Estimation of bias and variance of measurements made from tomography scans

OPEN ACCESS
IOP Publishing
Measurement Science and Technology
Meas. Sci. Technol. **27** (2016) 095402 (17pp)
doi:10.1088/0957-0233/27/9/095402

Estimation of bias and variance of measurements made from tomography scans

Robert S Bradley[1,2]

[1] School of Materials, The University of Manchester, Manchester, UK

E-mail: rob.bradley@geotek.co.uk

Received 22 December 2015, revised 22 February 2016
Accepted for publication 24 June 2016
Published 9 August 2016

Abstract

Tomographic imaging modalities are being increasingly used to quantify internal characteristics of objects for a wide range of applications, from medical imaging to materials science research. However, such measurements are typically presented without an assessment being made of their associated variance or confidence interval. In particular, noise in raw scan data places a fundamental lower limit on the variance and bias of measurements made on the reconstructed 3D volumes. In this paper, the simulation-extrapolation technique, which was originally developed for statistical regression, is adapted to estimate the bias and variance for measurements made from a single scan. The application to x-ray tomography is considered in detail and it is demonstrated that the technique can also allow the robustness of automatic segmentation strategies to be compared.

Keywords: computed tomography, CT, x-ray, noise, bias, SIMEX, segmentation

(Some figures may appear in colour only in the online journal)

1. Introduction

Tomography refers to the imaging of the interior of a specimen in sections or slices, from which a 3D representation of the specimen can be constructed. A key application of tomography is the measurement of characteristics of internal features. For example, x-ray computed tomography (XCT) is being increasingly used as the basis of non-destructive measurement schemes across a wide range of disciplines. Measurements include, for example, tumour size in medical CT scans [1, 2] and porosity, defect size and crack growth in materials science and engineering applications [3–6]. However, such measurements are typically presented without an assessment being made of their associated variance or confidence interval. In particular, noise in scan radiographs places a fundamental lower limit on the variance (precision) and bias (accuracy) of measurements made on the resulting volumetric data sets. The complete process from scan acquisition to quantification should therefore be viewed as an estimation process.

Measurement uncertainty in XCT has recently received much attention for industrial dimensional metrology [7, 8]. In this application, a representation of the surface of a sample is generated from XCT data and then compared to a computer-aided design (CAD) model or reference workpiece to determine if the sample falls within a given tolerance. The overall measurement uncertainty has contributions from [7] (1) the geometric calibration of the tomography system, (2) the measurement process itself (from acquisition to quantification) and (3) variations in the object, arising from surface roughness and thermal expansion for example. There are a number of approaches to estimate (1) as reviewed in [9], which can be carried out before scanning the object of interest. Furthermore, (3) can be minimised by maintaining a suitably controlled environment around the sample. However, the measurement process uncertainty is the most difficult to determine. Kruth *et al* [7] reviewed methods for assessing this uncertainty which are applicable to dimensional

[2] Present address: Geotek Limited, Daventry, UK

metrology and suggested that approaches based on simulations or repeated measurements were the most promising. The simulation approach relies on a representative virtual sample being specified in terms of its geometry and its spatially varying complex refractive index. Ray tracing and Monte Carlo methods are then used to generate synthetic CT data, for which an accurate model of the tomography system is required, including probability distributions of all random variables. Such a scanner model will be complex and difficult to specify accurately [10, 11], and the simulation process is time consuming. A simplified bootstrap approach was proposed by Hiller *et al* [12] to estimate certain aspects of the measurement uncertainty. The alternative, experimental, approach is to take repeated measurements of a calibrated test specimen, which is of similar geometry and material as the actual samples [13]. The external dimensions of the test specimen are measured by a 'gold standard' technique (such as a coordinate measurement machine), enabling both the bias and repeatability of the CT measurements to be estimated. A correction for systematic bias can then be applied in subsequent measurements. Simulations can be combined with experimental results in order to reduce the experimental effort [14].

For general tomography applications, taking repeated measurements is not practical for a number of reasons: (1) scans times can be prohibitively long for high resolution data (e.g. sub-micron resolution XCT scans can take many hours particularly for low contrast specimens [15, 16]) or for clinical scanning; (2) it may be important to minimise dose (e.g. for dose sensitive materials or for clinical and preclinical CT); and (3) specimens may change over time (e.g. *in situ* x-ray and neutron tomography scans of dynamic processes [6, 17, 18]). Furthermore a 'gold standard' calibration measurement is not available for bias assessment except in specific cases, such as for porosity measurements by mercury porosimetry [19]. The simulation approach is similarly not generally applicable as the geometry and composition of the sample as well as the scanner model will not be known to high accuracy.

The effects of random errors in data have been widely studied in the context of statistical inference. Errors-in-variable or 'measurement error' models are regression models that correct for random errors in the dependent variables [20, 21]. These errors, if uncorrected, lead to bias in the parameter estimates, loss of power in detecting signals and the masking of features of the data [21]. The simulation-extrapolation (SIMEX) technique is a general purpose measurement error correction technique developed in the 1990s [22–25]. The SIMEX method involves two stages; the first consists of simulations to add random noise (with known variance and probability distribution) to measurement error covariates. For different noise variances, the mean values of the resulting parameters estimates are calculated over a large number of simulation repeats. The second stage involves using the variation of the parameter estimates with noise level to extrapolate back to the values when there is no measurement error. Cook and Stefanski [22] have shown that the SIMEX method produces an approximately consistent estimator of the true values, where the approximation relates to the accuracy of the extrapolation step. SIMEX is applicable when the bias is a smooth monotonic function of noise level, and that the noise level and distribution are well understood. A key advantage of the SIMEX method is that it is general purpose as it makes no assumptions about the distribution of the unobserved true covariate (other than it is known).

SIMEX has previously been applied outside of its original use in regression. For example, it has been adapted for use in differential equation models [26], receiver-operator curves [27], miss-classification [28], probability density deconvolution [29, 30], and MRI diffusion tensor imaging [31, 32].

In this paper, a framework is presented for applying the SIMEX method to investigate the variance and bias in measurements made from volumetric data, as arising from noise in the tomography acquisition. The adaption of the original SIMEX framework involves replacing the solving of the estimating equation in stage 1 with a 'tomographic measurement process', which comprises tomographic acquisition followed by a 'volumetric measurement process' for making measurements from the reconstructed data. Detailed consideration is given to the specific application to XCT, through a simulated test case and two case studies, and it is demonstrated that the method can be used to estimate the bias and uncertainty of measurements made from a single scan. Consideration is also given to minimising systematic sources of bias.

2. Theory

2.1. Volumetric and tomographic measurement processes

A 'volumetric measurement process' takes as its input a 3D volume of scalar data, and outputs one or more values which measure certain characteristics of the data. The process will generally involves a number of steps such as filtering (noise reduction), segmentation of the volumetric data into regions (or materials) followed by a calculation made on the binary (or labelled) data. The process can be described by function f_V, such that a scalar measurement M of the volumetric data $V(\mathbf{r})$, for spatial coordinate $\mathbf{r}$, is given by $M = f_V(V(\mathbf{r}))$. f_V is taken to be deterministic such that the only sources of variation are from the volumetric data. Manual segmentation is therefore not considered. SIMEX could be applied directly to the volumetric data if the noise characteristics of each voxel are known (see section 2.2). However, the tomography process typically results in the noise between nearby voxels being correlated and the noise power spectrum varying with position $\mathbf{r}$ [33]. This can occur due to the reconstruction step as well as due to detector blurring. In the case of transmission tomography, analytical expressions for the noise power spectrum in parallel beam, fan beam, and cone beam reconstructions have been derived for Fourier-based reconstruction algorithms [34–36]. However in practice, it may be difficult to obtain a good estimate of the noise characteristics, particularly when image processing steps are applied before reconstruction. Furthermore, experimental determination of the 3D noise characteristics is highly involved and time consuming [33].

In this paper, a more general approach is taken which is applicable to any particular tomographic imaging technique (which features additive noise), geometry, reconstruction algorithm and pre-processing steps. The general approach

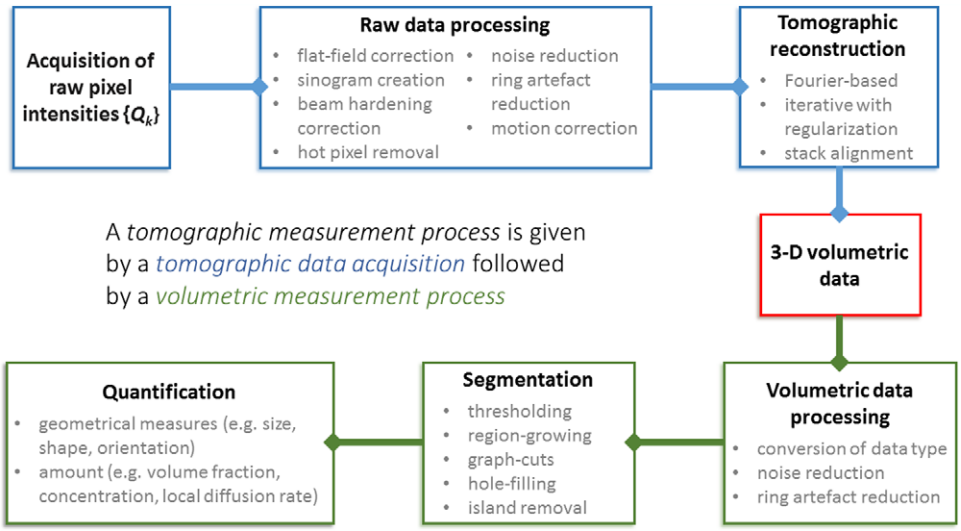

Figure 1. Schematic diagram illustrating the typical steps involved in a tomographic measurement process, which combines a tomographic data acquisition with a volumetric measurement process. At each step, examples are given of data processing procedures that could be employed.

involves incorporating together the tomographic acquisition with the volumetric measurement into a 'tomographic measurement process' as shown in figure 1. The process provides a mapping, f_T, from the ordered set of pixel intensities, Q_k, which comprise the sequence of scan radiographs or raw data, to the output measurement $M = f_T(\{Q_k\})$. Again, f_T is taken to be deterministic, such that any variation is due to noise in $\{Q_k\}$. In figure 1, examples are given of the types of algorithms which are typically applied at each step of the process. In XCT, corrections are often made to the raw projection data to reduce beam hardening artefacts, ring artefacts and scatter. However, the scheme is equally applicable to destructive tomography techniques, such as serial sectioning in an SEM, in which the reconstruction process would comprise alignment of the image stack and pre-processing steps may include correction for non-uniform illumination and perspective foreshortening [37]. Many of these processing step are non-linear so that noise in the acquisition will lead to bias in the measurement. In related work, Rajbhandary and Pelc [38] showed that the spectral range of polychromatic x-rays and the log transformation step taken during tomographic reconstruction introduce bias in material decomposition analysis, with the bias increasing with noise level.

