for reference, together with the 20 control point spline approximation of the CT solution.

Figure 3.24: Cortical thickness estimation error histograms of the template method, for the a) 20-view and b) 4-view experiments, when evaluating against the gold and bronze standards.

3.6.6 Evaluating performance in the region where estimation is expected to be best

Figure 3.25: The orientation of the 4 DRRs used in the clinical experiments is shown relative to the position of a femur (top view). Thickness estimation is expected to be best at the vertices whose normals are parallel to one of the views. This is because, for example, the APP sample s_i , which is between the clipping lines at locations x_i and x_{i+1} on the 0° view, contains information only about the cortex and not the trabecular compartment. Thus, the optimiser is only concerned with the cortical density and thickness, minimising the adverse effects of a wrong trabecular thickness and/or density estimation (recall the discussion about the ill-posed nature of the problem, and the thickness-density trade-off discussed in Section 3.4.2). The angle θ dictates the allowable misalignment of the vertex normals from the DRRs. In this example, the vertex with normal $\hat{n}$ is misaligned by θ' degrees from the 40° DRR. For this particular DRR arrangement, $\theta = 60^\circ$ will result in the inclusion of all vertices.

Section 2.3.3 associates the reconstruction resolution with the number of X-ray projections. Furthermore, it explains how sparse projections lead to gaps in the Fourier space that need to be interpolated, which in turn leads to artifacts. As an example, close inspection of Figure 2.4c reveals that the reconstruction of the bright oval “shell” is best at the top right and bottom left. In these regions, the “shell” normal is almost perpendicular to one of the four DRR casting directions, and hence maximum information about its thickness is obtained.

3. TEMPLATE-BASED CORTICAL THICKNESS ESTIMATION

This concept is explained in more detail in Figure 3.25, which presents how this affects cortical thickness estimation in the four DRR-view clinical experiments. In an effort to reveal the femoral regions where thickness estimation is best, Table 3.2 and Figure 3.27 breakdown the error statistics according to the maximum allowable vertex normal direction misalignment from the four DRR views. Similarly to what was observed in the filtered-backprojection simulations of Section 2.3.3, cortical thickness estimation for misalignments up to $\theta' = 10^\circ$ (i.e. all vertex normals that fall within the -30° to 50° range relative to the AP view) display a significantly better RMS error improvement.

On the downside, since by definition the 0° AP view presents the greatest area of the femur parallel to the DRR, this estimate sub-selection drastically reduces the examined femoral region. For an allowable misalignment up to 10° , 84.5% of the thickness measurements are ignored (see Table 3.2 and Figures 3.26 and 3.27c). Moreover, comparing Figures 3.12c and 3.28d reveals that the distribution has skewed towards the thicker side, which might partially explain the higher reduction in the RMS, as estimation is expected to be best in thicker regions where the blur's adverse effects are minimised. However, it is encouraging that this region is relevant to fracture (see Section 1.2). Error statistics for the restricted view experiments are collated in Table 3.2 and depicted in the colour maps of Figure 3.29.

Figure 3.26: The canonical morphology colour mapped according to the per-vertex angle misalignment from the four DRR views (-20° , 0° , 20° , 40°). Only vertices which are shaded red have a normal which deviates less than 10° from the views.

Table 3.2: Error statistics of restricted range clinical experiments. Accumulated values are shown in parenthesis, and estimation errors in the region -30° to 50° relative to the AP view are shown in bold. The *RMS change* values are calculated by comparing the RMS error of the optimised results with the RMS error of the blind estimator. Negative values correspond to improvement.

normal direction misalignment ($^\circ$)	mean error (mm)	std. deviation (mm)	RMS error (mm)	RMS change (%)	estimates (%)	
$0 - 5^\circ$	-0.08 (-0.08)	0.80 (0.80)	0.81 (0.81)	-0.25 (-0.25)	7.7%	(7.7%)
$5 - 10^\circ$	-0.09 (-0.08)	0.76 (0.78)	0.76 (0.79)	-0.24 (-0.25)	7.7%	(15.5%)
$10 - 15^\circ$	-0.05 (-0.08)	0.76 (0.78)	0.76 (0.78)	-0.17 (-0.24)	2.9%	(18.4%)
$15 - 20^\circ$	-0.03 (-0.07)	0.80 (0.78)	0.80 (0.78)	-0.18 (-0.23)	3.2%	(21.6%)
$20 - 25^\circ$	-0.05 (-0.07)	0.82 (0.79)	0.82 (0.79)	-0.18 (-0.22)	4.0%	(25.6%)
$25 - 30^\circ$	0.01 (-0.06)	0.78 (0.79)	0.78 (0.79)	-0.19 (-0.22)	4.9%	(30.5%)
$30 - 35^\circ$	-0.03 (-0.05)	0.71 (0.77)	0.72 (0.78)	-0.18 (-0.21)	6.6%	(37.1%)
$35 - 40^\circ$	-0.07 (-0.05)	0.72 (0.76)	0.72 (0.77)	-0.16 (-0.20)	8.5%	(45.6%)
$40 - 45^\circ$	-0.13 (-0.07)	0.73 (0.76)	0.75 (0.76)	-0.13 (-0.19)	11.2%	(56.9%)
$45 - 50^\circ$	-0.15 (-0.09)	0.70 (0.75)	0.72 (0.75)	-0.14 (-0.18)	13.4%	(70.3%)
$50 - 55^\circ$	-0.13 (-0.09)	0.66 (0.73)	0.67 (0.74)	-0.12 (-0.17)	14.8%	(85.1%)
$55 - 60^\circ$	-0.16 (-0.10)	0.64 (0.72)	0.66 (0.73)	-0.11 (-0.16)	14.9%	(100.0%)

3. TEMPLATE-BASED CORTICAL THICKNESS ESTIMATION

Figure 3.27: a) Cortical thickness estimation mean and standard deviation errors for the 4-view clinical experiments (green) and the template initialisation (cyan), according to the vertex normal misalignment (0° – 5° , 5° – 10° , ... , 55° – 60° bins). b) RMS error change when compared to the template initialisation (blind estimator), again as a function of the vertex normal misalignment. c) Percentage of estimates ignored against the maximum vertex normal misalignment.

3. TEMPLATE-BASED CORTICAL THICKNESS ESTIMATION

Figure 3.28: Cortical thickness measurement error statistics for the 4-view clinical experiments when only estimates whose vertex normal deviates less than 10° from the DRR casting directions are considered (-30° to 50° relative to the AP view).

3. TEMPLATE-BASED CORTICAL THICKNESS ESTIMATION

Figure 3.29: Cortical thickness estimation mean error (a), standard deviation error (b) and RMS change (negative represents improvement) compared to the template initialisation (c), for all estimates whose vertex normal deviates no more than 10° from the DRR casting directions (-30° to 50° relative to the AP view).

3.7 FEMCO results (QCT resolution only)

As a further validation, the algorithm was assessed on the remaining 120 clinical-resolution femoral scans of the FEMCO study, using the bronze-standard ground-truth estimates from the model-based CT method. Similarly to before, we measured the mean, RMS and percentage errors, the standard deviation and the ROE for the following configurations (all results are collated in Table 3.3 and visualised in Figures 3.30–3.34):

- **12 and 20 B-spline control point approximations** of the endocortical and periosteal surfaces according to the model-based CT method. This corresponds to the upper bound of performance. The ROE is calculated using a single value of trabecular density and imaging blur *per cross-section*, and a constant cortical density *per femur*, to promote a fair comparison with the DXA method's ROE which is calculated in a similar manner.
- **Initialisation (blind estimator)** error, when aligning the template to the DRR scans using the initialisation technique described in Section 3.3.1.
- **20-view experiments, 20 control point splines**, using a single value of cortical density *per femur*, which was deduced using information from the thick cortices near the femoral shaft. A separate value of imaging blur and trabecular density was deduced by the optimiser *per cross-section*. Results are presented both with and without thickness and shape regularisation by 1.5 mm.
- **4-view experiments, 12 control point splines**, using the same configuration for the cortical and trabecular densities, and imaging blur, as above. As expected, regularisation of shape and thickness by 1.5 mm has a great impact on these results (see Table 3.3), for the reasons explained in Section 3.4.2.

Table 3.3: Aggregated error statistics of FEMCO experiments. The first section presents the errors of the CT solution spline approximation when using different numbers of control points. The template initialisation/blind estimator’s errors follow, before presenting the 20-view and 4-view experimental results, both with and without a 1.5 mm constraint on the template’s maximum deformation. Finally, the last section tabulates the errors for the restricted angle range experiments, for a vertex normal deviation less than 10° (-30° to 50° relative to the AP view).

		mean error (mm)	std. deviation (mm)	RMS error (mm)	percentage error (%)	ROE (%)
CT solution approximation						
12 control point spline apr.		-0.02	± 0.36	0.36	10.9	72.3
20 control point spline apr.		-0.01	± 0.27	0.27	6.4	70.9
template initialisation (blind estimator)						
	2D/3D registration					
initialisation	affine	-0.08	± 0.77	0.77	29.3	100*
DXA method results	constrain					
20 views	-	0.07	± 0.64	0.65	22.2	40.3
20 views	shape & th. 1.5 mm	-0.03	± 0.59	0.59	22.2	43.7
4 views	-	0.04	± 0.94	0.95	38.7	27.7
4 views	shape & th. 1.5 mm	-0.05	± 0.67	0.67	27.4	39.6
restricted angle range	maximum vertex normal deviation					
4 views (shape & th. 1.5 mm)	10°	0.07	± 0.66	0.66	23.8	

*baseline ROE

3. TEMPLATE-BASED CORTICAL THICKNESS ESTIMATION

Figure 3.30: Template method thickness estimation errors for the a) 20-view and b) 4-view experiments, when regularising the shape and thickness not to deviate more than 1.5 mm from the initialisation. The blind estimator’s error can be seen in (c), whereas (d) presents the cortical thickness distribution across all 120 femurs. The blind estimator’s bias at low and high thicknesses is expected, as explained in Section 3.6.2: when the actual thickness of a specimen at a particular location differs significantly from the average (outlier), this estimator is bound to produce poor results.

3. TEMPLATE-BASED CORTICAL THICKNESS ESTIMATION

Figure 3.31: Cortical thickness estimation efficacy of the various configurations discussed in this section. The x-axis corresponds to the measurement bias, whereas the vertical to the precision. The size of the markers is proportional to the ROE; the blind estimator’s (initialisation) ROE is plotted for reference.

The errors are in agreement with the observations of the previous section, albeit marginally lower. Cortical thickness differences are smaller in this population, as indicated by the lower error of the blind estimator. Thus, the small performance improvement might be a result of the template initialisation better approximating the correct solution. These results once more demonstrate that the proposed template assumptions result in an optimisation cost function which, although it disagrees with the optimality of the CT solution (since its ROE is higher, see Table 3.3 and Figure 3.31), improves upon the blind estimator (the current best guess).

3.8 Discussion

Some important points relevant only to the template method are outlined below, as a more general discussion is presented in Chapter 5, after investigating the alternative statistical-based approach in the next chapter.

Overall performance

Figures 3.22 and 3.34, which depict the RMS change compared to the blind estimator per vertex location, are particularly useful for identifying the regions where the algorithm is accurate. As expected performance is best for thicker cortices, and away from highly

3. TEMPLATE-BASED CORTICAL THICKNESS ESTIMATION

Figure 3.32: a) Mean CT thickness, b) mean DXA thickness from the validation experiments (20 views) and c) mean DXA thickness from the clinical experiments (4 views).

3. TEMPLATE-BASED CORTICAL THICKNESS ESTIMATION

Figure 3.33: a) Mean DXA thickness estimation error from a) the validation experiments (20 views) and b) the clinical experiments (4 views).

3. TEMPLATE-BASED CORTICAL THICKNESS ESTIMATION

Figure 3.34: Standard deviation of error, and RMS change (negative represents improvement) compared to the blind estimator for the validation (20-views) and clinical (4-views) experiments.

3. TEMPLATE-BASED CORTICAL THICKNESS ESTIMATION

asymmetric and convoluted cross-sections. In particular, a common observation between all experiments is that the lesser trochanteric region is highly problematic. Investigation of individual reconstructions reveals that if the initialisation procedure fails to predict the position of the lesser trochanter accurately, there is a very low chance of a meaningful optimisation. Furthermore, it is apparent that this problem is amplified when only 4-views are used, as asymmetrical regions are almost impossible to interpret.

Close inspection of the standard deviation colour maps in Figure 3.14 reveals that most of the anterior, posterior and lateral regions of the femur have a population standard deviation of less than 0.5 mm. In other words, in these regions estimation can be improved by at most 0.5 mm. Therefore, it is not surprising that the overall improvement over the blind estimator seems marginal when looking at the aggregate results.

B-spline contours vs. surface models

Modelling the femur using a set of B-spline cross-sections is very computationally efficient and greatly reduces the dimensionality of the problem, yet comes with some limitations. First, all template cross-sections need to be perpendicular to the axis of rotation, which reduces the femoral surface area where cortical thickness estimation is possible (i.e. regions where the pelvis or the femoral head/acetabulum do not overlap with the projection of the femur). Second, the radially-defined B-spline control points place a constraint on the possible cross-sectional shapes, preventing them from forming “meanders”. Nonetheless, in the few cases when this was necessary — predominantly around the lesser trochanteric region — estimation was expected to be poor either way, especially when dealing with a small number of DRRs. Finally, a large proportion of the information contained in the DRRs is ignored, despite the fact that 20 contours resulted in a tightly packed stack of cross-sections.

Replacing the B-spline cross-sections with B-spline surface models is an alternative approach which could mitigate some of these limitations. Alternatively, the method considered in the next chapter can be used, which reduces the dimensionality of the surface models by exploiting the statistics of a femoral population.

Direction of thickness measurement

In all experiments of this chapter, cortical thickness estimation is performed along the DXA thickness direction, as defined in Section 3.5.3. This should be taken into account when comparing the colour maps presented herein with ones that plot the thickness along the vertex normal direction. A particular region where high discrepancies are expected is the medial part of the femoral neck, whereby the template’s cross sections

3. TEMPLATE-BASED CORTICAL THICKNESS ESTIMATION

intersect the femoral surface at very oblique angles.

Optimisation cost function and ROE

Figure 3.17 clearly depicts that the cost functions of the DXA and CT methods do not agree. This was expected, as the two approaches are driven by significantly different amounts of unambiguous information, and rely on different assumptions. Most notably, the CT methods assume a constant value of trabecular density *per thickness estimate* (which is only dependent on a few trabecular voxels nearby the endocortical surface), whereas the DXA method assumes a constant value of trabecular density *per femoral cross-section* (which averages out all trabecular voxels across the whole cross-section).

However, to completely reject the possibility of the optimiser being trapped in a local minimum, the template was initialised to the CT method solution and optimised as usual. In line with all other experiments, the optimiser still deduced a solution with a decreased ROE, but deteriorated efficacy, albeit slightly better than when starting from the template’s mean. For 20 views and comparing against the gold standard, the error was 0.11 ± 0.54 mm, RMS was 0.55 mm, and the ROE 60.9% (compared to 0.03 ± 0.60 mm, RMS 0.60 mm and ROE 48.1%).

Gold standard vs. bronze standard

Ideally, the template and statistical (next chapter) methods should be evaluated against the gold standard. However, this would severely restrict the available data to the 35 BERN femurs, as gold standard measurements can only be obtained from HRpQCT scans. As mentioned in the objectives of this thesis, the next chapter investigates whether cortical thickness estimation efficacy can be improved by tailoring the statistical model to a particular population. As it will be seen in Section 4.5, this requires partitioning the data to two complimentary groups. Therefore, subsequent experiments are going to be performed on the FEMCO dataset (120 femurs) to a) prevent poor error statistics, a consequence of using small test sets, and b) allow the construction of the statistical models from a reasonable number of femurs. The results of Section 3.6.5 indicate that the bronze standard’s uncertainty added approximately 0.1 mm to the standard deviation of the cortical thickness estimation error, an observation that should henceforth be considered.

Number of views

Apart from the 20-view and 4-view experiments presented herein, single-, stereo- and 6-view experiments were also investigated — the results are omitted in the previous

3. TEMPLATE-BASED CORTICAL THICKNESS ESTIMATION

sections for brevity because of the following reasons:

- We promptly abandoned the effort for **single-view** reconstructions, as the assumptions required to regularise the problem overconstrained the algorithm, to the point where no meaningful reconstructions were possible. Admittedly, the approach was relatively naive, as the template was initialised to an annulus of varying radius and “cortical thickness”. Specifically, the periosteal contours were initialised to ellipses, whose eccentricity and radius were determined by best-fitting them to a canonical model. The “thickness” was set to a constant value per cross-section, using similar means. Nevertheless, no matter what initialisation is chosen, a template method that encapsulates no statistics about the anatomy of the femur has to assume symmetry across the DXA plane, as no depth information is available. This poses a significant barrier to 3D reconstructions: in this particular case, the tissue volumes per SAP sample have to be deformed by an equal magnitude on both sides of the DRR. Thus, the cortical thickness of the anterior of the femur is directly affected by that of the posterior and *vice versa*, and there is no way of untangling the two. It is also worth mentioning that setting an evaluation protocol is harder, as registration along the DRR “depth” direction has to be performed heuristically.
- **Stereo** reconstructions were attempted using the AP and 50° DRRs. Results are omitted herein, as they are worse than the blind estimator’s. Section 3.6.6 explains how cortical thickness estimation is best at locations where the surface normals are parallel to the DRR’s orientation. Thus, a likely interpretation is that an angular separation of 50° is incapable of providing enough information for the proper interpolation of the cortex boundaries between these regions where estimation is expected to be best. In other words, the second view, despite introducing valuable depth information, has little practical use when it comes to cortical thickness estimation, as similarly to the single-view experiments an assumption of symmetry has to be made for the biggest part of the femur.
- Finally, for **6-view** reconstructions, the results indicated a 0.04 mm better RMS compared to the 4-view experiments. This error is merely an improvement, and might very well be within the noise window boundaries. Thus, it was deduced that an angular reduction of 9° (compared to the 4-view experiments) between successive DRRs would only result in a higher radiation dose. To this end, no experiments with even more views were attempted, as this directly opposes the motivation for a low X-ray dose.

