
High resolution cortical thickness measurement from clinical CT data

G.M. Treece, A.H. Gee, P.M. Mayhew and K.E.S. Poole

CUED/F-INFENG/TR 634

12th October 2009

Cambridge University Engineering Department
Trumpington Street
Cambridge CB2 1PZ
England

Corresponding e-mail: gmt11@eng.cam.ac.uk

Abstract

The distribution of cortical bone in the proximal femur is believed to be critical in determining fracture resistance. Current CT technology is limited in its ability to measure cortical thickness, especially in the sub-millimetre range which lies within the point spread function of today's clinical scanners. In this paper, we present a novel technique that is capable of producing unbiased thickness estimates down to 0.3 mm. The technique relies on a mathematical model of the anatomy and the imaging system, which is fitted to the data at a large number of sites around the proximal femur, producing around 17,000 independent thickness estimates per specimen. In a series of experiments on sixteen cadaveric femurs, estimation errors were measured as -0.01 ± 0.58 mm (mean \pm 1 std. dev.) for cortical thicknesses in the range 0.3 mm to 4 mm. This compares with 0.25 ± 0.69 mm for simple thresholding and 0.90 ± 0.92 mm for a variant of the 50% relative threshold method. In the clinically relevant sub-millimetre range, thresholding increasingly fails to detect the cortex at all, whereas the new technique continues to perform well. The many cortical thickness estimates can be displayed as a colour map painted onto the femoral surface. Computation of the surfaces and colour maps is largely automatic, requiring around 15 minutes on a modest laptop computer.

1 Introduction

The exponential increase in hip fractures with ageing imposes a substantial and increasing health burden. They are the most common reason for acute orthopaedic admission in older people (Parker and Johansen, 2006) and their incidence is projected to rise worldwide from 1.7 million in 1990 to 6.3 million in 2050 (Sambrook and Cooper, 2006). Current imaging methods for predicting an individual's hip fracture risk use bone mineral density (BMD) estimates which are specific (Johnell et al., 2005; Kanis et al., 2008) but have low sensitivity (Kanis et al., 2008; Kaptoge et al., 2008; Sanders et al., 2006). Regional structural weakness in the aged hip may predispose to fracture (Mayhew et al., 2005; Poole et al., 2009; de Bakker et al., 2009) and there have been numerous calls to develop novel three-dimensional (3D) methods of assessing hip structure, most notably with multi-detector computed tomography, MDCT (Bouxsein and Delmas, 2008).

The anatomical distribution of cortical bone in the aged proximal femoral metaphysis is critical in determining its resistance to fracture, with trabecular bone playing a lesser role (Holzer et al., 2009; Verhulst et al., 2008). Compressive cracking of cortical bone of the superior parts of the neck and trochanter is often the first stage of fracture (de Bakker et al., 2009; Carpenter et al., 2005; Mayhew et al., 2005). Current MDCT studies are anatomically limited by set regions of interest and technically limited by thresholding errors caused by relatively low resolution data sets. Ideally, we would like to be able to map the cortical bone thickness to a high degree of accuracy, from a clinical resolution data set, over the entire proximal femur. Since bone anabolic drugs are capable of stimulating new bone formation at structurally important sites (Keaveny et al., 2008; Uusi-Rasi et al., 2005), such a technique might help not only with identification of weak areas but also in assessing treatment response.

Thin laminar structures such as the femoral cortex are not accurately depicted in clinical CT because of the images' limited spatial resolution. This typically results in an underestimate of density and an overestimate of thickness. A number of studies (Dougherty and Newman, 1999; Hangartner and Gilsanz, 1996; Prevrhal et al., 1999) have investigated the influence of the imaging system's in-plane point spread function (PSF), while Prevrhal et al. (2003) go further in quantifying the effects of the out-of-plane PSF as well. Such studies explain why straightforward thickness estimation techniques, such as those based on thresholding (Buie et al., 2007; Hangartner, 2007) or the 50% relative threshold method (Prevrhal et al., 1999, 2003), are unreliable when the structure is thin compared with the imaging system's PSF. The consensus is that, with typical PSFs from normal bore, clinical CT scanners, straightforward thickness measurements start to become inaccurate below around 2.5 mm (Dougherty and Newman, 1999; Hangartner and Gilsanz, 1996), with errors exceeding 100% for sub-millimetre cortices (Prevrhal et al., 2003).

More sophisticated thickness estimation techniques are capable of superior accuracy and precision. The classical approach to image deblurring is to assume models of both the object being scanned and the image formation process, then attempt to fit these models to the data. Often, this inverse problem is ill-posed, in that no unique set of model parameters explains the data, especially in the presence of imaging noise. In such circumstances, prior knowledge may be used to constrain some of the model parameters. Such an approach was used by [Streekstra et al. \(2007\)](#) to estimate the thickness of cartilage layers in the sub-millimetre range, well below the nominal width of the CT system’s PSF. In this instance, the model parameters relating to the PSF were fixed, using information obtained by scanning a suitable phantom.

In this paper, we describe a model-based approach for measuring femoral cortical bone thicknesses in the supra- to sub-millimetre range, using standard, normal bore, clinical CT. We employ the anatomical and imaging models described by [Prevrhal et al. \(2003\)](#), and explore the efficacy of constraining various parameters when fitting these models to the CT data. In contrast to the approach taken by [Streekstra et al. \(2007\)](#), we find that constraining the peak CT density, rather than the imaging system’s PSF, produces the best cortical thickness estimates. Our model-fitting process is guided by a semi-automatic segmentation of the outer bone surface. This leads to an illuminating way of visualizing the results: cortical thicknesses can be displayed, to high spatial resolution, as a colour map painted on the bone surface.

The paper is organised as follows. In Section 2, we recap the anatomical and imaging models of [Prevrhal et al. \(2003\)](#), describe our novel approach for fitting these models to low resolution CT data, and introduce our methodology for evaluating the results. In Section 3, we present results for sixteen cadaveric human femurs, scanned using both high resolution, peripheral CT and conventional, low resolution CT. Cortical thicknesses measured by a 50% relative threshold technique on the high resolution data provide the gold standard against which estimates from the low resolution data are judged. By way of comparison, results are presented for the new method and also the thresholding and 50% relative threshold approaches. These results are discussed in Section 4, before we draw our conclusions and suggest some avenues for future research in Section 5.

2 Method

2.1 Modelling the scanning process

We use as our raw data for each cortical thickness estimate the CT values along a line in the imaging plane normal to the cortical layer, and of sufficient length to pass through both the cortical layer itself and any blurred prolongation caused by the imaging process. A convenient method for locating such lines is discussed in Section 2.3. We assume that the cortical layer is locally flat and of uniform thickness, at least within the extent of the imaging system’s PSF. We can then model the CT values along the line as a convolution of the ‘real’ density with an in-plane and an out-of-plane PSF. The real density variation y over distance x is assumed to be piecewise constant, with different values in the surrounding tissue, the cortex and the trabecular bone. This can be expressed as the summation of two step functions:

$$y(x) = y_0 + (y_1 - y_0)H(x - x_0) + (y_2 - y_1)H(x - x_1) \quad (1)$$

where y_0 , y_1 and y_2 are the CT values in the surrounding tissue, cortical and trabecular bone, respectively, x_0 and x_1 are the locations of the outer and inner cortical surfaces, and $H(x)$ is a unit step function.

Following the approach taken by [Prevrhal et al. \(2003\)](#), we model the in-plane PSF g_i as a normalised Gaussian function:

$$g_i(x) = \frac{e^{-\frac{x^2}{\sigma^2}}}{\sigma\sqrt{\pi}} \quad (2)$$

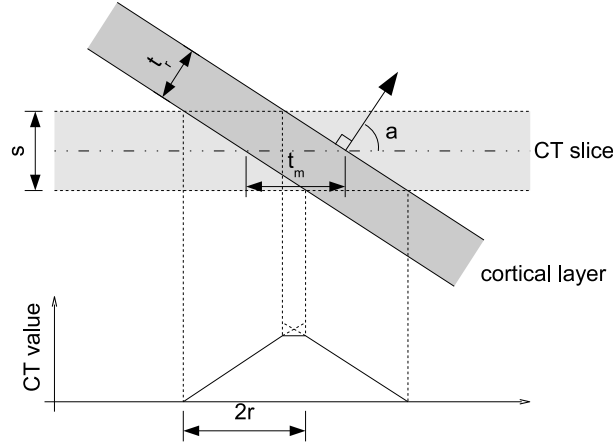


Figure 1: The effect of surface orientation and CT slice thickness on the apparent thickness and out-of-plane PSF. A finite CT slice width s causes cortical layers oblique to the slice to appear as if they have been blurred with a rectangular PSF, with extent dependent on surface orientation a , leading to a piecewise linear image. The true cortical thickness t_r can be expressed as a function of a and the measured thickness t_m .

where σ represents the extent of the blur. The out-of-plane PSF g_o is the result of the interaction between the CT slice thickness and the orientation of the cortical layer with respect to the imaging plane, as illustrated in Fig. 1. This can be expressed as a rectangular function:

$$g_o(x) = \frac{1}{2r} [H(x+r) - H(x-r)] \quad (3)$$

where the extent of blur $2r$ is calculated from s , the CT slice thickness, and a , the angle the cortical surface normal makes with the imaging plane:

$$r = \frac{s}{2} \tan a \quad (4)$$

Hence, the fully blurred CT values y_{blur} are given by¹:

$$y_{\text{blur}}(x) = y_0 + \iint \left(\frac{y_1 - y_0}{2r\sigma\sqrt{\pi}} \left[e^{-\frac{(x+r-x_0)^2}{\sigma^2}} - e^{-\frac{(x-r-x_0)^2}{\sigma^2}} \right] + \frac{y_2 - y_1}{2r\sigma\sqrt{\pi}} \left[e^{-\frac{(x+r-x_1)^2}{\sigma^2}} - e^{-\frac{(x-r-x_1)^2}{\sigma^2}} \right] \right) dx dx \quad (5)$$

except for the special case $r = 0$, when:

$$y_{\text{blur}, r=0}(x) = y_0 + \int \left(\frac{y_1 - y_0}{\sigma\sqrt{\pi}} e^{-\frac{(x-x_0)^2}{\sigma^2}} + \frac{y_2 - y_1}{\sigma\sqrt{\pi}} e^{-\frac{(x-x_1)^2}{\sigma^2}} \right) dx \quad (6)$$

The true cortical thickness t_r is then:

$$t_r = t_m \cos a = (x_1 - x_0) \cos a \quad (7)$$

Figure 2 shows y_{blur} for the case $\sigma = 1$ mm and $r = 0$, and with cortical thickness values of 5 mm, 1 mm and 0.5 mm. When the cortex is thin, as in Fig. 2(c), thickness estimation using a set threshold, or the 50% relative

¹These expressions are readily derived by differentiating $y(x)$ and $g_o(x)$ to obtain a sequence of δ -functions, convolving with $g_i(x)$, and integrating the result. y_{blur} can then be evaluated by numerical integration of eq. (5) or (6), which is more computationally efficient than the alternative approach of direct numerical convolution. This is particularly advantageous when y_{blur} has to be evaluated a large number of times, as in the Levenberg-Marquardt procedure described in Section 2.2.

