Independent measurement of femoral cortical thickness and cortical bone density using clinical CT

G.M. Treece and A.H. Gee

CUED/F-INFENG/TR 691

April 2014

Cambridge University Engineering Department
Trumpington Street
Cambridge CB2 1PZ
England

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

Abstract

The local structure of the proximal femoral cortex is of interest since both fracture risk, and the effects of various interventions aimed at reducing that risk, are associated with cortical properties focused in particular regions rather than dispersed over the whole bone. Much of the femoral cortex is less than 3 mm thick, appearing so blurred in clinical CT that its actual density is not apparent in the data, and neither thresholding nor full-width half-maximum techniques are capable of determining its width. Our work on cortical bone mapping showed how to produce more accurate estimates of cortical thickness by assuming a fixed value of the cortical density for each hip. However, although cortical density varies much less over the proximal femur than thickness, what little variation there is leads to errors in thickness measurement. In this paper, we develop the cortical bone mapping technique by exploiting local estimates of imaging blur to correct the global density estimate, thus providing a local density estimate as well as more accurate estimates of thickness. We also consider measurement of cortical mass per unit surface area and the density of trabecular bone immediately adjacent to the cortex. Performance is assessed with ex vivo clinical QCT scans of proximal femurs, with true values derived from high resolution HRpQCT scans of the same bones. We demonstrate superior estimation of thickness than is possible with alternative techniques (accuracy 0.12 ± 0.39 mm for cortices in the range 1–3 mm), and that local cortical density estimation is feasible for densities > 800 mg/cm^3.

1 Introduction

Hip fractures are the most common cause of acute orthopaedic hospital admission in older people (Parker and Johansen, 2006). An individual's fracture risk is currently assessed using bone mineral density (BMD). Although BMD is specific (Johnell et al., 2005; Kanis et al., 2008) it lacks sensitivity (Kanis et al., 2008; Kaptoge et al., 2008; Sanders et al., 2006), missing the majority who go on to fracture. There is now growing evidence that focal, structural weaknesses may predispose a hip to fracture (Mayhew et al., 2005; Poole et al., 2010; de Bakker et al., 2009). The distribution of both trabecular and cortical bone is critical in determining a femur's resistance to fracture (Holzer et al., 2009; Verhulp et al., 2008; Poole et al., 2012). Drug treatment and exercise regimes targeted at reducing fracture risk result in changes which are focused in particular regions rather than dispersed over the whole bone (Lang et al., 2014; Poole et al., 2011). Accurate measurement of local cortical structure is hence important in assessing fracture risk and monitoring interventions.

However, measurement of the thickness and volumetric density of the cortex is not straightforward. Thin laminar structures are not accurately depicted in clinical CT because of the images' limited spatial resolution. Consequently, simple thickness estimation techniques, like those based on thresholding (Buie et al., 2007; Hangartner, 2007) or some measure of full-width half-maximum (Prevrhal et al., 1999, 2003), are unreliable when the cortex is thin in relation to the imaging resolution. With normal bore, clinical CT scanners, such methods are increasingly inaccurate below 3 mm (Dougherty and Newman, 1999; Hangartner and Gilsanz, 1996), with errors exceeding 100% for sub-millimetre cortices (Prevrhal et al., 2003). In very thin cortices, the 'thickness' measured using these techniques is just the width of the imaging blur, and the 'density' (i.e. the maximum CT value) is in fact a measure of cortical mass per unit surface area (mg/cm^2, henceforth abbreviated as simply 'mass'), not cortical density (mg/cm^3).

Cortical bone mapping is a recently proposed technique which does allow accurate estimation of cortical thickness (Treece et al., 2010, 2012) based on an estimate of cortical density which is presumed to be constant for each hip. Density varies much less over the proximal femoral cortex than thickness, which explains the good results for thickness measurement when making this assumption. However, any deviation from the presumed constant density value results in an error in the thickness estimate. Hence, we would like to be able to estimate the variation of density over the cortex: this would improve the thickness estimates as well as provide an independent measure of density.

There is already a sizeable literature on cortical thickness and density estimation from clinical CT. To this body of knowledge, this paper adds refinements that are rather subtle but undoubtedly effective. The content of this paper is therefore of necessity both fairly broad (in order to perform a fair comparison with existing techniques) and technically detailed (in order to thoroughly explain the nature and reasoning behind the innovations). With an increasing number of studies reliant on cortical measurements, it is important to have a thorough understanding of the accuracy and limitations of the techniques that are used to make such measurements, and this is what this paper sets out to do. In Section 2, we motivate and describe two variations of cortical bone mapping that provide independent estimates of cortical density at each measurement point. The accuracy and precision of these estimates, as well as those for cortical thickness, mass and trabecular density, are assessed in Section 3 alongside other estimates from alternative techniques. We discuss our findings in Section 4 before drawing some conclusions in Section 5.

2 Method

It was shown in Treece et al. (2012) that the variation of CT data $y_{\text{blur}}(x)$ across the cortex can be modelled as:

$$y_{\text{blur}}(x) = y_0 + \frac{y_1 - y_0}{2}\left[1 + \text{erf}\left(\frac{x - x_0}{\sigma\sqrt{2}}\right)\right] + \frac{y_2 - y_1}{2}\left[1 + \text{erf}\left(\frac{x - x_1}{\sigma\sqrt{2}}\right)\right] \quad (1)$$

where x is the distance across the cortex, y_0, y_1 and y_2 are CT values in surrounding tissue, within the cortex, and for the trabecular compartment respectively, x_0 and x_1 are the locations of the periosteal and endocortical surfaces respectively, and σ is the standard deviation of the assumed Gaussian imaging blur.

Model-based cortical thickness estimation involves fitting this function to the actual CT data $c(x)$ at each measurement point, by optimisation of the model parameters $\{y_0, y_1, y_2, x_0, x_1, \sigma\}$ until $y_{\text{blur}}(x)$ matches $c(x)$. In Treece et al. (2010, 2012), good results were achieved by presetting the cortical density y_1 and leaving the other parameters to be optimised by the fitting algorithm. A more standard deconvolution approach would be to preset the imaging blur σ, but in this instance this is less successful as will be shown in Section 3. Others have used a similar model but allowed all parameters to be optimised (Pakdel et al., 2012), an approach which we will show in Section 3 gives low bias but unfortunately poor precision.

Most techniques in current use are not model-based. They tend to involve thresholding, either at some fixed level (Buie et al., 2007; Hangartner, 2007) or at half the local peak CT value, the latter resulting in a full-width half-maximum (FWHM) measurement (Prevrhal et al., 1999, 2003). Thresholding at a fixed value is known to be sensitive to the chosen level, and will fail to include the cortex at all if it is too thin. The FWHM approach (which can be implemented using the model-based approach but presetting y_1 to the observed CT peak value c_{peak}) overcomes this problem, but leads to massive overestimation of cortical thickness for thin cortices (Dougherty and Newman, 1999; Hangartner and Gilsanz, 1996).

We summarise these approaches (other than thresholding at a fixed value) in the left-most four columns of Fig. 1. For concision, they are henceforth referred to as:

nothing preset A model-based fit with all parameters found by optimisation.

FWHM A **F**ull-**W**idth **H**alf-**M**aximum model-based fit with $y_1 = c_{\text{peak}}$ and all other parameters optimised.

preset blur A model-based fit with $\sigma = \hat{\sigma_d}$, where $\hat{\sigma_d}$ is a prior estimate of the imaging blur, and all other parameters optimised.

CBM v1 The constant density **C**ortical **B**one **M**apping estimate described in Treece et al. (2010, 2012), which is a model-based fit with $y_1 = \hat{y_1}$ and all other parameters optimised. This technique is implemented in

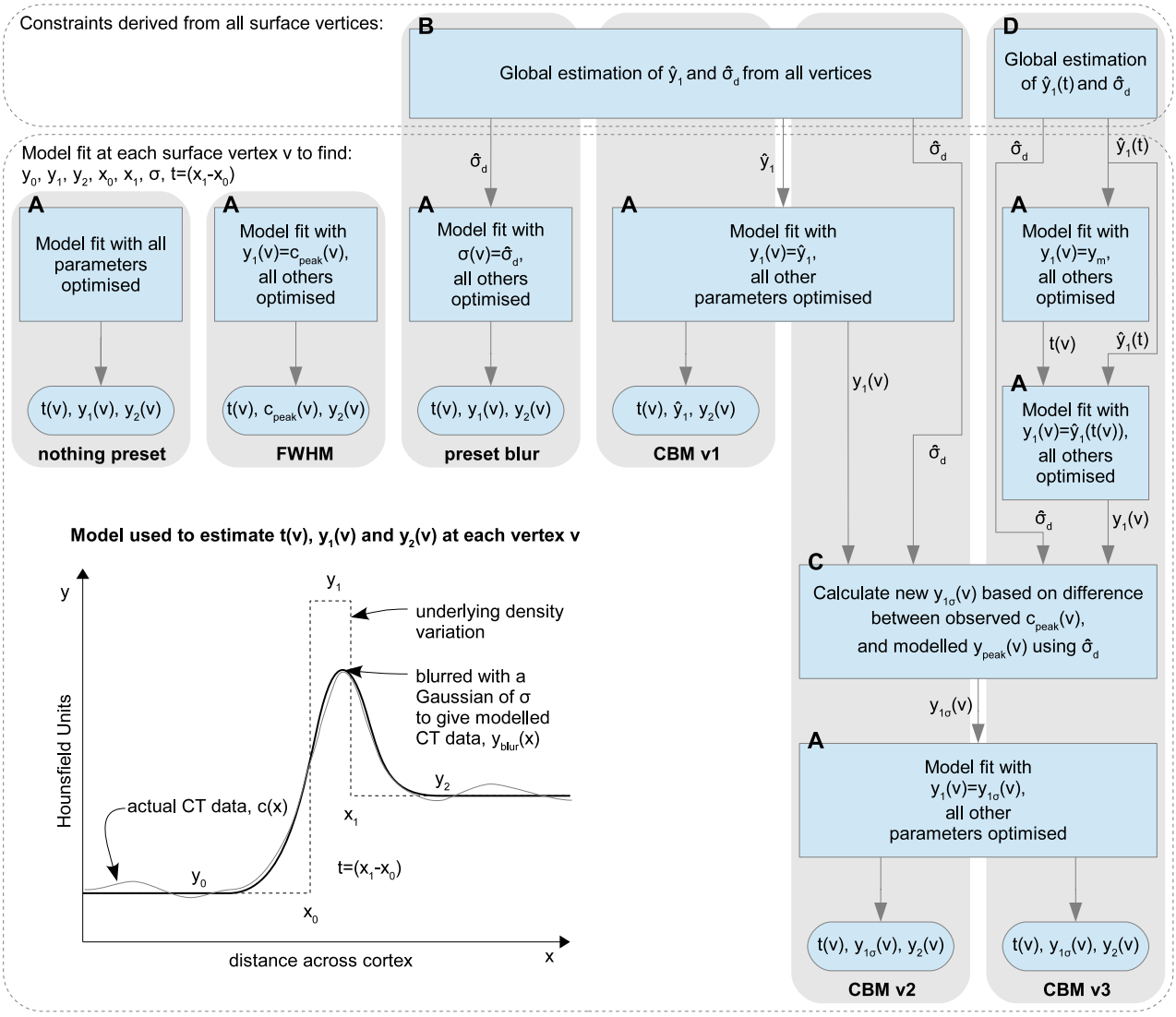
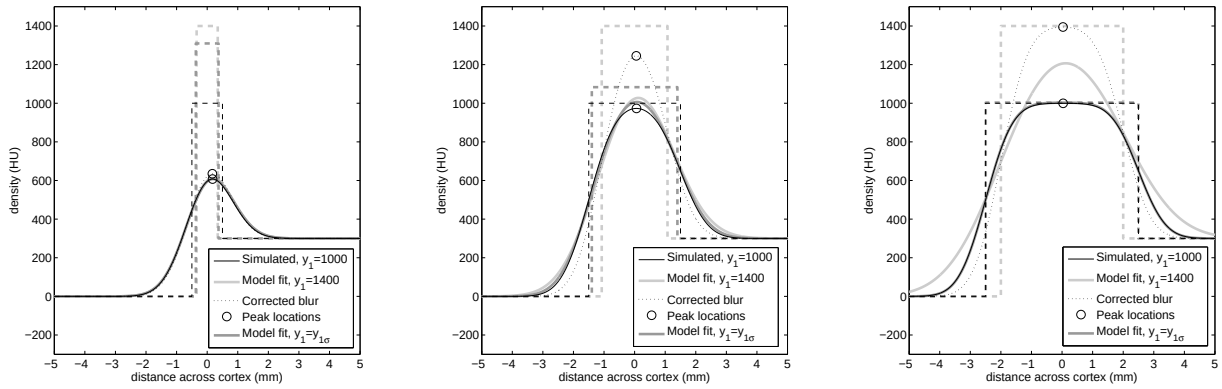