Computation time

3D cortical reconstruction of a single bone is performed in approximately one minute, single-threaded, on a Intel Core i5 661 @ 3.3GHz processor. Performance can be easily improved by exploiting the capabilities of modern multi-core processors or graphics cards, since this problem is prime territory for multi-threaded execution; all SAP's are independent of each other and can be synthesised in-parallel.

3.9 Summary

This chapter investigated the use of a template based method to extract multiple cortical thickness measurements across the proximal femur from a set of DRR scans. The following points summarise the overall findings.

- The blind estimator is already performing relatively well, since the cortical variation amongst individuals of our test sets is very small: in many locations across the femur, the RMS improvement against the blind estimator can be at most 0.5 mm.
- Reconstructing the shape and cortical thickness of the femur from DXA scans is a hard optimisation problem involving a high number of unknown parameters. Since the DRRs provide very limited, obscured information — especially when dealing with just four views — model complexity should be kept to a minimum, and constraints should be imposed to prevent overfitting.
- Cortical thickness and density are complimentary variables: an increase of thickness coupled with a decrease of density appears on the DRRs very similar to the opposite case. In addition, under the presence of blur, thin cortices might completely disappear. Many solutions explain the data to a similar degree, which might “trick” the search function of the optimisation algorithm to explore an erroneous direction.
- Error colour maps of cortical thickness show that the algorithm performs best in thicker regions of the bone. The overall RMS error was 0.73 mm in the four-view experiments, which implies that such an algorithm is unlikely to assist in the detection of cortical thinning, especially in the areas of interest (i.e. where the cortex is very thin).

To this end, the next chapter investigates whether cortical thickness estimation efficacy can be improved by exploiting femoral statistical variations captured from training cohorts.

Chapter 4

Model-based cortical thickness estimation

4.1 Introduction

This chapter is concerned with an alternative approach to cortical thickness estimation from multi-view DXA, whereby the B-spline template is replaced by the combination of a Statistical Shape Model (SSM) and a Statistical cortical Thickness Model (STM). It is in many respects similar to the template based method examined in the previous chapter as it shares the same 2D/3D registration principles and experimental framework. However, it differs by investigating whether cortical thickness estimation can be improved by exploiting known femoral statistical variations.

To this end, the following sections define the statistical models used and the methodology of their construction (Section 4.2), outline the slight modifications needed in the 2D/3D registration method (Section 4.3) and explain how the optimisation method is adapted (Section 4.4).

Apart from testing various levels of model complexity, we seek to investigate whether tailoring a model to a particular dataset — using a training cohort with similar demographics — improves estimation. To do so, we split the experiments into two groups: “homogeneous” and “heterogeneous”. Section 4.5 presents the demographics of the different training and test populations, and provides details of the experimental setup. All results are summarised in Section 4.6, before drawing conclusions in Section 4.7.

4.2 Model definition

In this study, we constructed statistical models of the proximal femur using Principal Component Analysis (PCA), encoding both the shape and the cortical thickness distribution. Each model is essentially a Combined Statistical Model (CSM), i.e. a combination of a SSM and a Statistical Appearance Model (SAM) (explained in Section 2.6.2). In our case, the SAM is represented by the combination of a Statistical cortical Thickness Model (STM) and a set of tissue density values: cortical, trabecular and background. As before, we opt to parametrise the model directly with the variables of interest (cortical thickness), so that the optimiser’s cost function is directly related to them.

The steps required to construct the SSM and the STM are summarised in Figure 4.1 and in the following sections. Finally, examples of how these models appear can be found in Figures 4.2 and 4.3.

4.2.1 Statistical Shape Model

To construct the SSM, each femur of the training datasets was first segmented manually using Stradwin [157]. Next, a canonical femur shape was spatially registered to them automatically, by means of a similarity transformation followed by a Locally Affine Deformation (LAD) using wxRegSurf [53]. We thus obtained one deformation field per training sample, all expressed in the canonical morphology, as shown in Figure 4.1.

The principal modes of variation were then identified by performing PCA on the deformation fields, which involves finding the eigenvectors and eigenvalues of the covariance matrix:

$$\mathbf{S} = \frac{1}{N-1} \sum_{i=1}^N (\mathbf{x}_i - \bar{\mathbf{x}})(\mathbf{x}_i - \bar{\mathbf{x}})^T \quad (4.1)$$

where N is the number of training samples (observations), $\mathbf{x}_i$ a vector representation of the deformed canonical morphology, and $\bar{\mathbf{x}}$ a vector representation of the mean morphology.

We thus obtained $\Phi_{1..N-1}$ PCs, their corresponding $e_{1..N-1}$ eigenvalues, and a matrix of $N \times (N-1)$ weight values, $w_{PC}^{observation}$, which can be used to perfectly reconstruct each training sample, i , as follows:

4. MODEL-BASED CORTICAL THICKNESS ESTIMATION

Figure 4.1: The following steps are performed on each training sample, in preparation for PCA:

- (a) Each training specimen is colour mapped with the thickness estimates from the CT method.
- (b) The similarity transformation that best registers the canonical morphology (red) to each sample (green) is calculated using an iterative closest point approach. This is achieved by minimising the sum of the squared distances between nearest neighbours from the two shapes.
- (c) The Locally Affine Deformation (LAD) that best fits the mean canonical shape to each specimen is calculated and applied.
- (d) The per-vertex deformation field that relates the mean and the LAD'ed canonical shapes is calculated by subtracting the initial (i.e. after the similarity transformation) from the final vertex locations.
- (e) Each vertex of the shape model is assigned the cortical thickness of its nearest specimen vertex neighbour. Hence, the cortical thickness is now expressed in the canonical morphology.
- (f) The cortical thickness is smoothed by averaging neighbouring values to remove outliers, noisy and invalid estimates which were inconsistent with their surroundings (the latter are shown as gray spots in the figure).

Figure 4.2: The first five PCs of the SSM, built using the 120 FEMCO femurs. The mean canonical morphology is rendered as the least transparent grey surface. The blue surfaces represent the result of superimposing each of the PCs at a magnitude of +5 standard deviations, whereas the red at -5 standard deviations. As can be seen, each of these modes happens to correspond to a meaningful femoral parameter (in order: overall size, femoral neck angle, femoral neck length, greater trochanter size and position, lesser trochanter size and position).

4. MODEL-BASED CORTICAL THICKNESS ESTIMATION

$$\mathbf{x}_i = \bar{\mathbf{x}} + \sum_{m=1}^{N-1} \mathbf{w}_m^i \Phi_m \quad (4.2)$$

The eigenvalues express the amount of variance captured by each PC.

4.2.2 Statistical Thickness Model

Cortical thickness was estimated at all vertices around the surfaces of the training population using the bronze standard CT method. The measurements were then projected onto the canonical morphology using the SSM deformation fields and a nearest neighbour approach, as shown in Figure 4.1. Once this mapping was complete, the thickness estimates were smoothed by performing a weighted average of neighbouring values, where the weights were proportional to the probability of an accurate CT measurement. Just enough smoothing was applied, so that no invalid estimates existed in the examined region, something necessary for the proper operation of the PCA algorithm. Before identifying the principal modes of thickness variation amongst the training population using PCA, a further round of smoothing was applied based on the assumptions and observations explained in the following paragraph.

Although the LAD deformation fits accurately each femur to the canonical morphology, the precise alignment of all anatomical regions is never guaranteed, and small variations are expected. Our experiments, which involved a training cohort of 60 femurs, show that applying a further round of smoothing before construction the models results in explaining approximately 5–10% more of the thickness variance when using fewer than 10 thickness PCs. However, there is an trade-off associated with this decision: perfect reconstructions when using the “true modes” are not possible, as the model is no longer built with the precise bronze-standard CT measurements. The upper performance bounds when using a limited number of PCs are examined in detail in Section 4.6.2. Nonetheless, this should not be of particular concern, as these errors are well within the uncertainty of the CT bronze-standard, and certainly a small fraction of the expected DXA error.

4.3 2D/3D correspondence

Correspondence between the CSM and each of the DRRs is performed as before, using APs and SAP (Section 3.3). Moreover, spatial initialization is achieved using the

4. MODEL-BASED CORTICAL THICKNESS ESTIMATION

Figure 4.3: Colour maps of the mean cortical thickness, and the first four PCs of the STM at a magnitude of +5 standard deviations. This model was built using the 120 FEMCO femurs.

4. MODEL-BASED CORTICAL THICKNESS ESTIMATION

same semi-automatic, silhouette-based registration procedure, whereby the template is replaced by the SSM.

In contrast to the template based approach which employs B-spline cross-sections to model the femur, the statistical method studied in this chapter is represented by a surface model. Hence, correspondence of each AP with a particular cross-section is not required, nor presents any immediate advantages, and thus APs do not need to be coplanar. The advantage is apparent in Figure 4.4: it is now possible to infer 3D information about the femoral neck from a subset of the DRRs, something not possible in the template-based method. The disadvantage is lack of computational efficiency, as generating the SAPs requires the additional step of calculating the cross-sectional shapes from the surface model on each iteration of the optimisation procedure, which is an expensive operation.

Figure 4.4: The positioning of AP lines (red) on each DRR is performed automatically, according to the location of the SSM (blue) in 3D space, after the semi-automatic initial 2D/3D registration. Care is taken to a) verify that no AP lines overlap with the projection of the femoral head and b) evenly distribute the AP lines across the whole unobstructed femoral view, making sure that AP lines intersect the femoral cortex almost perpendicularly. This is achieved by calculating the projection of the SSM's femoral head and shaft rim, onto each DRR. Once the semi-automatic initialisation of the model is complete, the AP line positions are fixed in space to ensure that errors are calculated for the same DRR pixels independently of the deformation.

4. MODEL-BASED CORTICAL THICKNESS ESTIMATION

Unfortunately, it is still not possible to perform cortical thickness estimation near the femoral head, as the overlapping projections of the acetabulum/pelvis still renders the present algorithm unsuitable. Finally, the examined surface region is also limited by the size of the SSM (see Figure 4.2), extending to just below the distal part of the lesser trochanter.

4.3.1 Registration of DXA and CT estimates and thickness direction

Registering DXA and CT cortical thickness estimates is simpler than before as a consequence of using a surface model instead of B-spline contours. The combined model (SSM & STM) expresses cortical thickness along the vertex normals, using exactly the same thickness definition as the CT method. Hence, cortical thickness estimates between the two approaches were coupled using a simple nearest neighbour algorithm, and no direction normalisation was necessary.

4.4 Model deformation

4.4.1 Optimisation

The combined model was fitted to multi-view DXA scans using the same search function as in the template method (Levenberg-Marquardt). In this case however, the optimiser's task was to deduce the weights of each shape and thickness PC since the model's deformation is now governed by them. The additional parameters required to model the appearance of the femur on the DRR scans remained unchanged; namely these were a single value of cortical density (d_c), trabecular density (d_t) and imaging blur (σ) per AP line (Section 3.3.3).

In summary, the steps performed on each iteration of the optimisation process are the following:

1. First, the 3D vertex positions and surface normals from the SSM, combined with the thickness estimates per vertex from the STM, are used to define the endocortical mesh.
2. The normal of each DRR (direction of X-rays), together with the position and orientation of the AP lines on the DRRs define the planes where all voxels of interest lie. Hence, the intersection between these planes and the endocortical and periosteal surfaces define the location where the meshes are sampled, and determine the cross-sections that should be examined. Technically, this translates to calculating the intersections between a plane and multiple triangles and expressing

4. MODEL-BASED CORTICAL THICKNESS ESTIMATION

the result as a set of vertices which define a closed contour. Such an algorithm was already available in Stradwin and is used as-is with minor modifications to adapt to this particular problem. The resulting contours are slightly post-processed to simplify them: any vertex whose removal results in a change of the poly-line within 0.05 mm of the original is ignored.

3. Using the cross-sections from the previous step, the SAPs are synthesised using the same artificial ray-casting procedure as before (Section 3.4).
4. The cost function is evaluated by comparing the APs and the SAPs, and the error gradient is used to guide the search direction, according to the Levenberg-Marquardt algorithm. This results in a new set of PC weights, density values and blur.
5. The PCs are scaled by their new weights and superimposed on the mean shape and thickness models.
6. The above steps are repeated until convergence.

In Step 4, the optimiser was allowed to modify both the shape and thickness PCs at the same time. Different optimisation strategies were also examined, such as optimising first the shape and then the thickness, or starting with a small number of PCs and gradually introducing more. In addition, we investigated the cortical thickness estimation performance when encoding both the shape and the thickness variation in a single model. However, as explained later in the discussion of this chapter (Section 4.7), no performance gain was observed using any of the alternatives and thus their results are omitted for brevity.

4.4.2 Regularisation

Since PCA encodes the principal modes of variation in decreasing importance, the problem's dimensionality can be reduced by retaining a subset of them (analogous to reducing the number of B-spline control points in the template method). In fact, the next section examines in detail the maximum number of PCs that it is reasonable to use, according to Horn's Parallel Analysis (HPA). Nevertheless, it might be necessary to even further reduce their number below this limit if the observed data is noisy or sparse. Despite this coming at the expense of model flexibility, it is a beneficial compromise as it prevents over-fitting. We thus retained a variable number of shape and thickness modes, accounting for various levels of the training population variance, to investigate how accurate cortical thickness estimation is in each case.

With respect to the density-thickness trade-off, the optimisation is still ill-posed, in that more than one set of parameters can explain the APPs to the same degree. Apart from this being obviously true when a small number of views is investigated (due to the sparseness of the data), it still poses a problem even with many views, as the trade-off between cortical density and imaging blur renders thin-dense cortices indistinguishable from thicker-less dense ones. To tackle this issue, we preserve the same regularisation approach as before, that is we assume a fixed cortical density, d_c , throughout the proximal femur.

4.5 Experimental data, setup and evaluation protocol

For the experiments of this chapter we used the 120 femurs of the FEMCO study (Section 3.5.5), both as a training cohort for the models and as the testing dataset. To allow for a fair comparison between the template and statistical methods, the validation and clinical experiments were performed using the same twenty- and four-view DRR configurations respectively.

As noted in Section 2.6, significant structural and anatomical variations are observed amongst people of different age, gender and ethnic origin. In order to investigate whether cortical thickness estimation can be improved by fitting distinct statistical models to different target groups, we designed a three-way experiment which is summarised in Table 4.1. To this end, a “homogeneous” model refers to one where all individuals of the training cohort have very similar demographics. Conversely, a “heterogeneous” model includes femurs with a broader background. If the above hypothesis is true, estimation efficacy should be best in case AC, that is when a homogeneous model is fitted to femurs akin to the ones of the training cohort, and worst in case AD. On the other hand, experiments involving the heterogeneous model are expected give better results than case AD, but not as good as in case AC. The following sections first explain how we partitioned the 120 FEMCO femurs into a homogeneous and a heterogeneous group, then explain the cross-validation scheme used, and finally discuss some considerations about the model size that need to be taken into account.

4.5.1 Homogeneous and heterogeneous models

Table 4.2 and Figure 4.5 summarise the demographics of the 120 FEMCO femurs. To build a homogeneous and a heterogeneous model, they were partitioned into two anatomically dissimilar groups. To do so, the following considerations were taken into account:

4. MODEL-BASED CORTICAL THICKNESS ESTIMATION

Table 4.1: Training and testing cohorts, and three-way experimental setup.

Training data	Testing data	Experiments
(A) homogeneous	femurs similar to those in A (C) and to some of those in B (but not in A or B)	AC, AD, BCD
(B) heterogeneous	femurs not like those in A (D) and similar to some in B (but not in B)	

- All of the femurs originate from elderly females, so they already represent a relatively narrow population group.
- Ideally we would like to split the 120 femurs into two groups of 60. As discussed shortly, model size is particularly important for problems with high dimensionality. Hence, to maximise a) the model size of the homogeneous group and b) the diversity of the heterogeneous group, partitioning in half is desirable.
- All femurs are categorised into three classes: fallers from standing height or less without (“*healthy*”) and with hip fracture (“*fractures*”)¹, and hemiplegic stroke patients (“*frail*”). This categorisation presents some possible ways of partitioning the data.
- As already mentioned in Section 1.1.3, age is probably the strongest predictor of cortical discrepancies amongst individuals, and therefore is an alternative way of partitioning the data. Table 4.3 collates the number of femurs falling under the “young” and “old” subsets.
- Figure 4.6 attempts to shed some light on how these five subsets (“young”, “old”, “healthy”, “frail” and “fractured”) of the FEMCO dataset compare to each other. The first four colour maps on the left column of Figure 4.6 (a,c,e,g), depict a simple metric to examine the heterogeneity between subsets, i.e. the difference of the means of each population. As seen, on average the “young” femurs display a thicker cortex than the “old” ones. Similarly, “healthy” femurs appear to have on average thicker cortices than “fractured” femurs. “Frail” bones seem to be somewhere in-between.

¹The “*fractures*” group consisted of the contralateral femur of each individual, not the fractured femur.

4. MODEL-BASED CORTICAL THICKNESS ESTIMATION

The remaining colour maps (b,d,f,h,i,j,k) examine how the homogeneity *within* each subset compares with the others, by contrasting the standard deviation of cortical thickness at each vertex. Between the “young” and the “old” groups the former seems to be the less diverse. In addition, the “healthy”, “frail”, and “old” groups appear to be equally homogeneous, and the “fractured” subset falls somewhere in between them and the “young” group.

- Before choosing a way to partition the data, along with the above observations one should consider the following. This study aims to assist in diagnostic imaging: that is, at the time of the examination it would not yet be known whether a bone is fragile due to cortical peculiarities. In that respect, although building separate models for “healthy” and “fractured” bones would satisfy the heterogeneous/homogeneous requirements discussed above, it makes little sense since at the time of diagnosis it wouldn’t be known which model to use. Moreover, this would require careful thought on how to categorise the “frail” bones, as they fall into the middle of the spectrum. It is likely that they would have been split between the two groups to make them equally sized, a necessary compromise due to the limited amount of data.