2.2. SIMEX applied to a tomographic measurement process

For simplicity, consider the case in which the set of measured intensities, $\{Q_k\}$, of the N pixels in the detector are subject to additive mutually independent noise ε_k with mean zero and standard deviation σ_k, so that $Q_k = P_k + \varepsilon_k$, where P_k are the 'true' pixel values, which are taken to be constant for a given specimen and imaging setup (including specimen stage position and rotation). The naïve approach (using terminology from the statistics community) would be to apply the tomographic measurement process directly to the measured pixel intensities, $M_{\mathrm{NAIVE}} = f_T(\{Q_k\})$. However the naïve value will be biased away from the true value, M_{TRUE}, with the magnitude varying with σ_k as shown in the second order Taylor series expansion:

$$E(M_{\mathrm{NAIVE}}) = M_{\mathrm{TRUE}} + \frac{1}{2}\sum_{k=1}^{N}\sigma_k^2\frac{\partial^2 f_T(\{P_k\})}{\partial P_k^2} + \ldots \quad (1)$$

The SIMEX method involves adding noise to the pixel intensities with the same form as ε_k, such that a new set of intensities is calculated as:

$$Q'_{bk}(\lambda) = Q_k + \sigma_k\sqrt{\lambda}Z_b \quad (2)$$

where Z_b is an independent random variable having the same distribution (PDF) as ε_k but with variance of 1. The expected variance of Q'_k over a large number of simulation repeats $b = 1,2,\ldots B$ is then $\sigma_k^2(1+\lambda)$. The expectation of the corresponding measurement M' over B simulation repeats is then:

$$\mathrm{E}(M'(\lambda)) = \frac{1}{B}\sum_{b=1}^{B}M'_b(\lambda) = \frac{1}{B}\sum_{b=1}^{B}f_T\left(\left\{Q_k + \sigma_k\sqrt{\lambda}Z_b\right\}\right) \quad (3)$$

Provided that $E(M')$ is a smooth and continuous function of λ, the true value of the measurement can be estimated by calculating $E(M')$ for a range of values of λ (typically between 0–2) and extrapolating to $\lambda = -1$ where the expected variance of Q'_k tends to zero. That is:

$$\lim_{\lambda=-1}E(M'(\lambda)) = f_T(\{P_k\}) = M_{\mathrm{TRUE}} \quad (4)$$

SIMEX therefore provides an approximately consistent estimator of the true measurement. The accuracy of the estimator, M_{SIMEX}, is limited by the extrapolation step. Further details can be found in references [21, 22, 24, 25]. SIMEX can therefore be used for bias estimation and hence bias reduction provided that σ_k are known or are well estimated. The method can be extended to account for any correlations in noise between pixels, by generating noise from the appropriate multivariate distribution using the known variance-covariance matrix. A noise model to estimate σ_k for XCT is developed

in section 2.4, and applied to two laboratory XCT systems featuring CCD based detectors in section 4.1. Consideration is given to correlations between pixels for the two systems.

A limitation of SIMEX is that it is not straightforward to calculate the standard error of M_{SIMEX}. Analytical methods to produce 'asymptotic' standard errors have been derived for homogeneous measurement errors in regression applications [23]. Otherwise, the variance can be estimated in a similar way to M_{SIMEX}, namely by calculating the variance of $\{M'_b(\lambda)\}$ for several λ and extrapolating to $\lambda = -1$. The extrapolated value, multiplied by -1 yields an approximate estimate for the variance, $\mathrm{var}(M_{\mathrm{SIMEX}})$ [24]. The variance $\mathrm{var}\{M'_b(\lambda)\} = 0$ when $\lambda = 0$, so the procedure can be simplified in the case of linear extrapolation by only using the variance calculated for the minimum $\lambda > 0$:

$$\mathrm{var}(M_{\mathrm{SIMEX}}) \approx \frac{1}{\lambda_{\min}} \mathrm{var}(\{M'_b(\lambda_{\min})\}_{b=1}^{B}) \tag{5}$$

2.3. Extension to estimate the naïve measurement variance

It is useful to estimate the variance of the naïve measurement, $\mathrm{var}(M_{\mathrm{NAIVE}})$, particularly when the estimated bias is low or negligible. A parametric bootstrap method can be used for this, in a variation of the approach taken to estimate the voxel variances in positron emission tomography (PET). In the sinogram-based bootstrap approach for PET [39, 40], new sinograms are generated by first drawing samples from a Poisson distribution with parameter equal to the corresponding bin values of the original sinogram. The voxel variances are calculated from the reconstructions of the resampled sinograms. A similar procedure can be applied for our purposes by generating resampled projections (or raw data) by adding noise to the 'noise-free' estimates of $\{P_k\}$. The complete tomographic measurement process is then applied to the resampled pixels, and the measurement process variance estimated. If the measured Q_k are taken, on average, to be equal the true (noise free) values, P_k, then this bootstrap approach is equivalent to the simulation step of SIMEX for $\lambda = 1$.

It can be anticipated, therefore, that the variance of M' for $\lambda = 1$ will provide a good estimate of the variance of M_{NAIVE}, when the bias in M_{NAIVE} is 'low'. It is reasonable to expect that the tomographic measurement process will show some robustness, which can be seen theoretically by expanding to second order M_b' and comparing its expected variance to that of M_{NAIVE} calculated over $s = 1, \ldots, S$ repeat scans. In the case of the pixel noise being independent for simplicity, and setting $\lambda = 1$:

$$E_s \,\mathrm{var}_b(M'_{sb}(1)) \approx \mathrm{var}_s(M_{\mathrm{NAIVE}}) + [f_T(P_k) - E_s(M_{\mathrm{NAIVE}})] \times \sum_{k=1}^{N} \sigma_k^2 \frac{\partial^2 f_T(P_k)}{\partial^2 P_k} \tag{6}$$

Therefore, the variance over SIMEX repeats of M' for $\lambda = 1$ is expected to provide a good estimate of the variance of M_{NAIVE}, provided that f_T is a smooth function and that the bias is low and slowly varying.

2.4. A noise model for XCT

Noise in radiographs will contain contributions from detector noise (including thermal, readout and quantization noise) as well as from counting statistics. A complete model of the detector noise characteristics can be developed via a cascaded systems approach, such as that presented in [41]. In this section, a simpler model is developed that is better suited for practical implementation. The signal in the kth pixel can be modelled as follows:

$$Q_k = \gamma_k + C_k^{\mathrm{dark}} + \varepsilon_k^{\mathrm{det}} \tag{7}$$

where γ_k is the contribution to the signal from the detected photons, $\varepsilon^{\mathrm{det}}$ is the total detector noise (with variance σ_k^{det}) and C^{dark} is the dark current (with variance σ_k^{dark}). The form of γ_k depends on the type of detector [42]: if the detector is photon counting γ_k will be proportional to the weighted sum over photon energy of the number of detected quanta and will follow Poisson statistics; if the detector is energy integrating (representative of most detectors) then the signal will be proportional to the total imparted energy and will follow compound Poisson statistics. The detector is assumed to have a linear response, and in either case the statistical noise (shot noise) variance of γ_k is proportional to the mean value γ_k [42, 43]. This has been reported to be the case for a range of detectors [42, 44]. The dark current is typically measured before the start of the CT scan, by averaging n^{dark} images with the x-ray source switched off, to yield an estimate $\widetilde{C}_k^{\mathrm{dark}}$. Flat field reference correction is then applied to calculate the transmission through the object, T_k. An average of n^{ref} reference images are taken with the object removed from the field of view, to give an estimate $\widetilde{Q}_k^{\mathrm{ref}}$, and the transmission is calculated as $T_k = (Q_k - \widetilde{C}_k^{\mathrm{dark}})/(\widetilde{Q}_k^{\mathrm{ref}} - \widetilde{C}_k^{\mathrm{dark}})$. In the case of pixel noise being independent, the variance of T_k is then given as follows, as shown in the appendix:

$$\sigma_k^2 = \frac{\alpha T_k^2}{\gamma_k^{\mathrm{ref}}} \left[\frac{1}{T_k} + \frac{1}{n^{\mathrm{ref}}} + \frac{\psi_k^{\mathrm{det}}}{\gamma_k^{\mathrm{ref}}} \left(\frac{1}{T_k^2} + \frac{1}{n^{\mathrm{ref}}} \right) \right] \tag{8}$$

where $\psi_k^{\mathrm{det}} = \alpha^{-1}(\sigma_k^{\mathrm{dark}\,2}/n^{\mathrm{dark}} + \sigma_k^{\mathrm{det}\,2})$. Therefore, resampled transmission images in the 1st stage of the SIMEX method can be generated by adding heteroscedastic noise of the above form. The practical application of the noise model for two microCT systems is considered in section 4.1.

3. Simulation test case

The feasibility of applying SIMEX to XCT data was assessed using a simulated parallel beam scan of a uniform cylinder over a wide range of noise levels. Two measurements were considered, namely the cylinder cross-sectional area and radius.

3.1. Description of the simulation

The cylinder had a radius of 10 voxels and the simulated scan was carried out using a 1D line detector of 201 pixels, with 316 radiographs being taken over 180°. The cylinder

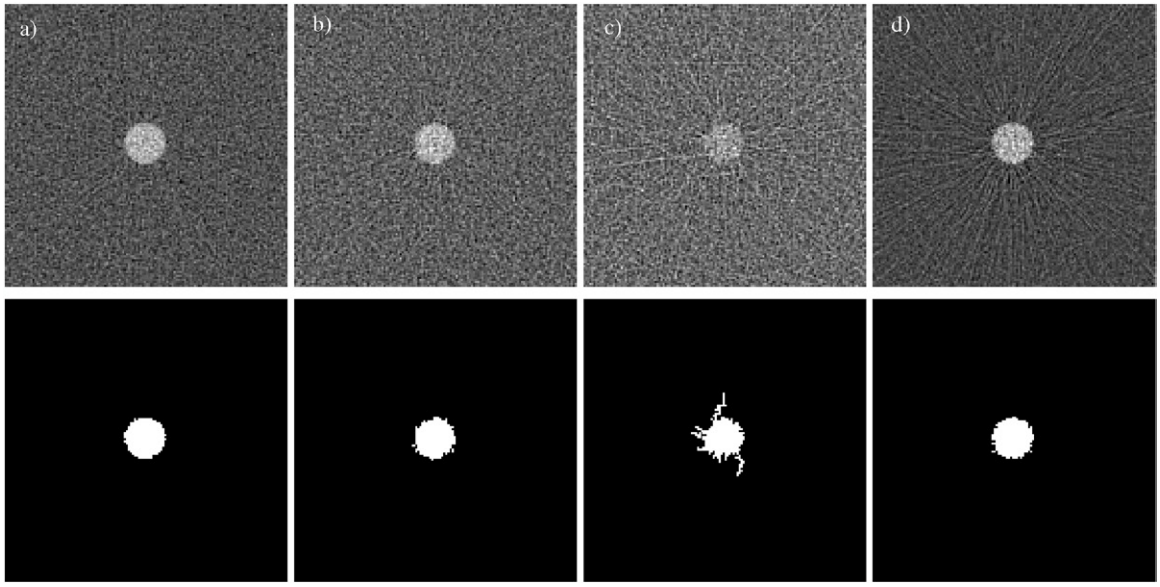

Figure 2. Simulated reconstructed slices of a cylinder (top row) and the corresponding binary segmentations (bottom row). For (a)–(c), the transmission through the centre of the cylinder was set to $T_c = e^{-0.5} \approx 61\%$, with the noise level, $\gamma^{ref-0.5}$, being (a) 0.15, (b) 0.2 and (c) 0.35. In (d) $T_c = e^{-2} \approx 13.5\%$, and $\gamma^{ref-0.5} = 0.35$.