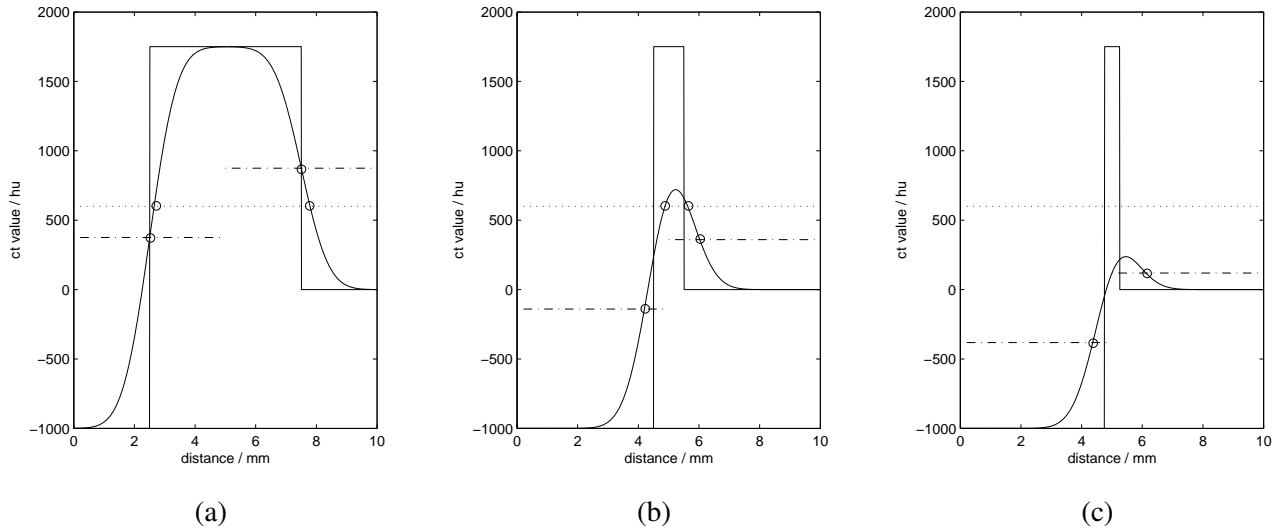


Figure 2: A simulation of the cortical imaging process. The CT value along a line passing through the cortex is simulated by blurring a piecewise constant density function (with CT values of -1000 , 1750 and 0) with a Gaussian kernel of standard deviation 1 mm. The solid lines show the underlying and blurred CT values, the dotted lines show a threshold of 600 , and the dashed lines show values half way between the appropriate CT baseline and the blurred peak (50% relative threshold). In (a), the cortex is sufficiently wide for thresholding and the 50% relative threshold method to work correctly. In (b), the threshold still gives approximately the correct thickness, but the 50% relative threshold method overestimates. In (c), no values exceed the threshold, and the 50% relative threshold method now reports the width of the PSF rather than that of the cortex.

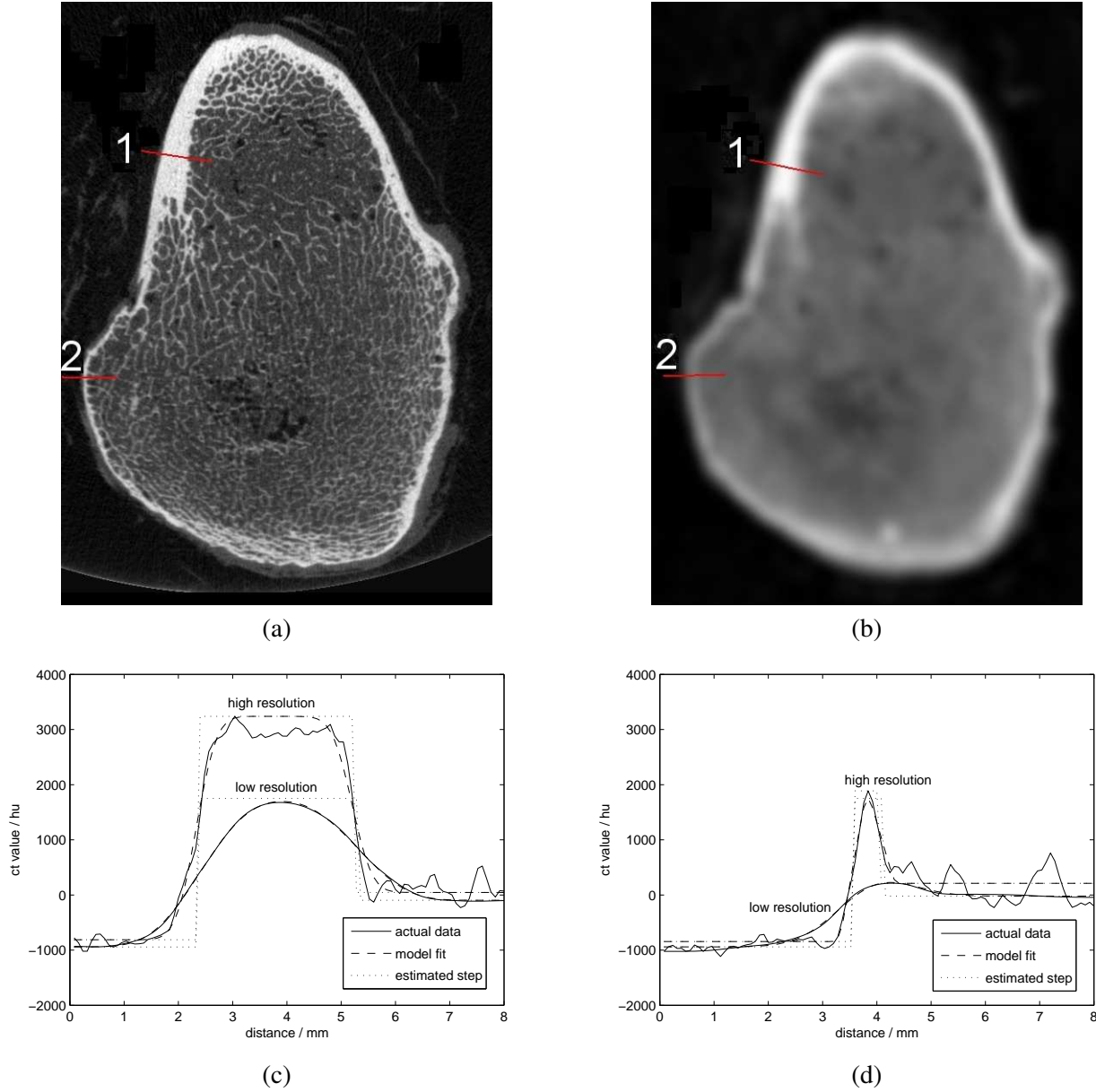


Figure 3: Cortical thickness in high and low resolution CT data. (a) and (b) show corresponding CT slices from a cadaveric femur scanned at high resolution (Xtreme pQCT, $82\ \mu\text{m}/\text{pixel}$, Scanco Medical AG, Brüttisellen, Switzerland) and low resolution (Siemens Somatom Sensation 64 MDCT, $589\ \mu\text{m}/\text{pixel}$, Siemens AG, Erlangen, Germany). (c) and (d) show the CT values along the lines labelled (1) and (2) in the images. In (c), the cortical layer is sufficiently thick to show a significant peak in the low resolution data. In (d), the thickness is somewhat less, and in this case there is only a very shallow peak in the low resolution data. Also shown in (c) and (d) are the best fit models y_{blur} , and the corresponding piecewise constant density functions.

threshold method, is highly error prone. [Prevral et al. \(2003\)](#) demonstrated that eq. (5) is a good model for what is typically observed when estimating cortical thickness in practice. The real examples in Fig. 3 confirm that the imaging process for low resolution clinical data does indeed follow the model fairly closely. The equivalent high resolution data indicates that the ‘real’ density variation is reasonably well modeled by the piecewise constant function of eq. (1), though clearly the CT values on either side of the cortex are not perfectly flat.

2.2 Thickness estimation

If we could fit the model of eq. (5) to real CT data, then we could derive the cortical thickness t_r using eq. (7) and knowledge of the surface orientation. This potentially allows thickness measurement well below the blurring level σ of the imaging process. Fitted models are shown as dashed lines in Figs. 3(c) and (d). For the high resolution data, the dashed lines follow the edges of the cortex well, despite the distracting density variations elsewhere. For the low resolution data, the dashed lines are nearly indistinguishable from the real data (solid lines), providing further evidence of the model’s validity.

In our proposed method, optimization of the model parameters is performed using the Levenberg-Marquardt method ([More, 1977](#)). The CT data is resampled at 100 points along the line in the image plane normal to the cortical surface, using the Mitchell-Netravali cubic spline ([Burger and Burge, 2009](#)) to interpolate between the original CT samples. Automatic positioning of this line is discussed in Section 2.3. The model is then fitted to the 100 new samples, with eq. (5) evaluated numerically for each candidate set of parameters. Model-data discrepancies at one end of the line (that closest to the outer cortical surface) are penalised slightly more than discrepancies elsewhere. This encourages the optimization algorithm to fit the model to the actual cortex, and not to secondary CT peaks within the interior that are especially common in high resolution data.

There are seven model parameters in total, $(y_0, y_1, y_2, x_0, x_1, \sigma, r)$. Of these, the out-of-plane blur r is assumed to be known from the surface orientation (see Section 2.3) and the slice thickness, following eq. (4). The edges of the cortical layer (x_0, x_1) are the values we are attempting to estimate. This leaves three parameters (y_0, y_1, y_2) describing the CT values, and one, σ , describing the in-plane PSF. The CT values of the surrounding tissue (y_0) and of the trabecular bone (y_2) vary with location around the proximal femur, so we leave the model free to discover these parameters for each individual line. This requires that the line penetrates each region sufficiently: we therefore set the line length to at least three times the maximum cortical thickness we expect to encounter, and position it so that one third lies outside the exterior cortical surface. For the experiments reported in Section 3, the line length was about 18 mm in all cases (6 mm outside, 12 mm inside).

The remaining two model parameters, y_1 , the cortical CT value, and σ , the in-plane blur, are more problematic. For thin cortical layers, it is possible to trade one off against the other to produce values of y_{blur} that are very similar. In other words, the problem is ill-posed. When both these parameters are left free, the Levenberg-Marquardt method frequently fails to converge, and there is much variability in the subsequent thickness estimates. To better condition the optimization problem, we could constrain either σ , as was done for the more complex cartilage model of [Streekstra et al. \(2007\)](#), or y_1 . Neither choice is ideal, since both σ and y_1 are expected to vary with location around the proximal femur. While it might seem more natural to constrain σ , it transpires that this barely improves the conditioning of the problem. In contrast, constraining y_1 is far more effective. The evidence for this assertion can be found in Section 3, where we subsequently assess two approaches for setting y_1 (a fixed value for all data, or automatically estimating a value for each femur), and establish how sensitive the thickness estimates are to errors in y_1 .

For comparison, we also tested thresholding, with levels as in ([Poole et al., 2008](#)), and a method similar to the 50% relative threshold of [Prevral et al. \(1999\)](#). For a fair comparison at equivalent resolutions, all methods used the same 100 sampled CT points. Additionally, our implementations of the various methods all involve fitting the model to the data in one way or another: this ensures consistent discrimination between

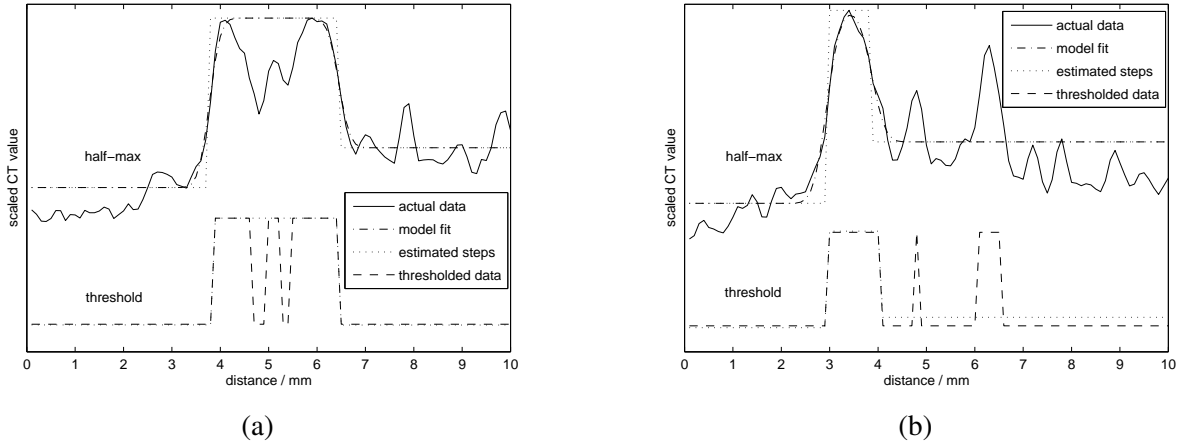


Figure 4: Discrimination of cortical and trabecular bone peaks for the half-max and threshold methods. (a) and (b) are two examples from the high resolution data showing how the model-fitting process establishes which peaks are within the cortex. The upper plots correspond to the half-max method, the lower plots to the threshold method. The model-fitting algorithm optimally (in a least squares sense) divides the data into spans of higher (cortical) and lower (trabecular, exterior) density values. For the half-max method, the boundaries tend to be positioned very close to the half-way point between the cortical and trabecular/exterior values. For the threshold method, the boundaries are positioned precisely on one of the threshold crossing points.