Figure 1: An overview of the six cortical parameter estimation algorithms tested in this paper. Many of these algorithms share similar stages. **A**: the local model fitting was described in Treece et al. (2010) and is summarised in the graph above. Each algorithm fits the same model, but starts with different parameters preset to specific values, the remaining being optimised. **B**: the calculation of a global cortical density $\hat{y_1}$ is described in Treece et al. (2012), and this procedure can be adapted to also estimate the global blur $\hat{\sigma_d}$ (see Appendix A). **C**: a new method for adjusting the presumed cortical density $\hat{y_1}$ to give a local estimate $y_{1\sigma}$ is described in Section 2.1. **D**: a new method for estimating a global density $\hat{y_1}(t)$ which varies with thickness is described in Section 2.2.



(a) (b) (c)

Figure 2: Estimating local cortical density $y_{1\sigma}$ using the global blur $\hat{\sigma_d}$. In all cases, the simulated actual cortical density is 1000 HU but the presumed cortical density $\hat{y_1}$ is set to 1400 HU. The simulated density variation is given by the black lines: solid lines show the blurred CT values whereas dashed lines show the underlying true variation. The light grey curves show the initial model fit $y_{\text{blur}}(x)$, with the dotted curve showing the fit after correcting σ to $\hat{\sigma_d}$. $y_{1\sigma}$ is then calculated using eq. (4) and the two CT values shown as circles. The final model fit with the new density value is shown in dark grey. (a) at very thin thickness ($t < 1$ mm) there is little difference between σ and $\hat{\sigma_d}$ and hence the correction has limited effect. (b) for thin cortices (1 mm $\leq t < 3$ mm) the correction improves the thickness estimate. (c) for thick cortices ($t > 3$ mm), the correction effectively results in the FWHM value, which is a good estimate in this domain.

Stradwin[1], our in-house software which is available for free download.

We are also interested in the peak CT value y_{peak} as predicted by the model $y_{\text{blur}}(x)$. We can find this by positioning the cortex, with thickness t, at $x = 0$ (i.e. $x_0 = -\frac{t}{2}$ and $x_1 = \frac{t}{2}$). Setting the differential of eq. (1) to zero then gives the location of the peak:

$$x_{\text{peak}} = \frac{\sigma^2}{t} \ln\left(\frac{y_1 - y_0}{y_1 - y_2}\right) \tag{2}$$

and the correct peak value is then $y_{\text{peak}} = y_{\text{blur}}(x_{\text{peak}})$ with $x_0 = -\frac{t}{2}$, $x_1 = \frac{t}{2}$.

For the purposes of estimating the global cortical density $\hat{y_1}$ in Treece et al. (2012), the peak is presumed to occur at $x = 0$, a reasonable simplification provided that the background levels $\{y_0, y_2\}$ are similar:

$$y_{\text{peak}} \approx y_{\text{zero}} = y_{\text{blur}}(0) = y_b + (y_1 - y_b)\,\text{erf}\left(\frac{t}{2\sqrt{2}\sigma}\right) \tag{3}$$

where $y_b = \frac{y_0 + y_2}{2}$. Estimation of $\hat{y_1}$ involves fitting eq. (3) to the data but starts with repeated fitting of $y_{\text{blur}}(x)$ to $c(x)$ (in order to estimate the values of t), each model-fit also providing a local estimate of the imaging blur σ. A global estimate $\hat{\sigma_d}$ is derived from these local estimates using the method described in Appendix A.

2.1 Local estimation of density

Although CBM v1 provides better thickness estimates than FWHM, it offers no local estimate of the cortical density, since y_1 is preset to a constant $\hat{y_1}$ for each bone. Hence there is an error in measuring thickness if

[1] http://mi.eng.cam.ac.uk/~rwp/stradwin

the real cortical density at the measurement point is not equal to $\hat{y_1}$. It would be of benefit to reduce this error while also measuring changes in density over the cortex. Whether this might be possible depends on the thickness of the cortex. For thick cortices as in Fig. 2(c) ($t > 4\sigma$, or about 3 mm for typical clinical resolution), cortical density and thickness can be resolved unambiguously: the cortex is thick enough for its actual density to be apparent in the CT data. For thinner cortices, although the cortical mass is preserved, the imaging blur makes it difficult to distinguish to what extent this mass is the result of a higher density or a higher thickness. For very thin cortices as in Fig. 2(a) (typically $t < 1$ mm), any cortex with the same mass, regardless of the relative contribution from thickness or density, will look the same in the blurred CT data.

Hence it is trivial to distinguish thickness and density in thick cortices, and impossible to do so (without prior information) in very thin cortices. However, in the mid-range, which is typical of cortical values in the proximal femur, it is possible to deduce the relative contributions of density and thickness by looking at the modelled imaging blur. If we overestimate the density ($\hat{y_1} > y_1$), as in Fig. 2(b), the modelled value of σ will compensate by being greater than the real imaging blur $\hat{\sigma_d}$. In contrast, if we underestimate the density, σ will compensate by being less than $\hat{\sigma_d}$. Having initially fitted a model using some presumed density $\hat{y_1}$, we can use these observations to give us a new *local* estimate of the actual density at each point.

We could simply compare σ from the CBM v1 estimate with the global imaging blur $\hat{\sigma_d}$, and adjust $\hat{y_1}$ accordingly. However, σ is not only sensitive to errors in the presumed cortical density, it is also affected by the presence of small pores in the cortex. These tend to be located around the cortical edges and once blurred by the imaging system result in shallower slopes in the CT data $c(x)$, which are modelled as larger values of σ. We hence choose to use peak CT values (which are as far as possible from the cortical edges) when adjusting for cortical density:

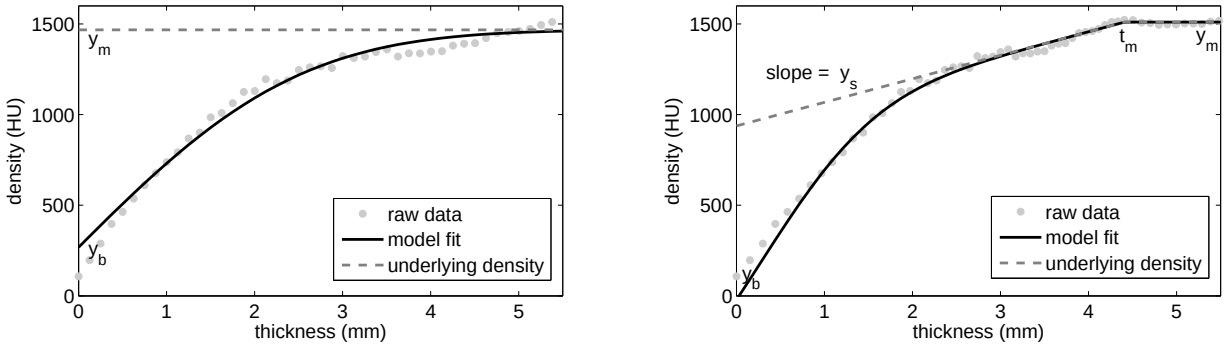
$$y_{1\sigma} = (\hat{y_1} - \max\{y_0, y_2\}) \frac{c_{\text{peak}} - \max\{y_0, y_2\}}{y_{\text{peak}} - \max\{y_0, y_2\}} + \max\{y_0, y_2\} \tag{4}$$

Equation (4) shows how we calculate the new cortical density estimate $y_{1\sigma}$: this is a well-motivated heuristic rather than a precise mathematical model. First, the CBM v1 algorithm is used with the preset $\hat{y_1}$ density value. Then this value is adjusted according to eq. (4), where c_{peak} is the maximum value observed in the CT data at this location, and y_{peak} is the predicted maximum value $y_{\text{blur}}(x_{\text{peak}})$ evaluated using the CBM v1 results for $\{y_0, \hat{y_1}, y_2, x_0 = -\frac{t}{2}, x_1 = \frac{t}{2}\}$ but replacing σ with the global estimate $\hat{\sigma_d}$. y_{peak} tells us what the peak CT value is if we use the correct imaging blur in the model: hence the ratio between this and the observed peak c_{peak} tells us by how much we need to adjust $\hat{y_1}$. Model-fitting is then repeated using $y_1 = y_{1\sigma}$.

We henceforth refer to this algorithm as **CBM v2** since it is a modification of the CBM v1 algorithm. An overview is given in the fifth column of Fig. 1, where stage **C** is the novel adjustment given in eq. (4). Examples are given in Fig. 2, demonstrating that a reasonable estimate of local density can indeed be achieved for relatively thin cortices. For thick cortices, the adjustment corrects the CBM v1 preset density to the actual density seen in the data.

2.2 Global estimation of density variation with thickness

While the CBM v2 algorithm has the potential to improve both density and thickness estimates for thick and relatively thin cortices, it offers no improvement over the CBM v1 algorithm for the very thin cortices of Fig. 2(a). In this case, there is no indication in the shape of the CT data $c(x)$ of what the density should be. However, it was noted in Treece et al. (2012) that there might be a trend for cortical density to increase with thickness in the human proximal femur. If such a trend exists, and we can see it in the data, then we can replace our initial constant density estimate $\hat{y_1}$ with a new estimate $\hat{y_1}(t)$ that varies with thickness. For thick and relatively thin cortices, the estimate can still be adjusted locally as in CBM v2, however for very thin cortices, this less biased initial estimate should result in more accurate measurements.



(a) (b)

Figure 3: Estimating global cortical density $\hat{y_1}(t)$ as a function of thickness. (a) shows estimation of $\hat{y_1}$ as explained in Treece et al. (2012), where the cortex is presumed to have a constant density which is blurred by the imaging system, resulting in an apparent reduction with thickness which is modelled by eq. (3). (b) The new technique explained in Section 2.2 allows the density to vary linearly with thickness up to a point t_m. The imaging blur in this case is fixed at $\hat{\sigma_d}$.

In order to estimate $\hat{y_1}(t)$, we return to the procedure for estimating $\hat{y_1}$ as outlined in Treece et al. (2012). In this procedure, FWHM measurements are made at many locations around the femur. After adjustment, these measurements provide many pairs of $\{t, y_{\text{zero}}\}$ data points. The function in eq. (3) is fitted to this data, giving global estimates of $\hat{y_1}$, $\hat{y_b}$ (the average background density) and $\hat{\sigma}$. This is summarised in Fig. 3(a). However, we have already noted that we can obtain an estimate $\hat{\sigma_d}$ of the imaging blur, not from the global model-fit in eq. (3), but by using eq. (1) on all the individual data measurements: see Appendix A for details.