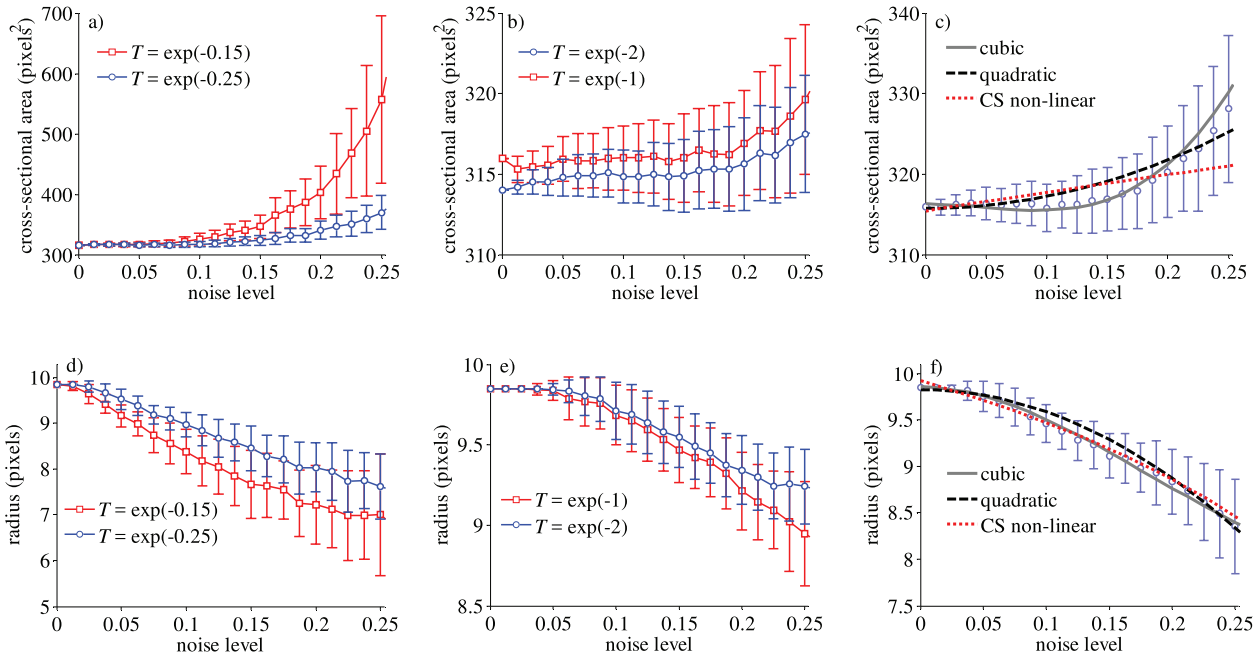

Figure 3. Variation of measured cross-sectional area (top row) and radius (bottom row) with noise level ($\gamma^{ref-0.5}$) for the simulation test case, with transmission (T) through the centre of the cylinder being (a) and (d) 77.9% and 86.1%, (b) and (e) 13.5% and 36.8%, (c) and (f) 60.7%. In (c) and (f) the variation with $1/\gamma^{ref}$ is modelled by quadratic and cubic functions as well as the non-linear function proposed by Cook and Stefanski (CS non-linear [22]), namely $y(x) = a + b/(c + dx)$ for parametric regression. Noise level has units of counts$^{-0.5}$.

was taken to be non-scattering and the transmission through the centre of the cylinder, T_c, using monochromatic illumination, was varied from 86% to 2.75% to cover a wide range of materials from soft tissues to highly absorbing super alloys. Gaussian noise was added to the simulated projections using equation (8), with α set to 1 and the noise level ($\gamma_k^{ref-0.5}$, which was set to be the same for all pixels, corresponding to uniform illumination of the field of view) being varied over a wide range from 0 to 0.25 (detector counts as low as 16), while n^{ref} was set to 21. The detector noise and dark current were set to zero so that the effects of shot noise could be considered in isolation. Filtered back-projection was used to reconstruct the volumetric data and the cylinder was segmented from background by applying a threshold set halfway between the grey levels of the object and background for the no-noise case. The largest connected set of pixels was then selected as being the cylinder and any internal holes were filled.

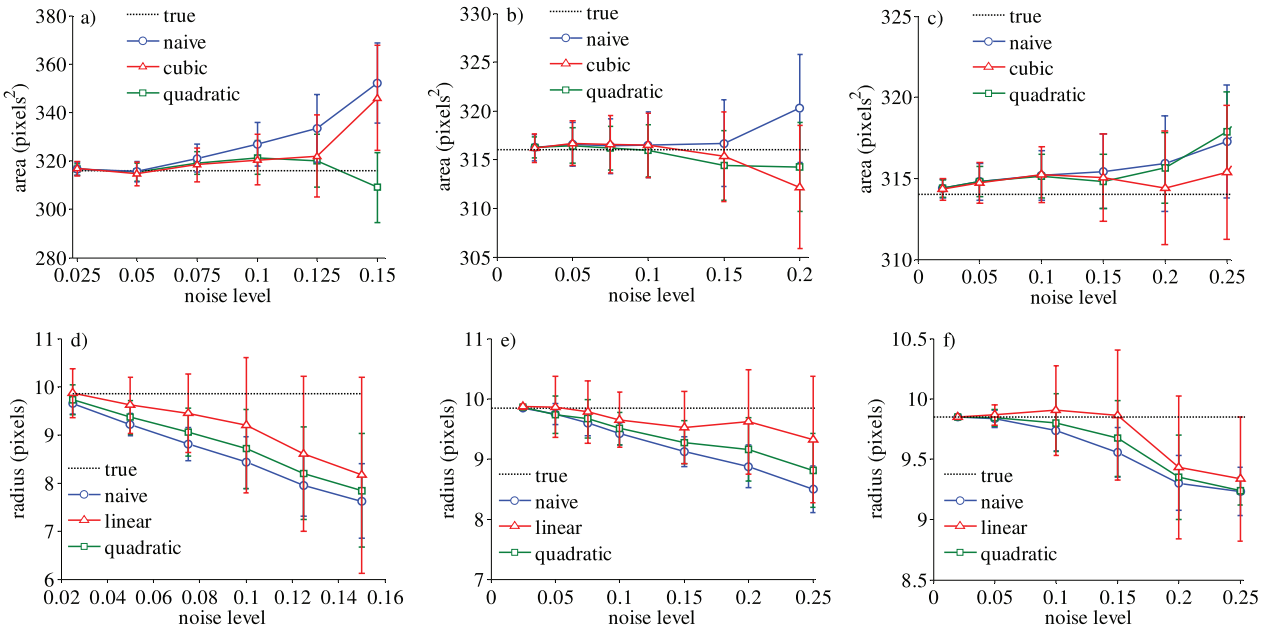

Figure 4. SIMEX estimate of cross-sectional area (top row) and radius (bottom row) for the simulated cylinder as a function of noise level ($\gamma^{\text{ref}-0.5}$). The extrapolation step was carried out either using a cubic, quadratic or linear function, with a restricted λ range of $\{0, 0.25$ and $0.5\}$ used for radius. The naïve (uncorrected) and true values are shown for comparison. The transmission through the centre of the cylinder, T_c, was (a) and (d) 86.1%, (b) and (e) 60.7% and (c) and (f) 13.5%. Noise level has units of counts$^{-0.5}$.

Examples of reconstructed slices and corresponding binary segmentations are shown in figure 2. Two measurements were considered, namely the cylinder cross-sectional area (with theoretical value of 314 pixels^2) and radius (nominal value of 10 pixels) as determined by the maximum of the distance transform within the segmented region. The process was repeated 200 times for each noise level. SIMEX was applied to each simulated scan, using equations (2) and (8) with λ taking the values 0, 0.25, 0.5, 1, 1.5, 2. For each $\lambda > 0$, 75 simulation repeats were made. The complete process was repeated 50 times for each combination of scan contrast and noise-level.

3.2. Variation of measurements with noise

The measurements of area and radius are shown in figure 3 as a function of transmission and noise level. Area generally increases with noise level as noisy background voxels become included in the segmentation (see figure 2(c)) while radius decreases as holes in the segmentation appear at the cylinder edges. Both measurements vary smoothly and show robustness to noise, as the gradients tend to zero at low noise. In the 'low bias' regime, the bias is close to zero (e.g. <1 pixels^2 for area and 0.05 pixels for radius) and slowly varying, but the standard deviation increases more rapidly with noise-level. More generally, the variation can be complex and not strictly monotonic, depending on both the contrast and noise level.

When the contrast of the object is relatively low ($T_c \sim$ 60% or greater), the variation for both measurements can be approximated well by polynomial functions up to high biases, as shown in figures 3(c) and (f). The gradients are close to zero in the low bias regime and are approximately constant at high bias. When the behaviour is more complex (figures 3(a) and (b)), simple extrapolants would still be expected to yield accurate predictions but over a restricted noise range (i.e. close to the low bias regime). In practice, scans are unlikely to be taken with noise levels greater than ~0.1 (see section 4), particularly if manufacturers recommendations are followed (e.g. in the Zeiss Xradia versa user manual recommends a minimum for detector counts of 2000 which corresponds to a noise level of ~0.016, see section 4.1).

3.3. Bias reduction

SIMEX was applied using the quadratic and cubic extrapolants that capture the variation of bias with noise level over a reasonable range (see figure 3). To ensure that the extrapolants were well-behaved, the gradients were set to zero at $\lambda = -1$. For radius, linear extrapolation was also considered, and a restricted range of λ, namely $\{0, 0.25, 0.5\}$, was found to provide more accurate estimates as this was less sensitive to the decrease in gradient at high biases (see figure 3(d), $\gamma^{\text{ref}-0.5} > 0.24$).

Figure 4 shows that SIMEX on average can achieve a useful amount of bias reduction over a wide range of contrasts and noise levels, as well as confirm when the bias is low or negligible. For example, the majority of the bias (over 50%) in both area and radius was eliminated at a noise level of $\gamma^{\text{ref}-0.5} = 0.1$ for $T_c = 86.1\%$. In general, the cubic function provided the better estimates for area, while the linear function was more accurate for radius. The trends in the estimates largely reflect how well the local fit of the function (over λ) extrapolates to capture the lower noise variation. For example, the performance was not as good when the gradient of bias with noise level decreases, as for $T_c = 13.5\%$ at noise levels

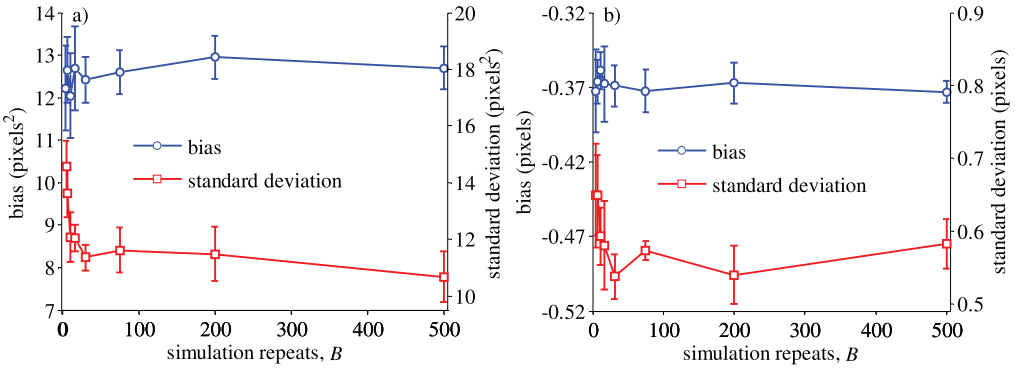

Figure 5. Mean and standard deviation of SIMEX estimates of bias for $T_c = 86.1\%$ and $\gamma^{\text{ref}-0.5} = 0.125$ as a function of number of simulation repeats, B, for (a) area and (b) radius. Quadratic extrapolation was used for area and linear extrapolation for radius. SIMEX was applied to 100 data sets and the complete processes was repeated 3–5 times for calculation of the average values and standard deviations shown.

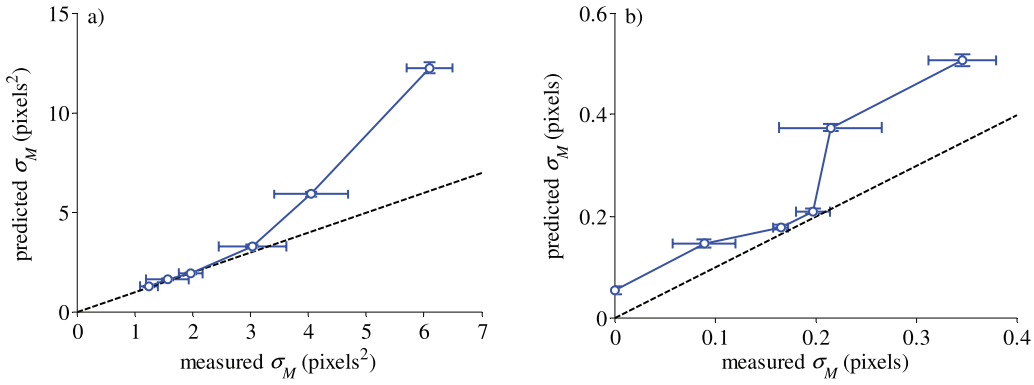

Figure 6. Comparison between the measured and predicted values of the naïve measurement standard deviation (σ_M) for (a) area and (b) radius of the simulated cylinder. T_c was set to 60.6%. SIMEX was applied to 40 data sets at each noise level and $B = 50$. The complete process was repeated 5 times to calculate the average values and standard deviations shown.

of ~0.075 for area and ~0.2 for radius (see figures 3(b), (d) and 4(c), (f)). In this case the simple extrapolation functions used here do not capture this more complex variation.