CT peaks considered to be in the cortex and those considered to be in the trabecular bone. In general, the low resolution data exhibits just a single peak along the length of the line, but the high resolution data resolves smaller features, including intra-cortical pores, leading to multiple peaks. In such circumstances, the model fitting algorithm implicitly determines which peaks lie within the cortex, as illustrated in Fig. 3 for the new method and in Fig. 4 for the comparison methods. Although they share this same methodology for cortical peak identification, the various methods otherwise differ in many respects:

threshold The sampled data is first converted to binary form using the threshold: values above the threshold are set to one, those below are set to zero. Both r and σ are fixed at zero, and y_1 is set to one. The model-fitting process then determines which peak or peaks are considered to be cortex: x_0 and x_1 are exactly aligned with 0-1 transitions in the binary data, as in Fig. 4, and these are taken to be the cortical edges. The chosen threshold was constant over all data sets scanned in the same way with the same scanner, corresponding to approximately 430 mg/cm^2 areal BMD.

half-max y_1 is set to the maximum CT value along the line. The model-fitting process then provides a robust estimate of the exterior and trabecular CT values (y_0, y_2) and also the location of the edges (x_0, x_1). The y_1 constraint tends to result in the edges being located on y_{blur} almost exactly half way between y_0 and y_1 , and y_2 and y_1 , as in Fig. 4. Since y_1 is the maximum observed CT value, this is very similar to the 50% relative threshold method.

new fixed The new model-fitting technique with an assumed cortical CT value (y_1) fixed over all data sets scanned in the same way with the same scanner. For the experiments reported in Section 3, a value of 1750 HU was used throughout.

new variable The new model-fitting technique with an assumed cortical CT value (y_1) chosen individually for each data set. A CT slice is selected far enough down the femur for the cortex to be sufficiently

thick, such that the blurred CT peak attains the same value as the unblurred peak (i.e. as in Fig. 2(a), but not (b) and (c)). The model is then fitted at multiple locations on this slice with both y_1 and σ unconstrained. The average estimated y_1 is then used to fix the cortical CT value for that data set.

There are, of course, many ways that the threshold and half-max methods could have been implemented. We chose implementations that allow a direct comparison with the new technique, and are sufficiently robust to provide independent estimates at a large number of locations on the proximal femur. While different implementations will undoubtedly vary in their nuances, they will nevertheless share the same high-level characteristics. Thresholding will sometimes miss the cortical layer entirely, and techniques similar to our half-max implementation, including the 50% relative threshold method, will estimate the PSF rather than the cortical thickness when the latter is relatively thin (Newman et al., 1998).

2.3 Mapping of thickness to the cortical surface

So far, we have restricted our discussion to finding a cortical thickness estimate at one location in the data, given prior knowledge of both the approximate cortical surface location and orientation. However, to probe any relationship between cortical thickness and hip fracture, and to assess the various thickness estimation techniques, we would like to be able to map cortical thickness across the entire surface of the proximal femur.

An approximate exterior cortical surface can be extracted from the CT data by thresholding at an appropriate value. After clicking on one seed point inside the femur, a 3D flood fill algorithm is used to label all connected CT voxels above the threshold. This results in a set of labelled pixels in each CT image, from which contours can be extracted to sub-pixel precision by making use of the threshold. Labelled regions frequently contain holes where the CT value in the trabecular bone drops below the threshold, but these are readily eliminated by discarding all contours that are circumscribed by another. This fully automatic process results in one or more contours on each slice. There are roughly 150 slices for the MDCT data sets with 1 mm slice thickness, and 50 for the 3 mm examples. For the high resolution Xtreme pQCT data, where the slice thickness is 0.082 mm, contours were extracted on every fourth slice, giving roughly 450 slices in total.

Following this automatic process, the contours are checked manually and edited where necessary. Figures 5(a) and (b) show the two most common causes of error. The entire set of contours is then converted to a discrete 3D shape-based representation (Treece et al., 2000), and an iso-surface extracted at the original contour locations (Treece et al., 1999). This creates a detailed, triangulated surface of the proximal femur, with sub-pixel precision except at those locations where manual editing was necessary. The whole process takes approximately 15 minutes for typical low resolution data, producing a surface with up to 20,000 vertices. Surfaces extracted from the high resolution data are far more detailed, with over 200,000 vertices and a total processing time of about an hour.

The vertices of these surfaces are used as locations for the thickness estimates, with the corresponding surface normals defining the orientation of the cortical layer. These are projected back onto the nearest CT plane: one such plane is shown in Figure 5(c). The cyan lines show the projected vertices and normals. The red points are the cortical edges (x_0, x_1) estimated at each vertex, using lines parallel to the projected normals and positioned such that each vertex lies exactly one third of the way along the corresponding line.

It should be emphasised that the segmented surface is used only to position and orient the lines along which the thickness estimates are made, and to calculate the appropriate value of a for eqs. (4) and (7). Actual cortical edges are found using the model-fitting process, and the exterior edges are generally not coincident with the projected vertices. The outcome is one thickness estimate per surface vertex. The thicknesses can be displayed collectively as a colour map painted onto the surface, as in Fig. 6(a).

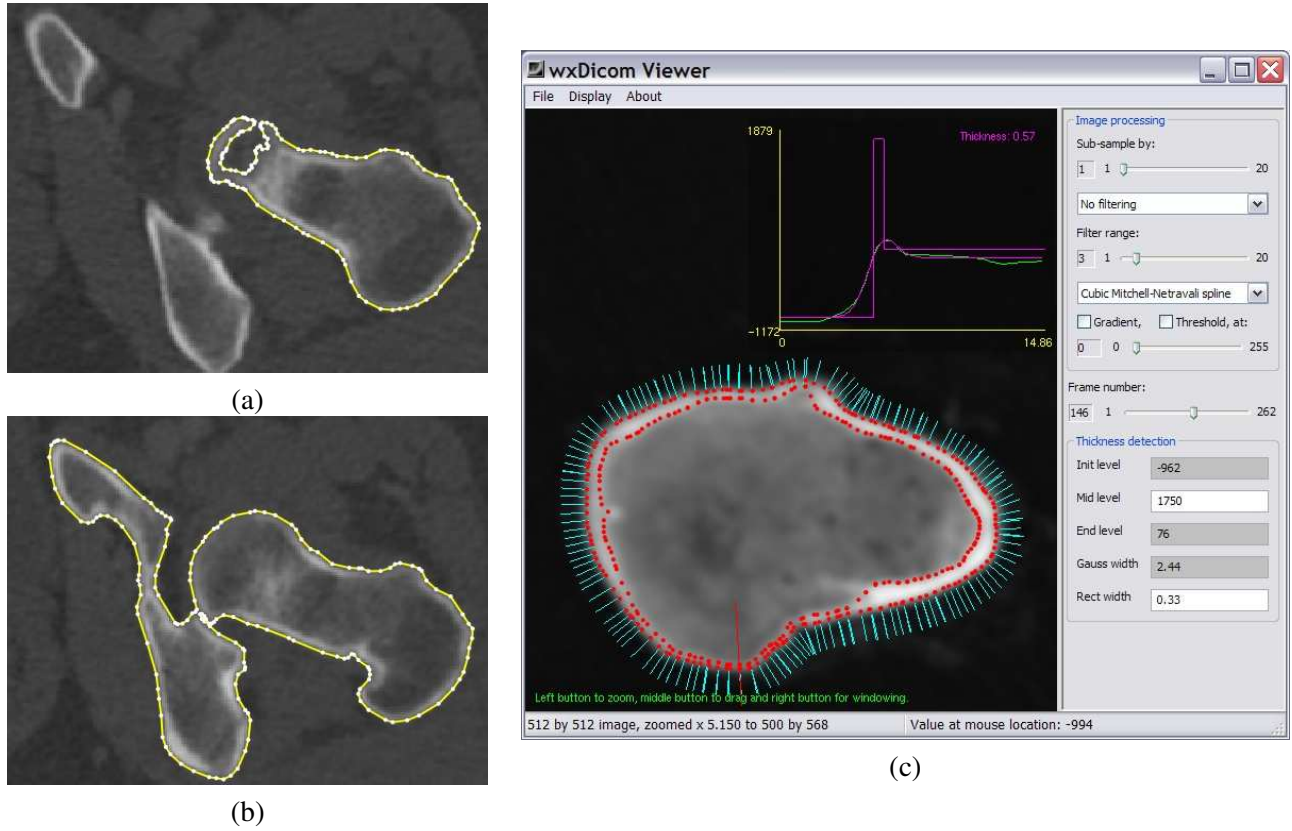


Figure 5: Creation of the surface and the use of surface normals for guiding thickness estimation. A surface is generated by thresholding the entire data set, then extracting contours in each plane to sub-pixel resolution. Contours are then locally edited (a) to correct erroneously excluded regions and (b) to remove any adjoining structures. A surface is interpolated through these contours, and the surface vertices and normals used to guide the in-plane thickness estimation. (c) The surface normals are shown in cyan, cortical edges as red dots, and the plot shows the thickness estimation process for the red line at the bottom.

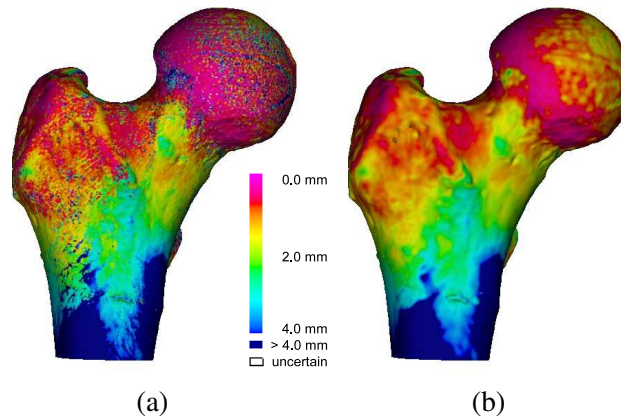


Figure 6: Mapping and filtering of high resolution thickness estimates. (a) Once cortical thickness has been estimated at each vertex, this is mapped back onto the surface as a colour. (b) The high resolution thickness map is filtered over the surface, such that the spatial resolution of the map is approximately equivalent to that of the low resolution data.

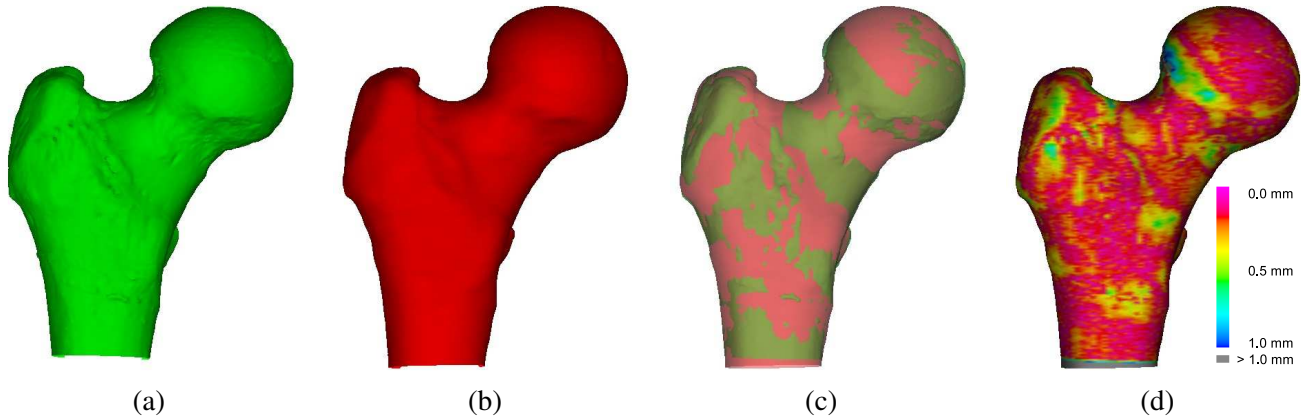


Figure 7: Registration of high and low resolution surfaces. High resolution (a) and low resolution (b) surfaces are aligned using manual registration followed by the ICP algorithm. (c) The registered surfaces. (d) Point-wise alignment discrepancies are generally well below 1 mm.