Therefore, we divided the femurs into two groups according to the age of the individuals, “young” and “old”, which led to the creation of two homogeneous models. Between them there was no age overlap, and their mean age difference was just short of two decades (Table 4.3). Finally, as seen in Figure 4.6, they seem to satisfy the requirements of the proposed three-way experimental setup: they appear to be heterogeneous between them, and homogeneous within them (at least compared to the other proposed partitioning schemes).

The heterogeneous model was trained using 30 “young” and 30 “old” bones, and is thus henceforth referred to as “*combined*”. Stratified partitioning ensured that the proportion of femurs from each of the six subcategories (young, old and healthy, frail, fractured) remained the same as in the homogeneous models, to eliminate any bias and preserve an even balance in the demographics.

Table 4.2: FEMCO demographics

	AGE (years)		BMD (mg/cm ³)		WEIGHT (kg)		HEIGHT (m)	
	mean	st. dev.	mean	st. dev.	mean	st. dev.	mean	st. dev.
young	67.3	±4.8	1,176	±49	68.8	±13.2	1.63	±0.07
	85.7	±3.0	1,171	±90	59.7	±11.3	1.59	±0.06
old								

Table 4.3: FEMCO partitioning of data. The combined dataset is split into two rows to show how stratified partitioning ensured that there was no bias towards the “young” or “old” femurs, nor towards fractured or healthy bones.

		min.	max.	HEALTHY		FRACTURES		FRAIL		TOTAL	
		age	age	n	age	n	age	n	age	n	age
young	old	53	74	14	64.3 ± 5.9	23	67.7 ± 4.1	23	66.9 ± 5.0	60	67.3 ± 4.8
		81	93	20	83.6 ± 1.5	13	85.5 ± 2.5	27	87.4 ± 3.2	60	85.7 ± 3.0
combined	(young)	60	74	7	64.1 ± 5.6	11	67.7 ± 4.1	12	66.8 ± 5.0	30	67.2 ± 4.7
	(old)	82	91	10	83.6 ± 1.5	6	85.4 ± 2.5	14	87.4 ± 3.2	30	85.7 ± 3.0

4. MODEL-BASED CORTICAL THICKNESS ESTIMATION

Figure 4.5: Age, BMD, weight and height demographics of the FEMCO dataset. Each individual is represented by a marker in the scatter plots. The crosses are centered at the mean values of the young and old populations, and their sizes are equal to one standard deviation.

4.5.2 Cross-validation testing

To properly evaluate the performance of the predictive statistical models and assert that they generalise well when fitted to independent data, we resorted to the cross-validation scheme summarised in Table 4.4. We avoided partitioning the available data to two complementary sets, i.e. one for testing and one for training the models, since doing so would severely limit the size of both sets. Instead, we opted to perform a Leave-One-Out (LOO) validation whenever required.

4.5.3 Model size

The statistical model is represented by a canonical surface with 5,580 vertices, each one of which is described by a 3D vector in Cartesian coordinates. In addition, each vertex is associated with a thickness value, leading to a total of 22,320 variables to be estimated. There seems to be no definite answer in the literature concerning the required size of the training cohort for problems dealing with a large number of degrees of freedom, coupled

4. MODEL-BASED CORTICAL THICKNESS ESTIMATION

Figure 4.6: Homogeneity/heterogeneity of the FEMCO dataset. The label of each sub-figure indicates which statistic (mean: [a,c,e,g], standard deviation: [b,d,f,h,i,j,k]) and which subsets of the data (old, young, healthy, fractured, frail) are compared. For example, (a) depicts the difference between the means of all “young” and all “old” femurs (independent of whether they are healthy/frail/fractured), and (d) illustrates the difference between the standard deviations amongst all “healthy” and all “fractured” femurs (independent of their age).

4. MODEL-BASED CORTICAL THICKNESS ESTIMATION

Table 4.4: Cross-validation experimental scheme. The numbers in parenthesis in the first column correspond to the number of models built. For an explanation of the last column, please refer to Table 4.1 which summarises the experimental setup.

Model	Model size	Testing data	Testing data size	Cross-validation scheme	Case
young (60)	59	young	60	LOO	AC
old (60)	59	old	60	LOO	
young(1)	60	old	60	independent	AD
old (1)	60	young	60	independent	
combined (30)	59	young	30	LOO	BCD
combined (1)	60	young	30	independent	
combined (30)	59	old	30	LOO	
combined (1)	60	old	30	independent	

with a relatively low sample size (HDLSS), such as the one examined herein [22]. The answer depends heavily on the type of the problem and the constraints that can be imposed to regularise it. A relatively small training sample that does not properly approximate the overall statistics of the population will result in a biased model: its predictive ability might be great for specimens akin to the training samples, but when fitted to unseen data it might not generalise well (overfitting).

It is often suggested that the number of training samples should be determined based on the number of variables. Ratios of observations to variables of 2:1 up to 20:1 have been proposed, but there is no theoretical rationale to support them [3; 29; 46; 86]. Instead, simulations and studies with real data have shown that this was not an important component in characterising pattern stability, whereas absolute sample size was [60]. Recommendations for a minimum of 100–200 samples have been proposed [29; 101], although such claims are based on empirical observations and have no mathematical foundation. In addition, Guadagnoli and Velicer [60] used a Monte-Carlo procedure to deduce that the most important factor for determining the training sample size was component saturation (i.e. the amount of correlation between the PCs and the observed variables). Yet, they state that the general consensus seems to agree on the use of as many training samples as possible.

The following experiment was devised to examine the stability of the present model and assess the amount of PC saturation. We created multiple models using a variable

4. MODEL-BASED CORTICAL THICKNESS ESTIMATION

number of training samples, ranging from 10 to 120. Fifty models were created for each model size (50 folds), whereby in each case stratified partitioning ensured that the proportions of young/old and healthy/frail/fractured femurs remained the same. Subsequently, we compared each n -sized model with the one created using all 120 femurs. This was achieved by pairwise comparing the mean and first four *shape* PCs between them (scaled at one standard deviation using the eigenvalues), and calculating the average RMS difference. In a similar manner we evaluated the mean *percentage*¹ RMS difference for the *thickness* mean and first four PCs. Figure 4.7 plots the averaged results over the 50-folds against the model size. In addition, we calculated the intra-fold standard deviation of the PC difference for equally-sized models, which is also plotted in the same figure. These findings are expanded in the colour maps of Figure 4.8 which depict the differences per vertex location.

Thickness-wise, results show an average difference of up to 6% between the models created using 60 and 120 femurs, although the intra-fold variation of the 60-bone models reached up to 11%. Shape-wise, the corresponding differences were approximately 0.1 mm and 0.25 mm respectively. These experiments demonstrate the amount of PC saturation and model stability, although it would be desirable in the future to repeat the experiments with a larger sample size.

4.5.4 Number of Principal Components and Horn’s Parallel Analysis

Selecting the wrong number of components to retain can have a severe impact on the results [28; 48]. The amount of variance captured by each PC is indicated by the eigenvalue associated it. Thus, it is possible to sort the PCs in decreasing order of importance and ignore the least significant ones, as they are dominated by noise.

Multiple techniques to deal with this problem have been proposed in the relevant literature, such as the Guttman-Kaiser (GK) eigenvalue greater than one rule, retaining enough components to explain 95% (or so) of the variance, Bartlett’s test for equality of eigenvalues, Cattell’s scree test, Velicer’s minimum average partial and Horn’s Parallel Analysis (HPA) [46; 56]. Amongst them, the latter is probably regarded as the best method [46].

Therefore, we applied HPA [69], a Monte-Carlo based simulation method, to our statistical shape and thickness models. To do so, PCA was performed multiple times on uncorrelated normally distributed data of size equal to the femoral training dataset.

¹In this case, PC differences were expressed as a percentage of the mean thickness value at each vertex location. In many cases, the 120-femur model PC thickness values were very close to zero at many vertices, thus not allowing the calculation of a meaningful percentage difference against them.

4. MODEL-BASED CORTICAL THICKNESS ESTIMATION

Figure 4.7: This figure illustrates how the size of the model affects the statistics of the (a) thickness and (b) shape models. The methods used to create it are explained in Section 4.5.3. The solid lines represent the average RMS difference between a model of size x and the model built using all 120 FEMCO femurs. The dotted lines depict the average standard deviation between models of the same size, but which were built using different, semi-randomly (stratified partitioning) selected femurs. They are drawn at ± 1 standard deviations.

4. MODEL-BASED CORTICAL THICKNESS ESTIMATION

Figure 4.8: Colour maps (a-d) expand upon Figure 4.7, showing the differences between the mean and the first four PCs of the thickness models created using 60 and 120 femurs. The methods used to create them are explained in Section 4.5.3. In addition, sub-figures (e-f) display the standard deviation of the intra-fold difference for the models created using 60 femurs. Please note that the colour scale limits vary amongst the figures.

4. MODEL-BASED CORTICAL THICKNESS ESTIMATION

Only the PCs of the actual shape and thickness models whose eigenvalues are bigger than the average of those obtained from the completely random data should be retained.

Figures 4.9a and 4.9b plot the thickness and shape eigenvalues respectively, demonstrating the amount of variance captured by each PC. In addition, the results of the HPA are overlaid, suggesting that a maximum of 10 thickness, and 6 shape modes should be retained. In addition, in Figures 4.9c and 4.9d, the y-axis represents the cumulative variance explained using only the first n PCs, depicting the effect of reducing the dimensionality; that is, ignoring the PCs with the smallest corresponding eigenvalues.

Inspection of the plots in Figure 4.9 reveals that the old models are able to capture approximately 10% more of the thickness variance using just the most significant PC. The first two rows of the colour maps in Figure 4.3 show that the distribution of the dominant mode is very similar to the mean thickness. Hence, the first PC corresponds approximately to a proportional thickening/thinning around the femur. The above observations lead to the conclusion that the overall femoral differences within the old population are better explained by overall thinning.

4.6 FEMCO Results

4.6.1 Blind estimators

Before deforming the statistical model to fit a set of DXA scans, one should establish a comparative baseline. The template method presented in the previous chapter is the first algorithm of its kind, i.e. able to extract thousands of localised femoral cortical thickness estimates from DXA scans. Hence, comparing against its performance would be the intuitive approach. However, direct comparison might be misleading for two reasons. First, thickness is measured along different directions. Second, slightly different femoral regions are examined: the template method neglects the femoral neck, and the model-based approach ignores the proximal region of the femoral shaft distal to the lesser trochanter. Nevertheless, Section 5.1 of the next chapter attempts to directly compare the two methods by examining the estimation performance in the overlapping regions, along a common thickness direction.

We therefore once more set as the comparative baseline the blind estimator, for the reasons explained in the previous chapter. In this case however, three distinct blind estimators were calculated using the homogeneous and heterogeneous cohorts from the FEMCO dataset: one corresponding to the 60 “young” femurs, one to the 60 “old” and one to the “combined” set (all 120 femurs). Figure 4.10 depicts the estimation efficacy of each one of them when fitted to the different test sets, along with the cortical thickness

4. MODEL-BASED CORTICAL THICKNESS ESTIMATION

Figure 4.9: Comparison of the amount of variance captured as a function of PCs used for each of the models tested. The “one out” results were calculated by averaging the eigenvalues of all the corresponding LOO models. The dotted lines in (a) and (b) represent the eigenvalues obtained by HPA.

4. MODEL-BASED CORTICAL THICKNESS ESTIMATION

distributions observed.

By definition, estimation is best when the training and test datasets are identical (Figures 4.10a, 4.10e and 4.10i). It is important to note that proper interpretation of Figures 4.10b to 4.10i requires their examination in conjunction with Figures 4.10j to 4.10l. Close inspection of the histograms clearly reveals how the tailored models manage to achieve the lowest errors: for each model, accuracy is best around the thickness mark where its corresponding distribution peaks, illustrating the advantage of exploiting the anatomical variations expected in the test set. When deviating away to less frequent thicknesses (outliers), the blind estimators' errors are ever increasing. This is an expected behaviour, as by definition they are only good at predicting cortices which are not unusually thick/thin. Finally, examination of these figures reveals an overestimation bias when fitting the “young” mean to the “old” test set, and vice-versa in the opposite case. This observation is in agreement with what is expected, as ageing is associated with cortical thinning.

The colour maps in Figure 4.11 present the mean and standard deviation of the cortical thickness distribution across the femur for the “young”, “old” and “combined” populations. Once more, it is apparent that the “young” cohort exhibits on average an overall thicker cortex, especially around the lesser trochanteric region. The last figure of this section (Figure 4.12) attempts to shed some additional light on the estimation efficacy of the blind estimators. However, in this case the results are categorised according to the experimental setup introduced in Section 4.5. The first row of colour maps presents the error statistics when fitting the homogeneous models (“young” and “old”) to their appropriate respective test sets. In contrast, the second row depicts the errors obtained when they are fitted to specimens from the wrong test set (i.e. the “young” model is fitted to the “old” test set and vice-versa). As expected errors are noticeably higher in the latter case. Finally, the third row examines the performance of the heterogeneous, “combined”, model, averaged across both the “young” and “old” test sets.

4.6.2 CT solution — validation and upper performance bound

As already discussed in Section 4.5.1, PCA provides us with the weights required to perfectly reconstruct the training samples using Equation 4.2. Two models were built using all available “young” and “old” femurs respectively. These can be used to investigate if the DXA optimisation cost function correlates with cortical thickness estimation

4. MODEL-BASED CORTICAL THICKNESS ESTIMATION

Figure 4.10: DXA thickness against the bronze-standard thickness, after fitting the mean “young” (a,d,g), “old” (b,e,h) and “combined” (c,f,i) SSM and STM models to each femur of the “young” (first row), “old” (second row) and “combined” (third row) test sets. These results are equivalent to the blind guess estimators. Moreover, (j), (k) and (l) illustrate the distribution of thickness observed in each of the training/test sets.

Figure 4.11: Mean and standard deviation colour maps of the cortical thickness distribution observed in the “young” (a,b), “old” (c,d) and “combined” (e,f) training/test sets.

4. MODEL-BASED CORTICAL THICKNESS ESTIMATION

Figure 4.12: Colour maps of the the standard deviation of the error when fitting the mean SSM and STM of the homogeneous (a,b) and heterogeneous (c) models to different test sets. Cases AC, AD and BCD were introduced in Section 4.5: in AC, the “young” and “old” mean models are used to reconstruct specimens in the “young” and “old” test sets respectively. In contrast, in AD, the wrong model is fitted to each femur. Finally, in BCD, the “combined” model is used to predict the thickness of all 120 femurs, from both the “young” and “old” cohorts. Please note that the inclusion of the mean error colour maps for each of these cases is omitted, as they exhibit a zero error across the whole femur. Although this is obvious for cases AC and BCD, it is also true for case AD: the underestimation observed when fitting the “old” model to the “young” femurs is cancelled out by the overestimation seen in the opposite case.

4. MODEL-BASED CORTICAL THICKNESS ESTIMATION

accuracy as required. In other words, we expect the ROE to be inversely correlated with the number of true modes used. Figure 4.13 and Table 4.5 summarise the average error statistics across all femurs for both the “young” and “old” models, and verify this hypothesis. However, the relatively small reduction in ROE indicates that, especially when dealing with the “old” population, the blind estimator is already a very good fit. In addition, Figure 4.14 expresses the same statistics as a function of cortical thickness, and expands upon the compact representation of Figure 4.13, to demonstrate how the upper-bound estimation accuracy is affected when reducing the problem dimensionality. For all of these experiments the correct DXA to CT registration matrices initialised the position of the statistical models, instead of employing the semi-manual, silhouette-based registration.

It is worth noting that smoothing the bronze-standard cortical thickness distribution of each femur prior to the PCA analysis (Section 4.2.2) is most probably responsible for the tiny, consistent cortical thickness overestimation bias of 0.03–0.05 mm. As a validation test, the non-smoothed model was also used to perfectly reconstruct each femur, whereby this bias vanishes. For brevity reasons, Figure 4.13 and Table 4.5 include only the said reconstruction when all 59 available shape and thickness modes were applied (blue marker). In that case, the 0.06 mm RMS (0.4%) error can be safely attributed to the precision of the MATLAB PCA algorithm. Smoothing of the bronze standard was performed on the assumption that it would not limit the predictive ability of the model when a small subset of the PCs are used, while at the same time it would allow the capture of a bigger variance percentage with fewer modes of variation. The (approx.) 0.3 mm RMS error observed when using 6 shape and 10 thickness PCs justifies this decision, since this upper-bound performance limit is well below the expected error.

4. MODEL-BASED CORTICAL THICKNESS ESTIMATION

Figure 4.13: Plot of the accuracy (x-axis), precision (y-axis) and ROE (size of markers, with 100% represented by the largest error, i.e. the mean models) for both the “young” (red hue) and “old” (green hue) models as a function of the number of “true modes” used (color saturation, explained in the legend). In addition, the result of the validation experiment, where the mapped thickness values were not smoothed prior to constructing the “old” model, is plotted for reference (blue). Exact values are collated in Table 4.5.

4. MODEL-BASED CORTICAL THICKNESS ESTIMATION

Figure 4.14: DXA thickness against true thickness for the “old” model with (green), and without (blue) smoothing after projecting to the canonical morphology, and the “young” (red) model, when different numbers of true modes are used (indicated in the label in the form [shape,thickness]).