The performance of SIMEX as a function of simulation repeats, B, was investigated for $T_c = 86.1\%$ and a noise level of $\gamma^{\text{ref}-0.5} = 0.125$, as shown in figure 5. The SIMEX bias estimates show little variation with the number of simulation repeats, B. However the standard deviation does increase rapidly as B is reduced to below ~17. Practical application of SIMEX will therefore be a compromise between computation time and variability of the SIMEX estimate. These results indicate that a reasonable trade off can be achieved with B as low as ~11.

3.4. Estimation of the naïve measurement variance

The standard deviation of the naïve measurement, σ_M, was estimated using equation (6), over a wide range of noise levels for $T_c = 60.7\%$ as shown in figure 6.

The predicted standard deviations for cross-sectional area were in good agreement with the actual values for noise levels up to 0.1 (measured σ_M being 3.0 ± 0.5 pixels^2), where the bias in the naïve measurement is low. Above this noise level, as the bias in the naïve measurement increases, the predicted standard deviations become an over estimate. The method is less accurate for the radius as this shows a more complex variation with noise level. However, the method can still provide a reasonable approximation for noise-levels up to 0.1 (corresponding to the measured σ_M being 0.20 ± 0.02 pixels). At higher noise levels, the method can still provide a good order of magnitude estimate, with the predicted standard deviations being within a factor of 2 of the actual values even at the highest noise level considered here.

3.5. Estimation of the SIMEX measurement variance

The variance of the SIMEX estimate, var(M_{SIMEX}), is shown in figure 7 as a function of increasing noise for $T_c = 60.6\%$. The variation is predominantly linear for both area and radius except at low noise levels. Estimates of the variance using equation (5) with $\lambda_{\text{min}} = \{0.25, 1\}$ are shown for comparison. Overall, better estimates are produced with $\lambda_{\text{min}} = 1$, which indicates that the majority of the variance is associated with that of the naïve measurement (see equation (6)). It is likely that the values for $\lambda_{\text{min}} = 0.25$ are affected by the decrease in the gradient at low noise levels (or equivalently when λ is small) as indicated by the curves in figure 7.

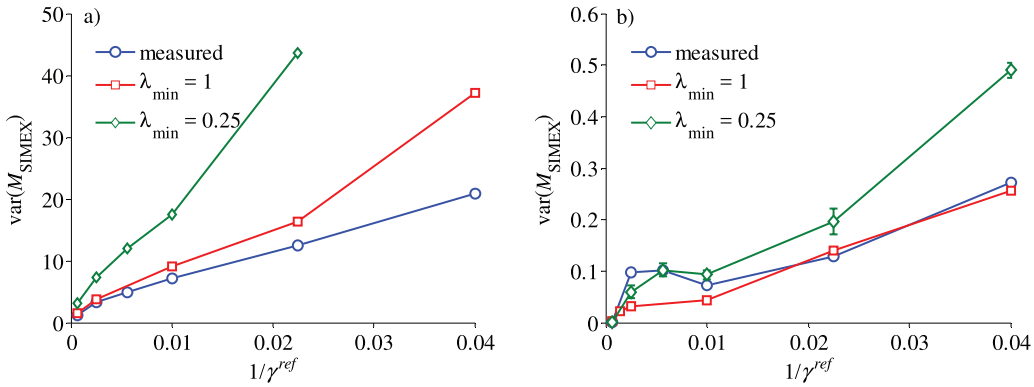

Figure 7. Variance of the SIMEX estimate of (a) area and (b) radius of the simulated cylinder as a function of noise level ($1/\gamma^{\text{ref}}$) for $T_c = 60.6\%$. The measured value is displayed alongside the estimates from equation (5) with $\lambda_{\min} = 0.25$ and 1.

4. Case studies

SIMEX was applied to measurements made on an aluminium cylinder and a mouse femur, which were scanned using two different cone-beam micro XCT systems, but of similar design, namely a Zeiss Xradia microXCT-400 and a Zeiss Xradia versa-520. The scan settings are shown in table 1 and a range of noise levels were considered by varying the exposure time per projection.

Repeat scans were taken to enable the variance of the measurements at each noise level to be calculated. However it was found that systematic changes in the scanners resulted in large variations in the measurements, above that attributable to noise in the scan radiographs. Therefore a segmentation strategy was developed to reduce these systematic effects, which is discussed in section 4.2. The application of the noise model to the two scanners is detailed in section 4.1. SIMEX was applied using code developed in Matlab 2013a (Mathworks, Natick, MA, USA) in which the FDK reconstructions were run automatically by calling the Zeiss Xradia XMreconstructor software from the command line. Code is available from www.github.com/rsbradley/tomotools. The results for the two samples are detailed in sections 4.3 and 4.4.

4.1. Application of the XCT noise model

The noise characteristics of the imaging system can be determined from the noise model given by equation (8), by measuring the noise variance, σ_k^2, as a function of pixel intensity for $\{T_k\} = 1$ (i.e. no object being imaged). However, there are likely to be systematic changes in pixel intensity over repeat radiographs (see section 4.2). Instead, the noise characteristics can be measured from the variance of a small region of a projection image, where Q_k^{ref} is approximately constant, provide the net detector noise, ψ_k^{det}, is constant or slowly varying over the detector and the correlations between pixels is small. This can be shown mathematically by making use of the fact that for a collection of random variables X_i ($i = 1, 2 \ldots n$) with finite means $E(X_i) = \mu_i$ and finite variances σ_i^2, the expectation of the sample variance (S^2) of X_i can be written as [45]:

$$E(S^2) = \frac{1}{n}\sum_{i=1}^{n}\sigma_i^2 + \frac{1}{n-1}\sum_{i=1}^{n}\left(\mu_i - \frac{1}{n}\sum_{i=1}^{n}\mu_i\right)^2 - \frac{2}{n}\sum_{i=2}^{n}\sum_{j=1}^{i-1}\rho_{ij}\sigma_i\sigma_j \quad (9)$$

Table 1. Scan settings used for the case studies.

Parameter	microXCT-400	versa-520
Source voltage (kV)	120	90
Source current (μA)	83	90
Source to sample distance (mm)	35	12
Sample to detector distance (mm)	7	12
Number of projections	271	721
Angular range	192°	360°
Reconstructed voxel size (μm)	5.49	3.40

Note: Beamharding correction was applied using a second order polynomial.

where the last term equals the average value of the off-diagonal elements of the covariance matrix with ρ_{ij} being the correlation coefficient between X_i and X_j. Zhang *et al* [44] found that for a Varian flat panel detector there were only non-zeros correlations between the nearest neighbour pixels.

75 repeat images were taken for a range of exposure times for the two microCT systems, operating under the same conditions as detailed in table 1. A central 7×7 image patch was taken for calculation of the pixel variances and the correlation coefficients. Figure 8 shows that for the versa-520 system, there were essentially no correlations between pixels, and for the microXCT-400 system that there were only substantial correlations between the 8 nearest neighbours, with correlation coefficients in the range 0.16–0.52. Therefore, the underestimation in the variance by not accounting for the correlations between pixels is expected to be less than 2% (assuming σ_i are equal) for the CT systems used in this study.

Figure 9 shows that equation (8) can accurately model the pixel standard deviations for both tomography systems used in this work, provided that T_k can be estimated. In the following work, 'noise-free' estimates of T_k were generated by filtering the transmission images with a median filter of size 5×5 pixels.

4.2. Systematic variation and segmentation strategy

Systematic changes in the x-ray source and detector over time can lead to changes, between scans, in the grey-scale histograms of the reconstructed volumes. Segmentation based on a constant threshold will therefore be particularly sensitive to these systematic changes.

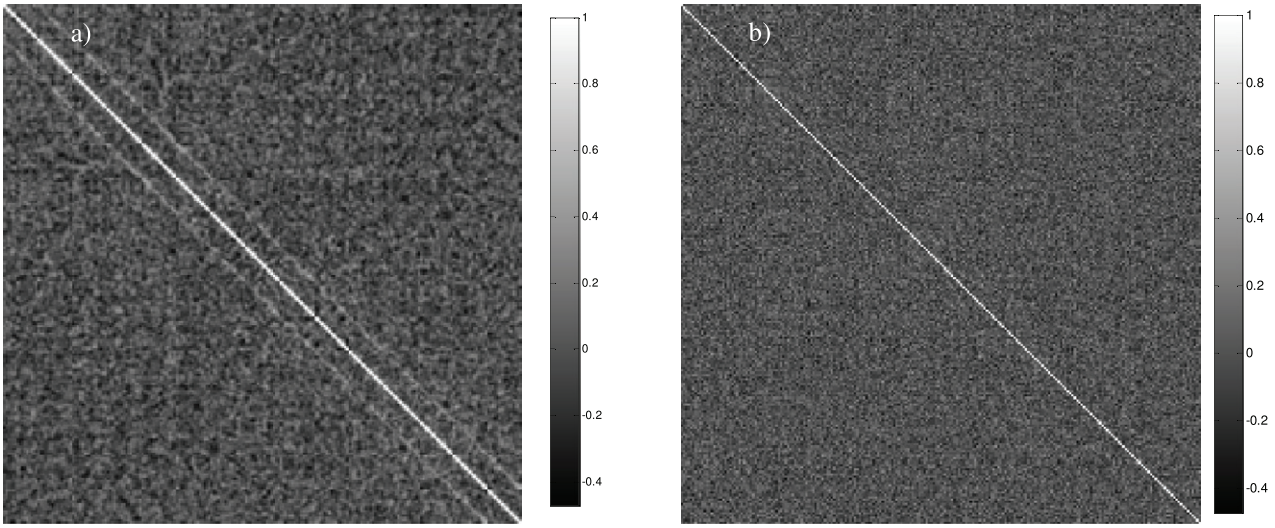

Figure 8. Matrix of correlation coefficients among 49 detector pixels obtained from 75 repeat measurements for a 7 × 7 patch. The measurements were made for (a) the Zeiss Xradia microXCT-400 and (b) the Zeiss Xradia versa-520, operating with the same conditions as specified in sections 4.3 and 4.4. The pixels are listed in a row-wise manner such that pixel $i + 1$ is vertically below pixel i, and pixel $i + 7$ is the right-hand neighbour of pixel i.

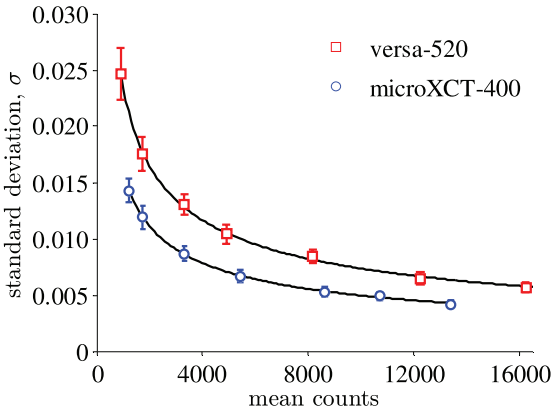

Figure 9. Standard deviation, σ, as a function of mean counts for two microCT systems. The black lines show the fit of equation (8) to the data. The fitted parameters were $\alpha = 0.242 \pm 0.004$, $\psi^{\text{det}} = 14 \pm 27$ for the microXCT-400, and $\alpha = 0.53 \pm 0.01$, $\psi^{\text{det}} = 10 \pm 35$ for the versa-520.