In the case of a cortex which is genuinely a constant density irrespective of thickness, $\hat{\sigma}$ and $\hat{\sigma_d}$ do indeed return the same value, as would be expected. However, we have seen in practice across many studies of the proximal femur that $\hat{\sigma}$ is usually greater than $\hat{\sigma_d}$, sometimes by up to a factor of two. To understand this difference, we note that the initial slope of the curve in Fig. 3, which is substantially an error function (erf), is largely dependent on the value of σ in eq. (3). However, if the cortical density is actually increasing with thickness, this would also affect the slope of this curve. Hence the larger $\hat{\sigma}$ value is the combination of the actual imaging blur $\hat{\sigma_d}$ and the trend we have previously noted for density to vary with thickness.

It is not possible to simultaneously estimate $\hat{\sigma}$ and a varying density function, since these both have similar effects on the model in eq. (3). However, if we preset $\sigma = \hat{\sigma_d}$, we can then afford to replace y_1 with a simple piecewise function of thickness, which is linear up to some thickness t_m and constant thereafter, giving the modified equation:

$$y_{\text{zero}} = y_b + (y_1(t) - y_b)\,\text{erf}\left(\frac{t}{2\sqrt{2}\hat{\sigma_d}}\right), \quad \text{where} \quad y_1(t) = \begin{cases} y_s\,(t - t_m) + y_m & \text{if } t < t_m \\ y_m & \text{otherwise} \end{cases} \tag{5}$$

In this case the model contains four free parameters $\{y_b, y_m, y_s, t_m\}$ rather than three in eq. (3) $\{y_b, y_1, \sigma\}$. A typical fit is shown in Fig. 3(b) on the same data as in Fig. 3(a).

Implementation of this algorithm, henceforth termed **CBM v3**, requires an additional model fit at each location, since we first have to find the approximate thickness at each measurement location (using the maximum density y_m), then we can adjust our presumed density to $\hat{y_1}(t)$ before continuing as in CBM v2. This procedure is summarised in the sixth column of Fig. 1, where stage **D** is the novel application of eq. (5).

3 Experiments and results

3.1 Simulations

We investigated the effect of the local density correction for CBM v2 in eq. (4) by using eq. (1) to simulate the CT data variation $c(x)$ through ideal cortices, i.e. with perfectly constant density and no pores, surrounded by material which is also at a constant (but lower) density. These cortices were blurred with a Gaussian kernel with $\sigma = 0.71\,\text{mm}$ (FWHM of 1.67 mm), which is reasonably typical of a good clinical CT system. The cortices were up to 6 mm thick.

We then obtained cortical thickness estimates using the CBM v1 and v2 algorithms. This was repeated multiple times, with different initial density estimates $\hat{y_1}$, ranging from an 80% underestimate to an 80% overestimate of the actual simulated density. The presumed imaging blur $\hat{\sigma_d}$ in the CBM v2 algorithm was set to the actual simulated blur. The results are shown in Fig. 4.

3.2 Comparing HRpQCT and QCT data

Moving on from simulations, we examined data from a study ethically approved by the Medical University of Vienna, comprising cadaveric proximal femurs from 18 females and 17 males of mean age 77 years (range 59–96 years). Both left and right femurs were stripped of soft tissue, submerged in a saline solution and, after vacuum to remove air bubbles, scanned using both HRpQCT and QCT. Further details of this data have been published previously (Dall'Ara et al., 2013b,a). The QCT data was scanned using a Brilliance64[2] scanner at 120 kV, with voxel size $0.33 \times 033 \times 1.0\,\text{mm}^3$, and converted from Hounsfield Units to density using a BDC calibration phantom[3]. The reference HRpQCT data was scanned using an XTremeCT[4] scanner, with voxel size $0.082 \times 0.082 \times 0.082\,\text{mm}^3$, and converted to density using the manufacturer-provided phantom with some adjustments as detailed in Appendix B. This gave a total of 70 matching high and low-resolution data sets from 35 subjects.

Each of the femurs was semi-manually segmented in the QCT data using in-house Stradwin software, and surface meshes created with uniform triangle size and shape (Treece et al., 1999). Cortical measurements as described in the previous section were performed at all vertices in these meshes, in directions consistent with the surface normal at that point. Cortical thickness t, density y_1 and mass $t \times y_1$ were recorded at each point, at a total of 772,054 separate locations, about 11,000 per femur. We also recorded the y_2 density value, which corresponds to the average trabecular density just inside the cortex.

Each QCT-derived surface was registered with the corresponding HRpQCT data by minimising the mean squared distance between matched points from the QCT and HRpQCT meshes after a rigid body transformation with separate scale factors in the three principal directions. This transformation was used to import the QCT-derived surface into the HRpQCT reference frame, so that measurements could be made in the HRpQCT data at exactly the same locations, and in exactly the same directions, as the measurements in the QCT data.

Reference HRpQCT measurements of cortical thickness, density and mass were made using the established FWHM algorithm, which is accurate on this data down to a cortical thickness of 0.3 mm. These measurements were adjusted slightly to compensate for the use of a peak value of cortical density rather than a mean value, as explained in Appendix B. At each location, HRpQCT data was averaged *parallel* to the cortex as described in Treece et al. (2012), such that each estimate's spatial localization was consistent with the QCT data, while preserving the HRpQCT resolution *through* the cortex.

[2] Philips, Germany
[3] QMR Gmbh, Germany
[4] Scanco Medical AG, Switzerland

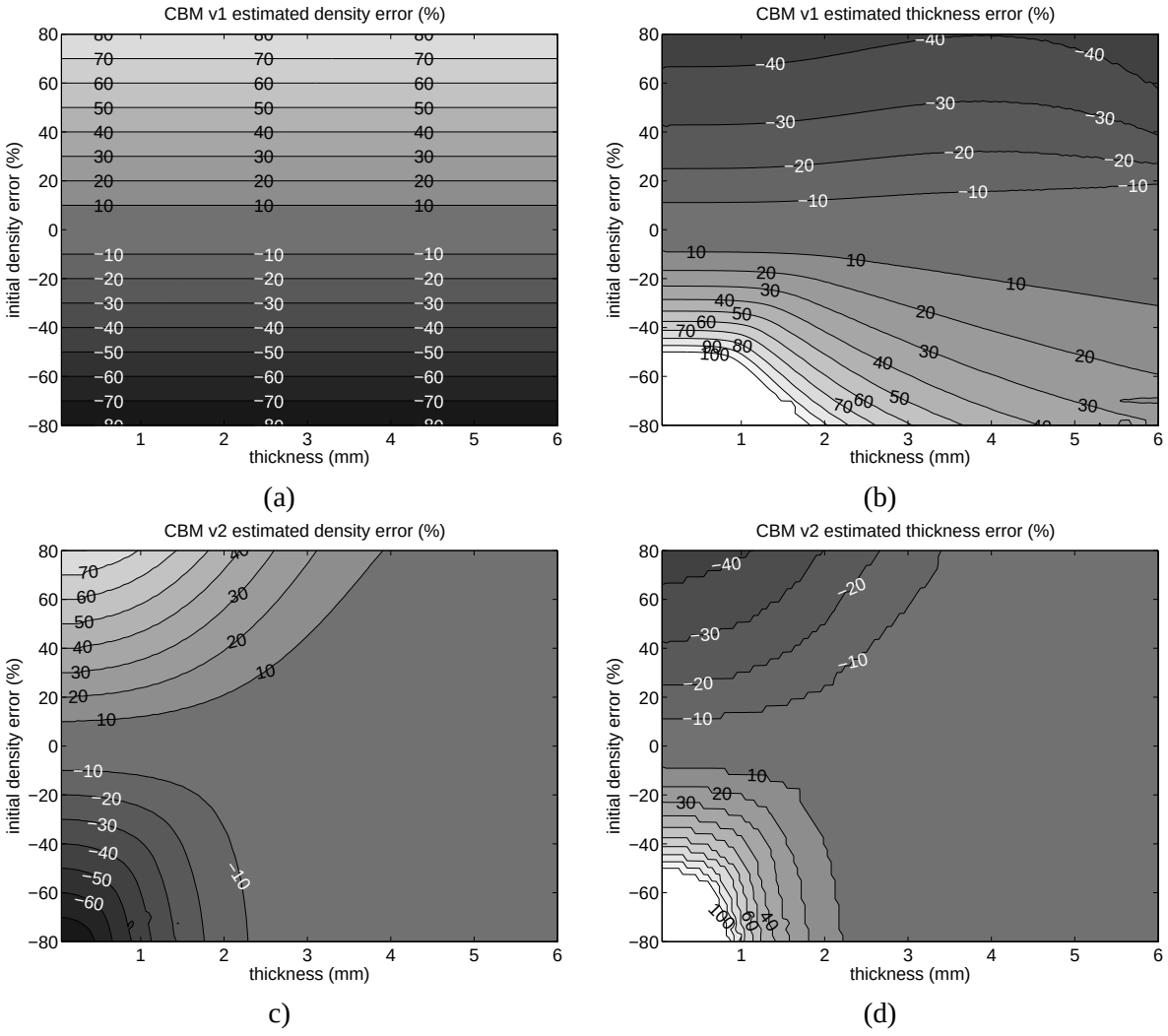


Figure 4: Thickness and density estimation errors with varying initial density $\hat{y_1}$. (a) and (b) show density and thickness estimation errors, respectively, for the CBM v1 algorithm described in Treece et al. (2010). (c) and (d) are the equivalent graphs for the CBM v2 algorithm described in Section 2.1. If $\hat{y_1}$ is correct (i.e. zero error on the vertical axis), then density and thickness estimates are correct for all thickness values in both cases. In CBM v1, the errors roughly scale with errors in $\hat{y_1}$: if $\hat{y_1}$ is too large, then the estimated density is (by definition) too large and the estimated thickness is too small. In CBM v2, eq. (4) is used to replace the preset $\hat{y_1}$ estimate with the local $y_{1\sigma}$ value. For thick cortices ($t > 3\,\text{mm}$) this results in accurate estimation whatever the initial value of $\hat{y_1}$. For thinner cortices ($1\,\text{mm} \leq t < 3\,\text{mm}$) estimation is significantly less dependent on $\hat{y_1}$ than in the CBM v1 algorithm. For very thin cortices ($t < 1\,\text{mm}$), the two algorithms are similar, with both relying on a correct initial density estimate.