2.4 Evaluation

Sixteen cadaveric femurs were scanned in air using both high resolution Xtreme pQCT ($82\text{ }\mu\text{m}/\text{pixel}$, $82\text{ }\mu\text{m}$ slice thickness) and low resolution Siemens Somatom Sensation 64 MDCT ($589\text{ }\mu\text{m}/\text{pixel}$, 1 mm slice thickness). Cortical surfaces were extracted from these independently. Thickness estimates were mapped to the high resolution surface using the half-max method, since this is known to be a good, unbiased estimator, as long as the cortical thickness is sufficiently large compared with the imaging PSF (Prevhal et al., 1999), which is almost certainly the case for the Xtreme data (we shall return to this point in Section 4). All four thickness estimation methods were used on each low resolution data set, and compared with the ‘gold standard’ high resolution measurements. To facilitate a meaningful comparison, the high resolution thickness estimates were smoothed over each surface using a filter weighted by the inverse of the residual model-fitting errors. The extent of the smoothing kernel was chosen to match the spatial resolution of the lower resolution data, as shown in Fig. 6(b). In effect, we assume that the best we can do with the low resolution data is provide a thickness estimate that accurately represents the average thickness over the extent of the imaging PSF.

To compare corresponding high and low resolution thickness maps, the two surfaces were first brought into approximate alignment by hand, using interactive visualization software that allows the pose of one surface to be varied while the other remains stationary. The registration was then refined using a fairly standard implementation of the iterative closest point (ICP) algorithm (Besl and McKay, 1992), as illustrated in Fig. 7. For each vertex on the low resolution surface, the closest vertex on the high resolution surface was found. Since the low and high resolution scans do not cover exactly the same areas of bone, some way of ignoring unmatched vertices is required: we chose to simply ignore any vertices on the low resolution surface that were more than 1 mm away from the closest vertex on the high resolution surface. An error metric was then defined as the sum of the squared distances from each remaining vertex on the low resolution surface to the closest vertex on the high resolution surface. This metric was minimized by adjusting the pose (six degrees of freedom, translation and rotation) and the scale of the low resolution surface. The scale parameter allows for any systematic discrepancy between the two surfaces caused by the segmentation technique². Optimization was performed using the Levenberg-Marquardt method (More, 1977). The ICP algorithm involves iterating

²Thresholding the blurred, low resolution CT data produces contours that are typically 1–2% larger than the corresponding high resolution contours. Thus, seven degree-of-freedom (DOF) registration gives a slightly better fit than six DOF registration. Exploratory investigations with six DOF registration showed that the subsequent thickness comparisons were not particularly sensitive to the nuances of the registration technique.

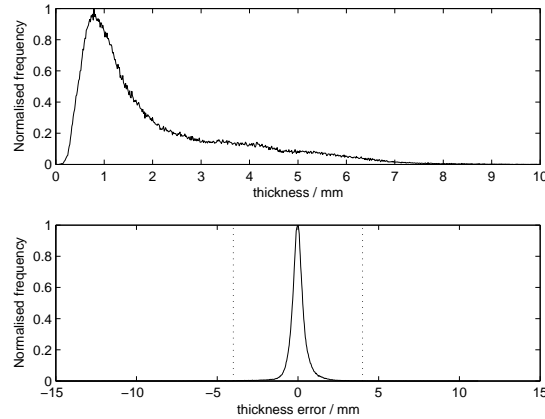


Figure 8: Distribution of cortical thickness and errors across all sixteen cadaveric data sets. The top graph shows the thickness distribution derived from the high resolution data. The bottom graph shows the error distribution for the ‘new variable’ measurement method. Despite the narrow peak, there are outlying errors up to ± 15 mm. The dotted lines indicate the 4 mm limits used to exclude poor matches from some of the subsequent error analysis.

this procedure, starting each iteration with a new search for the corresponding points, until the pose and scale parameters converge. For all sixteen femurs, convergence was typically within 100 iterations, requiring around 15 s of computation on a 2.4 GHz Intel Core 2 Duo 6600 processor.

Having registered the surfaces, each low resolution thickness estimate was compared to the geographically closest estimate on the high resolution surface. Where this closest estimate was located more than 1 mm away, the two estimates were ignored, since we do not wish alignment errors to unduly influence the results. Such misalignments occur when the high resolution surface follows small, superficial features that are not resolved in the low resolution data. In practice, the 1 mm tolerance excluded only a few percent of the low resolution thickness estimates. There remained, for each femur, up to 20,000 independent estimates, with corresponding high resolution ‘ground truth’, spanning the entire range of cortical thicknesses encountered on the proximal femur. From these, we calculated the mean (accuracy) and standard deviation (precision) of the error statistic $t_l - t_h$, where t_l is the low resolution thickness estimate and t_h is the corresponding high resolution estimate.

Figure 8(a) shows the average distribution of cortical thickness over the sixteen proximal femurs, which is well populated in the 0.25 mm to 6 mm range. Since we are more interested in thinner cortical regions, we use 0 mm to 4 mm colour maps for surface renderings such as those in Fig. 6. The error distribution for the ‘new variable’ thickness estimation technique is shown in Fig. 8(b). While there is a narrow peak, outlying errors extend up to ± 15 mm.

Many of these outliers can be attributed to the evaluation protocol as opposed to the thickness estimation technique. Figure 9 shows two situations where the protocol leads to inappropriate comparisons between thickness estimates. Figure 9(d), in particular, demonstrates the difficulty defining cortical thickness at two different resolutions when the cortex is porous. To eliminate such inappropriate comparisons from our results, we chose a threshold of 4 mm, above which we assume the error is more likely to be an artefact of the evaluation protocol than the thickness estimation technique. Clearly, this threshold might also remove genuine estimation errors, although Fig. 8(b) shows that these will be very few in number. It is also unlikely that an error exceeding 4 mm would be mistaken for a correct measurement in practice. Nevertheless, in Section 3 we present error statistics both with and without the 4 mm cut-off, so the reader can judge for themselves.

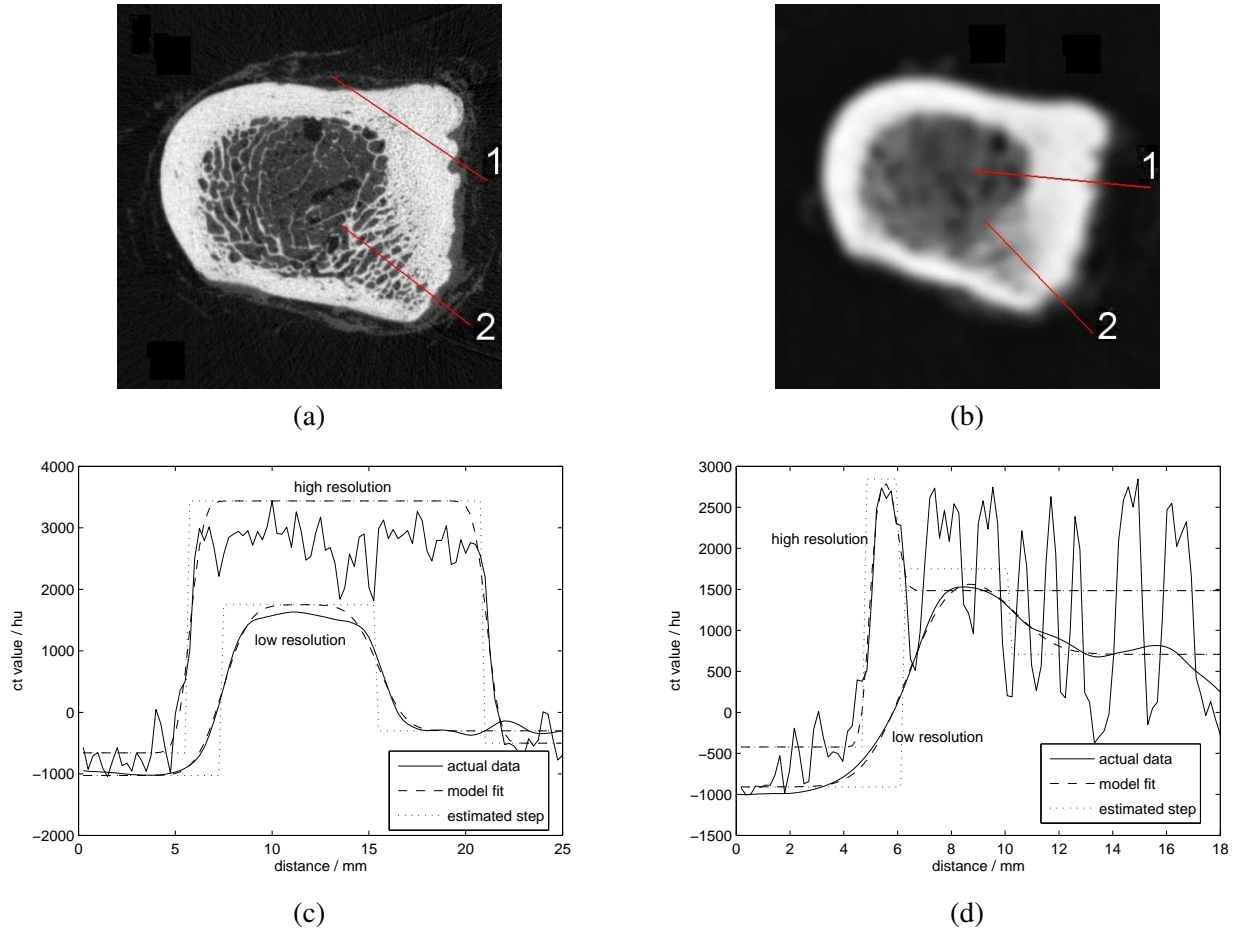


Figure 9: Difficulties comparing low and high resolution thickness measurements. (a) and (b) show corresponding high and low resolution slices through the CT data. (c) and (d) are plots of the CT values along the lines labelled (1) and (2) respectively. In (c), the surface normal in the high resolution data is not compatible with that of the low resolution data. In (d), there is ambiguity in the definition of cortical thickness: both estimators detect the outermost resolvable layer, but this layer is not the same at the two resolutions.