Characteristics of x-rays sources, such as x-ray flux, spectrum and focal spot position, can vary over a range of timescales. Modern x-ray tubes can achieve a relatively stable output flux, after an initial warm up period, over a typical length of a ~1 micron resolution scan (i.e. <1% variation over 8 h or so). By contrast, the focal spot position can take a few hours to stabilise [11, 46]. Over longer times Fukuda *et al* [47] found that the half-value layer (the thickness of material required to reduce the x-ray flux by half, and is dependent on the x-ray spectrum) varied for a particular x-ray source by ±2% over a 103 week period, while for another source the entrance air kerma increased by ~11%. Similarly, the output of x-ray detectors can vary over a range of timescales [48]. For example, flat panel detectors can show significant lag [49, 50]. Over longer times, for both direct and indirect CMOS flat panels, Han *et al* [51] found that the dark pixel signal increase approximately quadratically with dose, reducing the dynamic range of the detector.

In this work, an automatic strategy for binary segmentation is developed which is less affected by these systematic variations. This 'volumetric thresholding' strategy is based on setting the threshold value such that the segmented volume of object in a given ROI is kept constant across data sets. The ROI must contain object as well as background. This is appropriate to use if one or more regions of the object do not change between scans. If the object does change, or if different objects are scanned, then a 'volumetric standard' could instead be placed in the scanner field of view. The volumetric standard should have approximately the same composition as the objects and it's segmented volume should be the same across all data sets. The strategy is applied by using a threshold to segment the volumetric standard in the 1st data set. The threshold for subsequent data sets is then adjusted until the segmented volume equals that obtained for the 1st. The optimization of the threshold was carried out using the Nelder–Mead simplex method [52] in Matlab.

The robustness of the volumetric thresholding strategy was determined by taking repeat scans with identical scanner settings, as shown in figure 10 for the mouse femur and aluminium samples (see sections 4.3 and 4.4 for further details and scan settings). Care was taken to ensure that the same region of the sample was analysed for all repeats scans, as assessed visually to pixel accuracy. Two other common strategies are shown for comparison namely: (1) set a contrast threshold value for all scans, and (2) calculate the threshold for each data set as 0.5 × (mean grey level of object + mean grey level of surrounding material) from user specified region-of-interests (ROIs). Strategy 2 will be referred to as ROI thresholding.

For both samples, which were scanned on two different systems, there is evidence of systematic changes in the measurements with repeat scan, and hence time. For the constant and ROI thresholding strategies, there are low frequency trends in the measurements (which are almost cyclic in nature

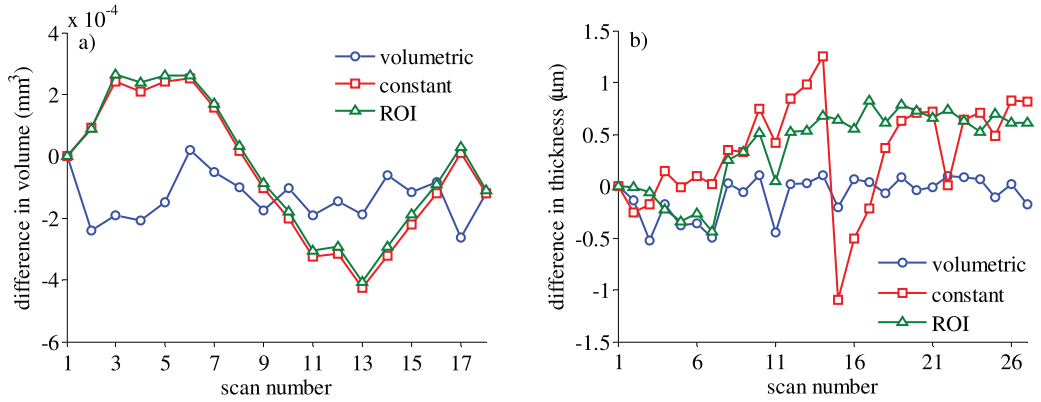

Figure 10. Variation of measurements over repeated scans for the three segmentation strategies, namely volumetric thresholding, constant thresholding, and ROI thresholding. The differences in the measurements relative to the first scan are shown. (a) Volume of the aluminium cylinder scanned with a Zeiss Xradia microXCT-400 system with 1.5 s exposure time. 18 scans were taken consecutively over ~17 h. (b) Mean local thickness of the mouse femur scanned with a Zeiss Xradia versa-520 system with a 1 s exposure time. The sample was scanned in 3 blocks: scans 1–7 (over ~5 h) followed 5 d later by scans 8–21 (over ~10 h), a gap of ~9 h, then scans 22–27 (over ~4 h).

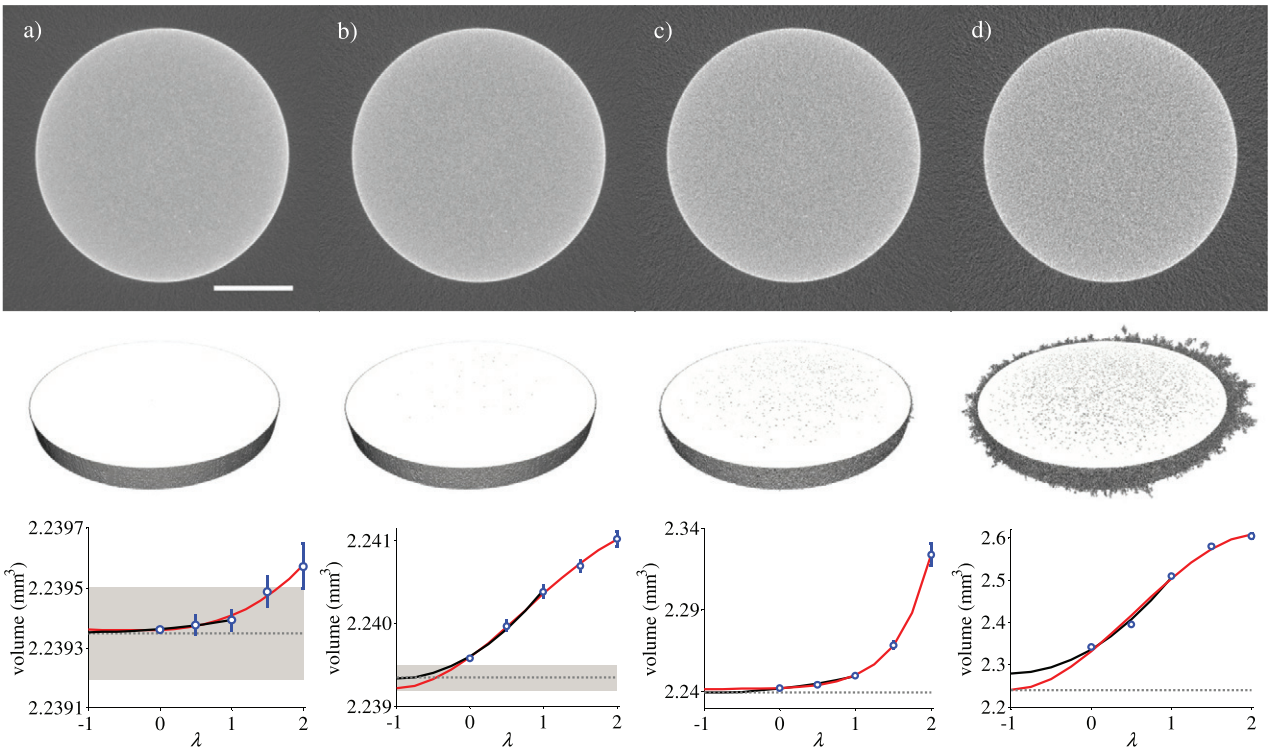

Figure 11. Reconstructed slices through the aluminium cylinder scanned with an exposure time of (a) 1.5 s, (b) 0.75 s, (c) 0.3 s, and (d) 0.15 s. The scale bar is 1 mm. 3D renderings of the corresponding segmented regions are shown in the middle row. The bottow row shows the corresponding variation of the segmented volume with SIMEX λ. The error bars show the standard deviation over $B = 14$ repeats, and the red and black lines show the fitted extrapolents for the full and restricted λ range, respectively. The dotted black line shows the mean value of the segmented volume over 17 repeat scans taken at 1.5 s exposure, and the grey rectangle denotes the corresponding 95% confidence interval.

for the aluminium sample), which are largely removed by the volumetric strategy.

The likely contribution from mechanical drift was assessed by retaking the measurements for the 1st scan after shifting the data by ± 1 pixel in the vertical direction. Both samples extended outside of the analysis ROI in the vertical direction only, making the measurements more sensitive to shifts in this direction. For the aluminium sample, the volume varied similarly for the 3 segmentation strategies and was in the range 2–4 $\times$ 10^{-5} mm^3, well below the variation evident in figure 10(a) which corresponds to the lowest noise data considered in section 4.3. For the mouse femur sample, the variation with pixel shift was in the range 0.3–0.9 μm. Consequently, any notable mechanical drift (not corrected for by the manual assessment to pixel accuracy) would therefore be evident as a trend in figure 10(b) for the volumetric strategy. It can be concluded therefore that mechanical drift is likely to have had only a minimal effect in both case studies presented in the following sections.

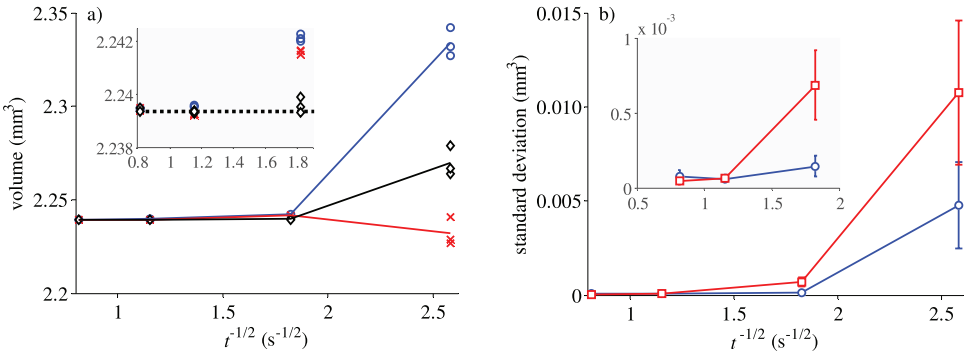

Figure 12. (a) Segmented volume of the aluminium cylinder as a function of noise level, which is proportional to $t^{-1/2}$. The blue circles show the naïve values for the 3 scans, and the corresponding SIMEX estimates are given by the black diamonds for the restricted λ range and the red crosses for the full λ range. The dashed line shows the mean value for 1.5 s exposure over 17 repeat scans. (b) Standard deviation of the segmented volume, determined using the naïve approach, as a function of noise level. The measured values over 18–19 repeat scans are show by the blue circles. The estimated standard deviation is show by the red squares, which are average values from SIMEX applied to 3 scans. The error bars show the 95% confidence intervals, which were calculated assuming normality for the measurements, and calculated for the estimated values from the standard deviation over the 3 scans.

4.3. Volume of an aluminium cylinder

4.3.1. Methods. A 3.2 mm diameter aluminium cylinder was scanned using a Zeiss Xradia microXCT-400 system with settings given in table 1. Scans were taken at a range of noise levels by varying the exposure time per projection from 0.15 s to 1.5 s, corresponding to contrast-to-noise ratios between aluminium (at the centre of the cylinder) and air being in the range 8.4–2.7. Reconstructed slices are shown in figure 11. The volume occupied by the aluminium cylinder in 51 slices was measured, by first applying a grey-level threshold to segment the cylinder. The threshold was set using the 'volumetric thresholding' strategy described above, by matching the volume measured over the upper most 5 slices to that determined for a scan taken with an exposure time of 1.5 s. Small isolated regions in the segmentation (islands) were removed if they were less than 1000 voxels in volume and holes were filled in they were less than 51 voxels. Care was taken to ensure that the same region of the sample was analysed for all repeats scans (to pixel accuracy). Volume renderings of the segmented volumes are presented in figure 11 (middle row). At high noise, segmented regions of the background connected to the cylinder are evident, increasing the volume measured, which is counteracted to some degree by holes within the cylinder.