Table 4.5: Cortical thickness estimation errors and variance captured from each training dataset for the “young” (Y) and “old” (O) models when reducing the model dimensionality. The first row corresponds to the blind estimator errors. In this case, when fitting the model to a femur each PC is multiplied by the weight that would lead to perfect reconstruction if all modes were used (as derived from Equation 4.2). Hence, these error statistics represent the upper-bound performance limits.

modes		variance (%)				mean error (mm)		std. deviation $\pm$ (mm)		RMS (mm)		percentage error (%)		ROE (%)	
shape	thickness	shape Y	shape O	thickness Y	thickness O	Y	O	Y	O	Y	O	Y	O	Y	O
0 (<i>mean</i>)	0 (<i>mean</i>)	0.0	0.0	0.0	0.0	0.04	0.04	0.56	0.60	0.56	0.60	31.2	43.6	100.0*	100.0*
1	1	58.9	58.0	33.0	45.5	0.03	0.05	0.38	0.42	0.38	0.42	22.7	32.7	97.7	98.8
2	2	75.8	72.4	44.9	52.2	0.03	0.05	0.36	0.39	0.36	0.39	21.2	30.7	95.2	98.2
3	3	80.9	80.9	51.0	58.0	0.04	0.05	0.35	0.37	0.35	0.37	20.5	29.6	94.7	98.0
5	5	87.0	87.2	60.5	65.5	0.04	0.05	0.32	0.35	0.32	0.35	19.1	28.1	91.0	96.5
7	7	90.4	90.0	67.5	71.5	0.04	0.05	0.30	0.33	0.30	0.33	18.2	26.4	90.4	95.2
5	9	87.0	87.2	72.4	75.7	0.03	0.05	0.28	0.31	0.29	0.32	17.2	25.1	90.0	96.3
6	9	89.2	88.8	72.4	75.7	0.03	0.05	0.28	0.31	0.29	0.32	17.3	25.2	89.8	95.3
9	9	92.4	91.9	72.4	75.7	0.04	0.05	0.29	0.31	0.29	0.32	17.4	25.2	89.2	93.9
5	10	87.0	87.2	74.4	77.5	0.03	0.05	0.28	0.30	0.28	0.31	16.9	24.6	89.9	96.5
6	10	89.2	88.8	74.4	77.5	0.03	0.05	0.28	0.30	0.28	0.31	16.9	24.7	89.7	95.5
12	12	94.4	94.0	77.9	80.7	0.03	0.05	0.27	0.29	0.27	0.29	16.2	23.5	88.4	93.4
15	15	95.6	95.3	82.2	84.6	0.03	0.05	0.25	0.27	0.25	0.28	15.1	22.0	88.0	93.3
20	20	97.0	96.9	87.4	89.2	0.03	0.05	0.23	0.25	0.23	0.25	13.7	20.0	87.3	92.4
59 (<i>all</i>) ⁺	59 (<i>all</i>) ⁺	100.0	100.0	100.0	100.0	0.00	0.00	0.06	0.07	0.06	0.07	0.4	0.5	86.9	91.9

* baseline ROE

⁺ no thickness-smoothing after mapping to canonical morphology

4.6.3 20-view and 4-view experiments

This section presents the experimental results of the independent testing. In other words, all reconstructions were performed using models which were built without using the examined femurs, as summarised in the cross-validation experimental scheme of Table 4.4. As in the template method, validation experiments were performed using 20 equiangular DRR views in the range 0° – 171° , whereas clinical experiments used just 4 views at -20° , 0° , 20° and 40° .

Estimation efficacy vs. number of PCs

Tables 4.6 and 4.7 present detailed statistics of the validation and clinical experiments respectively. Errors are tabulated according to the number of PCs used in the optimisation procedure. Figure 4.15 illustrates them, summarising the three most important error statistics: the mean estimation error, the standard deviation and the ROE. Its primary goal is to reveal whether the limited amount of information captured by the DXA scans requires a further reduction of the dimensionality of the models, below the limit set by HPA. In addition, it separates the predictive ability of each model based on the test set examined. Therefore, the hypothesis stating that a tailored model leads to minimal measurement errors can be ascertained.

Table 4.6: **20-view**, averaged DXA cortical thickness estimation errors when fitting the “young”, “old” and “combined” models to the “young” (Y) and “old” (O) test sets, as a function of the number of PCs used. The first row of each section corresponds to the blind estimator errors.

modes		mean error (mm)		std. deviation $\pm$ (mm)		RMS (mm)		percentage error (%)		ROE (%)	
shape	thickness	Y	O	Y	O	Y	O	Y	O	Y	O
COMBINED MODEL											
0	0	-0.09	0.15	0.55	0.61	0.56	0.63	29.2	49.1	100.0*	93.8
5	9	-0.02	0.01	0.50	0.53	0.50	0.53	25.1	35.0	70.4	62.5
5	10	-0.04	0.00	0.51	0.53	0.51	0.53	25.4	34.4	69.5	62.0
6	9	-0.02	0.01	0.50	0.53	0.50	0.53	25.3	34.8	69.5	62.2
6	10	-0.04	0.00	0.51	0.53	0.51	0.53	25.6	34.7	68.6	61.9
OLD MODEL											
0	0	-0.20	0.04	0.55	0.60	0.58	0.60	28.9	43.7	102.9	91.4
5	9	-0.03	0.01	0.52	0.53	0.52	0.53	28.1	35.9	76.1	64.4
5	10	-0.03	0.01	0.52	0.53	0.52	0.53	27.6	36.0	75.4	63.4
6	9	-0.01	0.03	0.51	0.52	0.51	0.52	27.4	35.8	70.9	61.6
6	10	0.00	0.02	0.51	0.53	0.51	0.53	26.9	35.9	70.4	61.0
YOUNG MODEL											
0	0	0.04	0.28	0.55	0.65	0.59	0.70	31.2	57.3	98.3	98.5
5	9	-0.01	0.02	0.48	0.54	0.48	0.54	24.1	35.1	70.3	66.0
5	10	0.00	0.02	0.47	0.54	0.47	0.54	24.2	35.6	70.1	65.1
6	9	-0.01	0.02	0.47	0.54	0.47	0.54	24.1	35.0	69.7	65.5
6	10	0.00	0.03	0.47	0.54	0.47	0.54	24.1	35.7	68.8	64.5

* baseline ROE

Table 4.7: **4-view**, averaged DXA cortical thickness estimation errors when fitting the “young”, “old” and “combined” models to the “young” (Y) and “old” (O) test sets, as a function of the number of PCs used. The first row of each section corresponds to the blind estimator errors.

modes		mean error (mm)		std. deviation $\pm$ (mm)		RMS (mm)		percentage error (%)		ROE (%)	
shape	thickness	Y	O	Y	O	Y	O	Y	O	Y	O
COMBINED MODEL											
0	0	-0.09	0.15	0.55	0.62	0.56	0.63	29.1	48.9	100.0*	92.0
4	8	-0.01	0.04	0.54	0.56	0.54	0.56	27.0	38.4	73.7	66.0
5	9	-0.03	0.04	0.53	0.56	0.53	0.56	26.8	38.2	70.6	62.9
5	10	-0.04	0.04	0.53	0.55	0.53	0.56	27.1	38.2	71.5	62.9
6	9	-0.03	0.04	0.52	0.55	0.52	0.56	26.3	38.1	69.8	62.7
6	10	-0.03	0.04	0.53	0.55	0.53	0.55	26.9	38.1	68.9	62.1
OLD MODEL											
0	0	-0.20	0.04	0.55	0.60	0.59	0.60	28.9	43.6	103.0	89.0
4	8	-0.03	0.04	0.55	0.54	0.55	0.54	29.1	36.7	75.1	67.1
5	9	-0.04	0.04	0.55	0.55	0.55	0.55	28.9	37.5	75.5	65.9
5	10	-0.04	0.05	0.55	0.55	0.55	0.55	29.1	37.9	75.1	65.0
6	9	-0.06	0.03	0.52	0.54	0.53	0.54	27.3	36.9	71.2	64.1
6	10	-0.05	0.03	0.53	0.54	0.54	0.54	27.8	36.8	70.7	64.0
YOUNG MODEL											
0	0	0.03	0.28	0.56	0.64	0.56	0.70	31.2	57.1	100.4	98.5
4	8	-0.01	0.04	0.49	0.55	0.46	0.55	24.4	36.9	71.3	65.8
5	9	-0.01	0.04	0.49	0.56	0.49	0.56	24.9	36.8	69.2	63.3
5	10	0.00	0.04	0.50	0.56	0.50	0.56	25.3	37.6	68.4	63.0
6	9	-0.01	0.05	0.49	0.56	0.49	0.56	24.8	37.8	68.5	63.7
6	10	-0.01	0.05	0.50	0.56	0.50	0.56	25.5	38.4	68.1	63.2

* baseline ROE

4. MODEL-BASED CORTICAL THICKNESS ESTIMATION

Figure 4.15: Thickness estimation accuracy (x-axis), precision (y-axis) and ROE (size of marker) as a function of PCs used. Brighter colors correspond to more shape and thickness PCs — the exact number of modes is shown in the legend. The colors of the markers identify the experimental setup: for example, the red ones were obtained by fitting the “combined” model to the “young” test population. All plotted values are collated in Tables 4.6 and 4.7.

Table 4.8: **20-view and 4-view**, averaged DXA cortical thickness estimation errors when fitting the homogeneous and heterogeneous models to different test sets, according to the cases defined in Table 4.1. The first row of each section corresponds to the blind estimator errors, whereas the second is obtained after optimising 10 thickness and 6 shape modes.

modes		mean error (mm)		std. deviation $\pm$ (mm)		RMS (mm)		percentage error (%)		ROE (%)	
shape	thickness	20 views	4 views	20 views	4 views	20 views	4 views	20 views	4 views	20 views	4 views
HOMOGENEOUS TO CORRECT (case AC)											
0	0	0.04	0.04	0.58	0.58	0.58	0.58	37.4	37.5	94.0	94.2
6	10	0.01	0.01	0.50	0.52	0.50	0.52	30.1	31.2	64.4	65.6
HOMOGENEOUS TO WRONG (case AD)											
0	0	0.04	0.04	0.64	0.64	0.64	0.65	43.1	43.2	100.0*	100.0*
6	10	0.01	0.00	0.52	0.55	0.52	0.55	31.4	34.4	67.0	66.5
HETEROGENEOUS TO ALL (case BCD)											
0	0	0.03	0.03	0.59	0.60	0.60	0.60	39.0	39.2	95.3	96.2
6	10	-0.02	0.01	0.52	0.54	0.52	0.54	30.2	31.6	64.7	65.0

* baseline ROE

Three-way experimental results

Subsequently, the experiments were split according to the three-way experimental setup presented in Table 4.1. The following three categories, which correspond to cases AC, AD and BCD respectively, were used to a) assess how the diversity in the training dataset affects estimation, and b) investigate the effect of having a good or bad fit between the training and test sets.

1. **Homogeneous to correct (AC):** Tailored model to a specific population, assuming a good fit between model and test set. Aggregate results obtained by fitting the “young” and “old” models to the “young” and “old” test sets using LOO cross-validation
2. **Homogeneous to wrong (AD):** Tailored model trained from a homogeneous population, but assuming a low fit between model and test set. Aggregate results obtained by fitting the “young” and “old” models to the “old” and “young” test sets, using independent validation (the training and test sets are by definition separate in this case).
3. **Heterogeneous (combined) to all (BCD):** Heterogeneous model targeted at a broader population, with higher diversity in the training set. Aggregate results obtained by fitting the “combined” model to the “young” and “old” tests. A combination of independent and LOO cross-validation is used, as defined in Table 4.4. Briefly, recall that the model was built using 30 “old” and 30 “young” femurs, chosen using stratified partitioning. Thus, half of the “young” and “old” test femurs which were not part of the training population were reconstructed using independent testing, whereas LOO cross-validation was used for the remaining femurs.

All error statistics of the 20- and 4-view experiments are collated in Table 4.8. In addition, results are illustrated in the usual ways in Figures 4.16–4.21. In order of appearance, these plot separately the following for each configuration: a summary of the mean, standard deviation and ROE errors (Figure 4.16), the ROE as a function of the initialisation ROE and the DRR orientation (Figure 4.16b) and the thickness estimation accuracy as a function of thickness (Figure 4.17). In addition, the following colour maps are included: mean thickness (Figure 4.17), mean error (Figure 4.19),

4. MODEL-BASED CORTICAL THICKNESS ESTIMATION

(a) mean error, standard deviation of error and ROE for cases AC, AD, BCD

(b) ROE error per DRR view for cases AC, AD, BCD

Figure 4.16: (a) Thickness estimation accuracy (x-axis), precision (y-axis) and ROE (size of marker) for the homogeneous and heterogeneous models, when tested on different populations. (b) ROE per DRR view. In both figures, the darker markers correspond to the blind estimators (initialisation), whereas the brighter to the errors obtained after optimising 6 shape and 10 thickness PCs. The colors of the markers identify the experimental setup and are explained in the legend. The “homogeneous to correct” errors are obtained by averaging the results of fitting the “young” model to the “young” test set, and the “old” model to the “old” test set. In contrast, the “homogeneous to wrong” ones are obtained when fitting the “young” model to the “old” dataset and vice-versa. Finally, the “heterogeneous” values were calculated after fitting the “combined” model to all 120 “young” and “old” specimens. All plotted values are collated in Table 4.8.

4. MODEL-BASED CORTICAL THICKNESS ESTIMATION

Figure 4.17: DXA thickness against the bronze-standard, after fitting to 20 and 4 DXA views the mean (a,b,c,g,h,i) and optimised (d,e,f,j,k,l) homogeneous and heterogeneous models, according to the experimental scheme described in Table 4.1. The first column corresponds to case AC, the second to AD and the third to BCD. The optimiser was allowed to deform the mean model using 6 shape and 10 thickness PCs.

4. MODEL-BASED CORTICAL THICKNESS ESTIMATION

Figure 4.18: Bronze-standard mean cortical thickness across all examined femurs (a), mean DXA thickness for the homogeneous (second and third row) and heterogeneous (fourth row) models for 20- (left column) and 4-views (right column).

4. MODEL-BASED CORTICAL THICKNESS ESTIMATION

Figure 4.19: Mean cortical thickness error across all examined femurs.

standard deviation of the error (Figure 4.20) and RMS change compared to the blind estimator (Figure 4.21).

Figure 4.20: Standard deviation error per vertex location across all examined femurs.

4.7 Discussion

Estimation errors and their interpretation

The 20-view results in Figure 4.16 and Table 4.8 establish the limits of performance of the DXA statistical method. Approximately 0.4 mm of the error may be originating from the uncertainty of the bronze standard CT measurements, while the remainder can be reasonably ascribed to the necessary constraints of our approach.

An important observation relates to the way errors should be reported and interpreted. As revisited in the next chapter, a common observation in the relevant literature is the presentation of reconstruction accuracies only in absolute terms, and without a

4. MODEL-BASED CORTICAL THICKNESS ESTIMATION

Figure 4.21: Change of RMS error per vertex location (negative represents improvement) across all examined femurs. Note that (a–d) are compared against the homogeneous mean models, whereas (e) and (f) against the heterogeneous model.

4. MODEL-BASED CORTICAL THICKNESS ESTIMATION

reference to a comparative baseline. Judging from the present results, stating that the estimation efficacy is 0.01 ± 0.54 mm (sub-millimeter), without a reference to the blind estimator errors, might incorrectly be interpreted as a significant accomplishment. Knowing that the blind estimator has an efficacy of 0.03 ± 0.60 mm provides a completely different picture.

The colour maps verify that aggregating estimation performance into a few statistical error values may be misleading, and most likely not the best way to evaluate results. Figure 4.21, which depicts if the RMS error improves or deteriorates per vertex location, identifies the femoral regions where estimation is improved. In particular, and as expected, thicker cortices tend to be predicted best, whereas problems arise when measuring the medial aspect of the lesser trochanter. As mentioned before, due to the convoluted cross-sectional shape of this region, multiple layers of cortex are frequently superimposed in the projections. Hence, multiple solutions might similarly explain the underlying data, as the optimiser may trade thicknesses between different layers. Otherwise, in the vast majority of the femoral surface the optimiser is unable to reduce the error, perfectly illustrating the difficulty of the task at hand: the thickness variation amongst all femurs is very small and the blind estimator manages to perform remarkably well, leaving a very small margin of measurable improvement.

As far as the ROE is concerned, the results obtained are in line with what was expected: as seen in Figure 4.15 and Tables 4.6 and 4.7, the optimiser manages to further reduce it when provided with more PCs, except for very few cases (especially in the 4-view configuration). The latter is observed when the dimensionality of the problem is relatively high compared to the amount of observable information, which results in the search function exploiting the parameter space in the wrong direction. These observations lead to the following two conclusions: a) The DXA cost function does not always agree with its CT counterpart, and b) the poor 4-view results are generally not a consequence of the optimiser not sufficiently exploiting the parameter space (i.e. getting trapped into a local minimum in the very flat energy landscape), but rather the outcome of wrongly interpolating the sparse data due to inadequate information. If a solution to these problems existed, then there would be an ≈ 0.2 mm RMS improvement margin, as the upper-bound performance limit when using 6 shape and 10 thickness PCs is approximately 0.30 mm RMS (Section 4.6.2).

Homogeneous vs. heterogeneous models

As expected, the experimental results demonstrate that exploiting the anatomical variations within the test sets leads to an improved performance. Although likely, it is

4. MODEL-BASED CORTICAL THICKNESS ESTIMATION

unclear whether this improvement can be completely attributed to the smaller errors of the tailored blind estimators, or whether some of the amelioration comes from the tailored PCs.

This observation touches on a point discussed in the literature review of model based reconstructions (Section 2.6). Reported reconstruction errors found in the literature should be interpreted with caution, as they might not reveal the full truth about their applicability. If the training/test sets are obtained from a homogeneous population, it is likely that the errors will not generalise well when applied to more diverse data. Relevant studies should always report the diversity between the test and training sets, and the fit between the two populations, something which is frequently omitted (Section 2.6).

Number of principal components

The use of a SSM and a STM is primarily justified by the desire to reduce the dimensionality of the problem. Especially when using only four views, the observable information guiding the optimiser is very sparse. Thus, to prevent over-fitting, it is necessary to trade some of the model complexity for more assumptions.

For all models examined herein, HPA determined that up to six shape and ten thickness PCs can be retained. However, it is possible that these are too many degrees of freedom, principally when using only four views. This is why Tables 4.6 and 4.7 present the errors when using a variable number of PCs, up to the aforementioned limit.

Yet, it was deduced that error statistics between them were marginally different. Having a greater model flexibility did not translate to better or worse thickness estimation, regardless of whether the optimiser was able to find slightly better solutions in terms of the ROE. Therefore, all subsequent experiments were performed using six shape and ten thickness PCs for consistency reasons.