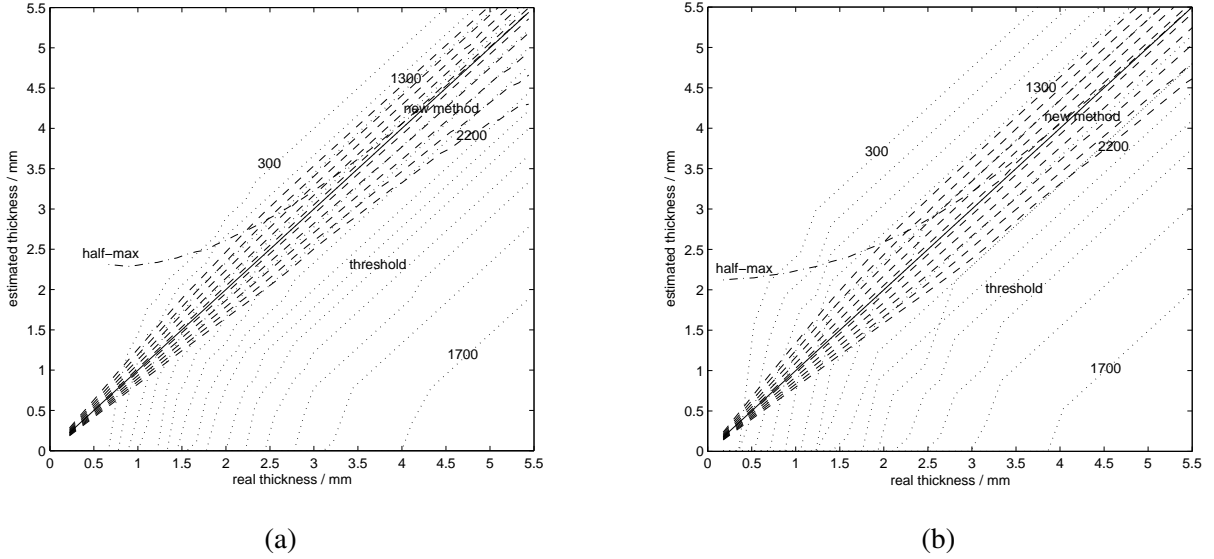


Figure 10: Thickness estimation on simulated data with varying cortical thickness. For a range of cortical thicknesses from 0.2 mm to 5.5 mm, the CT value on a line passing through the cortical layer was simulated by blurring a piecewise constant density function with a Gaussian kernel of standard deviation 1.25 mm. The thickness was then estimated using the half-max method, thresholding at CT values from 300 to 1700, and the new method assuming a cortical CT value (y_1) of 1300 to 2200. (a) shows a simulated scan in air (density goes from -1000 to 1750 to 0) and (b) shows an *in vivo* simulation (density goes from 0 to 1750 to 0). In both cases, the half-max method overestimates at low thicknesses. Thresholding gives better results for carefully chosen thresholds: note how the optimal threshold varies significantly from (a) to (b). The new method gives consistently good results, with relatively little sensitivity to the assumed cortical CT value.

3 Results

3.1 Simulated data

The thickness estimation methods described in Section 2.2 were first tested on simulated data, in order to investigate performance and parameter sensitivity in an ideal, noise-free environment. The model of eq. (5) was used to simulate the CT values along lines passing through cortices of varying thickness in the range 0.2 mm to 5.5 mm, with an in-plane blur σ of 1.25 mm and no out-of-plane blur (i.e. the cortical layer was assumed to lie orthogonal to the CT imaging plane). These values were chosen to match approximately those of the low resolution cadaveric data sets. The half-max, threshold and new methods were then used to estimate the original, known, cortical thickness. The latter two methods were evaluated across a range of thresholds and assumed cortical CT values (y_1).

Figure 10 shows results for simulated scans in air (a) and *in vivo* (b). As reported by Newman et al. (1998), the half-max method is accurate for large thicknesses but increasingly biased for smaller thicknesses. The thresholding technique is potentially more accurate below 3 mm, provided the threshold is chosen carefully. For the air simulation, a threshold of 600 HU (for the low resolution data, this corresponds to an areal BMD of approximately 430 mg/cm^2) produces reasonable thickness estimates down to 1.3 mm. For the *in vivo* simulation, the ideal threshold is higher, at around 800 HU. Below 1.3 mm, thresholding starts to ‘miss’ the cortex and becomes increasingly unreliable, as previously noted by Poole et al. (2008). The new method generates accurate results down to the simulated thickness limit of 0.2 mm. This in itself is not surprising:

Table 1: Summary statistics for the sixteen cadaveric femur data sets. N is the number of vertices on the low resolution surfaces for which a corresponding ‘ground truth’ thickness is available. ‘Percentage success’ indicates how frequently each method produced any sort of thickness estimate. The final column shows the assumed cortical CT value used for the ‘new variable’ method. The ‘new fixed’ method assumed a value of 1750 HU throughout.

ID	Sex	Age	Wt. (kg)	Ht. (cm)	N	Percentage success				y_1 (HU)
						Half-max	Threshold	New fix.	New var.	
443	F	74	69	160	17794	99.92	76.72	99.97	99.97	1750
456	F	68	53	167	16732	99.93	78.53	99.92	99.90	1800
457	F	75	69	153	16267	99.70	54.60	99.63	99.50	1760
458	F	66	77	167	18989	99.95	76.47	99.95	99.95	1680
464	M	19	66	164	17873	99.47	73.87	99.13	99.13	1820
465	F	77	56	150	17233	99.76	63.31	99.53	99.47	1710
466	F	63	62	148	15154	99.85	46.21	99.95	99.97	1900
467	F	41	104	179	17194	99.92	86.73	99.71	99.74	1720
470	F	80	69	157	19302	99.90	73.73	99.77	99.77	1780
473	F	42	58	173	16612	99.95	81.73	99.72	99.82	1820
474	M	21	74	182	20324	99.92	69.13	99.94	99.94	1750
477	F	47	83	163	16798	99.76	72.22	99.54	99.77	1810
483	F	40	42	152	11722	99.90	86.86	99.85	99.86	1850
485	F	40	63	167	18537	99.76	60.35	99.88	99.83	1720
493	M	70	79	161	19436	99.92	72.93	99.91	99.89	1720
494	M	83	60	157	19012	99.95	67.42	99.88	99.92	1660
all	-	-	-	-	278979	99.85	71.18	99.77	99.78	-

we have, after all, used exactly the same model to simulate as to estimate thickness. More noteworthy is the low degree of sensitivity to the assumed cortical CT value, especially at thicknesses below 2 mm. Such thin cortices are precisely those likely to be of most clinical interest.

3.2 Cadaveric data

Summary statistics for the sixteen cadaveric data sets can be found in Table 1. For the half-max and new methods, the percentage success rate indicates how frequently the model-fitting algorithm managed to converge and hence produce a result. For the threshold method, the number indicates how frequently there were *any* CT values above the threshold along the line used for thickness estimation. Neither convergence nor the existence of supra-threshold values implies anything about the accuracy of the resulting thickness estimate. The processing time was similar for all methods. Loading the CT data, estimating cortical thickness at approximately 17,000 locations and saving the results to disk (but not calculating or registering the surfaces) took around five minutes on a 2.16 GHz Intel Core2 CPU, with the software running single-threaded.

Table 2 shows estimation accuracy and precision for the entire range of errors, i.e. not making any allowance for likely evaluation artefacts like those in Fig. 9. While it is immediately apparent that both variants of the new method are less biased and more precise than either of the alternative methods, we must at this point take a brief diversion, to test the claim in Section 2.2 that constraining y_1 is the best strategy for conditioning the model-fitting process. To this end, we performed some supplementary experiments on femur 477, chosen since it is an average specimen in terms of the estimation errors in Table 2. In Section 2.2, we considered three possibilities: constraining the assumed cortical CT value y_1 ; constraining the in-plane blur σ ; or

Table 2: Cortical thickness estimation error, considering the entire range of errors and thicknesses. The tabulated quantity is the mean and standard deviation of $t_l - t_h$, for all successful low resolution thickness estimates for which a high resolution match could be found.

ID	Mean error \pm 1 std. dev. (mm)			
	Half-max	Threshold	New fixed	New variable
443	1.17 ± 1.29	0.36 ± 1.18	0.19 ± 0.82	0.19 ± 0.82
456	0.92 ± 1.31	0.67 ± 2.07	0.00 ± 0.89	-0.05 ± 0.87
457	1.18 ± 2.39	0.14 ± 0.64	0.07 ± 0.74	0.07 ± 0.83
458	1.23 ± 1.32	0.36 ± 1.20	0.12 ± 0.74	0.21 ± 0.78
464	1.10 ± 1.05	0.47 ± 0.75	0.10 ± 0.54	0.05 ± 0.55
465	0.95 ± 1.18	0.24 ± 0.96	0.02 ± 0.68	0.07 ± 0.71
466	1.28 ± 3.91	0.18 ± 0.81	0.07 ± 0.52	-0.02 ± 0.53
467	0.74 ± 1.52	0.40 ± 1.80	-0.15 ± 1.19	-0.14 ± 1.15
470	1.16 ± 1.47	0.30 ± 1.16	0.05 ± 0.77	0.05 ± 0.77
473	1.02 ± 0.86	0.47 ± 0.94	0.13 ± 0.47	0.07 ± 0.45
474	1.09 ± 1.21	0.40 ± 1.34	-0.09 ± 0.58	-0.09 ± 0.58
477	0.77 ± 1.06	0.22 ± 0.84	0.00 ± 0.72	-0.08 ± 0.67
483	0.96 ± 1.32	0.35 ± 1.32	0.12 ± 0.88	0.01 ± 0.81
485	1.12 ± 1.80	0.03 ± 0.50	-0.12 ± 0.52	-0.08 ± 0.48
493	1.14 ± 1.35	0.29 ± 1.05	-0.03 ± 0.62	0.00 ± 0.62
494	0.94 ± 2.07	0.68 ± 2.22	0.00 ± 1.39	0.02 ± 1.17
all	1.05 ± 1.70	0.36 ± 1.31	0.03 ± 0.80	0.02 ± 0.77

using no constraints at all. With femur 477, no constraints resulted in an estimation error of 0.90 ± 4.04 mm, constraining σ (to a sensible median value obtained from the previous experiment) but not y_1 gave an error of 0.66 ± 3.75 mm, whereas fixing y_1 at 1750 HU produced by far the best results of 0.00 ± 0.72 mm. This is typical of our experience with all the data sets we have examined. There was very little difference in the percentage success rates for each of these variants of the new technique.

Table 3 presents the results after accounting for likely evaluation artefacts using the ± 4 mm error cut-off suggested in Section 2.4. The order of merit of the various methods is the same as in Table 2, though clearly all the standard deviations are reduced. These results may better reflect the actual measurement error across the entire surface of the proximal femur. Finally, Table 4 presents results confined to areas where the cortex is relatively thin, in the range 0.3 mm to 2 mm. Such areas are likely to be the focus of interest when assessing fracture risk. For thin cortices, the differences between the methods are even more pronounced.

Figure 11 expands the results in Table 3 to show performance as a function of true cortical thickness. Since the techniques produce so many independent thickness estimates in the range 0.3 mm to 4 mm (Fig. 8(a)), splitting the data for any one femur into bins 0.1 mm wide produces over 100 values per bin, and well over 1000 values when the data for all femurs is combined. This is sufficient to analyse performance as a function of thickness in the aforementioned range. The conglomerated results in Fig. 11(c) have more in common with the results for the best case (femur 473, Fig. 11(a)) than those for the worst case (femur 467, Fig. 11(b)), suggesting that the latter is the exception.