SIMEX was applied to 3 scans at each exposure time for $B = 14$ repeats. Two ranges for λ were considered, namely the 'standard range' $\{0, 0.5, 1, 1.5, 2\}$ and the 'restricted range' $\{0, 0.5, 1\}$. Extrapolation was carried out using the quadratic function for the restricted λ range and either a cubic or exponential function for the standard range, with the extrapolent giving the best fit chosen based on the maximum R^2 value.

4.3.2. Results. Figure 11 (bottom row) shows the variation of volume with λ for a scan at each exposure time. The curves show that no bias is evident for the 1.5 s exposure time, but at shorter scan times, positive bias becomes apparent which increases with noise level. The predicted volumes given by SIMEX are shown in figure 12(a), together with the naïve values. A reduction in bias is achieved for all exposure times less than 1.5 s (for which no bias is evident). For example, the

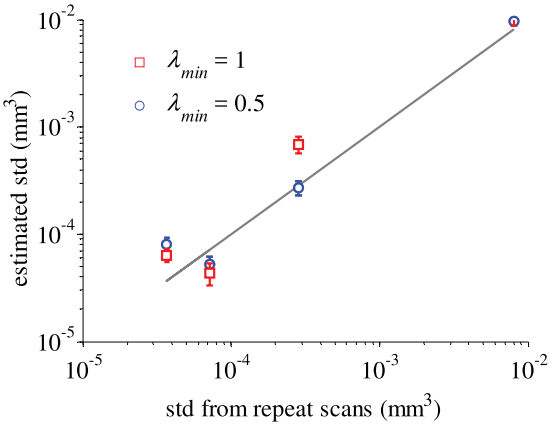

Figure 13. Comparison between the measured and estimated standard deviations of the cylinder volume determined by SIMEX. The measured values were calculated from 3 repeat scans, while the estimated values were calculated from SIMEX repeats with $\lambda_{\min} = 0.5$ or 1 using equation (8). Equality between the 2 sets of values is shown by the line. A linear function was used in the extrapolation step of SIMEX.

average bias for the naïve approach (relative to the average naïve value at 1.5 s exposure) is $0.12 \pm 0.06\%$ at 0.3 s exposure time compared with $0.009 \pm 0.013\%$ for SIMEX with the restricted λ range. Cubic extrapolation with the full λ range was most accurate for the noisiest data (0.15 s exposure time), with the mean bias being only $-0.3 \pm 0.3\%$, compared with $1.4 \pm 0.4\%$ for quadratic extrapolation with the restricted λ range and $4.2 \pm 0.3\%$ for the naïve approach.

The estimated naïve measurement standard deviations (from SIMEX with $\lambda = 1$) are in good agreement with the experimentally determined values over 18–19 repeat scans when the bias is low (see figure 12(b)). As the bias increases above ~0.1%, the SIMEX values are an overestimation. However even for the largest bias, the SIMEX technique can still provide a reliable order of magnitude estimate. Similarly, the standard deviation of the SIMEX estimate was found to be reasonably well predicted by equation (5) for both $\lambda_{\min} = 0.5$ and 1 (as shown in figure 13) with all predicted values being within factor of 2.2 of the measured values.

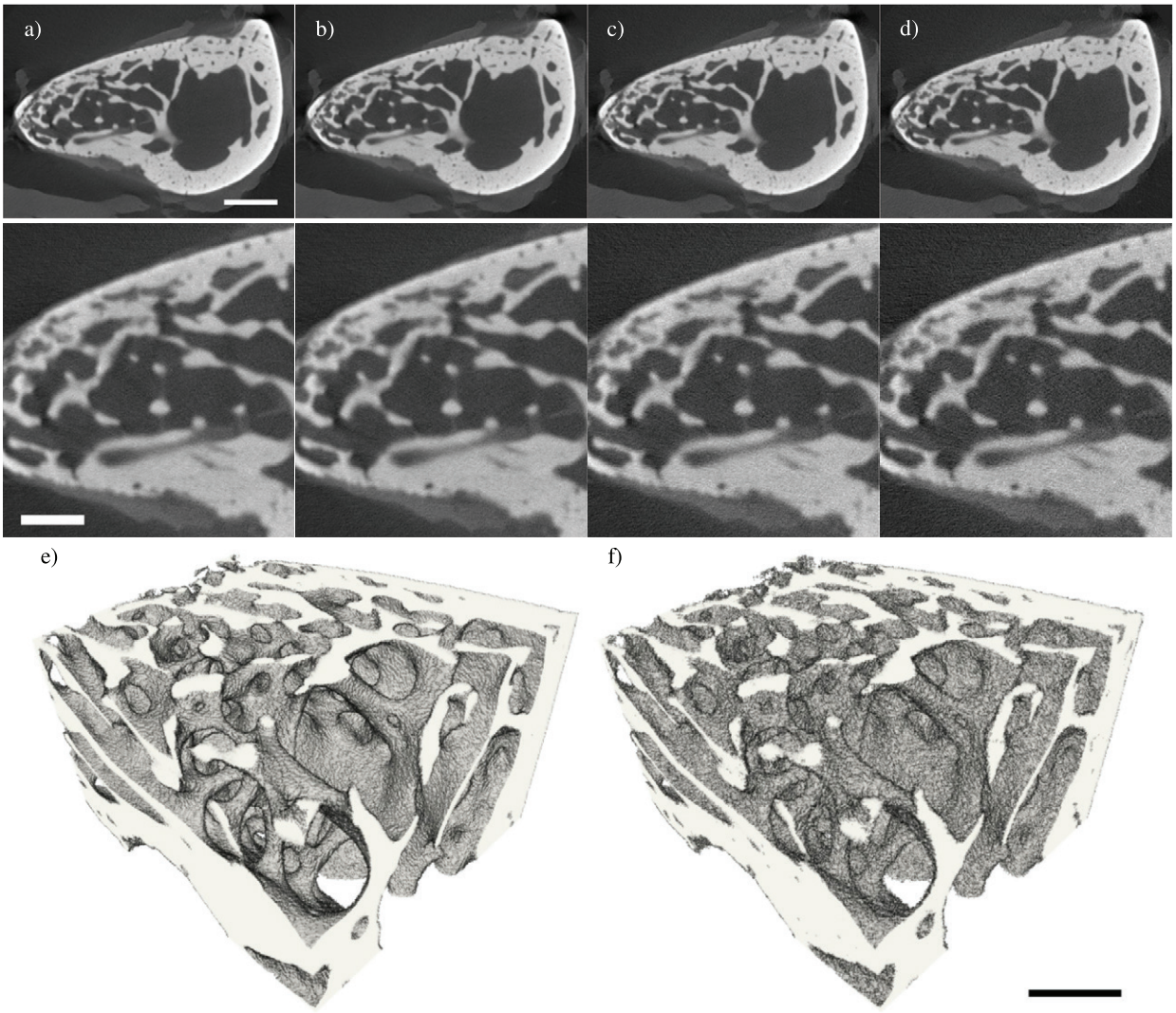

Figure 14. Reconstructed slices through the mouse femur scanned with an exposure time of (a) 1.25 s, (b) 1 s, (c) 0.45 s, and (d) 0.26 s. The scale bar is 0.5 mm. Corresponding enlarged regions are shown in the second row, where the scale bar is 0.2 mm. The bottom row shows 3D renderings of a selected portion of the segmentation for the scans at (e, left) 1.25 s and (f, right) 0.26 s. The scale bar is 0.2 mm.

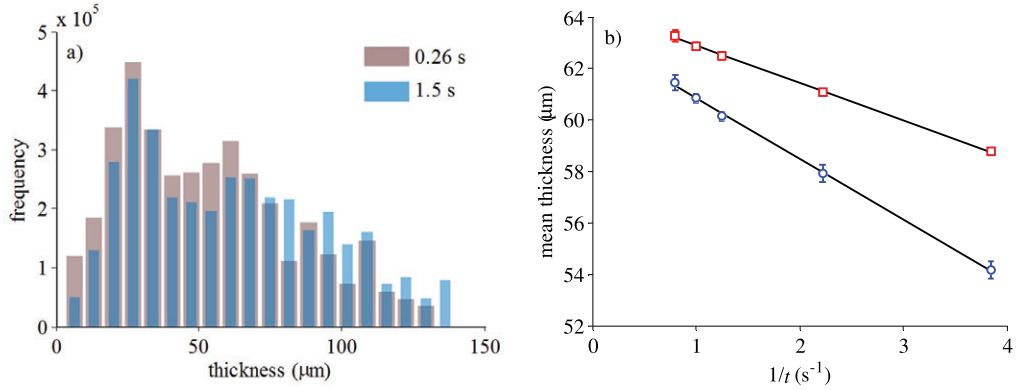

Figure 15. (a) Histograms of bone thickness determined by the local thickness method for scans taken with an exposure time of 1.25 and 0.26 s. No hole-filling was applied. (b) Variation of mean bone thickness as a function of $1/t$, with (red squares) and without (blue circles) hole-filling. The values are averages over 7 scans taken at each exposure time and the error bars show the standard deviations. The black lines show the linear fit to the data.

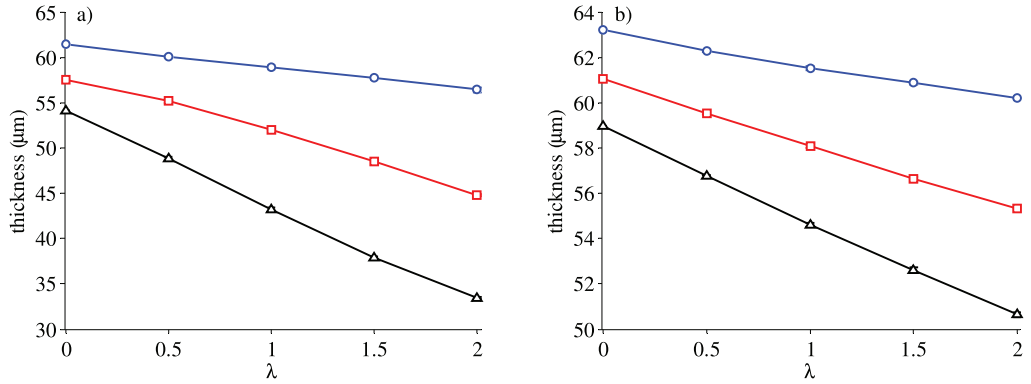

Figure 16. Variation of the mean local thickness with SIMEX λ for scans taken with exposure times of 1 s (blue circles), 0.45 s (red squares) and 0.26 s (black triangles). The thickess is shown when holes are (a) unfilled and (b) filled.

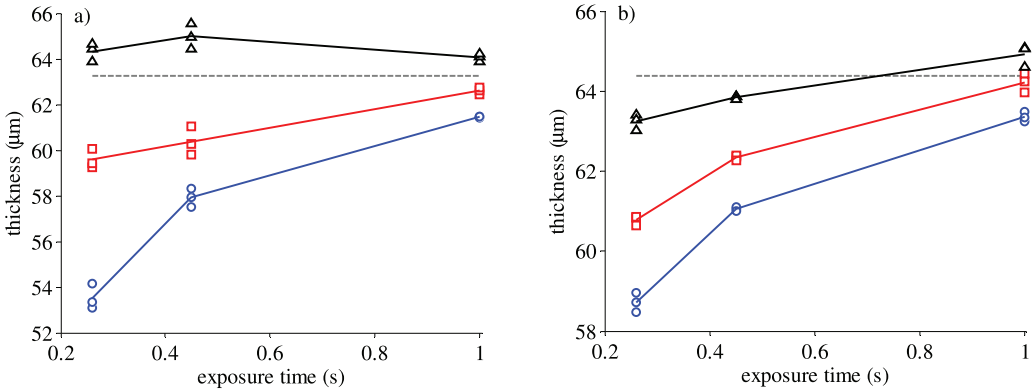

Figure 17. SIMEX estimates of noise-free mean thickness as a function of exposure time when holes are (a) unfilled and (b) filled. The estimates for shown for linear extrapolation (black triangles) and quadratic extrapolation (red squares), and the naïve estimates are shown by the blue circles. Linear extrapolation was applied using a restricted λ range of $\{0, 0.5\ 1\}$. The grey dashed line shows the noise-free estimate from linear extrapolation of the measured values shown in figure 15(b).