Semi-manual initialisation of model position

Section 4.2 explains the process of creating the SSMs. The first step involves an automatic registration of a canonical femur to each training sample by means of a similarity transformation, followed by a Locally Affine Deformation (LAD). PCA is then performed on the resulting deformation fields, which are all expressed in the canonical morphology. Thus, by definition, all PCs are expressed relative to the mean model, and proper reconstructions rely on its correct positional initialisation.

For this reason, the statistical method is vulnerable to a poor semi-manual initialisation. Even if the optimiser deduces the correct weights for all PC, if the mean model's starting position is not exactly the same as the one deduced by the similarity transfor-

4. MODEL-BASED CORTICAL THICKNESS ESTIMATION

mation during the training phase, then the nearest neighbour algorithm between CT estimates and DXA estimates is going to produce mismatches. Above all, if the semi-manual initialisation is far from correct, then the PCs will deform the model towards not statistically derived directions.

A possible solution to this problem would be to allow the optimiser to modify the seven parameters which define the spatial initialisation of the model (three for translation, three for rotation and one for scale). However, this would greatly increase the problem dimensionality, and would call for additional regularisation strategies, especially when one considers the very limited amount of available information from the DRR scans.

Yet, before trying to do so, it is important to quantify the potential performance gain. Table 4.9 presents the cortical thickness estimation efficacy of the statistical method when, instead of performing a semi-manual model initialisation, the proper registration matrices are used. The results show that the differences are minute, and not always positive. Therefore, we restrained from extending the optimisation parameters to include the initialisation of the model.

Table 4.9: **20-view and 4-view**, averaged DXA cortical thickness estimation errors when fitting the homogeneous and heterogeneous models to different test sets, according to the cases defined in Table 4.1. These results were obtained by initialising the position of the model using the proper registration matrices (i.e. the ones used during the training of the models), and not the semi-manual initialisation. The first two rows of each section correspond to the blind estimator errors, whereas the next two are obtained after optimising 10 thickness and 6 shape modes. The values in red (second and fourth row of each section) indicate how these results compare to the ones obtained when semi-manually initialising the position.

modes		mean error (mm)		std. deviation $\pm$ (mm)		RMS (mm)		percentage error (%)		ROE (%)	
sh.	th.	20 views	4 views	20 views	4 views	20 views	4 views	20 views	4 views	20 views	4 views
HOMOGENEOUS TO CORRECT (case AC)											
0	0	0.01	0.01	0.58	0.58	0.58	0.58	36.2	36.2	93.5	92.2
		-0.03	-0.03	0.00	0.00	0.00	0.00	-1.2	-1.3	-0.5	-2.0
6	10	0.01	0.01	0.50	0.53	0.50	0.53	30.5	31.7	64.5	65.4
		0.00	0.00	0.00	0.01	0.00	0.01	0.4	0.5	0.1	-0.2
HOMOGENEOUS TO WRONG (case AD)											
0	0	0.02	0.02	0.63	0.63	0.63	0.63	41.1	41.1	100.0 *	100.0 *
		-0.02	-0.02	-0.01	-0.01	-0.01	-0.02	-2.0	-2.1	0.0	0.0
6	10	0.01	0.01	0.52	0.57	0.52	0.57	33.2	36.2	66.4	68.0
		0.00	0.01	0.00	0.02	0.00	0.02	1.8	1.8	-0.6	1.5
HETEROGENEOUS TO ALL (case BCD)											
0	0	0.01	0.01	0.59	0.59	0.59	0.59	38.7	38.7	94.2	94
		-0.02	-0.02	0.00	-0.01	-0.01	-0.01	-0.3	-0.5	-1.1	-2.2
6	10	-0.03	0.02	0.53	0.54	0.53	0.54	30.6	31.5	63.1	64.6
		-0.01	0.01	0.01	0.00	0.01	0.00	0.4	-0.1	-1.6	-0.4

* baseline ROE

4. MODEL-BASED CORTICAL THICKNESS ESTIMATION

Optimiser variants

The performances of the following optimiser variants were investigated:

1. Optimising both the shape and thickness PCs simultaneously (results presented in Section 4.6.3)
2. Optimising first the size (1st shape PC), and then shape and thickness simultaneously
3. Optimising first all shape PCs and then shape and thickness simultaneously

All three scenarios performed almost identically to each other leading to, apart from very similar error statistics, very similar solutions in terms of the ROE. Case 1 was significantly faster than its alternatives and thus preferred. For brevity, the performance characteristics of cases 2 and 3 are omitted in this report.

Number of views

Although shape reconstruction studies have shown that adding a second view significantly improves reconstruction accuracy but adding further has little effect [143], we found that four views performed better than two views. In fact, when using only two views estimation efficacy deteriorated when compared to the blind estimators.

Computation time

Reconstruction times were significantly longer compared to the template method. They ranged from a couple of seconds per femur (4 views), up to *circa* 10 minutes (20 views). The primary reason was the necessity to calculate the intersection cross-sections between the planes of the APPs and the model endocortical and periosteal surfaces (15 APPs $\times$ number of views), which was computationally expensive. It is worth noting that this was the case despite performing the above step, and calculating each SAP independently in different threads (thus using both hyper-threaded CPU cores).

Chapter 5

Conclusions and future work

5.1 Comparison of Template and Statistical methods

The following two sections discuss the obstacles which prevent a straightforward comparison between the template and the statistical methods, together with a presentation of the proposed solutions. Subsequently, the error statistics of the algorithms are expressed in a common framework, compared and discussed.

Region of estimation

The femoral regions examined by each of the two methods differ (see Figures 5.1b and 5.1c). Specifically, the template method ignores the vast majority of the femoral neck (except its distal part) and the proximal part of the greater trochanter. On the other hand, the statistical method includes all of the femoral neck and greater trochanter. Differences exist also on the distal part of the proximal femur, where the template experiments include some of the femoral shaft, whereas the statistical method extends to just below the lesser trochanter.

This difference is expected to result in an unfair direct comparison, as one of the prevalent regions of high thickness and variability is the proximal part of the femoral shaft. The opposite holds for the femoral neck. For an illustration of the above points, please refer to the cortical thickness distribution plots (Figures 3.30d and 4.10l for the template and statistical methods respectively) and the blind estimator's error colour maps (Figures 3.34 and 4.11f).

To allow for a fair comparison, two masks were created — one for the template method and one for the statistical — which spatially filtered the results. In that way, the error statistics were calculated by examining only the regions common to both methods.

Figure 5.1: In (a), the template method's canonical morphology (red) is registered to that of the statistical method's (green) by means of a similarity transformation followed by a LAD. This establishes a mapping between the vertices of the two surfaces, which allows the identification of the femoral regions for which estimation is performed using both methods. In (b) and (c), the common regions of evaluation are displayed in blue, whereas the regions unique to the statistical and template experiments respectively are shown in red.

To do so, the mean template shape was registered to the mean statistical model using a similarity transformation followed by a LAD (Figure 5.1a), and the masks were created according to the established vertex correspondences.

Thickness measurement direction

A straightforward comparison is still not possible even after spatially filtering the results, since each method defines thickness in a different direction. For the reasons explained in Section 3.5.3, the template method defines thickness on planes which are perpendicular to the DRR axis of rotation (by sampling the B-spline cross-sections). On the other hand, the statistical method measures thickness along the vertex normals (Section 4.3.1). Thus, they should be first projected along the same direction.

Estimating the template's results along the vertex normals would require an interpolation between successive B-spline cross-sections, or an algorithm to nonlinearly deform the mean template according to the optimised B-spline contours. On the other hand, projecting the statistical method's estimates to the correct direction can be simply achieved using the algorithm presented in Section 3.5.3 and Figure 3.10. Briefly, the intersections between the endocortical and periosteal surfaces of the deformed statistical model and the planes of the APPs are calculated, and the resulting cross-sections are sampled in the required directions.

Comparison

Table 5.1 summarises the aggregated results of both methods: *global* errors are obtained before spatially filtering the results, *masked* errors are calculated only in the common regions of estimation, and *masked and projected* errors are also expressed along the same directions. In the latter case, the mean bronze-standard thickness is 2.18 mm for the template method which, as expected, is directly comparable to its 2.24 mm equivalent from the statistical approach. This discrepancy is not surprising, since, despite only averaging CT measurements lying on the common regions, the sampled cross-sections are unique to each method.

Measurement bias differences were evaluated using a two-tailed, two-sample *t*-test with unequal variances. Precision differences were tested using an *F*-test. Each approach was compared against its corresponding blind estimator and against all other methods. Significance was set at the conservative level of $p < 0.001$ to account for multiple comparisons. Strong significance ($p < 0.00001$) was obtained in all comparisons, apart from the bias difference between the template and statistical (homogeneous)

methods ($p = 0.0687$) when reconstructing from twenty views, and the bias difference between the homogeneous and heterogeneous experiments ($p = 0.0011$) when using four views.

In the light of these results, the following two conclusions can be made. First, both algorithms significantly improve both the accuracy and precision of their initialisations. Second, it is concluded that the template method is more precise than the statistical, irrespective of whether a homogeneous or a heterogeneous model is used. It is also shown that using a tailored model significantly improves estimation efficacy. However, it is worth noting the differences are very small — statistical significance is obtained due to the very high number of point thickness estimates compared.

One advantage of the template method, which might partially explain the above observations, is that it is free to deform any B-spline control point independently of all others. On the contrary, the statistical method’s objective is to find a combination of weighted PCs which best fits the data. A training cohort of 60 femurs is not guaranteed to capture the correct principal modes of variation of a broad target audience. A model built from a large set of femurs might reveal that capturing most of the intra subject variance requires an excessive number of PCs, thus rendering the reconstruction problem highly dimensional.

Furthermore, the statistical method is vulnerable to a poor semi-manual initialisation: the PCs rely on a proper fit of the mean model since they are expressed relative to it. This problem is exacerbated in the clinically relevant experiments, as fitting the model to just four DRRs is susceptible to errors. The template method can cope better with a poor initialisation, since each cross-section is independently deformed to fit the data.

Finally, one factor influencing the results is intra-operator variability, as the initialisation of both algorithms is performed using the semi-manual, silhouette-based registration. For example, this is why the blind estimator errors are not always identical between the twenty- and four-view experiments: initialisation in the latter case is more prone to errors.

Table 5.1: Comparison of template and statistical methods before (*global*) and after (*masked*) filtering the results to include only estimates in the common regions (blue areas in Figure 5.1). In addition, in the case of the statistical methods, the results are also tabulated as measured in the template DXA direction (*masked & projected*), to allow for a direct comparison with the template method. The column title abbreviations stand for: mean $\pm$ standard deviation thickness of bronze standard (CT), mean thickness of DXA method (DXA), mean thickness error (MEAN), standard deviation of error (STD), root mean squared error (RMS) and percentage error (PER).

HETEROGENEOUS MODEL		CT	20 VIEWS					4 VIEWS				
			DXA	MEAN	STD	RMS	PERC	DXA	MEAN	STD	RMS	PERC
global	blind est.	1.63 ± 1.02	1.66	0.03	± 0.59	0.60	39.0	1.66	0.03	± 0.60	0.60	39.2
	optimised	1.63 ± 1.02	1.61	-0.02	± 0.52	0.52	30.2	1.64	0.01	± 0.54	0.54	31.6
masked	blind est.	1.90 ± 1.09	1.93	0.03	± 0.61	0.61	32.0	1.91	0.01	± 0.64	0.64	33.5
	optimised	1.90 ± 1.09	1.89	-0.01	± 0.55	0.55	29.1	1.93	0.03	± 0.59	0.60	31.5
masked & projected	blind est.	2.24 ± 1.15	2.26	0.02	± 0.68	0.68	26.8	2.24	0.00	± 0.69	0.69	26.8
	optimised	2.24 ± 1.15	2.23	-0.01	± 0.62	0.62	21.2	2.24	0.00	± 0.62	0.62	21.8
HOMOGENEOUS MODEL												
global	blind est.	1.63 ± 1.02	1.67	0.04	± 0.58	0.58	37.4	1.67	0.04	± 0.58	0.58	37.5
	optimised	1.63 ± 1.02	1.64	0.01	± 0.50	0.50	30.1	1.64	0.01	± 0.52	0.52	31.2
masked	blind est.	1.90 ± 1.09	1.92	0.02	± 0.60	0.60	31.6	1.92	0.02	± 0.60	0.60	31.6
	optimised	1.90 ± 1.09	1.96	0.06	± 0.53	0.54	27.6	1.92	0.02	± 0.57	0.58	30.2
masked & projected	blind est.	2.24 ± 1.15	2.25	0.03	± 0.65	0.65	25.1	2.27	0.01	± 0.67	0.67	25.8
	optimised	2.24 ± 1.15	2.24	0.00	± 0.59	0.59	20.6	2.23	-0.01	± 0.61	0.61	21.0
TEMPLATE												
global	blind est.	2.46 ± 1.28	2.38	-0.08	± 0.77	0.77	29.3	2.32	-0.14	± 0.79	0.80	30.7
	optimised	2.46 ± 1.28	2.47	0.01	± 0.59	0.59	22.2	2.41	-0.05	± 0.67	0.67	27.4
masked	blind est.	2.18 ± 1.12	2.13	-0.05	± 0.65	0.65	27.0	2.11	-0.07	± 0.68	0.68	28.4
	optimised	2.18 ± 1.12	2.18	0.00	± 0.50	0.50	20.1	2.13	-0.05	± 0.54	0.55	23.2

5.2 Conclusions

The present work’s objective is to investigate the extent to which cortical thickness of the proximal femur can be measured from multi-view DXA scans. Strong evidence suggests that focal structural defects, such as cortical bone thinning, are a critical component in characterizing hip fragility. However, current risk assessment tools ignore them. Considering that hip fracture is the leading cause of acute orthopaedic hospital admission amongst the elderly, it is of paramount importance to develop new tools which can assist in patient assessment.

Cortical thickness can be measured using CT, but this is expensive and involves a significant radiation dose. The suggested methods herein work with DXA scans, the current standard in clinical practice. In that respect, the following sections present the contributions of this work, the primary challenges faced and some noteworthy observations relating to the relevant literature.

5.2.1 Contributions and main findings

Two alternative techniques, capable of extracting thousands of localised cortical thickness estimates across the proximal femur from multi-view DXA scans, were proposed, implemented and tested. To the author’s best knowledge, such algorithms have never been proposed in the past. Both were realised as an extension to the medical imaging software Stradwin. They were assessed on DRRs derived from CT data of 120 females as a surrogate for DXA scans: this allows their direct evaluation against the current state-of-the-art in cortical thickness estimation which works with CT data.

Template method

The first method relies on a data-driven optimiser which deforms a femoral template to fit an individual’s DXA scans, and operates in the trochanteric regions and the proximal part of the femoral shaft. The femur is modeled using a set of B-spline cross-sections which can be deformed independently of each other.

In a series of experiments involving 120 femurs, estimation errors were 0.01 ± 0.59 mm and -0.05 ± 0.67 mm for the validation (twenty DXA scans) and clinical (four DXA scans) experiments respectively. The blind estimator’s error was -0.08 ± 0.77 mm, which grants the algorithm a RMS error improvement of 0.18 mm and 0.11 mm respectively. The theoretical upper-bound performances, as derived from the models’ complexity (twenty and twelve control-point splines respectively), were -0.02 ± 0.26 mm and -0.03 ± 0.41 mm.

Statistical method

The second algorithm achieves the same task by optimising simultaneously a statistical shape and a cortical thickness model. It operates on the whole proximal femur, excluding the femoral head and extending to just below the lesser trochanter. We assessed the estimation efficacy of three separate models, trained using sixty femurs each, to evaluate (a) the effect of the structural diversity in the training set and (b) the possibility of improving performance by building a tailored model to a particular population. (a) was achieved by building a homogeneous and a heterogeneous model (selection was based on age, 18.4 years mean age difference between the groups), and (b) by fitting the homogeneous model to two populations: one akin to the training set, and one dissimilar. Six shape and ten thickness PCs were used throughout, according to HPA.

In a series of cross-validation experiments, estimation was best in terms of precision when the homogeneous model was fit to a population akin to the training set (0.01 ± 0.50 mm and 0.01 ± 0.52 mm for the twenty- and four-view experiments respectively), followed by the heterogeneous model (-0.02 ± 0.52 mm and 0.01 ± 0.54 mm). Estimation was worse when the homogeneous model was used to reconstruct a population dissimilar to the training set (0.01 ± 0.52 mm and 0.00 ± 0.55 mm). The corresponding blind estimator errors were: 0.04 ± 0.58 mm (homogeneous to similar), 0.03 ± 0.60 mm (heterogeneous) and 0.04 ± 0.64 mm (homogeneous to dissimilar). Thus, the algorithm achieved a RMS error improvement of 0.08 mm/0.06 mm, 0.08 mm/0.04 mm, and 0.12 mm/0.10 mm respectively. The upper performance bound based on the model complexity was calculated as 0.04 ± 0.29 mm.

Two-sample, two-tailed t -tests and F -tests revealed that both algorithms are significantly better than their blind estimator counterparts, both for the twenty-, and the four-view experiments. Similar tests show that the template method is significantly more precise than the statistical method, irrespective of which model is used. Finally the homogeneous model outperforms the heterogeneous. Note that the above errors for the two methods are not directly comparable as they are expressed in different thickness directions and correspond to distinct femoral regions. Please refer to Section 5.1 for a direct comparison.

A further contribution of this study relates to the presentation of the error statistics of both techniques using colour bone mapping. This provides a much better insight to the performance characteristics, since estimation can be examined as a function of location. It also reveals that assessing the algorithms based on the aggregated statistics alone

is not ideal, as the colour maps reveal regions where estimation is both improved and worsened compared to the blind estimator. Specifically, the following observations are made: estimation is best for thicker cortices, it is particularly troublesome in the medial aspect of the lesser trochanter where the cross-section shape is highly asymmetric, and remains unaltered (compared to the initialisation) for the vast majority of the remaining femur.