Figure 11 also shows percentage success rates as a function of thickness. The threshold method starts to ‘miss’ estimates as the true cortical thickness drops below 1.5 mm, and is largely incapable of producing any reliable results below 0.75 mm. For the carefully chosen threshold, the estimates are reasonably unbiased down to 1 mm. The half-max and ‘new variable’ methods both generate cortical thickness estimates over the full 0.3 mm to 4 mm range. However, the half-max method overestimates the thickness below 3 mm, whereas

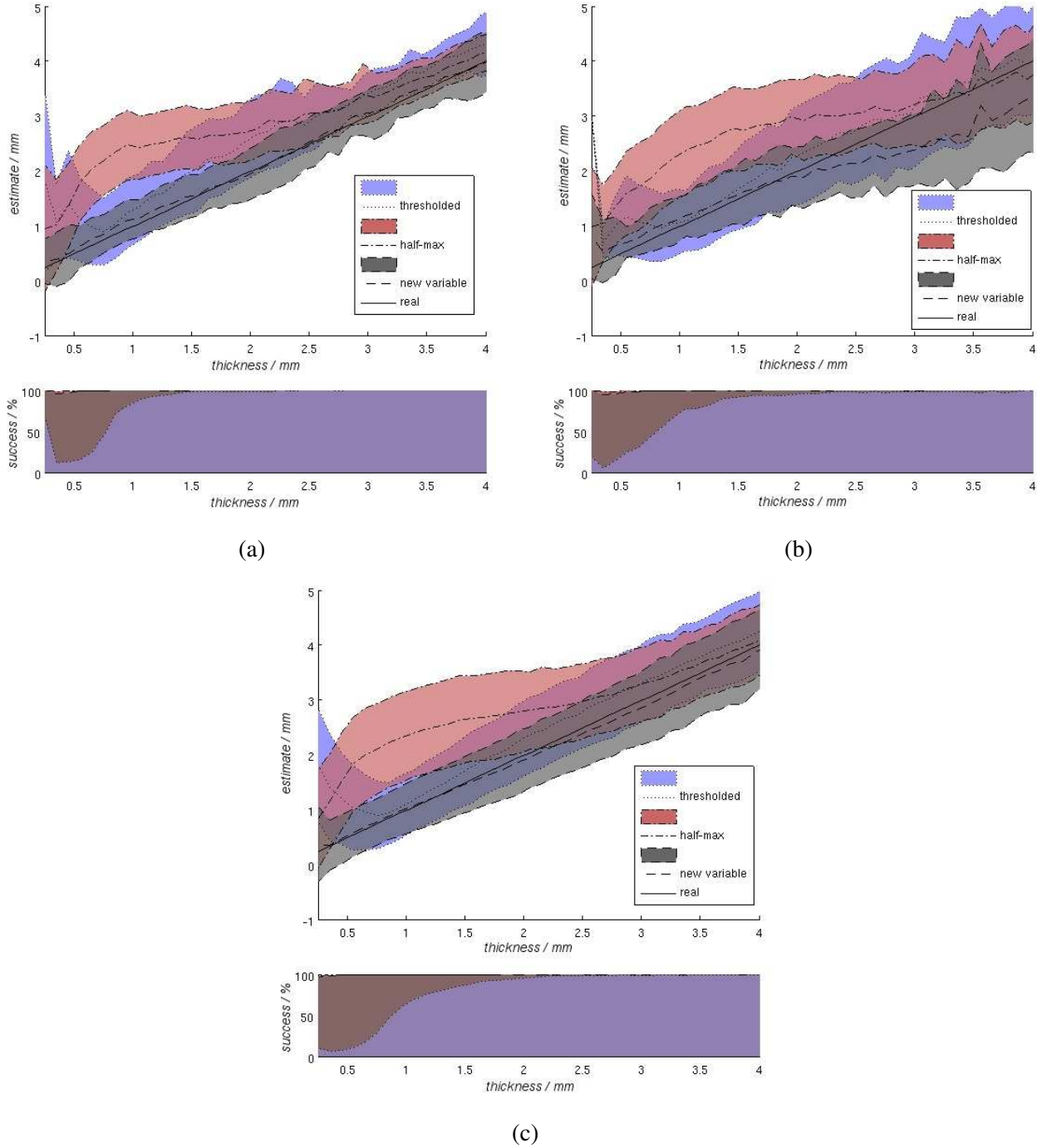


Figure 11: Cadaveric cortical thickness estimation as a function of thickness. The upper graph in each case shows the mean ± 1 standard deviation of t_l on the y -axis, for each 0.1 mm band of t_h on the x -axis. This is similar to the presentation used for the simulated data in Fig. 10. The lower graph indicates the percentage of successful thickness estimates in each band. (a) shows results for the best case (473), (b) for the worst (467) and (c) the combined results across all sixteen femurs. The three colours have been rendered with some transparency, so the ± 1 standard deviation ranges remain distinguishable where they overlap.

Table 3: Cortical thickness estimation error, for errors below 4 mm and the entire range of thicknesses. The tabulated quantity is the mean and standard deviation of $t_l - t_h$. By ignoring the very small number of errors above 4 mm, the results might better reflect the merits of the various techniques and not the evaluation shortcomings identified in Fig. 9.

ID	Mean error \pm 1 std. dev. (mm)			
	Half-max	Threshold	New fixed	New variable
443	1.02 ± 0.90	0.28 ± 0.71	0.16 ± 0.61	0.16 ± 0.61
456	0.78 ± 0.91	0.38 ± 0.77	-0.06 ± 0.61	-0.11 ± 0.60
457	1.04 ± 0.83	0.14 ± 0.56	0.03 ± 0.48	0.03 ± 0.47
458	1.05 ± 0.83	0.29 ± 0.65	0.10 ± 0.57	0.17 ± 0.58
464	1.02 ± 0.85	0.43 ± 0.54	0.09 ± 0.48	0.04 ± 0.47
465	0.84 ± 0.87	0.20 ± 0.68	0.02 ± 0.59	0.06 ± 0.60
466	1.09 ± 0.90	0.17 ± 0.56	0.06 ± 0.46	-0.02 ± 0.44
467	0.59 ± 1.06	0.17 ± 0.93	-0.21 ± 0.84	-0.18 ± 0.85
470	0.94 ± 0.94	0.21 ± 0.64	0.02 ± 0.56	0.02 ± 0.56
473	0.97 ± 0.74	0.40 ± 0.61	0.12 ± 0.43	0.07 ± 0.42
474	1.00 ± 0.94	0.29 ± 0.68	-0.09 ± 0.51	-0.09 ± 0.51
477	0.68 ± 0.77	0.18 ± 0.65	-0.01 ± 0.57	-0.09 ± 0.55
483	0.83 ± 0.96	0.30 ± 0.82	0.05 ± 0.59	-0.04 ± 0.58
485	1.00 ± 1.00	0.02 ± 0.46	-0.12 ± 0.46	-0.08 ± 0.45
493	0.96 ± 0.93	0.23 ± 0.61	-0.04 ± 0.55	-0.01 ± 0.54
494	0.48 ± 0.93	0.23 ± 0.79	-0.18 ± 0.70	-0.09 ± 0.73
all	0.90 ± 0.92	0.25 ± 0.69	-0.01 ± 0.58	-0.01 ± 0.58

the new method provides relatively unbiased estimates at all thicknesses. In general, the new method also compares favourably with the others in terms of precision, again across the entire range of thicknesses.

Finally, the results are presented as a function of location on the proximal femur, for the best and worst cases in Figs. 12 and 13 respectively. These images are stills from a user interface which displays the cortical thickness as a colour map, as described in Section 2.3. The surface can be examined from any viewpoint. The light grey colour, mostly evident for the threshold method in the second column, indicates regions in which no estimate is available. Away from these regions, the threshold method produces thickness maps that appear similar to the high resolution ‘ground truth’ in the first column. The half-max method (third column) generates few thickness estimates below 2 mm (green). The ‘new variable’ method (fourth column) produces thickness maps that are similar to the high resolution maps, with absolute errors generally less than 0.5 mm (pink in the error maps).

3.3 In vivo data

The various methods for cortical thickness estimation were assessed using cadaveric femurs scanned in air, since the human proximal femur cannot be imaged *in vivo* at the high resolutions required to establish the ‘ground truth’. In clinical CT scans, the femur is of course surrounded mostly by soft tissue, not air. While the new thickness estimation technique makes no assumption about the surrounding material, except that it is reasonably uniform³, it is nevertheless reasonable to question how performance might vary between cadaveric

³It should be noted that the cadaveric femurs were not cleaned before scanning and there was some residual tissue attached to the outer cortical surface. The combination of residual tissue and air may in fact have made the estimation process more difficult, since the exterior CT value y_0 would have been estimated in an area not well characterized by a constant value.

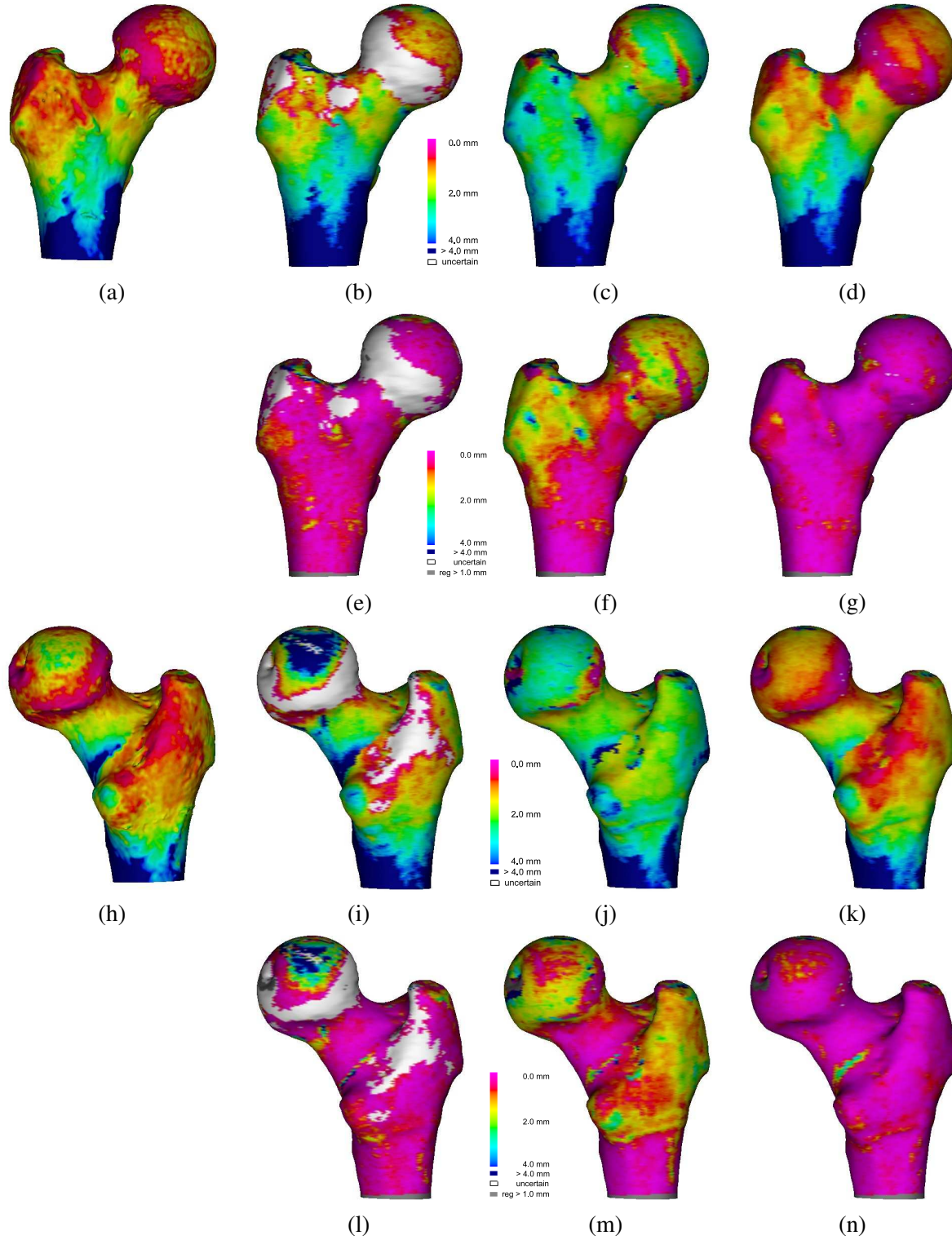


Figure 12: Thickness estimation for the best case (473). (a) to (d) are thickness maps from the front, (h) to (k) from the back. (e) to (g) and (l) to (n) are absolute thickness error maps. The first column (a, h) shows the high resolution 'ground truth' measurements. The second column (b, e, i, l) shows the thresholding results, the third column (c, f, j, m) half-max, and the fourth column (d, g, k, n) the 'new variable' results.