Table 1: Cortical thickness, cortical density, cortical mass and trabecular density estimation errors for very thin ($0.3\,\text{mm} \leq t < 1.0\,\text{mm}$), thin ($1.0\,\text{mm} \leq t < 3.0\,\text{mm}$) and thick ($3.0\,\text{mm} \leq t < 6.0\,\text{mm}$) cortices. Values are expressed as bias $\pm$ precision (std), with positive bias indicating overestimation. Success rate is expressed as the overall percentage of measurements for which the model-fit procedure was successful and hence some parameters were estimated, even if these estimates were not correct. Processing times are for all measurements from a single femur ($\approx 11,000$ measurements), running on a single core of a 2.67 GHz Intel Core i7 processor.

quantity	**thickness range (mm)**	**algorithm**					
		FWHM	**preset blur**	**nothing preset**	**CBM v1**	**CBM v2**	**CBM v3**
cortical thickness (mm)	$3.0 \leq t < 6.0$	0.04 ± 0.25	0.08 ± 0.26	0.08 ± 0.25	-0.09 ± 0.32	0.04 ± 0.25	0.04 ± 0.25
	$1.0 \leq t < 3.0$	0.48 ± 0.37	0.51 ± 0.46	-0.01 ± 0.88	-0.24 ± 0.32	0.12 ± 0.39	0.26 ± 0.40
	$0.3 \leq t < 1.0$	1.06 ± 0.37	2.35 ± 1.46	0.58 ± 1.07	-0.24 ± 0.14	-0.15 ± 0.23	0.04 ± 0.31
cortical density (mg/cm^3)	$3.0 \leq t < 6.0$	-29 ± 110	-47 ± 119	-44 ± 105	11 ± 124	-26 ± 109	-27 ± 109
	$1.0 \leq t < 3.0$	-170 ± 137	-165 ± 159	-59 ± 244	195 ± 217	-26 ± 178	-89 ± 160
	$0.3 \leq t < 1.0$	-378 ± 166	-373 ± 182	-241 ± 300	462 ± 259	193 ± 331	-30 ± 264
cortical mass (mg/cm^2)	$3.0 \leq t < 6.0$	-5.3 ± 59.2	-7.5 ± 62.0	-8.2 ± 59.5	-9.7 ± 64.9	-6.8 ± 60.6	-6.5 ± 61.2
	$1.0 \leq t < 3.0$	8.6 ± 26.5	9.5 ± 29.0	8.0 ± 28.3	7.5 ± 25.3	8.0 ± 25.8	8.7 ± 26.1
	$0.3 \leq t < 1.0$	11.8 ± 16.3	58.9 ± 60.3	18.4 ± 28.9	0.6 ± 10.4	1.4 ± 11.1	2.9 ± 11.7
trabecular density (mg/cm^3)	$3.0 \leq t < 6.0$	-50 ± 76	-57 ± 81	-53 ± 76	-51 ± 83	-51 ± 75	-48 ± 74
	$1.0 \leq t < 3.0$	-20 ± 61	-24 ± 67	-25 ± 64	-20 ± 60	-19 ± 58	-20 ± 61
	$0.3 \leq t < 1.0$	2 ± 31	-24 ± 60	-2 ± 35	3 ± 30	3 ± 30	3 ± 30
success rate (%)	$0.3 \leq t < 6.0$	97.8	66.7	92.7	97.6	97.8	98.2
processing (secs)	$0.3 \leq t < 6.0$	34	60	56	71	85	93

3.3 Aggregate ex vivo results

Figure 5 shows aggregate results for the measurement of cortical thickness (t) using all six algorithms. These plots summarise a very large number of measurements expressed as probability density functions: the different shades of grey show the expected distribution of thickness measurements for a particular confidence interval. Hence the 'p=0.68' region is similar to a one-standard-deviation range, though these distributions are clearly not normal, nor even unimodal. Another way of expressing this is that the plots show the density of points on a scatter plot of all pairs of true vs. estimated measurements from the HRpQCT and QCT data. An ideal estimator would be a thin diagonal line.

The bias in thickness estimation (the most probable thickness estimate for each actual thickness) for all algorithms is summarised in Fig. 8(a), with Fig. 8(b) giving the distribution of the (HRpQCT) thickness values that were successfully estimated in the QCT data (i.e. the model-fit was successful, even if the parameters were wrong). Note that there are more thickness estimates for lower thicknesses, but none at all below 0.3 mm.

Figure 6 summarises cortical density (y_1) estimation in exactly the same way as in Fig. 5, and the bias of each algorithm is again summarised in Fig. 8(c), with the distribution of true density values in Fig. 8(d). As would be expected, cortical density is generally high, with relatively few measurements at lower values.

It was noted in Section 2 that density estimation is highly dependent on cortical thickness, and hence it is important to look at how the density estimation error varies with true cortical thickness. This is presented in Fig. 7, with the true (HRpQCT) minus the estimated (QCT) density on the y-axis, so an ideal estimator would be a thin horizontal line at $y = 0$. Once again the bias is summarised for all algorithms in Fig. 8(e).

For the sake of concision, we report only a few of the cortical mass ($t \times y_1$) results in Fig. 9. All algorithms estimate cortical mass relatively well, with the two extremes of performance shown in Figs. 9(a) and (b). The bias of all algorithms is presented in Fig. 9(c) and the mass distribution in Fig. 9(d).

Estimation of trabecular density (y_2) is straightforward and very similar results are obtained with all algorithms, summarised in Fig. 10. As with cortical density, we include a plot of trabecular density error against thickness in Fig. 10(c), to see to what extent this measure is dependent on cortical thickness.

Numerical summaries of all these results are collated in Table 1. Following on from the discussion in Section 2, we are expecting different behaviour for thick ($t > 3$ mm), thin (1 mm $\leq t < 3$ mm) and very thin ($t \leq 1$ mm) cortices, and hence estimation errors are presented for each of these domains separately.

3.4 Spatial distribution of ex vivo results

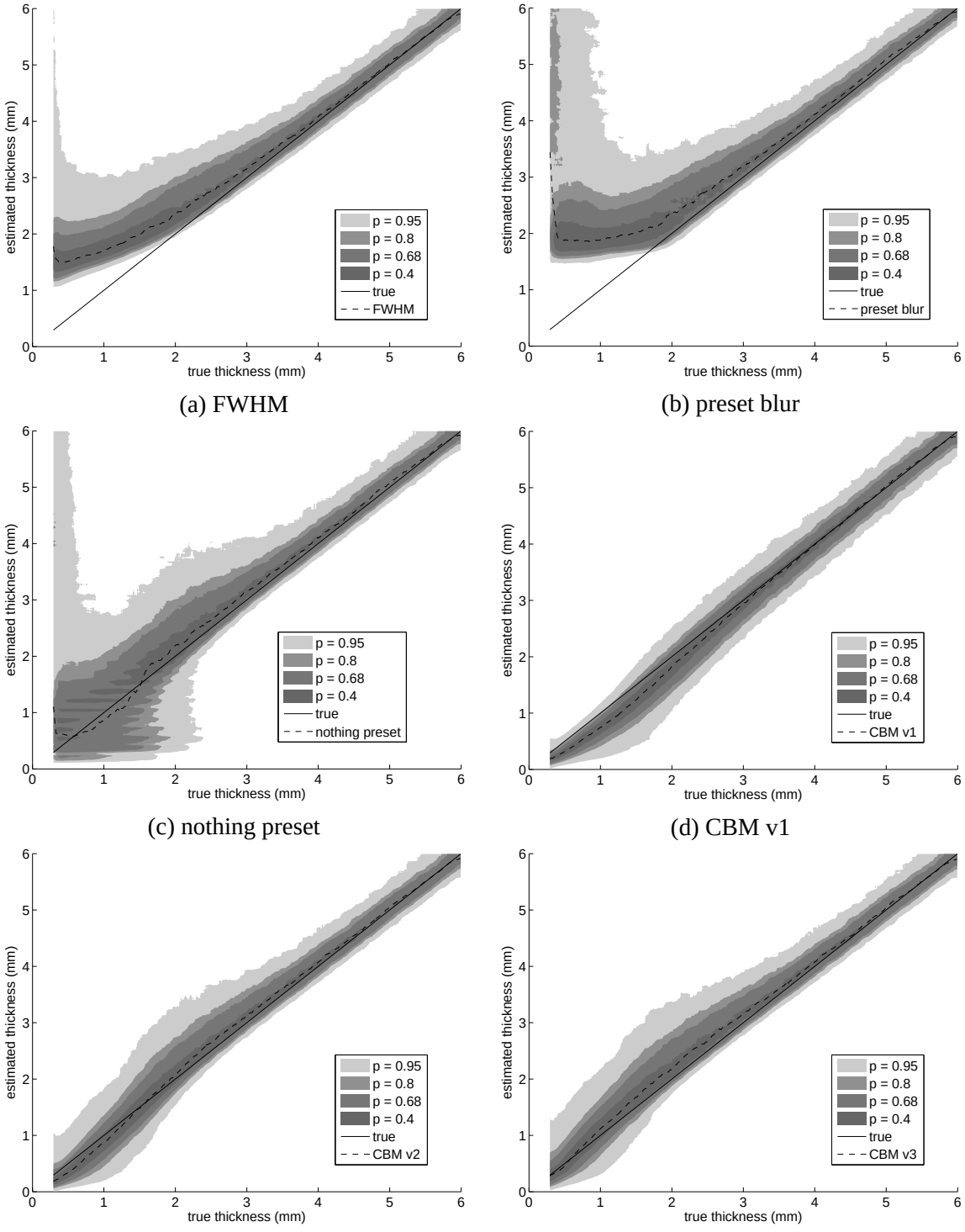
The previous section examines performance as a function of cortical thickness, but it does not give a sense of how the errors vary over a typical proximal femur. To elucidate this, we registered each of the QCT-derived surfaces to a canonical femur, using a nonrigid registration technique embodied in our in-house wxRegSurf[5] software and described in Gee and Treece (2014). This allowed us to map the cortical parameters from each subject to the same surface, and hence present estimation bias as a colour map over this surface.

Cortical thickness estimation bias is shown in Fig. 11, with cortical density in Fig. 12. The mean thickness and density are also shown for context. As previously noted, cortical mass and trabecular density errors were similar for the various algorithms, so a reduced set of results is presented in Fig. 13.

4 Discussion

Looking first at cortical thickness, Fig. 5 is consistent with previous results (Treece et al., 2012) in that FWHM dramatically overestimates small thicknesses while CBM v1 only slightly underestimates these. We