5.2.2 Challenges and noteworthy observations

A brief summary of the principal challenges faced during the course of this study follows. Where relevant, a short comment describes how other researchers have dealt with them in the past.

- The first challenge relates to the definition of cortex per se. Cortical porosity renders the endocortical surface ambiguous, if not impossible to uniquely identify, especially in femurs of the elderly. For this reason, some researchers have even proposed to define separately a transitional zone between the cortex and the trabeculum [169]. For good measure, defining the cortical boundaries when imaging in clinical resolution is even harder. Under the presence of imaging blur, thin, dense cortices appear almost indistinguishable from less dense, thicker ones. Sub-millimeter cortices may completely disappear. This has an effect on the results of this thesis: the bronze standard CT estimates are derived from clinical resolution CT scans, and hence some of the error in the techniques proposed herein may be attributed to its uncertainty. Additionally, defining the direction along which thickness should be measured is a non-trivial task and may not produce consistent results. Although measuring along the surface vertex normals is a convenient approach, it is possible for a line along these directions to never cross the endocortical surface.
- The second challenge relates to the proper presentation and interpretation of the results. As explained previously, resorting to a single value that encompasses the global performance characteristics might be misleading. In addition, it is imperative to provide a reference against which the results should be assessed. This can either be in the form of a naive estimator, such as the population mean, or at least by stating the variance of the test set. It is remarkable that from the relevant studies discussed in Section 2.6, only Kurazume et al. [88] compare against the blind estimator, and Gamage et al. [51] report the intra-test set variance. All others provide no reference whatsoever [1; 70; 71; 84; 91; 93; 94; 152; 164; 165]. It

is incorrect to claim good reconstruction accuracies without providing a baseline, since the blind estimator’s performance might be unexpectedly good, especially when examining a small, homogeneous population.

- A problem almost always mentioned in relevant studies is the limited amount of data available for evaluation and model construction. This leads to many adverse effects. First, examining a small population may induce results with high uncertainty margins. Second, as already said, there is a high chance of the intra-set anatomical diversity being small: if that is the case, results should be interpreted with caution, as they might not generalise properly. There have been numerous studies which reveal significant anatomical and structural discrepancies between people of different gender [27; 73; 139; 147], race [4; 105; 113; 114] and age [32; 73; 74; 107; 139; 147; 153] (for example in terms of geometry/size [105; 114; 147], BMD [73; 76; 105; 114; 147; 153] and cortical thickness [73; 105; 114; 147; 153]). Third, reconstructions employing statistical models are particularly prone to bias for two reasons. On one hand, if the training set is small it might not capture properly the statistics of the target population leading to poor results. On the other hand, if the training and test sets are drawn from a homogeneous population, there is a high chance of observing low, but unrealistic, errors. The authors should quantify the fit between the training and test sets, for the readers to critically assess how well the algorithms would generalise to unseen data. Once more, this is rarely the case in the studies examined herein [1; 70; 71; 88; 152; 164; 165] (see Section 2.6).
- Finally, the present work proves that measuring the cortex of the proximal femur from multi-view DXA scans is an extremely hard task, especially when a small number of projections are used. Many factors contribute to this conclusion. First, as examined in Section 2.3.3, there are large gaps in the Fourier space that are difficult to infer. In other words, the optimiser is faced with a very flat energy landscape which contains many ambiguous solutions that explain the data to the same degree. Second, a problem unique to this approach (and not present in algorithms operating on CT data) is the thickness and density trade-offs between the cortical and trabecular compartments. Projections of the femur superimpose these two structures, and it was deduced that arriving at a single plausible solution requires a regularisation of a) the cortical density, b) the imaging blur and c) the maximum deviation from the blind estimator. Third, the relatively small intra-subject variance renders the blind estimator a good enough fit over much of the proximal femur, allowing a very small margin of improvement only in the thicker

cortical regions.

5.2.3 Clinical usefulness

The clinical usefulness of the proposed algorithms should be examined in relation to individual diagnosis or cohort studies. Possible applications of the latter include monitoring change in response to treatment, or comparing groups for structural differences.

As far as individual diagnosis is concerned, to detect focal cortical thinning the accuracy of the reconstruction algorithm should be less than the variability in the population. In our experiments, the variability in the population can be directly inferred from the errors of the blind estimator. This is because, by definition, the blind estimator assumes that all test specimens can be modelled by the mean shape and cortical distribution obtained by “averaging” a large number of femurs. Our experiments show that both the statistical and template methods significantly improve upon its estimation efficacy. However, before reaching a positive conclusion about the usefulness of these algorithms with respect to individual diagnosis, one should consider that the variability in fracture prevalent regions, i.e. regions where the cortex is particularly thin, is way less than the accuracy of either proposed methods. The primary reason for which they outperform the blind estimator is because they are capable of correctly measuring thicker cortices. Hence, we believe that the proposed techniques are unlikely to assist in individual diagnosis.

On the other hand, cohort studies benefit from the improved statistics obtained by examining a large number of femurs. In that way, cortical thickness differences between groups which are below the accuracy of the algorithms can be established with statistical significance. Estimation error statistics can be also improved by examining the overall cortical thickness of small patches, by averaging neighbouring point thickness estimates. This technique might be particularly useful when trying to detect cortical thinning in fracture prone regions.

5.3 Future work

- In the light of our results, measuring the cortex of the femoral neck and trochanteric regions from multi-view DXA scans is unlikely to assist in fracture risk assessment. However, the proposed algorithms might find applicability in the sub-trochanteric regions. The cortex is much thicker and remains unambiguous under the presence of imaging blur. Moreover, the cross-sectional shape is relatively symmetric and less convoluted, and thus it might be possible to accurately reconstruct it from

a small number of DXA views. It would be very interesting to examine whether such measurements can predict the ever-increasing occurrences of atypical fractures: these are thought to be associated with cortical thickening in the shaft region in response to pharmaceutical treatments [55; 98; 144; 145].

- To that end, the template method can be improved by employing surface models instead of a set of coplanar B-spline contours. This would require the development of more advanced algorithms evaluating the similarity between DXA scans and template DRRs in two dimensions. Doing so would feed the optimiser with more information. However, even if the optimisation remains driven by the similarity between APPs and SAPs, the added benefit of constraining the shape of neighbouring contours might reduce the amount of required regularisation.
- According to the conclusions of this thesis, a statistical method is less likely to help. Nevertheless, if one still wanted to investigate its performance on the femoral shaft, the following suggestions are made. First, models should be trained from larger populations: this might reveal that capturing enough of the variance requires a large number of PCs, thus rendering the problem highly complex. Second, model homogeneity and the effects of building tailored models should be further evaluated using populations of different ethnic, age and gender groups.
- These algorithms could be compared against an alternative technique based on the relatively novel field of compressed sensing. This might be possible provided that a suitable basis can be conceptualised, whereby the modeling of the femoral cortex can be performed using sparse components.
- Finally, it is suggested that some previously published model-based reconstruction methods should be re-evaluated, particularly those claiming good performance from a single DXA projection. It would be interesting to see whether they improve significantly on the blind estimator, and to quantify any performance gain.

Appendices

A Projection Theorem (aka. central-slice theorem)

The projection theorem states that “the Fourier transform of the projection of an N -dimensional function $f(\mathbf{r})$ onto an M -dimensional linear submanifold is equal to an M -dimensional slice of the N -dimensional Fourier transform of that function, consisting of an M -dimensional linear submanifold through the origin in the Fourier space which is parallel to the projection submanifold” [151; 167] (see Figure A.1 for a graphical illustration of the definition in two-dimensions). Hence, the inverse radon transform can be evaluated by simply calculating the inverse N -dimensional Fourier transform of the M -dimensional Fourier transform of the projections.

The mathematical proof for a 2D function $f(\mathbf{r})$ is straightforward, and can be easily generalised for higher dimensions. Without loss of generality, we work with the projection of $f(\mathbf{r})$ which is parallel to the x -axis [167]:

$$p(x) = \int_{-\infty}^{\infty} f(x, y) dy \quad (.1)$$

The Fourier transform of $f(x, y)$ is

$$F(k_x, k_y) = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} f(x, y) e^{-2\pi i(xk_x + yk_y)} dx dy \quad (.2)$$

The slice $s(k_x)$, passing through the origin of the Fourier space and parallel to the direction of projection is defined by

$$\begin{aligned} s(k_x) &= F(k_x, 0) \\ &= \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} f(x, y) e^{-2\pi i x k_x} dx dy \\ &= \int_{-\infty}^{\infty} \left[\int_{-\infty}^{\infty} f(x, y) dy \right] e^{-2\pi i x k_x} dx \\ &= \int_{-\infty}^{\infty} p(x) e^{-2\pi i x k_x} dx \end{aligned} \quad (.3)$$

which by definition is the 1-dimensional Fourier transform of $p(x)$, as required.

Figure A.1: Illustration of the projection theorem in two-dimensions. $F(k_x, k_y)$ is the 2D Fourier transform of $f(x, y)$. $p(x)$ is the 1D projection of $f(x, y)$ parallel to the x -axis. $s(k_x)$ is a 1D slice of $F(k_x, k_y)$, passing through the origin of the Fourier space and perpendicular to the projection direction. The projection theorem states that $s(k_x)$ is the 1D Fourier transform of $p(x)$.

B Digitally Reconstructed Radiographs (DRRs)

We constructed DRRs from MDCT as a surrogate for DXA, as no femoral dataset scanned using both techniques was available during the course of this study. This approach is favoured by most researchers faced with the same problem [104]. The synthesised DXAs were constructed by performing an average compounding of all voxels between two parallel reslice planes through each femoral MDCT scan, in a direction perpendicular to them (see Figure B.1). DXAs differ from DRRs in their ability to subtract the contribution of any soft tissue by using two different X-ray energy beams. Hence, in an effort to minimize the discrepancy between them, all background voxels of the MDCT scans within the “compounded” region were replaced by their collective average value. Voxels were classified as background if they lay more than 3mm from the segmented femoral surface, first to account for segmentation errors, and second to allow for blurred cortices.

Figure B.1: a) Each pair of green planes corresponds to a different DRR. Average compounding of voxels for the pair shown in red is performed in the direction of the arrow. All background voxels between the planes are replaced by a constant value, to better simulate DXA scans which subtract the contribution of soft tissue.

b) An example of a typical DRR: this is a 0° (standard DXA) view — notice how the lesser trochanter is barely visible.

The orientation of all DRRs was specified using their relative rotation compared to the routinely performed AP scan, which is the optimal scanning direction. Its proper acquisition requires the patient to fully flex the leg straight on a table, so that the shaft is parallel to the vertical axis of the picture. Moreover, a $15^\circ - 25^\circ$ internal rotation presents the long axis of the femoral neck perpendicular to the X-ray beam, resulting in the greatest area. Usually, an optimally imaged femur is associated with the lesser

trochanter being on the verge of unobstructed visibility [45; 97; 115]. Care was taken to synthesise the DRRs according to this specification.

References

- [1] OM AHMAD, KR, KE WILSON, K ENGELKE, RL PRINCE, AND RH TAYLOR. Volumetric DXA (VXA): A new method to extract 3D information from multiple in vivo DXA images. *J. Bone Miner. Res.*, **25**[12]:2744–51, 2010. [3](#), [12](#), [13](#), [14](#), [28](#), [37](#), [41](#), [62](#), [148](#), [149](#)
- [2] OM AHMAD, K RAMAMURTHI, KE WILSON, K ENGELKE, M BOUXSEIN, AND RH TAYLOR. 3D structural measurements of the proximal femur from 2D DXA images using a statistical atlas. **7260**. SPIE, 2009. [1](#), [12](#), [13](#), [14](#), [31](#), [41](#)
- [3] AR BAGGALEY. Deciding on the ratio of the number of subjects to number of variables in factor analysis. *Multivariate Experimental Clinical Research*, **6**:81–85, 1982. [110](#)
- [4] E BARRETT-CONNOR, ES SIRIS, LE WEHREN, PD MILLER, TA ABBOTT, ML BERGER, AC SANTORA, AND LM SHERWOOD. Osteoporosis and fracture risk in women of different ethnic groups. *Journal of Bone and Mineral Research*, **20**[2]:185–194, 2005. [40](#), [149](#)
- [5] TJ BECK. Extending DXA beyond bone mineral density: Understanding Hip Structure Analysis. *Current Osteoporosis Reports*, **5**[2]:49–55, June 2007. [3](#), [4](#), [12](#)
- [6] TJ BECK. Hip Structure Analysis: Using DXA scans to measure bone strength, presentation of the HSA method, implemented in most Hologic DXA scanners. Technical report, Division of Medical Imaging Physics, Department of Radiology, The John Hopkins University School of Medicine, 2007. [42](#)
- [7] TJ BECK, AC LOOKER, CB RUFF, H SIEVANEN, AND HW WAHNER. Structural trends in the aging femoral neck and proximal shaft: Analysis of the Third National Health and Nutrition Examination Survey Dual-energy X-ray Absorptiometry data. *J. Bone Miner. Res.*, **15**[12]:2297–2304, 2000. [3](#), [42](#)

REFERENCES

- [8] TJ BECK, TL ORESKOVIC, KL STONE, CB RUFF, K ENSRUD, MC NEVITT, HK GENANT, AND SR CUMMINGS. Structural adaptation to changing skeletal load in the progression toward hip fragility: The study of osteoporotic fractures. *Journal of Bone and Mineral Research*, **16**[6]:1108–1119, 2001. [3](#), [42](#)
- [9] C BERGOT, V BOUSSON, A MEUNIER, M LAVAL-JEANTET, AND JD LAREDO. Hip fracture risk and proximal femur geometry from DXA scans. *Osteoporosis International*, **13**[7]:542–50, 2002. [3](#)
- [10] M BHANDARI, PJ DEVEREAUX, MF SWIONTKOWSKI, P TORNETTA, W OBREMSKEY, KJ KOVAL, S NORK, S SPRAGUE, EH SCHEMITSCH, AND GH GUYATT. Internal fixation compared with arthroplasty for displaced fractures of the femoral neck: a meta-analysis. *J Bone Joint Surg[Am]*, **85-A**:1673–1681, 2003. [10](#)
- [11] GM BLAKE AND I FOGELMAN. Dual energy X-ray Absorptiometry and its clinical applications. *Seminars of Musculoskeletal Radiology*, **6**[3]:207–18, 2002. [12](#)
- [12] FL BOOKSTEIN. Principal warps: thin-plate splines and the decomposition of deformations. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, **11**[6]:567–85, 1989. [33](#), [36](#)
- [13] ML BOUXSEIN, AC COURTNEY, AND WC HAYES. Ultrasound and densitometry of the calcaneus correlate with the failure loads of cadaveric femurs. *Calcified Tissue International*, **56**:99–103, 1995. [2](#)
- [14] CA BRAUER, M COCA-PERRAILLON, DM CUTLER, AND AB ROSEN. Incidence and mortality of hip fractures in the United States. *JAMA*, **302**[14]:1573, 2009. [1](#)
- [15] RA BROOKS AND G DE CHIRO. Principles of Computer Assisted Tomography (CAT) in radiographic and radioisotopic imaging. *Phys. Med. Biol.*, **21**[5]:689–732, 1976. [28](#), [29](#)
- [16] LC BRUNNER, L ESHILIAN-OATES, AND TY KUO. Hip fractures in adults. *Am Fam Physician*, **67**[3]:537, 2003. [2](#), [7](#), [11](#)
- [17] HR BUIE, GM CAMPBELL, RJ KLINCK, JA MACNEIL, AND SK BOYD. Automatic segmentation of cortical and trabecular compartments based on a dual threshold technique for in vivo micro-CT bone analysis. *Bone*, **41**[4]:50515., 2007. [56](#)

REFERENCES

- [18] EJ CANDES, J ROMBERG, AND T TAO. Robust uncertainty principles: Exact signal reconstruction from highly incomplete frequency information. *IEEE Transactions of Information theory*, **52**[2]:489–509, 2006. [30](#), [31](#)
- [19] EJ CANDES, JK ROMBERG, AND T TAO. Stable signal recovery from incomplete and inaccurate measurements. *Communications on pure and applied mathematics*, **59**[8]:1207–1223, 2006. [30](#), [31](#)
- [20] L CAPONETTI AND AM FANELLI. 3D bone reconstruction from two X-ray views. In *Engineering in Medicine and Biology Society, 1990., Proceedings of the Twelfth Annual International Conference of the IEEE*, pages 208–210, nov 1990. [32](#), [33](#)
- [21] L CAPONETTI AND AM FANELLI. Computer-aided simulation for bone surgery. *IEEE Computer Graphics and Applications*, **13**[6]:86–92, 1993. [32](#), [33](#)
- [22] J CATES, PT FLETCHER, AND R WHITAKER. A hypothesis testing framework for high-dimensional shape models. *2nd MICCAI Workshop on Mathematical Foundations of Computational Anatomy*, pages 170–81, 2008. [110](#)
- [23] NATIONAL CLINICAL GUIDELINE CENTRE. Hip fracture - the management of hip fracture in adults. <http://www.nice.org.uk/nicemedia/live/13489/54921/54921.pdf>, Jun 2011. [5](#), [10](#), [14](#)
- [24] R CHANA, A NOORANI, N ASHWOOD, U CHATTERJI, J HEALY, AND P BAIRD. The role of MRI in the diagnosis of proximal femoral fractures in the elderly. *Injury*, **37**:185–9, 2006. [14](#)
- [25] R CHARTRAND. Fast algorithms for nonconvex compressive sensing: MRI reconstruction from very few data. In *Biomedical Imaging: From Nano to Macro, 2009. ISBI'09. IEEE International Symposium on*, pages 262–265. IEEE, 2009. [31](#)
- [26] XG CHENG, G LOWET, S BOONEN, PH NICHOLSON, P BRYNS, J NIJS, AND J DEQUEKER. Assessment of the strength of proximal femur in vitro: relationship to femoral bone mineral density and femoral geometry. *Bone*, **20**[3]:213–8, 1997. [2](#), [5](#)
- [27] DD CODY, GW DIVINE, K NAHIGIAN, AND M KLEEREKOPER. Bone density distribution and gender dominate femoral neck fracture risk predictors. *Skeletal Radiology*, **29**[3]:151–61, 2000. [13](#), [40](#), [149](#)
- [28] AL COMREY. Common methodological problems in factor analytic studies. *Journal of Consulting and Clinical Psychology*, **46**:648–659, 1978. [111](#)