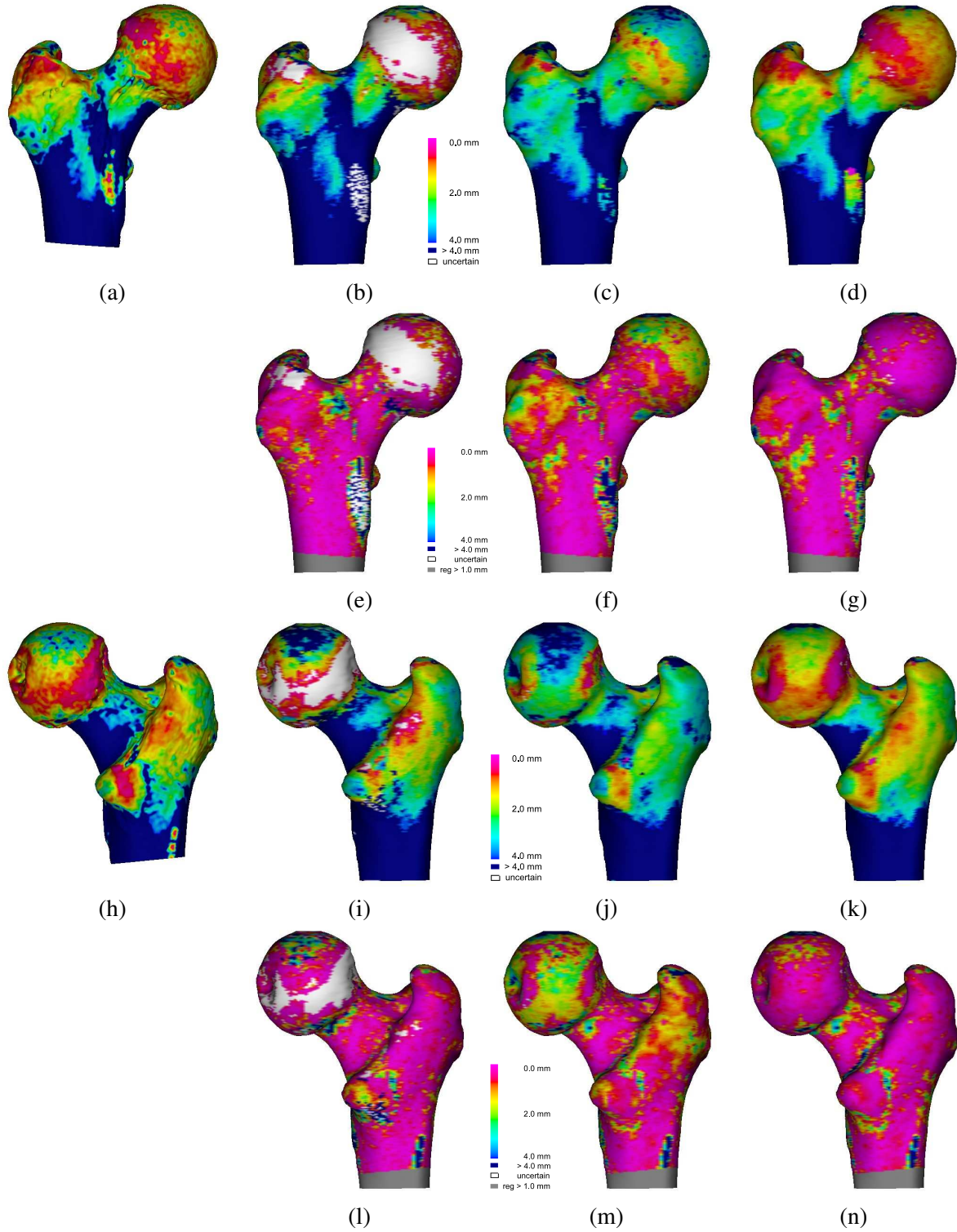


Figure 13: Thickness estimation for the worst case (467). (a) to (d) are thickness maps from the front, (h) to (k) from the back. (e) to (g) and (l) to (n) are absolute thickness error maps. The first column (a, h) shows the high resolution 'ground truth' measurements. The second column (b, e, i, l) shows the thresholding results, the third column (c, f, j, m) half-max, and the fourth column (d, g, k, n) the 'new variable' results.

Table 4: Cortical thickness estimation error, for errors below 4 mm and thicknesses in the range 0.3 mm to 2 mm. The tabulated quantity is the mean and standard deviation of $t_l - t_h$. The areas of most clinical interest are likely to be those where the cortex is particularly thin. Significantly, it is in precisely these areas that the various estimation techniques differ most in terms of performance.

ID	Mean error \pm 1 std. dev. (mm)			
	Half-max	Threshold	New fixed	New variable
443	1.43 ± 0.74	0.19 ± 0.58	0.14 ± 0.48	0.14 ± 0.48
456	1.16 ± 0.83	0.26 ± 0.73	-0.04 ± 0.51	-0.07 ± 0.49
457	1.29 ± 0.73	0.06 ± 0.44	0.06 ± 0.43	0.05 ± 0.43
458	1.39 ± 0.70	0.24 ± 0.60	0.12 ± 0.48	0.17 ± 0.50
464	1.24 ± 0.84	0.35 ± 0.53	0.06 ± 0.48	0.03 ± 0.47
465	1.10 ± 0.83	-0.06 ± 0.47	-0.02 ± 0.45	0.00 ± 0.47
466	1.37 ± 0.83	0.00 ± 0.41	0.03 ± 0.41	-0.02 ± 0.39
467	1.19 ± 0.76	0.18 ± 0.72	0.07 ± 0.48	0.10 ± 0.49
470	1.33 ± 0.82	0.19 ± 0.65	0.07 ± 0.51	0.07 ± 0.51
473	1.26 ± 0.65	0.39 ± 0.67	0.13 ± 0.41	0.09 ± 0.39
474	1.29 ± 0.88	0.17 ± 0.61	-0.07 ± 0.43	-0.07 ± 0.43
477	0.98 ± 0.67	0.14 ± 0.54	-0.04 ± 0.42	-0.07 ± 0.38
483	1.32 ± 0.81	0.36 ± 0.73	0.13 ± 0.54	0.07 ± 0.52
485	1.32 ± 0.95	-0.11 ± 0.40	-0.12 ± 0.44	-0.10 ± 0.44
493	1.29 ± 0.89	0.19 ± 0.65	-0.01 ± 0.50	0.00 ± 0.50
494	0.71 ± 0.93	0.17 ± 0.70	-0.22 ± 0.54	-0.19 ± 0.54
all	1.24 ± 0.83	0.19 ± 0.62	0.01 ± 0.48	0.01 ± 0.47

and *in vivo* examples.

Figure 14 shows cortical thickness maps produced using the ‘new variable’ method, working with *in vivo* MDCT data at 1.25 mm slice thickness. There is no ground truth for this data, and it is just a single example, so it cannot be used to draw any firm conclusions. However, there is reasonably good agreement between the independent maps on the fractured and contralateral hip; the new method produces estimates across most of the proximal femur; and the maps are not dissimilar to those in Figs. 12 and 13. The feasibility of applying this technique to fractured hips *in vivo* has potential clinical significance. Compared with intact femurs, the surface extraction stage was less straightforward (more manual editing was required), but the thickness estimation stage was no different, since all measurements are mutually independent.

Clinical data is, of course, typically acquired at slice thicknesses exceeding 1 mm. Figure 15 shows cortical thickness maps estimated using the ‘new variable’ method, starting from the same *in vivo* scan but with different CT reconstructions at 1 mm and 3 mm slice thickness. Again, we can draw no firm conclusions from a single example, but the 0.01 ± 0.30 mm discrepancy between the two maps is remarkably low. This gives reasonable grounds for optimism regarding the applicability of the method to 3 mm slice CT data.

The light grey regions in Figs. 14 and 15 indicate areas where the new method failed to produce a thickness estimate. This was due to the close proximity of the acetabulum to the femoral head, preventing the successful estimation of y_0 . The cadaveric maps are free of such regions because, although the femurs were supported by a pelvis phantom, the acetabulum was not as close to the femoral head as would be expected *in vivo*.

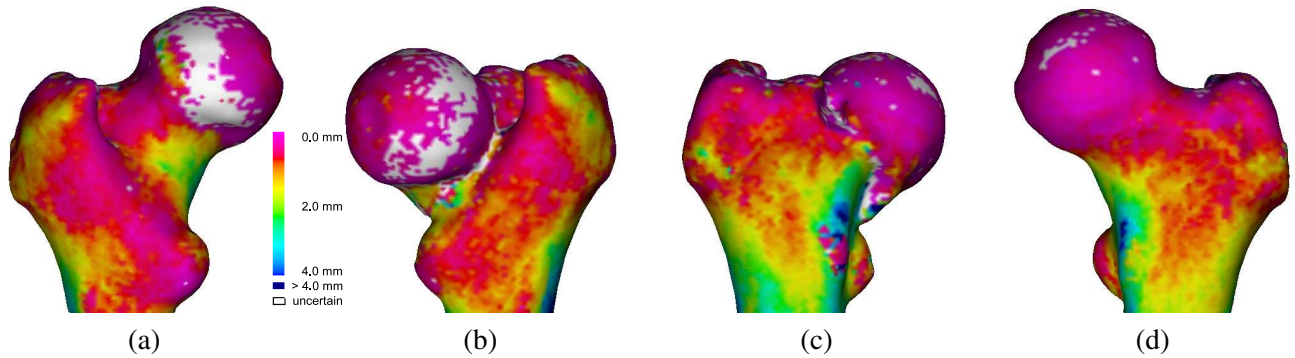


Figure 14: Cortical thickness estimation for an 87 year old male volunteer with a sub-capital femoral neck fracture. The maps were produced using the ‘new variable’ method. (a) and (b) show the contralateral and fractured femurs from the back, (c) and (d) from the front. This demonstrates the feasibility of estimating cortical thickness maps *in vivo*, even for fractured femurs. The fractured and contralateral maps are reassuringly similar.

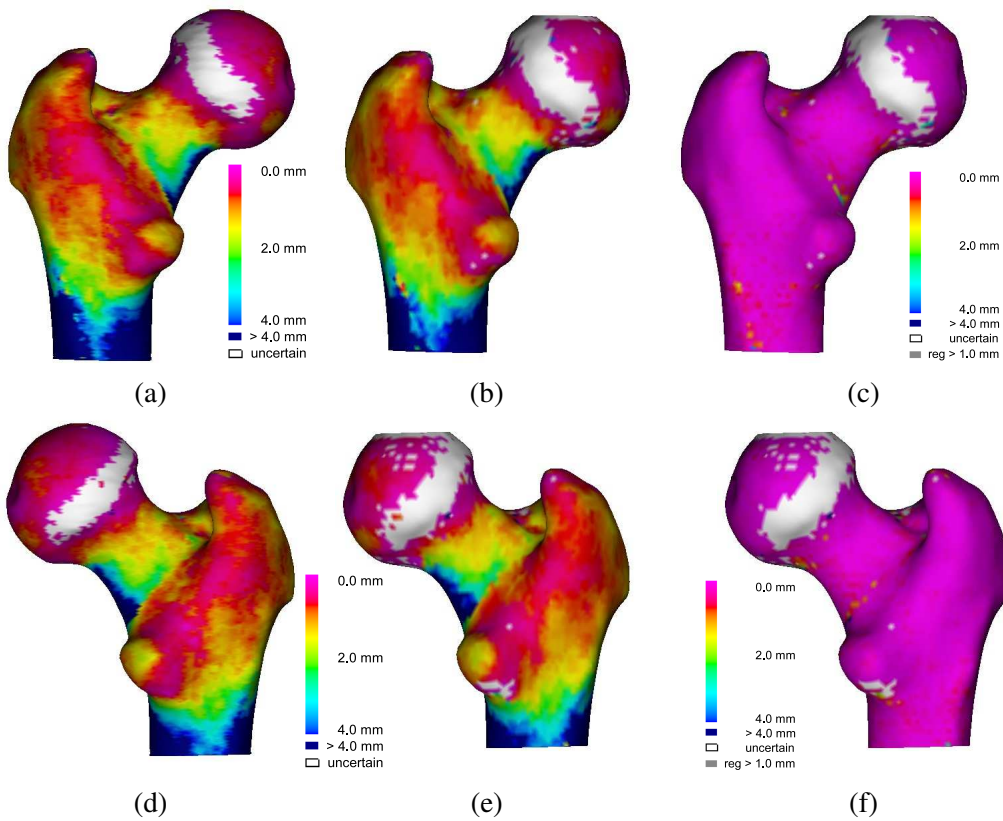


Figure 15: Cortical thickness estimation at 1 mm and 3 mm CT slice thickness for a 39 year old female volunteer. The maps were produced using the ‘new variable’ method. (a) and (d) are estimates from CT data reconstructed at 1 mm slice thickness, (b) and (e) from the same data reconstructed at 3 mm slice thickness. (c) and (f) are absolute difference maps, revealing very few discrepancies between the two sets of results.