[5] http://mi.eng.cam.ac.uk/~ahg/wxRegSurf/

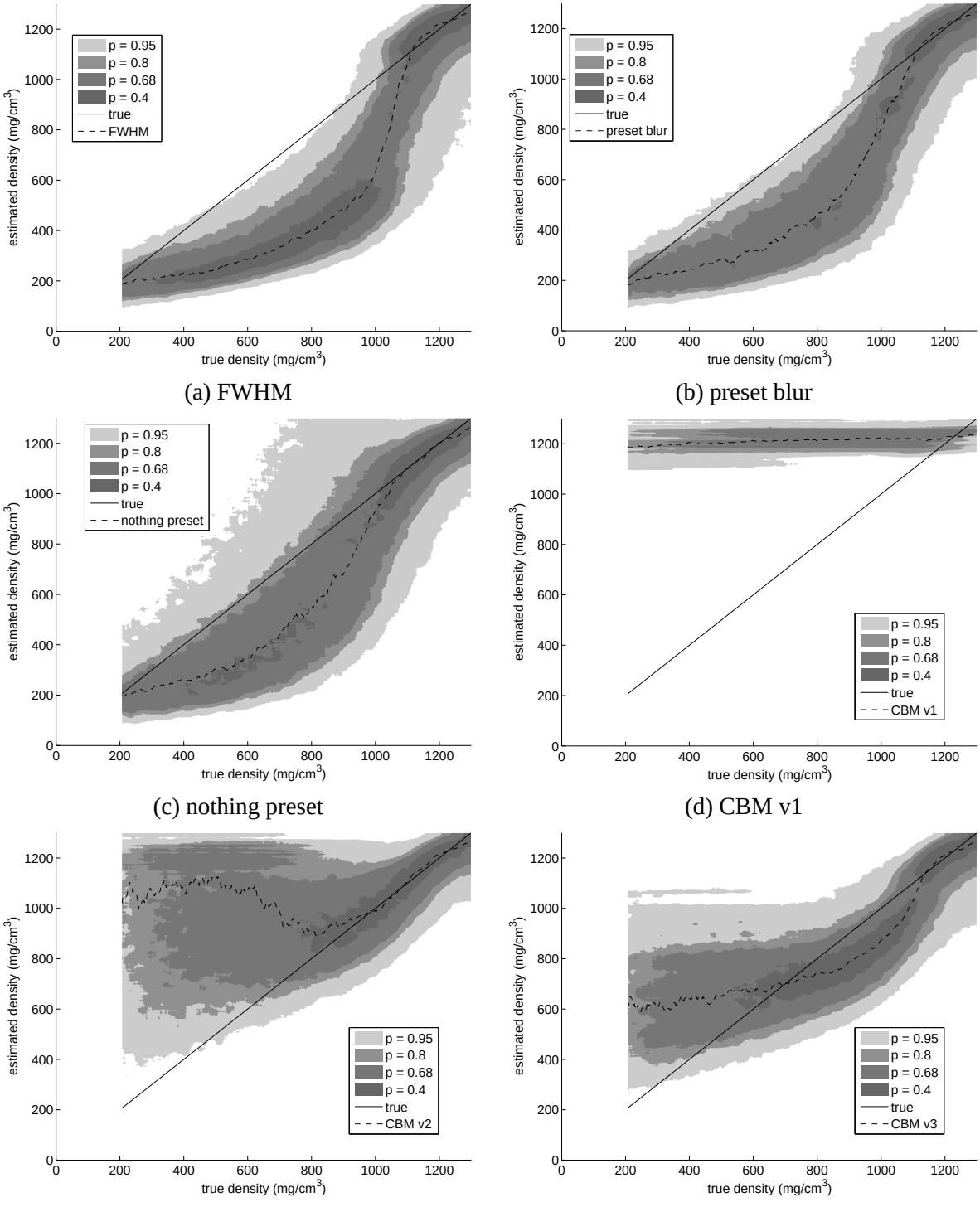


(a) FWHM (b) preset blur

(c) nothing preset (d) CBM v1

(e) CBM v2 (f) CBM v3

Figure 5: Cortical thickness estimation errors. The graphs show the distribution of estimated QCT thickness values plotted against the HRpQCT thickness values, hence the solid diagonal line represents perfect estimation. These results are aggregates across all the measurement locations on all the femurs.

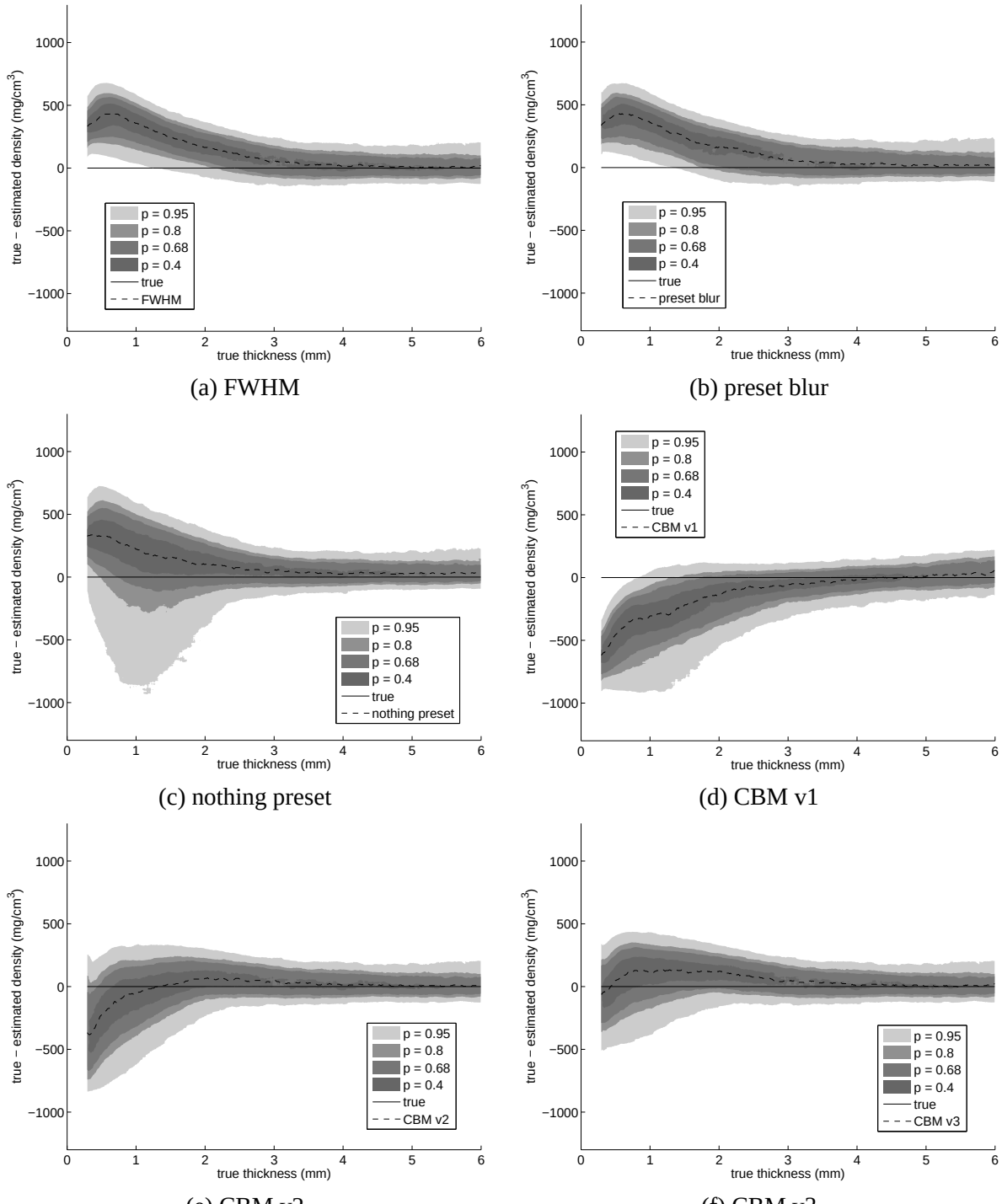


(a) FWHM (b) preset blur

(c) nothing preset (d) CBM v1

(e) CBM v2 (f) CBM v3

Figure 6: Cortical density estimation errors. The graphs show the distribution of estimated QCT density values plotted against the HRpQCT density values, hence the solid diagonal line represents perfect estimation. These results are aggregates across all the measurement locations on all the femurs.

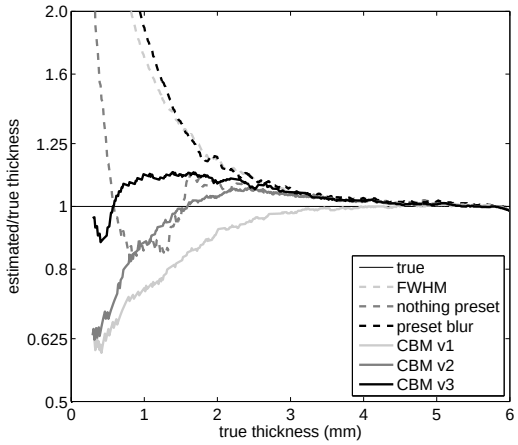


(a) FWHM (b) preset blur

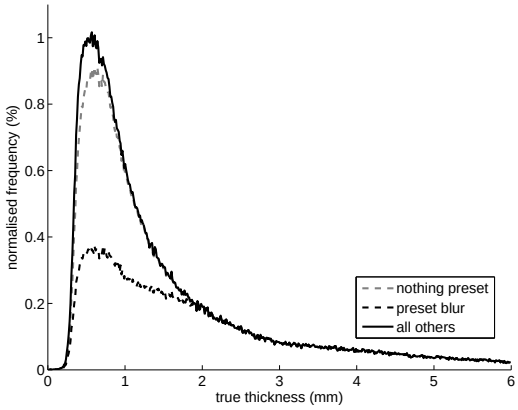
(c) nothing preset (d) CBM v1

(e) CBM v2 (f) CBM v3

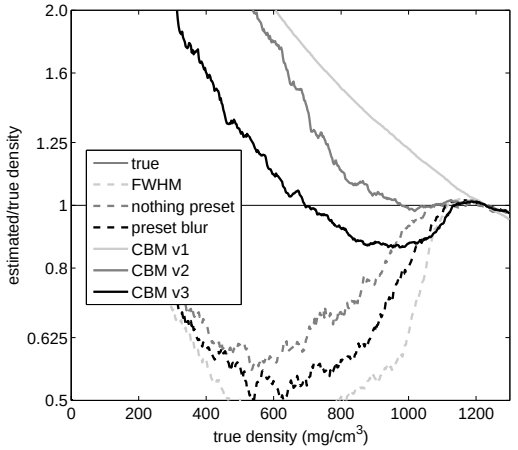
Figure 7: Cortical density estimation errors against thickness. Complementary to Fig. 6, here density estimation error is plotted against the HRpQCT thickness values.