4.4. Mean local thickness of a mouse femur

4.4.1. Methods. The femur was scanned using a Zeiss Xradia versa-520 system with settings given in table 1. A range of noise levels were considered by varying the exposure times between 1.25 s and 0.26 s per radiograph, with corresponding contrast-to-noise ratios between bone and air varying from ~13.5 to ~6.3. Reconstructed slices are presented in figure 14. 201 slices were selected for analysis of bone thickness, and care was taken to ensure that the same region of the sample was analysed for all repeats scans (as determined visually to pixel accuracy). The volumetric thresholding strategy was applied by using the upper 10 slices as the volumetric standard, so that the total bone volume (of the largest connected region) in these slices was kept constant across data sets. Small isolated regions in the segmentation (islands) were removed if they were less than 1000 voxels in volume and the effects of hole-filling were considered by removing holes less than 50 voxels in volume. The segmentation was down-sampled by a factor of 2 in each dimension for calculation of the mean trabeculae thickness via the local thickness technique [53], implemented in the ImageJ plugin [54]. SIMEX was applied to 3 repeat scans taken at each of exposure time of 1 s, 0.45 s and 0.26 s with $B = 14$ simulation repeats and $\lambda = \{0\ 0.5\ 1\ 1.5\ 2\}$. Linear extrapolation was applied using $\lambda = \{0\ 0.5\ 1\}$.

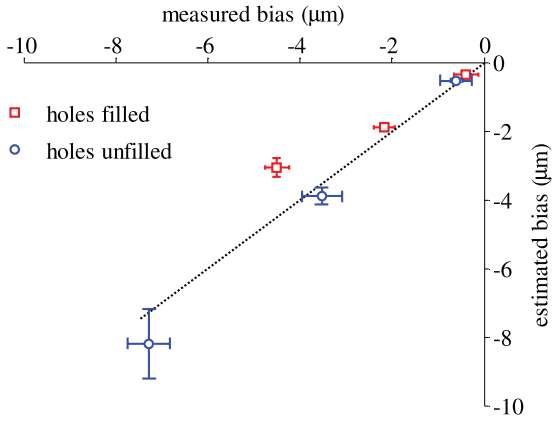

Figure 18. Comparison between the measured and estimated bias relative to the mean thickess measured for the 1.25 s exposure time when holes in the segmentation are unfilled and filled. The data points correspond to exposure times of 1 s, 0.45 s and 0.26 s, which show increasing bias. The measured data are average values over 7 scans. The estimated values are averages over SIMEX applied to 3 scans at each exposure time with linear extrapolation. The error bars show the standard deviations. Equality between measured and estimated bias is shown by the dashed line.

4.4.2. Results. Figure 14 (bottom row) clearly shows the effects of noise on the segmentation, leading to high spatial frequency surface texture and the appearance of holes within

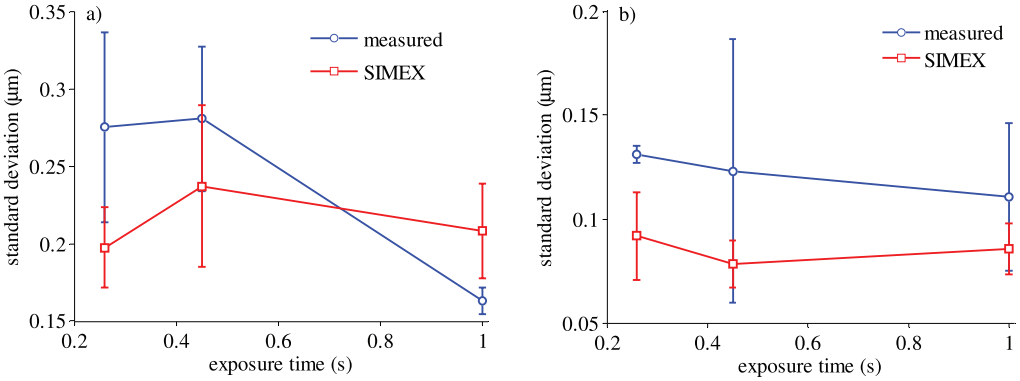

Figure 19. Comparison between the measured and estimated standard deviation of the naïve mean local thickness when holes were (a) unfilled and (b) filled. The measured values were determined from 27 to 28 repeats taken in 3 groups, and the error bars show the standard deviations of the average values of the 3 groups.

the structure. The noise leads to a general reduction in thickness measured throughout the bone ROI as indicated by the histograms in figure 15(a). The naïve mean thickness is shown in figure 15(b) and decreases essentially linearly with $1/t$ and hence noise level. This indicates that even at 1.25 s exposure time negative bias is evident. Extrapolating to $t = \infty$, gives an estimate of the bias-free mean thickness of 63.24 ± 0.09 μm without hole-filling and 64.38 ± 0.05 μm when hole-filling is applied. These values may be overestimates as it can be anticipated that the gradient of the variation should tend to zero as $t \rightarrow \infty$ (as found in section 3). However they do suggest that hole-filling introduces a small positive bias (on the order of 2%) by removing fine features; although hole-filling also introduces greater robustness to noise as demonstrated by the gradient in figure 15(b) being lower. A similar approximately linear relationship is found for SIMEX between mean thickness and λ, as shown in figure 16. Again, the simulation step confirms that hole-filling introduces some robustness to added noise.

The SIMEX estimates of noise-free mean thickness for linear and quadratic extrapolation are show in figure 17. Both extrapolation methods lead to more consistent results with decreasing exposure time, in comparison to the naïve measurements, indicating a clear reduction in relative bias. The values derived from linear extrapolation are also in good agreement with the corresponding values obtained from the direct linear extrapolation of the naïve measurements presented in figure 15(b). However, since the naïve (measured) value for the longest exposure time of 1.25 s is likely to contain bias, the absolute accuracy of the SIMEX extrapolation cannot be verified. Alternatively, SIMEX can be used to estimate the mean thickness for 1.25 s exposure time by extrapolating to $\lambda^* = t/1.25 - 1$ for $t \leqslant 1.25$ s (see equation (8) and taking ψ^{det} to be negligible). The estimated and measured bias relative to the naïve value at 1.25 s are shown in figure 18. Overall, SIMEX leads to a large reduction in bias, over 85% in almost all cases.

The standard deviations of the naïve measurements were determined experimentally from 27 to 28 scans at exposure times of 1.5, 0.45 and 0.26 s. The predicted values from SIMEX (with $\lambda = 1$) were calculated for 3 scans at each exposure time, and a comparison between the predicted and experimentally determined values is given in figure 19. Overall, the SIMEX technique produces reasonable estimates (to within a factor of 1.6) for all exposure times considered, although the agreement is better for the lowest noise data as would be expected. Both the predicted and measured values show that hole-filling leads to reduced variation as well as reduced bias. However, the predicted values are typically underestimates, which may indicate that there is an additional source of variation in the measured data. For example, movement of the sample by 1 pixel was found in section 4.2 to cause a change in the mean thickness of between 0.15–0.45 μm. While no sample drift was evident, random sub-pixel movement may increase the measured standard deviation noticeably.

The standard deviation of the SIMEX estimate was found to be reasonably well predicted by equation (5) for $\lambda_{\text{min}} = 0.5$. With quadratic extrapolation and no hole-filling for instance, the measured standard deviations (over the repeat 3 scans) were in the range 0.15–0.63 μm for exposure times between 1.5 s and 0.26 s. The corresponding predicted values were in the range 0.21–0.32 μm. Similar values were also obtained with $\lambda_{\text{min}} = 1$, in the range 0.20–0.24 μm. Again, the larger measured values may be attributed to sub-pixel movement of the sample between scans. The measured standard deviations were greater when linear extrapolation was used; by on average a factor of ~1.75. This may be due to greater variation when fitting the linear extrapolant using fewer λ values, which could be counteracted by increasing the number of simulation repeats.

5. Discussion and conclusions

SIMEX has been successfully adapted to tomographic measurement processes for estimation of bias and variance arising from noise in the acquisition. SIMEX is flexible in that any particular processing steps can be accommodated, including for example different reconstruction and automatic segmentation strategies. The technique is reliant on the distribution of the noise being known and the variation of the measurement with increasing noise level being monotonic and approximated well by simple functions (e.g. low order polynomials). The latter requirement was considered in detail for application to x-ray tomography scans and was found to be the case for volume and thickness measurements over a wide range of contrasts and noise levels in a simulation study, and also in

practice for typical scan settings. However for some combinations of contrast and noise level, the simulation study showed that the variation can be more complex. All measurements showed some robustness to noise, as denoted by the 'low bias' regime in which the variance increases much more rapidly than the bias. The SIMEX based estimates of both bias and variance were most accurate in or close to this regime. The key findings from the case studies can be summarised as follows:

- Bias estimation: Considerable reduction is bias is achievable, typically over 80% for the case studies.
- Estimation of the naïve measurement variance: equation (6) can provide an order of magnitude or better estimate of the standard deviation of the original naïve measurement. The estimates were typically within a factor of 2 of the measured standard deviations.
- Estimation of the SIMEX estimate variance: equation (5) can provide an order of magnitude or better estimate of the standard deviation of the original naïve measurement. The estimates were typically within a factor of 2 of the measured standard deviations.

Together these estimates facilitate the statistical comparison of measurements made on different specimens or on the same sample over time. Other key applications include:

- Confirmation of low bias. The plot of $M'(\lambda)$ versus λ can reveal when the bias in the naïve estimate is low or well within the measurement variance.
- Comparison of the robustness of reconstruction and segmentation strategies. The plot of $M'(\lambda)$ versus λ can be used to compare the robustness introduced by differing processing steps. For instance, hole-filling was clearly found to increase robustness in the femur case study. Furthermore the robustness to systematic variations can be investigated by comparing the variance of the naïve measurement over repeat scans to the SIMEX estimate.
- Faster and lower dose scanning. The approach can enable scan times to be minimised but still achieve a given accuracy or precision. The accuracy of the extrapolation can be verified taking a scan with low noise. A key application would be to time-lapse XCT studies, where a low noise scan is typically taken first before monitoring changes within a sample (e.g. crack growth or compression/tension experiments) using shorter scans.

The wider applicability of the approach to different measurements and/or tomographic imaging modalities would need to be established on a case by case basis, by for example simulation studies; however it seems reasonable to assume that a simple extrapolent would generally be suitable at least over a limited range of noise levels. For higher noise, low dose scanning more appropriate extrapolents could be determined for specific applications from simulation studies or from measurements made on a suitable 'calibration' sample. The main disadvantage of SIMEX is that is highly computationally expensive. However, both the simulation and the case studies showed that a relatively low number of repeats (e.g. $B \leqslant 17$) were required to achieve good estimates, and for some measurements (e.g. thickness) only 2 values of λ were required. Furthermore tomographic reconstruction times are reducing with the use of computers with multiple GPUs. For example, reconstruction times in this study were ~10 min for a 2000^3 voxel volume. The process could also be speeded up by applying SIMEX to a limited (but representative) ROI within the scan volume.

Acknowledgments

The author would like to acknowledge fruitful discussions with Prof Phil Withers and Dr Tristan Lowe of the Henry Moseley x-ray imaging facility, as well as funding for Facility under EPSRC grants EP/M010619, EP/K004530, EP/F007906, EP/F028431.