4 Discussion

In general, the various thickness estimation techniques were observed to perform in accordance with the assumed model. However, close inspection of the graphs in Fig. 11(c) reveals two clear differences from the simulations in Fig. 10. While the half-max estimates do, as predicted, attain a plateau of around 2.5 mm (consistent with an estimated in-plane blur of $\sigma = 1.25$ mm), at true cortical thicknesses below 1 mm they drop below this plateau, in a manner not explained by the model. Similarly, the threshold estimates inexplicably show a sharp increase at low cortical thicknesses. To understand these phenomena, we must first rule out any systematic inaccuracies in the ‘gold standard’ thicknesses. Recall, these were obtained by applying the half-max method to the Xtreme pQCT data, which is approximately seven times higher resolution than the Siemens MDCT data. Figure 11(c) shows the half-max method starting to produce biased thickness estimates below 3 mm, this being with the low resolution data. It follows, therefore, that we should not trust our ‘gold standard’ much below $3/7 = 0.43$ mm. This is sufficiently close to the lower limit of 0.3 mm in Fig. 11(c) that it is unlikely to have a significant bearing on the results.

Instead, explanations for these two phenomena become apparent on close examination of the data where the cortex is very thin. For the half-max method, the cortical CT value y_1 becomes poorly defined and is often very similar to the trabecular value y_2 . Consequently, the inner cortical edge x_1 can be positioned almost arbitrarily, resulting in spurious thickness estimates that do not necessarily reflect the width of the imaging PSF. The strange behaviour of the threshold graph is driven by a relatively small number of thickness estimates: in the vast majority of cases, the imaged cortex does not exceed the threshold and the threshold method fails. However, occasionally the trabecular bone, at the y_2 end of the estimation line, has sufficiently high density to exceed the threshold. Thresholding then does produce a thickness estimate, but not one of the thin cortex, rather one of a thicker region of trabecular bone. This is why the threshold graph in Fig. 11(c) turns upwards.

The thickness maps in Figs. 12 and 13 show that the errors are not homogeneously distributed across the surface. Instead, they are clustered at small focal patches. We observed similar behaviour in all sixteen data sets. The most likely cause of such errors is an ambiguous cortical-trabecular interface in these regions, suggesting different interpretations in the low and high resolution data. Certainly, the ‘worst’ data set (467) has a particularly dense matrix structure in the trabecular bone. The image in Fig. 9 is from this data set: at certain locations, such as around line 2, it is not at all clear where the cortex ends and the trabecular bone begins. While we attempted to eliminate such artefacts from our results by ignoring outlying errors, cortical ambiguity will not always result in a discrepancy exceeding the 4 mm cut-off, and the variances in Table 3 might therefore be seen as upper limits on the true estimation errors. Where the cortex is well defined, estimation errors are likely to be well within these bounds: witness the pink regions in the surface error maps (Figs. 12 and 13, (g) and (n)), where the error is everywhere less than 0.5 mm.

5 Conclusions

We have presented a novel method for estimating cortical thickness from clinical CT data. In a series of cadaveric experiments using standard MDCT (589 $\mu\text{m}/\text{pixel}$ resolution, in-plane blur well approximated by a Gaussian PSF of standard deviation 1.25 mm), the method was able to produce unbiased estimates down to 0.3 mm. Errors were observed to be -0.01 ± 0.58 mm over the full range of thicknesses, and 0.01 ± 0.47 mm for thin cortices in the range 0.3 mm to 2 mm. This is a clear improvement over the alternative thresholding and 50% relative threshold techniques, which are particularly poor when the cortex is thin. Thin cortices are precisely those likely to be of most clinical interest.

The new method is highly automated and typically produces 17,000 independent thickness estimates over

the surface of the proximal femur. These can be displayed as a colour map painted onto the surface, and visualized from any angle. Workflow is already in the realms of clinical practicality: around 10 minutes to segment the surface, then around 5 minutes to calculate the thicknesses, using just a modest laptop computer. Faster segmentation might be possible with the use of anatomically specific tools.

Future work will seek to further validate the new technique *in vivo*, using CT slice thicknesses of 3 mm and above, though Figs. 14 and 15 give grounds for optimism. In the longer term, it is anticipated that cortical thickness maps, like the ones in Figs. 12–15, will become a useful tool for both medical research and clinical practice. With inter-subject registration, statistical atlases could be constructed showing cortical thickness over the proximal femur for different groups: e.g. young vs. old, trochanteric vs. femoral neck fractures, cases vs. controls. By comparing such atlases, regions of interest might emerge to serve as biomarkers indicative of fracture risk. Cortical thickness maps might also inform longitudinal studies of patients undergoing anabolic therapy, by determining precisely where any thickening occurs.

Acknowledgments

K. Poole is supported by the Arthritis Research Campaign, the Evelyn Trust and the NIHR Cambridge Biomedical Research Centre. The cadaveric femurs were from the *Melbourne Femur Collection Research Tissue Bank of the Victorian Institute of Forensic Medicine*, with kind permission of Professor John Clement, Australia.

References

- Besl, P. J., McKay, N. D., Feb. 1992. A method for registration of 3-D shapes. *IEEE Transactions on Pattern Analysis and Machine Intelligence* 14 (2), 239–256.
- Bouxsein, M. L., Delmas, P. D., Aug. 2008. Considerations for development of surrogate endpoints for anti-fracture efficacy of new treatments in osteoporosis: A perspective. *Journal of Bone and Mineral Research* 23 (8), 1155–1167.
- Buie, H. R., Campbell, G. M., Klinck, R. J., MacNeil, J. A., Boyd, S. K., Oct. 2007. Automatic segmentation of cortical and trabecular compartments based on a dual threshold technique for *in vivo* micro-CT bone analysis. *Bone* 41 (4), 505–515.
- Burger, W., Burge, M. J., 2009. *Principles of Digital Image Processing: Core Algorithms*. Springer-Verlag.
- Carpenter, R. D., Beaupre, G. S., Lang, T. F., Orwoll, E. S., Carter, D. R., 2005. New QCT analysis approach shows the importance of fall orientation on femoral neck strength. *Journal of Bone and Mineral Research* 20 (9), 1533–1542.
- de Bakker, P. M., Manske, S. L., Ebacher, V., Oxland, T. R., Cripton, P. A., Guy, P., Aug. 2009. 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.
- Dougherty, G., Newman, D., Jul. 1999. Measurement of thickness and density of thin structures by computed tomography. *Medical Physics* 26 (7), 1341–1348.
- Hangartner, T. N., Mar. 2007. Thresholding technique for accurate analysis of density and geometry in QCT, PQCT and μ CT images. *Journal of Musculoskeletal and Neuronal Interactions* 7 (1), 9–16.

- Hangartner, T. N., Gilsanz, V., 1996. Evaluation of cortical bone by computed tomography. *Journal of Bone and Mineral Research* 11 (10), 1518–1525.
- Holzer, G., von Skrbensky, G., Holzer, L. A., Pichl, W., 2009. Hip fractures and the contribution of cortical versus trabecular bone to femoral neck strength. *Journal of Bone and Mineral Research* 24 (3), 468–474.
- Johnell, O., Kanis, J. A., Oden, A., Johansson, H., Laet, C. D., Delmas, P., Eisman, J. A., Fujiwara, S., Kroger, H., Mellstrom, D., Meunier, P. J., 3rd, L. J. M., O'Neill, T., Pols, H., Reeve, J., Silman, A., Tenenhouse, A., 2005. Predictive value of bmd for hip and other fractures. *Journal of Bone and Mineral Research* 20 (7), 1185–1194.
- Kanis, J. A., Burlet, N., Cooper, C., Delmas, P. D., Reginster, J. Y., Borgstrom, F., Rizzoli, R., 2008. European guidance for the diagnosis and management of osteoporosis in postmenopausal women. *Osteoporosis International* 19 (4), 399–428.
- Kaptoge, S., Beck, T. J., Reeve, J., Stone, K. L., Hillier, T. A., Cauley, J. A., Cummings, S. R., 2008. Prediction of incident hip fracture risk by femur geometry variables measured by hip structural analysis in the study of osteoporotic fractures. *Journal of Bone and Mineral Research* 23 (12), 1892–1904.
- Keaveny, T. M., Hoffmann, P. F., Singh, M., Palermo, L., Bilezikian, J. P., Greenspan, S. L., Black, D. M., Dec. 2008. Femoral bone strength and its relation to cortical and trabecular changes after treatment with pth, alendronate, and their combination as assessed by finite element analysis of quantitative CT scans. *Journal of Bone and Mineral Research* 23 (12), 1974–1982.
- Mayhew, P. M., Thomas, C. D., Clement, J. G., Loveridge, N., Beck, T. J., Bonfield, W., Burgoyne, C. J., Reeve, J., 2005. Relation between age, femoral neck cortical stability, and hip fracture risk. *The Lancet* 366 (9480), 129–135.
- More, J. J., 1977. The Levenberg-Marquardt algorithm: Implementation and theory. In: Watson, A. (Ed.), *Numerical Analysis. Lecture Notes in Mathematics* 630, Springer-Verlag, pp. 105–116.
- Newman, D. L., Dougherty, G., Obaid, A. A., Hajrasy, H. A., Mar. 1998. Limitations of clinical CT in assessing cortical thickness and density. *Physics in Medicine and Biology* 43 (3), 619–626.
- Parker, M., Johansen, A., 2006. Hip fracture. *British Medical Journal* 333 (7557), 27–30.
- Poole, K. E. S., Mayhew, P. M., Rose, C. M., Brown, J. K., Bearcroft, P. J., Loveridge, N., Reeve, J., 2009. Changing structure of the femoral neck across the adult female lifespan. *In press for Journal of Bone and Mineral Research*.
- Poole, K. E. S., Rose, C. M., Mayhew, P. M., Brown, J., Clement, J. G., Thomas, C., Reeve, J., Loveridge, N., 2008. Thresholds for the measurement of cortical thickness in-vivo using computed tomography (the 100 women study). *Calcified Tissue International* 83 (1), 33.
- Prevrhal, S., Engelke, K., Kalander, W. A., Mar. 1999. Accuracy limits for the determination of cortical width and density: the influence of object size and CT imaging parameters. *Physics in Medicine and Biology* 44 (3), 751–764.
- Prevrhal, S., Fox, J. C., Shepherd, J. A., Genant, H. K., Jan. 2003. Accuracy of CT-based thickness measurement of thin structures: Modeling of limited spatial resolution in all three dimensions. *Medical Physics* 30 (1), 1–8.

- Sambrook, P., Cooper, C., 2006. Osteoporosis. *The Lancet* 367 (9527), 2010–2018.
- Sanders, K. M., Nicholson, G. C., Watts, J. J., Pasco, J. A., Henry, M. J., Kotowicz, M. A., Seeman, E., 2006. Half the burden of fragility fractures in the community occur in women without osteoporosis. when is fracture prevention cost-effective? *Bone* 38 (5), 694–700.
- Streekstra, G. J., Strackee, S. D., Maas, M., ter Wee, R., Venema, H. W., Sep. 2007. Model-based cartilage thickness measurement in the submillimeter range. *Medical Physics* 34 (9), 3562–3570.
- Treece, G. M., Prager, R. W., Gee, A. H., 1999. Regularised marching tetrahedra: improved iso-surface extraction. *Computers and Graphics* 23 (4), 583–598.
- Treece, G. M., Prager, R. W., Gee, A. H., Berman, L., Nov. 2000. Surface interpolation from sparse cross-sections using region correspondence. *IEEE Transactions on Medical Imaging* 19 (11), 1106–1114.
- Uusi-Rasi, K., Semanick, L. M., Zanchetta, J. R., Bogado, C. E., Eriksen, E. F., Sato, M., Beck, T. J., 2005. Effects of teriparatide [rhPTH (1-34)] treatment on structural geometry of the proximal femur in elderly osteoporotic women. *Bone* 36 (6), 948–958.
- Verhulp, E., van Rietbergen, B., Huiskes, R., 2008. Load distribution in the healthy and osteoporotic human proximal femur during a fall to the side. *Bone* 42 (1), 30–35.