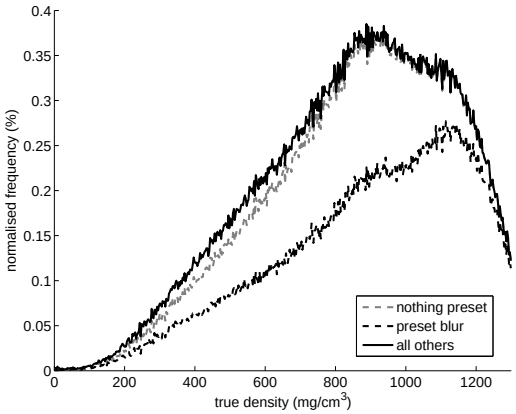
(a) thickness estimation error



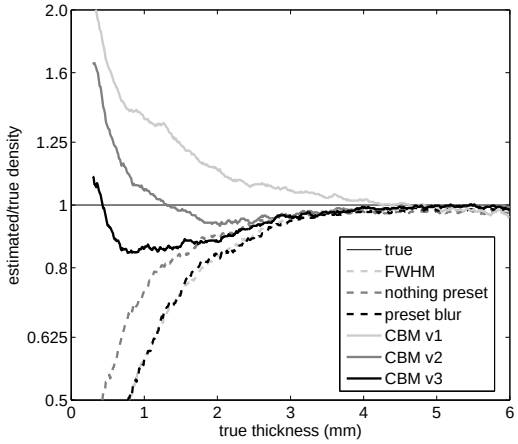
(b) HRpQCT thickness distribution



(c) density estimation error

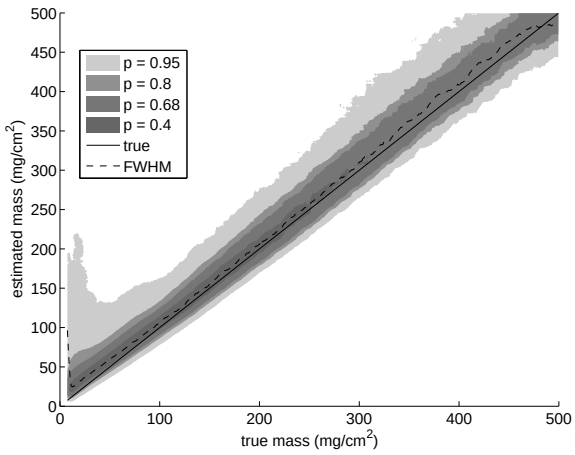


(d) HRpQCT density distribution

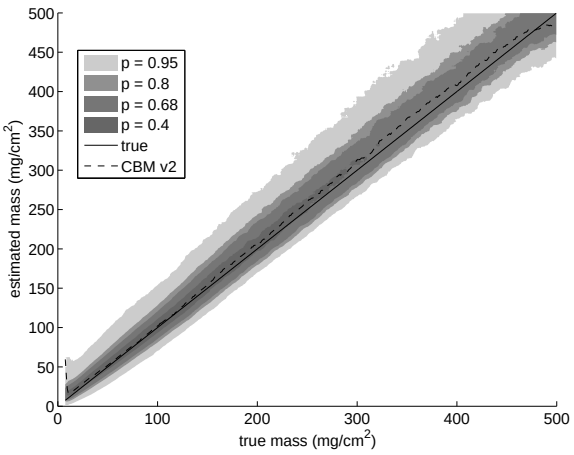


(e) density error with thickness

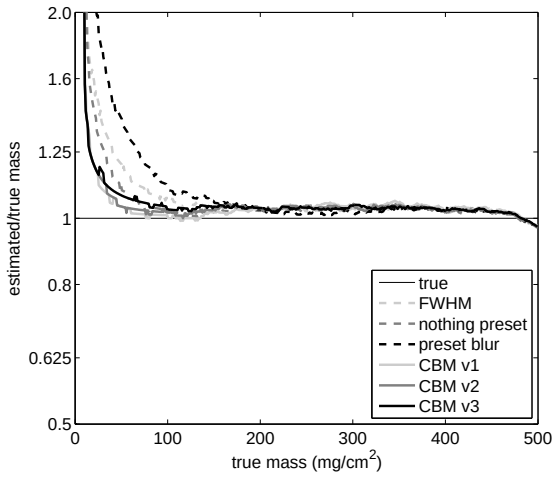
Figure 8: Summary results for cortical thickness and density. (a), (c) and (e) show the estimation bias for all algorithms. (b) and (d) show the distribution of the HRpQCT thickness and density values: most of the cortex is less than 2 mm thick, with density greater than 800 mg /cm^3.



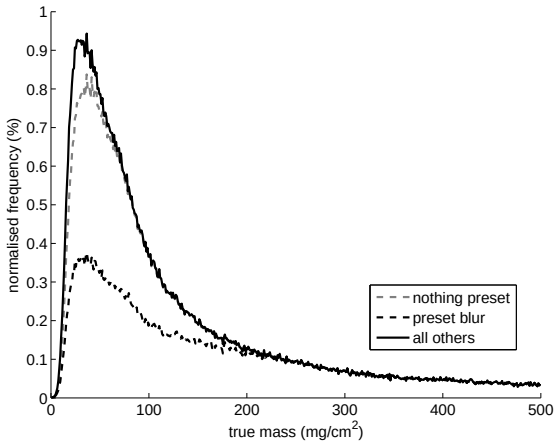
(a) FWHM mass estimation



(b) CBM v2 mass estimation

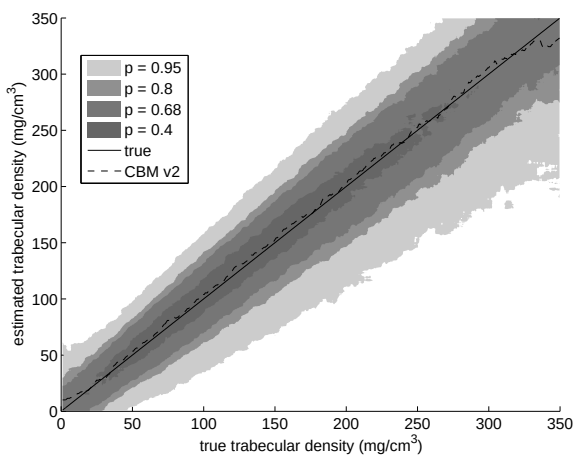


(c) mass estimation error

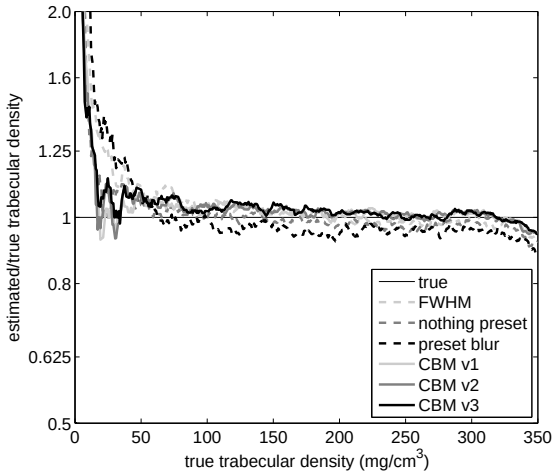


(d) HRpQCT mass distribution

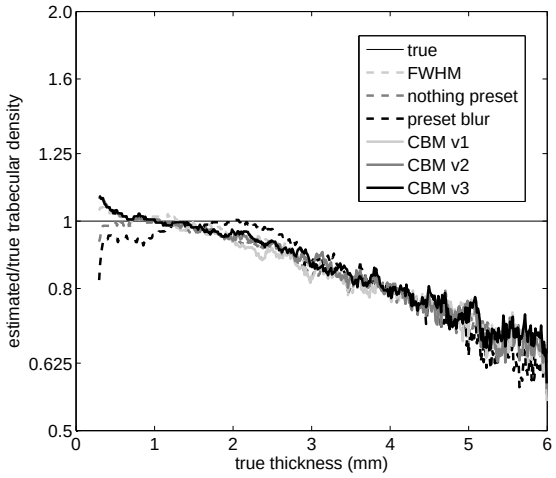
Figure 9: Cortical mass estimation errors. In general, estimation of mass is much more straightforward and there is less difference between algorithms. (a) and (b) show the two extreme sets of results, summarised in (c) and with the underlying distribution in (d).



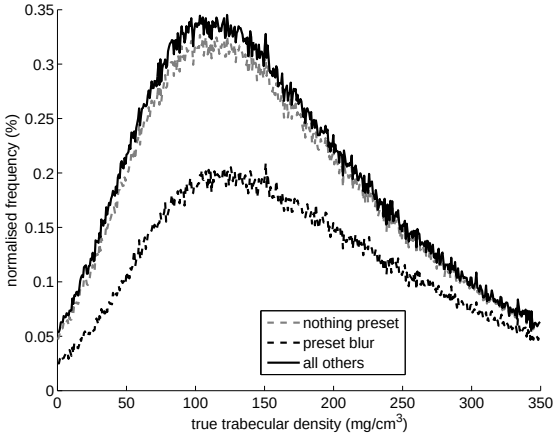
(a) CBM v2 trabecular density



(b) trabecular density error



(c) density error with thickness



(d) HRpQCT density distribution

Figure 10: Trabecular density estimation errors. Estimation of trabecular density is nearly identical for all algorithms, with very little bias with density, as shown in (a) and (b). (c) plots the errors against thickness, showing that there is some bias with increasing thickness. The trabecular density distribution is shown in (d).

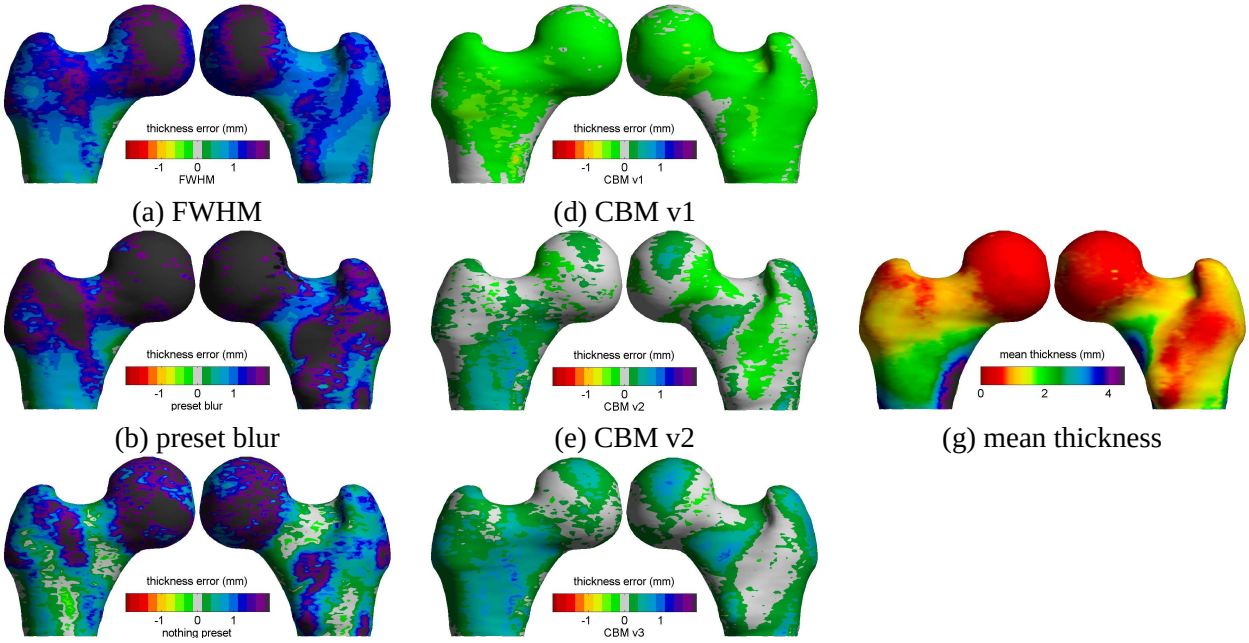


Figure 11: Spatial distribution of cortical thickness errors on the proximal femur. (a) to (f) show the bias in thickness estimation for the various algorithms. (g) shows the variation in the mean cortical thickness.