Appendix

Whiting *et al* [42] showed that when the x-ray detection is modelled as a Poisson or compound Poisson process, the variance of γ_k is proportional to its mean value. With the proportionality constant being set to α, the signal after background subtraction $S_k = Q_k - \widetilde{C}_k^{\text{dark}}$ has variance:

$$\text{var}(S_k) = \alpha\gamma_k + \sigma_k^{\text{dark}\,2}/n^{\text{dark}} + \sigma_k^{\text{det}\,2}$$

The variance of the transmission can then be calculated using the delta method as:

$$\begin{aligned}\frac{\sigma_k^2}{T_k^2} &= \frac{\text{var}(S_k)}{\gamma_k^2} + \frac{\text{var}(S_k^{\text{ref}})}{n^{\text{ref}}\gamma_k^{\text{ref}\,2}} = \frac{\alpha T_k\gamma_k^{\text{ref}} + \sigma_k^{\text{dark}\,2}/n^{\text{dark}} + \sigma_k^{\text{det}\,2}}{T_k{}^2\gamma_k^{\text{ref}\,2}} \\ &\quad + \frac{\alpha\gamma_k^{\text{ref}} + \sigma_k^{\text{dark}\,2}/n^{\text{dark}} + \sigma_k^{\text{det}\,2}}{n^{\text{ref}}\gamma_k^{\text{ref}\,2}} \\ &= \frac{\alpha}{\gamma_k^{\text{ref}}}\left[\frac{1}{T_k} + \frac{1}{n^{\text{ref}}} + \frac{\psi_k^{\text{det}}}{\gamma_k^{\text{ref}}}\left(\frac{1}{T_k^2} + \frac{1}{n^{\text{ref}}}\right)\right]\end{aligned}$$

where $\psi_k^{\text{det}} = \alpha^{-1}(\sigma_k^{\text{dark}\,2}/n^{\text{dark}} + \sigma_k^{\text{det}\,2})$. Rearrangement then gives equation (8). For ease of implementation, Poisson and compound Poisson PDFs can be approximated well by the normal distribution for means as low as ~25 [42, 55]. However, at very low counts, the delta method will also become inaccurate. In which case the PDF of T, and hence σ^2, can be derived from the full numerical implementation of the compound Poisson PDF in the polychromatic case [42]. Such an approach could also be used to account for non-linearities in the detector response.

References

[1] Arvold N D, Niemierko A, Mamon H J, Fernandez-del Castillo C and Hong T S 2011 Pancreatic cancer tumor size on CT scan versus pathologic specimen: implications for radiation treatment planning *Int. J. Radiat. Oncol. Biol. Phys.* **80** 1383–90

[2] Zhao B *et al* 2009 Evaluating variability in tumor measurements from same-day repeat CT scans of patients with non-small cell lung cancer *Radiology* **252** 263–72

[3] Beckmann F *et al* 2007 *In situ* synchrotron x-ray microtomography studies of microstructure and damage evolution in engineering materials *Adv. Eng. Mater.* **9** 939–50

[4] Withers P J 2015 Fracture mechanics by three-dimensional crack-tip synchrotron x-ray microscopy *Phil. Trans. R. Soc.* A **373** 20130157

[5] Eastwood D S *et al* 2014 The application of phase contrast x-ray techniques for imaging Li-ion battery electrodes *Nucl. Instrum. Methods* B **324** 118–23

[6] Maire E and Withers P J 2014 Quantitative x-ray tomography *Int. Mater. Rev.* **59** 1–43

[7] Kruth J P, Bartscher M, Carmignato S, Schmitt R, De Chiffre L and Weckenmann A 2011 Computed tomography for dimensional metrology *CIRP Ann. Manuf. Technol.* **60** 821–42

[8] Carmignato S 2012 Accuracy of industrial computed tomography measurements: experimental results from an international comparison *CIRP Ann. Manuf. Technol.* **61** 491–4

[9] Ferrucci M, Leach R K, Giusca C, Carmignato S and Dewulf W 2015 Towards geometrical calibration of x-ray computed tomography systems—a review *Meas. Sci. Technol.* **26** 092003

[10] Hiller J and Reindl L M 2012 A computer simulation platform for the estimation of measurement uncertainties in dimensional x-ray computed tomography *Measurement* **45** 2166–82

[11] Hiller J, Maisl M and Reindl L M 2012 Physical characterization and performance evaluation of an x-ray micro-computed tomography system for dimensional metrology applications *Meas. Sci. Technol.* **23** 085404

[12] Hiller J, Genta G, Barbato G, De Chiffre L and Levi R 2014 Measurement uncertainty evaluation in dimensional x-ray computed tomography using the bootstrap method *Int. J. Precis. Eng. Manuf.* **15** 617–22

[13] Müller P, Hiller J, Dai Y, Andreasen J L, Hansen H N and De Chiffre L 2014 Estimation of measurement uncertainties in x-ray computed tomography metrology using the substitution method *CIRP J. Manuf. Sci. Technol.* **7** 222–32

[14] Schmitt R and Niggemann C 2010 Uncertainty in measurement for x-ray-computed tomography using calibrated work pieces *Meas. Sci. Technol.* **21** 054008

[15] Walton L A *et al* 2015 Morphological characterisation of unstained and intact tissue micro-architecture by x-ray computed micro- and nano-tomography *Sci. Rep.* **5** 10074

[16] Withers P J 2007 X-ray nanotomography *Mater. Today* **10** 26–34

[17] Shearing P R *et al* 2011 Using Synchrotron x-ray nano-CT to characterize SOFC electrode microstructures in three-dimensions at operating temperature *Electrochem. Solid State Lett.* **14** B117–20

[18] Kardjilov N, Manke I, Hilger A, Strobl M and Banhart J 2011 Neutron imaging in materials science *Mater. Today* **14** 248–56

[19] Weber E, Fernandez M, Wapner P and Hoffman W 2010 Comparison of x-ray micro-tomography measurements of densities and porosity principally to values measured by mercury porosimetry for carbon–carbon composites *Carbon* **48** 2151–8

[20] Guolo A 2008 Robust techniques for measurement error correction: a review *Stat. Methods Med. Res.* **17** 555–80

[21] Carroll R J 2006 *Measurement Error in Nonlinear Models: a Modern Perspective* 2nd edn ed R J Carroll *et al* (Boca Raton, FL: Chapman & Hall/CRC)

[22] Cook J R and Stefanski L A 1994 Simulation-extrapolation estimation in parametric measurement error models *J. Am. Stat. Assoc.* **89** 1314–28

[23] Carroll R J, Kuchenhoff H, Lombard F and Stefanski L A 1996 Asymptotics for the SIMEX estimator in nonlinear measurement error models *J. Am. Stat. Assoc.* **91** 242–50

[24] Stefanski L A and Cook J R 1995 Simulation-extrapolation: the measurement error jackknife *J. Am. Stat. Assoc.* **90** 1247–56

[25] Carroll R J, Maca J D and Ruppert D 1999 Nonparametric regression in the presence of measurement error *Biometrika* **86** 541–54

[26] Liang H and Wu H L 2008 Parameter estimation for differential equation models using a framework of measurement error in regression models *J. Am. Stat. Assoc.* **103** 1570–83

[27] Kim J and Gleser L J 2000 SIMEX approaches to measurement error in ROC studies *Commun. Stat. - Theory Methods* **29** 2473–91

[28] Kuchenhoff' H, Mwalili S M and Lesaffre E 2006 A general method for dealing with misclassification in regression: the misclassification SIMEX *Biometrics* **62** 85–96

[29] Stefanski L A and Bay J M 1996 Simulation extrapolation deconvolution of finite population cumulative distribution function estimators *Biometrika* **83** 407–17

[30] Wang X F and Wang B 2011 Deconvolution estimation in measurement error models: the R package decon *J. Stat. Softw.* **39** 1–24

[31] Lauzon C B, Asman A J, Crainiceanu C, Caffo B C and Landman B A 2011 Assessment of bias for MRI diffusion tensor imaging using SIMEX *Med. Image Comput. Comput. Assist. Interv.* **14** 107–15

[32] Lauzon C B, Crainiceanu C, Caffo B C and Landman B A 2013 Assessment of bias in experimentally measured diffusion tensor imaging parameters using SIMEX *Magn. Reson. Med.* **69** 891–902

[33] Pineda A R, Tward D J, Gonzalez A and Siewerdsen J H 2012 Beyond noise power in 3D computed tomography: the local NPS and off-diagonal elements of the Fourier domain covariance matrix *Med. Phys.* **39** 3240–52

[34] Baek J and Pelc N J 2010 The noise power spectrum in CT with direct fan beam reconstruction *Med. Phys.* **37** 2074–81

[35] Baek J and Pelc N J 2011 Local and global 3D noise power spectrum in cone-beam CT system with FDK reconstruction *Med. Phys.* **38** 2122–31

[36] Hiriyannaiah H P, Snyder W E and Bilbro G L 1990 Noise in reconstructed images in tomography parallel, fan and cone beam projections *3rd Annual IEEE Symp. on Computer-Based Medical Systems* pp 81–8

[37] Maco B, Cantoni M, Holtmaat A, Kreshuk A, Hamprecht F A and Knott G W 2014 Semiautomated correlative 3D electron microscopy of *in vivo*-imaged axons and dendrites *Nat. Protocols* **9** 1354–66

[38] Rajbhandary P L and Pelc N J 2015 Statistical bias in material decomposition in low photon statistics region *Proc. SPIE* **9412** 94124W

[39] Haynor D R and Woods S D 1989 Resampling estimates of precision in emission tomography *IEEE Trans. Med. Imaging* **8** 337–43

[40] Lartizien C, Aubin J B and Buvat I 2010 Comparison of bootstrap resampling methods for 3D PET imaging *IEEE Trans. Med. Imaging.* **29** 1442–54

[41] Siewerdsen J H, Antonuk L E, El-Mohri Y, Yorkston J, Huang W and Cunningham I A 1998 Signal, noise power spectrum, and detective quantum efficiency of indirect-detection flat-panel imagers for diagnostic radiology *Med. Phys.* **25** 614–28

[42] Whiting B R, Massoumzadeh P, Earl O A, O'Sullivan J A, Snyder D L and Williamson J F 2006 Properties of preprocessed sinogram data in x-ray computed tomography *Med. Phys.* **33** 3290–303

[43] Huang S Y, Yang K, Abbey C K and Boone J M 2012 A semiempirical linear model of indirect, flat-panel x-ray detectors *Med. Phys.* **39** 2108–18
[44] Zhang H, Ouyang L, Ma J H, Huang J, Chen W F and Wang J 2014 Noise correlation in CBCT projection data and its application for noise reduction in low-dose CBCT *Med. Phys.* **41** 031906
[45] Joarder A and Omar M H 2015 The mathematical expectation of sample variance: a general approach *Bull. Malays. Math. Sci. Soc.* **38** 635–46
[46] Limodin N, Rethore J, Adrien J, Buffiere J Y, Hild F and Roux S 2011 Analysis and artifact correction for volume correlation measurements using tomographic images from a laboratory x-ray source *Exp. Mech.* **51** 959–70
[47] Fukuda A, Matsubara K and Miyati T 2015 Long-term stability of beam quality and output of conventional x-ray units *Radiol. Phys. Technol.* **8** 26–9
[48] Goertzen A L, Nagarkar V, Street R A, Paulus M J, Boone J M and Cherry S R 2004 A comparison of x-ray detectors for mouse CT imaging *Phys. Med. Biol.* **49** 5251–65
[49] Samei E 2003 Performance of digital radiographic detectors: quantification and assessment methods *Advances in digital radiography-categortical course in diagnostic radiology physics* ed E Samei and M J Flynn (Oak Brook, IL: RSNA)
[50] Siewerdsen J H and Jaffray D A 1999 A ghost story: spatio-temporal response characteristics of an indirect-detection flat-panel imager *Med. Phys.* **26** 1624–41
[51] Han J C *et al* 2010 An aging study of the signal and noise characteristics in large-area CMOS detectors *Proc. SPIE* **7622** 76223Y
[52] Lagarias J C, Reeds J A, Wright M H and Wright P E 1998 Convergence properties of the Nelder–Mead simplex method in low dimensions *SIAM J. Optim.* **9** 112–47
[53] Hildebrand T and Ruegsegger P 1997 A new method for the model-independent assessment of thickness in three-dimensional images *J. Micros.* **185** 67–75
[54] Doube M *et al* 2010 BoneJ Free and extensible bone image analysis in ImageJ *Bone* **47** 1076–9
[55] Daly F 1995 *Elements of Statistics* (Boston, MA: Addison-Wesley)