REFERENCES

- [29] AL COMREY AND HB LEE. *A first course in factor analysis (2nd ed.)*. Hillsdale, NJ: Erlbaum, 1992. [110](#)
- [30] TF COOTES, GJ EDWARDS, AND CJ TAYLOR. Active appearance models. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, **23**[6]:681–5, 2001. [36](#)
- [31] J CORDEY, M SCHNEIDER, C BELENDEZ, WJ ZIEGLER, BA RAHN, AND SM PERREN. Effect of bone size, not density, on the stiffness of the proximal part of normal and osteoporotic human femora. *Journal of Bone and Mineral Research*, **7**[2]:437–44, 1992. [5](#)
- [32] AC COURTNEY, EF WACHTEL, ER MYERS, AND WC HAYS. Age-related reductions in the strength of the femur tested in a fall-loading configuration. *Journal of Bone Joint Surgery*, **77A**:387–95, 1995. [40](#), [149](#)
- [33] SR CUMMINGS, SM RUBIN, AND D BLACK. The future of hip fractures in the United States. Numbers, costs, and potential effects of postmenopausal estrogen. *Clin Orthop Relat Res*, 1990. [1](#)
- [34] E DALL’ARA, B LUISIER, R SCHMIDT, F KAINBERGER, P ZYSSET, AND D PAHR. A nonlinear QCT-based finite element model validation study for the human femur tested in two configurations in vitro. *Bone*, **52**[1]:2738, Jan 2013. [61](#)
- [35] E DALL’ARA, B LUISIER, R SCHMIDT, M PRETTERKLIEBER, F KAINBERGER, P ZYSSET, AND D PAHR. DXA predictions of human femoral mechanical properties depend on the load configuration. *Medical Engineering & Physics*, **35**[11]:15641572, Nov 2013. [61](#)
- [36] RI DAVIDOVITCH, CJ JORDAN, KA EGOL, AND MS VRAHAS. Challenges in the treatment of femoral neck fractures in the nonelderly adult. *J Trauma*, **68**[1]:236–42, Jan 2010. [7](#), [10](#)
- [37] JN DAVISON, SJ CALDER, GH ANDERSON, G WARD, C JAGGER, WM HARPER, AND ET AL. Treatment for displaced intracapsular fracture of the proximal femur: a prospective randomised trial in patients aged 65 to 79 years. *J Bone Joint Surg Br*, **83**:206–12, 2001. [10](#)
- [38] PM DE BAKKER, SL MANSKE, V EBACHER, TR OXLAND, PA CRIPTON, AND P GUY. During sideways falls proximal femur fractures initiate in the superolateral cortex: evidence from high-speed video of simulated fractures. *Journal of Biomechanics*, **42**[12]:1917–1925, 2009. [4](#)

REFERENCES

- [39] M DE LA FUENTE, E SCHKOMMODAY, P LUTZ, M NEUSS, DC WIRTZ, AND K RADERMACHER. 3D reconstruction and navigated removal of femoral bone cement in revision THR based on few fluoroscopic images. *Int. Congr. Ser.*, **1268**:626–31, 2004. 1
- [40] DK DHANWAL, EM DENNISON, NC HARVEY, AND C COOPER. Epidemiology of hip fracture: Worldwide geographic variation. *Indian Journal of Orthopaedics*, **45**[1]:15–22, January 2011. 1
- [41] DL DONOHO. Compressed sensing. *IEEE Transactions on Information Theory*, **52**[4]:1289–1306, 2006. 30
- [42] LD DORR, R GLOUSMAN, AL HOY, R VANIS, AND R CHANDLER. Treatment of femoral neck fractures with total hip replacement versus cemented and noncemented hemiarthroplasty. *J Arthroplasty*, **1**:21–8, 1986. 10
- [43] G DOUGHERTY AND D NEWMAN. Measurement of thickness and density of thin structures by computed tomography. *Med. Phys.*, **26**[7]:134148, 1999. 56
- [44] CJ DY, KE MCCOLLISTER, DA LUBARSKY, AND JM LANE. An economic evaluation of a systems-based strategy to expedite surgical treatment of hip fractures. *J Bone Joint Surg Am*, **93**[14]:1326–34, Jul 2011. 1
- [45] A EL MAGHRAOUI AND C ROUX. DXA scanning in clinical practice. *Quarterly Journal of Medicine*, **101**[8]:605–17, 2008. 12, 156
- [46] LR FABRIGAR, DT WEGENER, RC MACCALLUM, AND EJ STRAHAN. Evaluating the use of exploratory factor analysis in psychological research. *Psychological Methods*, **4**:272299, 1999. 110, 111
- [47] KG FAULKNER, SR CUMMINGS, D BLACK, L PALERMO, CC GLUER, AND HK GENANT. Simple measurement of femoral geometry predicts hip fracture: the study of osteoporotic fractures. *Journal of Bone and Minearl Research*, **8**[10]:1211–7, 1993. 2, 3
- [48] JL FAVA AND WF VELICER. The effects of overextraction on factor and component analysis. *Multivariate Behavioral Research*, **27**:387–15, 1992. 111
- [49] AGENCY FOR HEALTHCARE RESEARCH AND QUALITY. Healthcare cost and utilization project. www.ahrq.gov/data/hcup, Aug 2007. 1

REFERENCES

- [50] K FRITSCHER, B SCHULER, T LINK, F ECKSTEIN, N SUHM, M HNNI, C HENGG, AND R SCHUBERT. Prediction of biomechanical parameters of the proximal femur using statistical appearance models and support vector regression. *Medical Image Computing and Computer-Assisted Intervention*, **11**[Pt 1]:568–75, 2008. 2, 13, 14
- [51] P GAMAGE, SQ XIE, P DELMAS, AND WL XU. Diagnostic radiograph based 3D bone reconstruction framework: Application to the femur. *Computerized Medical Imaging and Graphics*, **35**[6]:427–37, 2011. 32, 33, 34, 35, 148
- [52] JS GEBHARD, HC AMSTUTZ, DM ZINAR, AND FJ DOREY. A comparison of total hip arthroplasty and hemiarthroplasty for treatment of acute fracture of the femoral neck. *Clin Orth*, **282**:123–31, 1992. 10
- [53] AH GEE. wxRegSurf software. <http://mi.eng.cam.ac.uk/~ahg/wxRegSurf/>, 2015. 96
- [54] GE GESLIEN, JH THRALL, JL ESPINOSA, AND RA OLDER. Early detection of stress fractures using 99mTc-polyphosphate. *Radiology*, **121**[(3 pt 1)]:683–7, 1976. 14
- [55] CM GIRGIS, D SHER, AND MJ SEIBEL. Atypical femoral fractures and bisphosphonate use. *New England Journal of Medicine*, **362**[19]:1848–1849, 2010. 151
- [56] LW GLORFELD. An improvement on Horns parallel analysis methodology for selecting the correct number of factors to retain. *Educational and Psychological Measurement*, **55**:377393, 1995. 111
- [57] S GNUDI, C RIPAMONTI, G GUALTIERI, AND N MALAVOLTA. Geometry of proximal femur in the prediction of hip fracture in osteoporotic women. *British Journal of Radiology*, **72**:729–33, 1999. 3
- [58] CG GOMEZ, MD CURIEL, FH CARRANZA, RP CANO, AND AD PEREZ. Femoral bone mineral density, neck-shaft angle and mean femoral neck width as predictors of hip fracture in men and women. Multicenter project for research in osteoporosis. *Osteoporosis International*, **11**:714–20, 2000. 3
- [59] JC GOWER. Generalised procrustes analysis. *Psychometrika*, **40**[1]:33–51, 1975. 32
- [60] E GUADAGNOLI AND WF VELICER. Relation of sample size to the stability of Component Patterns. *Psychological Bulletin*, **103**[2]:265–75, 1988. 110

-
- [61] CA GUANCHE, SH KOZIN, AS LEVY, AND LA BRODY. The use of MRI in the diagnosis of occult hip fractures in the elderly: a preliminary review. *Orthopedics*, **17**:327–30, 1994. [14](#)
- [62] E GUILLEY, F HERRMANN, CH RAPIN, P HOFFMEYER, R RIZZOLI, AND T CHEVALLEY. Socioeconomic and living conditions are determinants of hip fracture incidence and age occurrence among community-dwelling elderly. *Osteoporos Int.*, **22**[2]:647, 2011. [2](#)
- [63] TN HANGARTNER. Thresholding technique for accurate analysis of density and geometry in QCT, PQCT and ICT images. *J. Musculoskelet. Neuronal. Interact.*, **7**[1]:916, 2007. [56](#)
- [64] TN HANGARTNER AND V GILSANZ. Evaluation of cortical bone by Computed Tomography. *Journal of Bone Mineral Research*, **11**[10]:1518–25, 1996. [56](#)
- [65] SL BRENNAN AND MJ HENRY, MA KOTOWICZ, GC NICHOLSON, Y ZHANG, AND JA PASCO. Incident hip fracture and social disadvantage in an Australian population aged 50 years or greater. *Bone*, **48**[3]:607, 2011. [2](#)
- [66] MR BAUMGAERTNER AND TF HIGGINS. *Rockwood and Green’s Fractures in Adults*, chapter Femoral neck fractures, page 1579. Lippincott Williams & Wilkins, Philadelphia, 2002. [2](#)
- [67] LE HOLDER, C SCHWARZ, PG WERNICKE, AND RH MICHAEL. Radionuclide bone imaging in the early detection of fractures of the proximal femur (hip): multifactorial analysis. *Radiology*, **174**:509–15, 1990. [14](#)
- [68] G HOLZER, G VON SKRBENSKY, LA HOLZER, AND W PICHL. Hip fractures and the contribution of cortical versus trabecular bone to femoral neck strength. *Journal of Bone and Mineral Research*, **24**[3]:468–474, 2009. [4](#)
- [69] JL HORN. A rationale and technique for estimating the number of factors in factor analysis. *Psychometrika*, **30**:179–185, 1965. [111](#)
- [70] L HUMBERT, T WHITMARSH, M DE CRAENE, LM DEL RIO BARQUERO, K FRITSCHER, R SCHUBERT, F ECKSTEIN, T LINK, AND AF FRANGI. 3D reconstruction of both shape and Bone Mineral Density distribution of the femur from DXA images. *I. S. Biomed. Imaging*, pages 456–459, 2010. [1](#), [2](#), [13](#), [14](#), [36](#), [148](#), [149](#)

REFERENCES

- [71] A HURVITZ AND L JOSKOWICZ. Registration of a CT-like atlas to fluoroscopic X-ray images using intensity correspondences. *Int J Comput Assist Radiol Surg*, **3**:493504, 2008. [39](#), [148](#), [149](#)
- [72] P JALOVAARA AND H VIRKKUNEN. Quality of life after primary hemiarthroplasty for femoral neck fracture. 6-year follow-up of 185 patients. *Acta Orthopaedica Scandinavica*, **62**:208–17, 1991. [10](#)
- [73] F JOHANNESDOTTIR, T ASPELUND, J REEVE, KES POOLE, S SIGURDSSON, TB HARRIS, VG GUDNASON, AND G SIGURDSSON. Similarities and differences between sexes in regional loss of cortical and trabecular bone in the mid-femoral neck: The AGES-Reykjavik longitudinal study. *Journal of Bone and Mineral Research*, **28**[10]:2165–2176, 2013. [40](#), [149](#)
- [74] F JOHANNESDOTTIR, KES POOLE, J REEVE, K SIGGEIRSDOTTIR, T ASPELUND, B MOGENSEN, BY JONSSON, S SIGURDSSON, TB HARRIS, VG GUDNASON, ET AL. Distribution of cortical bone in the femoral neck and hip fracture: a prospective case-control analysis of 143 incident hip fractures; the AGES-REYKJAVIK study. *Bone*, **48**[6]:1268–1276, 2011. [4](#), [40](#), [149](#)
- [75] A JOHANSEN, R WAKEMAN, C BOULTON, F PLANT, J ROBERTS, AND A WILLIAMS. National Hip Fracture Database, National report 2013. Technical report, Clinical Effectiveness and Evaluation Unit at the Royal College of Physicians, 2013. [1](#)
- [76] O JOHNNELL, JA KANIS, A ODEN, H JOHANSSON, C DE LAET, P DELMAS, JA EISMAN, S FUJIWARA, H KROGER, D MELLSTROM, PJ MEUNIER, LJ MELTON III, T O’NEILL, H POLS, J REEVE, A SILMAN, AND A TENENHOUSE. Predictive value of BMD for hip and other fractures. *Journal of Bone and Mineral Research*, **20**[7]:1185–94, 2005. [2](#), [40](#), [149](#)
- [77] PM JOSHEPH AND RA SCHULZ. View sampling requirements in fan beam computed tomography. *Medical Physics*, **7**[6]:692–702, 1980. [25](#)
- [78] M KAMANO, S NARITA, Y HONDA, K FUKUSHIMA, AND Y YAMANO. Contrast enhanced magnetic resonance imaging for femoral neck fracture. *Clin Orthop*, **350**:179–86, 1998. [14](#)
- [79] J KANIS, E MCCLOSKEY, H JOHANSSON, A ODEN, O STRÖM, AND F BORGSTRÖM. Development and use of FRAX in osteoporosis. *Osteoporosis Int.*, **21**[2]:407–413, 2010. [2](#), [3](#), [14](#)

REFERENCES

- [80] JA KANIS, A ODEN, H JOHANSSON, F BORGSTRM, O STRM, AND E MC-CLOSKEY. FRAX and its applications to clinical practice. *Bone*, **44**:734–43, 2009. 2, 14
- [81] KM KARLSSON, I SERNBO, KJ OBRANT, I REDLUND-JHONELL, AND O JHONELL. Femoral neck geometry and radiographic signs of osteoporosis as predictors of hip fracture. *Bone*, **18**:327–30, 1996. 3
- [82] WN KELLEY. *Kelleys Textbook of rheumatology*. Saunders, New York, 6th edition, 2001. 14
- [83] T KENNY AND C TIDY. Hip fracture. <http://www.patient.co.uk/health/Hip-Fracture.htm>, Sep 2011. 5, 10
- [84] S KOLTA, A LE BRAS, D MITTON, V BOUSSON, JA DE GUISE, J FECHTENBAUM, JD LAREDO, C ROUX, AND W SKALLI. Three-dimensional X-ray absorptiometry (3D-XA): a method for reconstruction of human bones using a Dual X-ray Absorptiometry device. *Osteoporosis Int.*, **16**[8]:969–76, 2005. 3, 13, 28, 32, 33, 34, 35, 41, 148
- [85] T KONISHIIE, E MAKIHATA, H TAGO, T SATO, AND H INOUE. Acute fracture of the neck of the femur. An assessment of perfusion of the head by dynamic MRI. *J Bone Joint Surg*, **81**:596–9, 1999. 14
- [86] JT KUNCE, WD COOK, AND DE MILLER. Random variables and correlational overkill. *Educational and Psychological Measurement*, **35**:529–534, 1975. 110
- [87] HO KUOKKANEN, PK SUOMINEN, AND OL KORKALA. The late outcome of femoral neck fractures. *Int Orthop*, **14**:377–80, 1990. 10
- [88] R KURAZUME, K NAKAMURA, T OKADA, Y SATO, N SUGANO, T KOYAMA, AND ET AL. 3D reconstruction of a femoral shape using a parametric model and two 2D fluoroscopic images. *Comput Vision Image Understanding*, **113**:20211., 2009. 39, 148, 149
- [89] PR KUZYK, M BHANDARI, MD MCKEE, AND ET AL. Intramedullary versus extramedullary fixation for subtrochanteric femur. *J Orthop Trauma*, **23**[6]:465–70, Jul 2009. 12
- [90] T LANG AND J DURYEA. Peripheral bone mineral assessment of the axial skeleton. In ERICS. ORWOLL AND MICHAEL BLIZIOTES, editors, *Osteoporosis*, Contemporary Endocrinology, pages 83–104. Humana Press, 2003. 13

REFERENCES

- [91] C LANGTON, S PISHARODY, AND J KEYAK. Generation of a 3D proximal femur shape from a single projection 2D radiographic image. *Osteoporosis Int.*, **20**[3]:455–61, 2009. [13](#), [14](#), [32](#), [33](#), [35](#), [40](#), [148](#)
- [92] S LAPORTE, W SKALLI, JA DE GUISE, F LAVASTE, AND D MITTON. A biplanar reconstruction method based on 2D and 3D contours: Application to the distal femur. *Comput Methods Biomech Biomed Engin*, **6**:16, 2003. [33](#)
- [93] A LE BRAS, S KOLTA, P SOUBRANE, W SKALLI, C ROUX, AND D MITTON. Assessment of femoral neck strength by 3-dimensional X-ray absorbiometry. *Journal of Clinical Densitometry*, **9**[4]:425–30, 2006. [32](#), [33](#), [34](#), [41](#), [148](#)
- [94] A LE BRAS, S LAPORTE, V BOUSSON, D MITTON, JA DE GUISE, JD LAREDO, AND W SKALLI. Personalised 3D reconstruction of proximal femur from low-dose digital biplanar radiographs. *Int. Congr. Ser.*, **1256**:214–9, 2003. [32](#), [33](#), [34](#), [41](#), [148](#)
- [95] ES LEBLANC, TA HILLIER, KL PEDULA, JH RIZZO, PM CAWTHON, HA FINK, JA CAULEY, DC BAUER, DM BLACK, SR CUMMINGS, AND WS BROWNER. Hip fracture and increased short-term but not long-term mortality in healthy older women. *Arch Intern Med.*, **171**[20]:1831–7, Nov 2011. [1](#)
- [96] BP LEE, DJ BERRY, WS HARMSSEN, AND FH SIM FH. Total hip arthroplasty for the treatment of an acute fracture of the femoral neck: long term results. *J Bone Joint Surg Am*, **80**:70–5, 1998. [10](#)
- [97] S LEKAMWASAM AND RS LENORA. Effect of leg rotation on hip bone mineral density measurements. *Journal of Clinical Densitometry*, **6**[(4)]:331–6, 2003. [156](#)
- [98] BA LENART, DG LORICH, AND JM LANE. Atypical fractures of the femoral diaphysis in postmenopausal women taking alendronate. *New England Journal of Medicine*, **358**[12]:1304–1306, 2008. [151](#)
- [99] EM LEWIECKI, N BINKLEY, AND SM PETAK. DXA quality matters. *Journal of Clinical Densitometry*, **9**[4]:388–92, 2006. [12](#)
- [100] BF LOGAN AND LA SHEPP. Optimal reconstruction of a function from its projections. *Duke Mathematical Journal*, **42**:645–59, 1975. [26](#)
- [101] R LOO. Caveat on sample sizes in factor analysis. *Perceptual and Motor Skills*, **56**[371-374], 1983. [110](#)