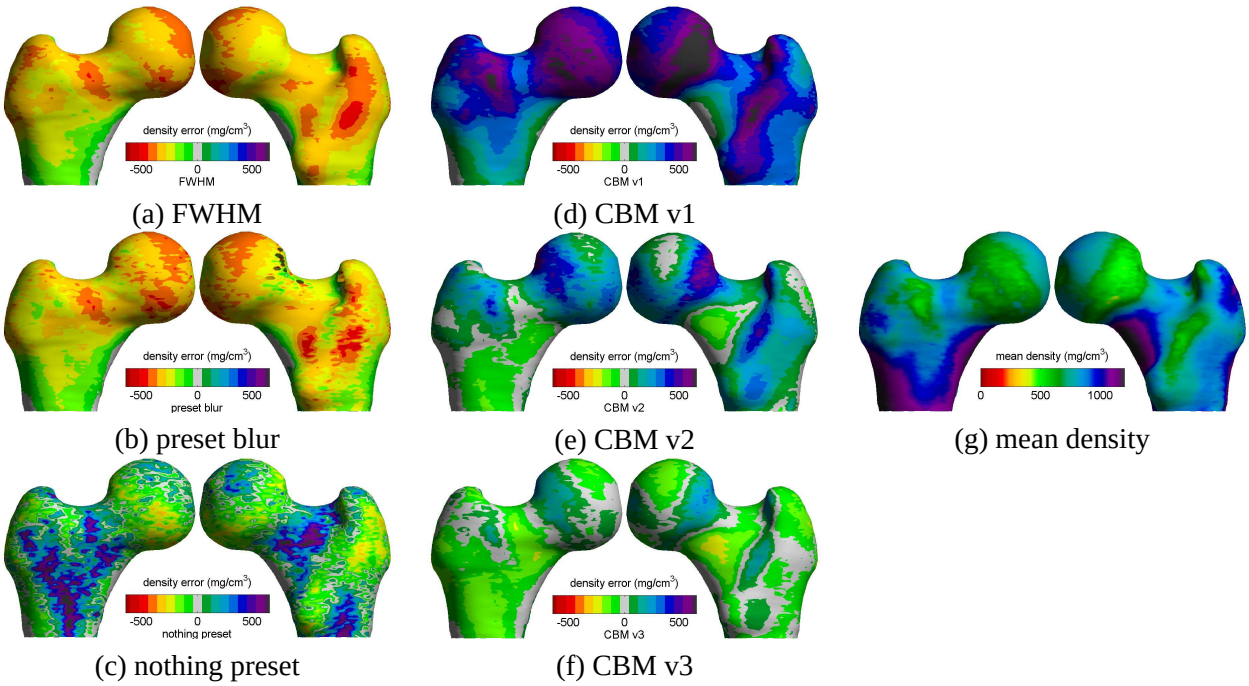
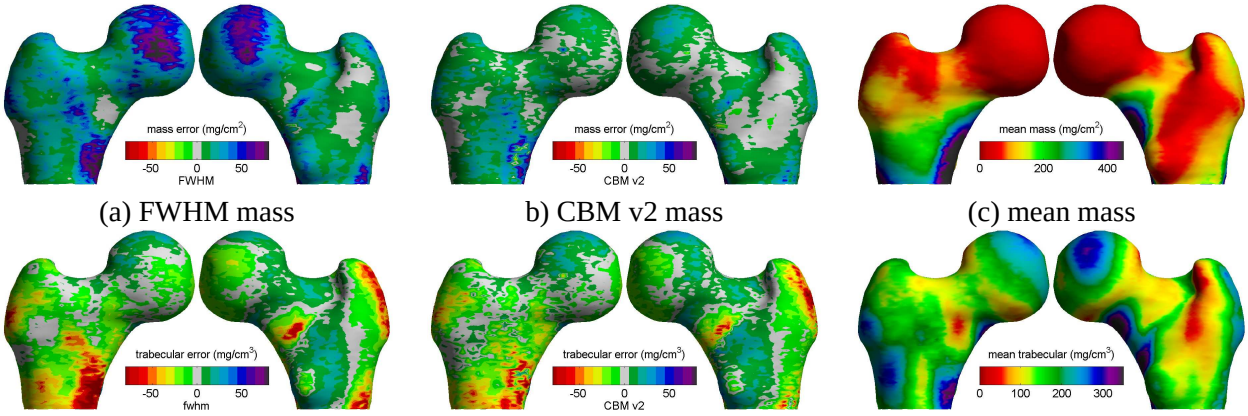


Figure 12: Spatial distribution of cortical density errors on the proximal femur. (a) to (f) show the bias in density estimation for the various algorithms. (g) shows the variation in the mean cortical density.



(a) FWHM mass b) CBM v2 mass (c) mean mass

(d) FWHM trabecular density (e) CBM v2 trabecular density (f) mean trabecular density

Figure 13: Spatial distribution of cortical mass and trabecular density errors on the proximal femur. (a) to (c) show results for cortical mass estimation, (d) to (f) for trabecular density estimation.

previously stated (without providing specific evidence) that presetting the imaging blur in the model was not very successful, and the results in Fig. 5(b) support this claim. Indeed, this technique often fails to fit the model at all, as can be seen in Figs. 8(b) and (d) and the very low success rate in Table 1. The most likely explanation for this is that the model needs to adjust the imaging blur to account for density errors or the presence of pores, and hence will often fail to fit when it is prevented from doing so.

The results for the nothing preset algorithm also agree with previous work (Pakdel et al., 2012), in that the bias is considerably reduced compared to FWHM and also compared to CBM v1. However, when applied to real data the precision of the method is poor, as is evident in Fig. 5(c), particularly at low thickness. This is due to the well understood ambiguity between thickness and density for thin cortices (Treece et al., 2012). Consequently, the success rate (Table 1) is low compared with alternative algorithms. For very thin cortices, Table 1 shows that both the accuracy and precision of this technique are poor.

The two new algorithms behave much as expected from the discussion in Section 2. Both CBM v2 and CBM v3 improve slightly the precision over CBM v1 for thick cortices, returning results which, from Table 1, are seen to be very similar to FWHM in this domain. For thinner cortices, both new methods reduce the bias present in CBM v1, at a cost of marginally worse precision, though CBM v2 performs the best. For very thin cortices, CBM v2 behaves largely as CBM v1, since it is unable to refine the density estimates in this domain, whereas CBM v3 does manage to improve the thickness bias even for these extreme cases.

The spatial distribution of thickness errors in Fig. 11 shows that, for the proximal femur, measurements are dominated by the performance on thin cortices. Here FWHM and preset blur dramatically overestimate thickness, and even the nothing preset algorithm shows large regions of overestimation around the femoral head. Errors in CBM v1 are fairly stable underestimates. CBM v2 gives qualitatively the best performance, with the most grey (very low bias) regions and a reasonable balance between underestimation and overestimation.

Turning to cortical density estimation, it is immediately clear from Fig. 6 that these are considerably less precise estimates than for thickness. For the highest densities (which tend also to be at thicker cortices), estimates are fairly good for all techniques, but at all other densities the FWHM and preset blur algorithms underestimate the density significantly. The nothing preset algorithm does slightly better, giving relatively unbiased results above $1000\,\mathrm{mg/cm^3}$, although once again the precision of this technique is noticeably less good. CBM v1 makes the assumption that the density is constant, though the value is different for each tested femur. Fig. 6(d) underlines that this value lies towards the top of the densities encountered in the data.

In contrast to CBM v1, the two new algorithms make some attempt to estimate varying density. CBM v2

is relatively unbiased down to about $800\,\mathrm{mg/cm^3}$ which, considering the distribution in Fig. 8(d), accounts for most of the cortex. However, it increasingly tends to overestimate when the real density is low. The CBM v3 results are better in this regard, though there is a tendency to underestimate higher densities. It is clear from Fig. 8(c) that both these algorithms show fairly low bias, but for different density ranges.

These results are easier to understand by looking at the distribution of density errors with cortical thickness in Fig. 7. For thick cortices, all algorithms estimate density successfully, and Fig. 7(d) shows that the actual cortical density does indeed tend to be fairly constant in this domain[6]. As the cortex becomes thinner, the actual density tends to reduce and this causes overestimation in CBM v1, while the nothing preset, CBM v2 and CBM v3 algorithms are still able to provide reasonable estimates. For very thin cortices, the nothing preset algorithm tends to mimic FWHM, but with considerably poorer precision, while CBM v2 behaves similarly to CBM v1, as anticipated in Section 2.

The premise of the CBM v3 algorithm was that there is a trend for linear density increase with thickness: looking at Fig. 7(f), the reality is no doubt a little more complex. The linear trend has successfully corrected the bias for very thin cortices, but it has also introduced a slight underestimation for thin cortices. This is supported by the summary results in Table 1, from which the CBM v3 algorithm in general seems to perform better than the alternatives in estimating density. The spatial distribution of errors over the proximal femur in Fig. 12 also supports this view: again, CBM v3 seems the least biased for this range of densities, with the greatest area of low bias (grey) and a more constrained bias range in general.

Cortical mass estimation ($t \times y_1$) is much more consistent between algorithms, with Fig. 9 and Table 1 showing precise and unbiased results for thick and thin cortices. However, for very thin cortices there is some difference in performance, with all of the CBM algorithms showing reduced bias and better precision in this domain. This improvement also makes a noticeable difference when viewed over the surface of the proximal femur in Figs. 13(a) and (b). The CBM v2 result is clearly less biased than that for FWHM, particularly over the femoral head where the cortex is very thin.

All algorithms estimate endocortical trabecular density (y_2) fairly accurately, as seen in Fig. 10 and Table 1. Fig. 10(c) reveals that there is, however, a clear bias with cortical thickness. This is unexpected: there is no obvious reason why a cortex of different thickness should affect the estimate of the trabecular density neighbouring the cortex. A possible explanation is that beam hardening and scatter cause the 'true' HRpQCT data to appear more dense than it actually is. We have attempted to correct for HRpQCT beam hardening within the cortex itself (see Appendix B), but the correction does not apply to trabecular estimation immediately adjacent to the cortex. It is well known that areas surrounded by a ring of high density material appear too dense as a result of beam hardening and scatter (Boas and Fleischmann, 2012). The QCT data is less susceptible to this phenomenon since the high density cortices occupy far less of the field of view.

4.1 Processing time

Processing times for all techniques are relatively fast, taking typically just over a minute to complete all sets of measurements (cortical thickness, mass, density and endocortical trabecular density) on a typical femur with 11,000 measurement locations. These times were measured on a laptop with the software running single-threaded on a 2.67 GHz Intel Core i7 processor, and do not include the time to perform the initial segmentation, which is typically an additional 15 minutes. The processing time increase for the CBM v2 and CBM v3 techniques is less than might be expected from the additional model-fitting steps outlined in Fig. 1. This is because for these additional steps, the model is initialised with the previously optimised values, and hence fewer iterations are required to find the new optima.

[6] Since CBM v1 presumes a constant density, this graph actually shows the density distribution with cortical thickness in the HRpQCT data, albeit offset so that the maximum density is at zero.

5 Conclusions

CBM v2 and CBM v3 both produce better estimates of cortical thickness and density in the proximal femur than previously published alternatives. CBM v2 is the preferred choice for most applications. It is not predicated on any supposed relationship between density and thickness, nevertheless achieving thickness accuracy of 0.12 ± 0.39 mm and density accuracy of -26 ± 178 mg/cm^3 for cortices in the range 1–3 mm. For very thin (< 1 mm) cortices, CBM v3 offers superior accuracy of 0.04 ± 0.31 mm for thickness and -30 ± 264 mg/cm^3 for density, though this assumes a particular relationship between thickness and density which appears to hold for the human proximal femur but may not generalise to other targets. All techniques are capable of measuring cortical mass and endocortical trabecular density fairly reliably, though the CBM techniques do offer more precise mass estimation for very thin cortices.

Compared with CBM v1, the two new variants offer reduced thickness bias and also local estimates of cortical density which are relatively unbiased above 800 mg/cm^3. The imprecision of the cortical density estimates suggests that density is fundamentally more difficult to estimate than thickness.

Acknowledgments

The authors extend their sincere thanks to Dieter Pahr, of the Vienna University of Technology, for providing the HRpQCT and QCT scans of the ex vivo femurs.

References

Boas, F. E., Fleischmann, D., Apr. 2012. CT artifacts: causes and reduction techniques. Imaging in Medicine 4 (2), 229–240.

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.

Dall'Ara, E., Luisier, B., Schmidt, R., Kainberger, F., Zysset, P., Pahr, D., Jan. 2013a. A nonlinear QCT-based finite element model validation study for the human femur tested in two configurations *in vitro*. Bone 52 (1), 27–38.

Dall'Ara, E., Luisier, B., Schmidt, R., Pretterklieber, M., Kainberger, F., Zysset, P., Pahr, D., Nov. 2013b. DXA predictions of human femoral mechanical properties depend on the load configuration. Medical Engineering & Physics 35 (11), 1564–1572.

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.

Fajardo, R. J., Cory, E., Patel, N. D., Nazarian, A., Laib, A., Manoharan, R. K., Schmitz, J. E., DeSilva, J. M., MacLatchy, L. M., Snyder, B. D., Bouxsein, M. L., Jan. 2009. Specimen size and porosity can introduce error into μCT-based tissue mineral density measurements. Bone 44 (1), 176–184.

Gee, A. H., Treece, G. M., Feb. 2014. Systematic misregistration and the statistical analysis of surface data. Medical Image Analysis 18 (2), 385–393.

Hangartner, T. N., Mar. 2007. Thresholding technique for accurate analysis of density and geometry in QCT, PQCT and μCT 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.

Lang, T. F., Saeed, I. H., Streeper, T., Carballido-Gamio, J., Harnish, R. J., Frassetto, L. A., Lee, S. M. C., Sibonga, J. D., Keyak, J. H., Spiering, B. A., Grodinsky, C. M., Bloomberg, J. J., Cavanagh, P. R., accepted in 2014. Spatial heterogeneity in the response of the proximal femur to two lower-body resistance exercise regimens. Journal of Bone and Mineral Research.

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.

Pahr, D. H., Dall'Ara, E., Varga, P., Zysset, P. K., Jul. 2012. HR-pQCT-based homogenised finite element models provide quantitative predictions of experimental vertebral body stiffness and strength with the same accuracy as models. Computer Methods in Biomechanics and Biomedical Engineering 15 (7), 711–720.

Pakdel, A., Robert, N., Fialkov, J., Maloul, A., Whyne, C., Dec. 2012. Generalized method for computation of true thickness and X-ray intensity information in highly blurred sub-millimeter bone features in clinical CT images. Physics in Medicine and Biology 57 (23), 8099–8116.

Parker, M., Johansen, A., 2006. Hip fracture. British Medical Journal 333 (7557), 27–30.

Poole, K. E. S., Mayhew, P. M., Collette, M., Brown, J. K., Bearcroft, P. J., Loveridge, N., Reeve, J., 2010. Changing structure of the femoral neck across the adult female lifespan. Journal of Bone and Mineral Research 25 (3), 482–491.

Poole, K. E. S., Treece, G. M., Mayhew, P. M., Vaculik, J., Dungl, P., Horák, M., Štěpán, J. J., 2012. Cortical thickness mapping to identify focal osteoporosis in patients with hip fracture. PLoS ONE 7 (6), e38466.

Poole, K. E. S., Treece, G. M., Ridgway, G. R., Mayhew, P. M., Borggrefe, J., Gee, A. H., 2011. Targeted regeneration of bone in the osteoporotic human femur. PLoS ONE 6 (1), e16190.

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.

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.

Treece, G. M., Gee, A. H., Mayhew, P. M., Poole, K. E. S., Jun. 2010. High resolution cortical bone thickness measurement from clinical CT data. Medical Image Analysis 14 (3), 276–290.

Treece, G. M., Poole, K. E. S., Gee, A. H., Jul. 2012. Imaging the femoral cortex: thickness, density and mass from clinical CT. Medical Image Analysis 16 (5), 952–965.

Treece, G. M., Prager, R. W., Gee, A. H., 1999. Regularised marching tetrahedra: improved iso-surface extraction. Computers and Graphics 23 (4), 583–598.

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.

A Estimation of imaging blur in QCT data

In order to make use of the CBM v2 and CBM v3 algorithms, we need a reasonable estimate of the imaging blur σ_d for each QCT data set. This blur varies in all dimensions as well as with spatial location, the major variation being between the blur within each image slice (in-plane blur, σ_i) and that orthogonal to the slice (out-of-plane blur, σ_o). Some pertinent information is usually stored in DICOM files, though this is not sufficient to completely describe the in-plane/out-of-plane variation. In-plane blur is determined by the reconstruction kernel rather than the pixel size, but kernels are only recorded by name rather than extent. Also, although the recorded slice thickness does relate to the full-width half-maximum slice *selectivity*, further image processing in the reconstruction stage may lead to out-of-plane blurs which are larger than this.

We therefore need to estimate σ_d in each data set, and we can do this at the same time as the cortical density is estimated by looking at σ as well as y_1 as we fit the model in eq. (1) to each point on the femoral cortex. This parameter tells us the apparent imaging blur at each location, which we can record along with the angle α that the measurement line made with the imaging planes. The variation of σ with α, for three different QCT data sets, is shown as a distribution of measurement points in Fig. 14. These measurements are clearly very noisy, but the in-plane and out-of-plane differences are apparent. To estimate σ_i and σ_o from this data, we need a model of how σ_d depends on these parameters. One which we have found fits well is:

$$\sigma_d = \sqrt{(\sigma_i \cos\alpha)^2 + (\sigma_o \sin\alpha)^2} \tag{6}$$

This model is fitted to the measured data using a least-squares regression, resulting in estimates $\hat{\sigma_i}$ and $\hat{\sigma_o}$. Since this measurement process uses the FWHM algorithm (we do not yet know the cortical density so

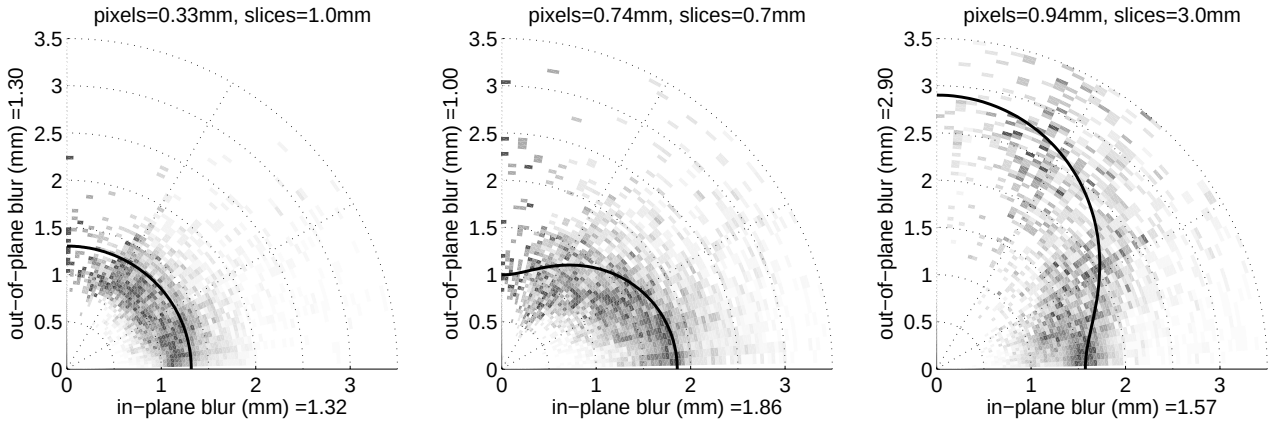


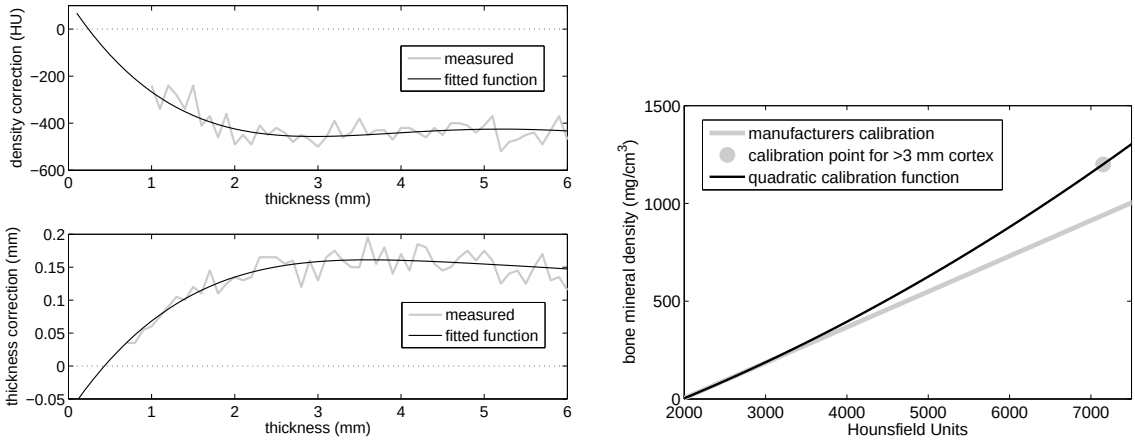
Figure 14: Estimation of the imaging blur in QCT data. The graphs show estimated Gaussian blurs from fitting the model in eq. (1) to multiple locations across the surface of the femur in three sample QCT data sets, with DICOM pixel sizes and slice thicknesses given at the top of each figure. Blur is plotted radially, with the angle corresponding to the measurement direction with respect to the image slices, such that the horizontal axis shows in-plane blur and the vertical axis shows out-of-plane blur. The solid line is the result of fitting the model in eq. (6) to these individual measurements. The values in the axis labels are the corresponding FWHM estimates of σ_i and σ_o.

cannot use one of the CBM algorithms), particular weight is given to σ_d estimates which come from cortices with large apparent peak density values c_{peak}, since the σ estimate is most accurate at these points. At other locations, σ will be underestimated as explained in Section 2.1. Hence the fitted models seen in Fig. 14 are towards the outer extents of the point clouds. Subsequently, whenever a blur estimate $\hat{\sigma_d}$ is required, we simply apply eq. (6) with the known measurement angle α and the estimated values of $\hat{\sigma_i}$ and $\hat{\sigma_o}$.

B Calibration of gold standard HRpQCT data

We need to pick an established algorithm as a gold standard against which to compare the various techniques introduced in this paper, and FWHM is the obvious candidate since it is known to perform well as long as the cortex is sufficiently thick in comparison to the imaging blur. For the HRpQCT data used in this paper, this corresponds to a thickness greater than 0.3 mm. However, our implementation of FWHM involves setting y_1 to the apparent peak in the CT data $c(x)$. At this resolution, the cortex is not observed as having uniform density: there are variations due to very small pores and also due to noise inherent in the imaging process. Hence the peak CT value is an overestimate of y_1 and this leads to a consequent underestimate of the cortical thickness. This bias can be removed by using the nothing preset algorithm: for HRpQCT data, this effectively sets y_1 to the average of the observed CT data within the cortex. However, this algorithm has more parameters to estimate and is therefore less precise than FWHM. We therefore use the FWHM algorithm as our HRpQCT gold standard, but correct the bias by calibration against the nothing preset algorithm, as shown in Fig. 15(a).

Conversion between Hounsfield Units and cortical density is largely achieved by the calibration phantom provided by the manufacturer of the scanner. However, the calibration densities are low compared to typical cortical densities. It has previously been noted that calibration in such scenarios is problematic: extrapolation of the low density calibration curve is imprecise, and there are also problems with beam hardening artefacts at higher densities (Fajardo et al., 2009). This was indeed the case in our study, with low trabecular densities showing a good match between QCT and HRpQCT, whereas high cortical densities did not. Since all



(a) (b)

Figure 15: Correction of HRpQCT FWHM thickness and density values. (a) On high resolution data, since we cannot use the CBM algorithms as a gold standard, the most reliable alternative is FWHM, whereas the least biased is nothing preset. By running both algorithms on all the data points from the high resolution data, calibration curves are fitted which allow the bias to be removed from the FWHM results. Note that these adjustments are very small, less than nearly all the precision values in Table 1. (b) The calibration provided by the manufacturer for the HRpQCT data is based on low densities around 300 mg /cm^2. Calibration for higher cortical densities is achieved by matching the data with the QCT values for thick cortices ($t > 3$ mm), for which the imaging blur is not problematic and all algorithms agree. The calibration curve was designed to match the manufacturer's calibration at low densities, and the single calibration point for thick cortices.

algorithms tested could very easily measure density in QCT data for thick cortices, and all returned the same results in this domain, the > 3 mm cortical densities were matched between QCT and HRpQCT, yielding an additional calibration point. Previous studies have also had to calibrate low and high densities separately (Pahr et al., 2012). The final calibration curve, shown in Fig. 15(b), matched the manufacturer's calibration at low density and the additional calibration point at high density. It must be acknowledged, however, that this is unlikely to correct for all beam hardening issues, and some imprecision will remain.