REFERENCES

- [102] MR KARAGAS AMD GL LU-YAO, JA BARRETT, ML BEACH, AND JA BARON. Heterogeneity of hip fracture: age, race, sex, and geographic patterns of femoral neck and trochanteric fractures among the US elderly. *Am J Epidemiol*, **143**[7]:677, 1996. [2](#)
- [103] LM MARCH, AC CHAMBERLAIN, ID CAMERON, RG CUMMING, AJ BRNABIC, TP FINNEGAN TP, AND ET AL. How best to fix a broken hip. Fractured neck of femur health outcomes project team. *Med J Aust*, **170**:489–94, 1999. [10](#)
- [104] P MARKELJ, D TOMAZEVIC, B LIKAR, AND F PERCUS. A review of 3D/2D registration methods for image-guided interventions. *Medical Image Analysis*, **16**[3]:642–61, 2012. [33](#), [34](#), [155](#)
- [105] LM MARSHALL, JM ZMUDA, BKS CHAN, E BARRETT-CONNOR, JA CAULEY, KE ENSRUD, TF LANG, AND ES ORWOLL. Race and ethnic variation in proximal femur structure and BMD among older men. *Journal of Bone and Mineral Research*, **23**[1]:121–130, 2008. [40](#), [149](#)
- [106] M MASSON, MJ PARKER, AND S FLEISCHER. Internal fixation versus arthroplasty for intracapsular proximal femoral fractures in adults (Cochrane Review). In: *The Cochrane Library*, Issue 2, Chichester, UK: John Wiley & Sons, Ltd., 2003. [10](#)
- [107] PM MAYHEW, CD THOMAS, JG CLEMENT, N LOVERIDGE, TJ BECK, W BONFIELD, CJ BURGOYNE, AND J REEVE. Relation between age, femoral neck cortical stability, and hip fracture risk. *The Lancet*, **366**[9480]:129–35, 2005. [4](#), [40](#), [149](#)
- [108] SIMON MEARS. Fixing hip fractures. http://www.hopkinsmedicine.org/gec/series/fixing_hip_fractures#top, Nov 2013. [7](#), [10](#), [11](#), [14](#)
- [109] LJ MELTON 3RD. Who has osteoporosis? A conflict between clinical and public health perspectives. *J Bone Miner Res*, **15**[12]:2309, 2000. [2](#)
- [110] C QUAH AMD C BOULTON AMD C MORAN. The influence of socioeconomic status on the incidence, outcome and mortality of fractures of the hip. *J Bone Joint Surg Br*, **93**[6]:801, 2011. [2](#)
- [111] JJ MORE. *The Levenberg-Marquardt algorithm: Implementation and theory*, pages 105–116. Springer-Verlag, 1977. [48](#), [51](#), [56](#)
- [112] RS MORRISON, MR CHASSIN, AND AL SIU. The medical consultant’s role in caring for patients with hip fracture. *Ann Intern Med*, **128**[12 Pt 1]:1010, 1998. [1](#)

REFERENCES

- [113] T NAKAMURA, CH TURNER, T YOSHIKAWA, CW SLEMENDA, M PEACOCK, DB BURR, Y MIZUNO, H ORIMO, Y OUCHI, AND C JR JOHNSTON C. Do variations in hip geometry explain differences in hip fracture risk between Japanese and white Americans? *Journal of Bone Mineral Research*, **9**:1071–6, 1994. [3](#), [40](#), [149](#)
- [114] DA NELSON, DA BARONDESS, SL HENDRIX, AND TJ BECK. Cross-sectional geometry, bone strength, and bone mass in the proximal femur in black and white postmenopausal women. *Journal of Bone and Mineral Research*, **15**[10]:1992–1997, 2000. [40](#), [149](#)
- [115] CF NJEH, T FUERST, D HANS, GM BLAKE, AND HK GENANT. Radiation exposure in bone mineral density assessment. *Applied Radiation and Isotopes*, **50**[1]:215–36, 1977. [12](#), [13](#), [14](#), [156](#)
- [116] SL NYCCION, DM HUNTER, AND GAM FINEMAN. *Orthopaedic Sports Medicine*, **2nd**, chapter Hip and pelvis, page 1443. WB Saunders, Philadelphia, 2003. [2](#)
- [117] A ODACN, E MCCLOSKEY, H JOHANSSON, AND J KANIS. Assessing the impact of osteoporosis on the burden of hip fractures. *Calcified Tissue International*, **92**[1]:42–49, 2013. [1](#)
- [118] J PANULA, H PIHLAJAMKI, VM MATTILA, P JAATINEN, T VAHLBERG, P AARNIO, AND SL KIVEL. Mortality and cause of death in hip fracture patients aged 65 or older: a population-based study. *BMC Musculoskelet Disord*, **12**:105, May 2011. [1](#)
- [119] RF PAPANDREA AND MI FROMSON. Total hip arthroplasty after acute displaced femoral neck fractures. *Am J Orthop*, **25**:85–8, 1996. [10](#)
- [120] M PARKER AND A JOHANSEN. Hip fractures. *Br. Med. J.*, **333**[7557]:27–30, 2006. [1](#), [10](#), [11](#)
- [121] MJ PARKER AND C BLUNDELL. Choice of implant for internal fixation of femoral neck fractures - meta analysis of 25 randomised trials including 4925 patients. *Acta Orthop Scand*, **69**:138–43, 1998. [10](#)
- [122] MJ PARKER AND K GURUSAMY. Internal fixation versus arthroplasty for intra-capsular proximal femoral. *Cochrane Database Syst Rev. 2006 Oct*, **18**[4], 2006. [10](#)

REFERENCES

- [123] MJ PARKER, JW MYLES, JK ANAND, AND R DREWETT. Cost-benefit analysis of hip fracture treatment. *J Bone Joint Surg Br*, **74**:261–4, 1992. 7, 11
- [124] MJ PARKER AND GA PRYOR. Internal fixation or arthroplasty for displaced cervical hip fractures in the elderly: a randomised controlled trial of 208 patients. *Acta Orthop Scand*, **71**:440–6, 2000. 10
- [125] G PATHAK, MJ PARKER, AND GA PRYOR. Delayed diagnosis of femoral neck fractures. *Injury*, **28**:299–301, 1997. 14
- [126] A PAWEL AND PENCZEK. Fundamentals of three-dimensional reconstruction from projections. *Methods Enzymol.*, **482**:133, 2010. 28
- [127] M PEACOCK, CH TURNER, G LIU, AK MANATUNGA, L TIMMERMAN, AND CC JR JOHNSTON. Better discrimination of hip fracture using bone density, geometry and architecture. *Osteoporosis International*, **5**:167–73, 1995. 3
- [128] W PISTOIA, B VAN RIETBERGEN, AND P REGSEGGER. Mechanical consequences of different scenarios for simulated bone atrophy and recovery in the distal radius. *Bone*, **33**[6]:937–945, 2003. 4
- [129] KES POOLE. The FEMCO study: Regional thinning of the FEMoral neck COrtex in hip fracture; a case-control study LREC07/H0305/61 ARC17822 v3.6 Cambridge research ethics committee 4. <http://public.ukcrn.org.uk/search/StudyDetail.aspx?StudyID=5290>, 2007-2017. 61
- [130] KES POOLE, GM TREECE, AH GEE, JP BROWN, MR MCCLUNG, A WANG, AND C LIBANATI. Denosumab rapidly increases cortical bone in key locations of the femur: A 3D bone mapping study in women with osteoporosis. *J Bone Miner Res.*, **30**[1]:46–54, 2015. 5, 6
- [131] KES POOLE, GM TREECE, PM MAYHEW, J VACULIK, P DUNGL, M HORAK, JJ STEPAN, AND AH GEE. Cortical thickness mapping to identify focal osteoporosis in patients with hip fracture. *PLoS One*, **7**[6]:e38466, 06 2012. 5
- [132] KES POOLE, GM TREECE, GR RIDGWAY, PM MAYHEW, J BORGGREFE, AND AH GEE. Targeted regeneration of bone in the osteoporotic human femur. *PLoS One*, **6**[1]:e16190, 2011. 5
- [133] S PREVRHAL, K ENGELKE, AND WA KALANDER. Accuracy limits for the determination of cortical width and density: the influence of object size and CT imaging parameters. *Phys. Med. Biol.*, **44**[3]:751764, 1999. 54, 56

REFERENCES

- [134] S PREVRHAL, JC FOX, JA SHEPHERD, AND HK GENANT. Accuracy of CT-based thickness measurement of thin structures: Modeling of limited spatial resolution in all three dimensions. *Med. Phys.*, **30**[1]:18, 2003. 54, 56
- [135] S PRICE. Metabolic bone diseases: Mapping therapy-induced changes to osteoporotic bone. *Nat. Rev. Rheumatol.*, **7**[4]:194, 2011. 5
- [136] SF QUINN AND JL MCCARTHY. Prospective evaluation of patients with suspected hip fracture and indeterminate radiographs: use of T1-weighted MR images. *Radiology*, **187**:469–71, 1993. 14
- [137] EL RAAHYMAAKERS AND RK MARTI. Non-operative treatment of impacted femoral neck fractures. A prospective study of 170 cases. *J Bone Joint Surg Br*, **73**:950–4, 1991. 7
- [138] RADIOLOGY MASTERCLASS. Trauma X-ray - lower limb - hip fracture. http://radiologymasterclass.co.uk/tutorials/musculoskeletal/x-ray_trauma_lower_limb/hip_fracture_x-ray.html, November 2013. 7, 8, 9, 10, 11, 14
- [139] BL RIGGS, LJ MELTON, RA ROBB, JJ CAMP, EJ ATKINSON, JM PETERSON, PA ROULEAU, CH MCCOLLOUGH, ML BOUXSEIN, AND S KHOSLA. Population-based study of age and sex differences in bone volumetric density, size, geometry, and structure at different skeletal sites. *Journal of Bone and Mineral Research*, **19**[12]:1945–1954, 2004. 40, 149
- [140] LZ RUBENSTEIN AND KR JOSEPHSON. The epidemiology of falls and syncope. *Clin Geriatr Med.*, **18**[2]:141, 2002. 2
- [141] D RUECKERT, P ALJABAR, RA HECKEMANN, JV HAJNAL, AND A HAMMERS. Diffeomorphic registration using B-splines. In *9th International Conference on Medical Image Computing and Computer-Assisted Intervention (MICCAI 2006)*, October 2006. 36
- [142] P SAMBROOK AND C COOPER. Osteoporosis. *The Lancet*, **367**[9527]:2010–8, 2006. 1
- [143] N SARKALKAN, H WEINANS, AND AA ZADPOOR. Statistical shape and appearance models of bones. *Bone*, **60**:129–40, 2013. 32, 36, 140
- [144] E SHANE, D BURR, B ABRAHAMSEN, RA ADLER, TD BROWN, AM CHEUNG, F COSMAN, JR CURTIS, R DELL, DW DEMPSTER, ET AL. Atypical

- subtrochanteric and diaphyseal femoral fractures: second report of a task force of the American Society for Bone and Mineral Research. *Journal of Bone and Mineral Research*, **29**[1]:1–23, 2014. 151
- [145] E SHANE, D BURR, PR EBELING, B ABRAHAMSEN, RA ADLER, TD BROWN, AM CHEUNG, F COSMAN, JR CURTIS, R DELL, ET AL. Atypical subtrochanteric and diaphyseal femoral fractures: report of a task force of the American Society for Bone and Mineral Research. *Journal of Bone and Mineral Research*, **25**[11]:2267–2294, 2010. 151
- [146] SCOTTISH INTERCOLLEGIATE GUIDELINES NETWORK SIGN. Management of hip fracture in older people. <http://www.sign.ac.uk/guidelines/fulltext/111/index.html>, Jun 2009. 7, 10, 11, 12
- [147] G SIGURDSSON, T ASPELUND, M CHANG, B JONSDOTTIR, S SIGURDSSON, G EIRIKSDOTTIR, A GUDMUNDSSON, TB HARRIS, V GUDNASON, AND TF LANG. Increasing sex difference in bone strength in old age: The age, gene/environment susceptibility-Reykjavik study (AGES-REYKJAVIK). *Bone*, **39**[3]:644–651, 2006. 40, 149
- [148] DL SNYDER AND JR COX. *Reconstruction tomography in diagnostic radiology and nuclear medicine*. University Park Press, Baltimore, 1977. 26
- [149] SM SNYDER AND E SCHNEIDER. Estimation of mechanical properties of cortical bone by computed tomography. *Journal of Orthopaedic Research*, **9**:422–31, 1991. 5
- [150] B SQUIRES AND G BANNISTER. Displaced intracapsular neck of femur fractures in mobile independent patients: Total hip replacement or hemiarthroplasty? *Injury*, **30**:345–8, 1999. 10
- [151] P SUETENS. *Fundamentals of medical imaging*, chapter X-ray computed tomography, pages 35–53. Cambridge medicine. Cambridge University Press, 2009. 25, 26, 153
- [152] J THEVENOT, J KOIVUMAKI, V KUHN, F ECKSTEIN, AND T JAMSA. A novel methodology for generating 3D finite element models of the hip from 2D radiographs. *J Biomech.*, **47**[2]:438–44, 2014. 32, 38, 148, 149
- [153] CL THOMAS, PM MAYHEW, J POWER, KES POOLE, N LOVERIDGE, JG CLEMENT, CJ BURGOYNE, AND J REEVE. Femoral neck trabecular bone:

- loss with aging and role in preventing fracture. *Journal of Bone and Mineral Research*, **24**[11]:1808–1818, 2009. 40, 149
- [154] C TIDY. Femoral fractures. <http://www.patient.co.uk/doctor/femoral-fractures>, Sep 2011. 7, 11, 12
- [155] GM TREECE AND AH GEE. Independent measurement of femoral cortical thickness and cortical bone density using clinical CT. *Medical Image Analysis*, **20**[1]:249 – 264, 2015. 5, 6, 56, 66
- [156] GM TREECE, AH GEE, PM MAYHEW, AND KES POOLE. High resolution cortical bone thickness measurement from clinical CT data. *Med. Image Anal.*, **14**[3]:276–290, 2010. 1, 4, 5, 38, 53, 54, 56
- [157] GM TREECE, AH GEE, AND R PRAGER. Stradwin software. <http://mi.eng.cam.ac.uk/~rwp/stradwin/>, 2014. 54, 96
- [158] GM TREECE, KES POOLE, AND AH GEE. Imaging the femoral cortex: thickness, density and mass from clinical CT. *Medical Image Analysis*, **16**[5]:952–65, Jul 2012. 5, 17, 38, 54, 56
- [159] F TROCHU. A contouring program based on dual Kriging Interpolation. *Engineering with Computers*, **9**:160–77, 1993. 34
- [160] J TRZASKO AND M ARMANDO. Highly undersampled magnetic resonance image reconstruction via homotopic l_0 -minimization. *IEEE Transactions on Medical imaging*, **28**[1]:106–121, 2009. 31
- [161] E VERHULP, B VAN RIETBERGEN, AND R HUISKES. Load distribution in the healthy and osteoporotic human proximal femur during a fall to the side. *Bone*, **42**[1]:30–35, 2008. 4
- [162] PM WATERS AND MB MILLIS. Hip and pelvic injuries in the young athlete. *Clin Sports Med.*, **7**[3]:513, 1988. 2
- [163] JM HUDDLESTON AND KJ WHITFORD. Medical care of elderly patients with hip fractures. *Mayo Clin Proc*, **76**[3]:295, 2001. 1
- [164] T WHITMARSH, L HUMBERT, M DE CRAENE, LM DEL RIO BARQUERO, AND AF FRANGI. Reconstructing the 3D shape and bone mineral density distribution of the proximal femur from Dual-energy X-ray Absorptiometry. *IEEE Trans. Med. Imag.*, **30**[12]:2101 –2114, 2011. 36, 37, 40, 148, 149

- [165] T WHITMARSH, L HUMBERT, M DE CRAENE, LM DEL RIO BARQUERO, K FRITSCHER, R SCHUBERT, F ECKSTEIN, T LINK, AND AF FRANGI. 3D bone mineral density distribution and shape reconstruction of the proximal femur from a single simulated DXA image: an in vitro study. In *SPIE 7623, 76234U*, doi:10.1117/12.844110, 2010. [36](#), [37](#), [40](#), [148](#), [149](#)
- [166] T WHITMARSH, GM TREECE, AH GEE, AND KES POOLE. Mapping bone changes at the proximal femoral cortex of postmenopausal women in response to Alendronate and Teriparatide alone, combined or sequentially. *Journal of Bone and Mineral Research*, 2015. [5](#)
- [167] WIKIPEDIA. Projection-slice theorem. Online, March 2013. Accessed 22 Jan 2014. [153](#)
- [168] FD WOLINSKY, JF FITZGERALD, AND TE STUMP. The effect of hip fracture on mortality, hospitalization, and functional status: a prospective study. *Am J Public Health*, **87**[3]:398, 1997. [1](#)
- [169] R ZEBAZE AND S SEEMAN. Cortical bone: A challenging geography. *Journal of Bone and Mineral Research*, **30**[1]:24–29, 2015. [148](#)