Tina Memo No. 2016-011
Internal.

Assessing MR Image Data Quality for Automated Volumetric Analysis

S. V Notley and N. A. Thacker

Last updated
13 / 5 / 2016

Imaging Science and Biomedical Engineering Division,
Medical School, University of Manchester,
Stopford Building, Oxford Road,
Manchester, M13 9PT.

Abstract

Quantitative analysis of the anatomical structures seen in MR images requires the assumptions inherent to the analysis method to be matched to data quality. If this is not done then inaccuracies and biases are likely to be incurred, restricting the practical value of summary variables. In our previous work we have explained how accurate segmentation of MR images will require models of data density which include partial volume contributions. In accordance with this the data must be seen to be consistent with these distributions. There are many factors affecting image formation which may lead to problems.

We are interested in developing a quantitative understanding in pathology (particularly tumour heterogeneity), but first we must be able to model normal tissue. Checks may be necessary on individual datasets in order to confirm that data meets our requirements (Quality Assurance). Constructing a fully automated approach to do this is seen as a challenging long term task. As a first step we present here a non-quantitative method which allows us to visually compare the complex density structures seen in data, as a check of self-consistency. We apply this to datasets obtained with a range of MR acquisitions, including the use of parallel imaging (SENSE) and much older bird cage head coils. We show how characteristic problems in data quality can be identified. We conclude that good quality image acquisition, suiatble for quantitative analysis, poses a significant challenge to modern scanner designs.

Assessing MR Image Data Quality for Automated Volumetric Analysis

The conventional model of MR image formation describes the majority of contrast seen in image data as due to only three effective parameters, conventionally T1, T2 and proton density. However, modern scanner sequences are often modified by secondary physics processes, which generate a greater range of possible variations in data. Non-the less, biological structures can often be seen to have a large degree of consistency within and between images. This greater variation can be captured in the form of multi-spectral MR signatures (the multi-dimensional distribution of signals at a single location from multiple images). We wish to test the hypothesis that consistency of these signatures has value for understanding the microstructure environment of tissues within a machine learning framework.

The authors of this documents are concerned with making quantitative measurements from MR images of brain tissue. Specifically we are interested in segmentation of brain images with the goal of detecting tumour tissue and, ultimately the quantification of tumour heterogeneity in terms of multi-spectral MR signatures. One issue with MR images is that the resultant grey levels of any pixels are not defined on any absolute and repeatable scale across images. This makes the assessment of tumour tissue heterogeneity and/or correlation of the MR signatures with longitudinal disease development and histology problematic. With regards to this, the fitting of normative tissue distribution parameters may provide a means of relative characterisation.

The first stage of the analysis involves identifying the voxels most likely to be composed of tumour tissue. Since the tumour tissue is composed of an unknown number of multi-spectral intensity distributions we cannot simply fit a fixed mixture model. The presence of unknown tissue classes also poses a problem for the unbiased fitting of any normative tissue models. One approach being developed to fit normative tissue models in the presence of tumour tissue is based on making use of spatial information in the form of tissue masks derived from the data (See **??**). This method allows the fitting of a multi-spectral tissue distribution model with higher parameter accuracies than from a single model fit and reduces bias from unknown tissues. With a normative model fit to the data the tumour tissue as a whole may be identified as outliers to the model. Further analysis of the data may then be restricted to the areas already identified as being tumour data. In this approach, accuracy of the normative model parameters is of primary concern since any errors in the model distribution will propagate into any heterogeneity analysis of the tumour tissue.

The starting point for any quantitative analysis of this form is the assumption that the measured grey levels are unambiguously and meaningfully related to the underlying tissue. Under this assumption it is possible to use the measured data to estimate the parameters of a grey-level probability distribution model. The estimated model allows statistically based and meaningful interpretations of the data to be made. Clearly this requires an appropriate model of the relationship between the underlying tissue and the data distributions. If the data distributions deviate from the assumed model then any interpretations of the data based on this model are suspect.

With the above comments in mind, this document describes a methodology using 2D scatterplots to assess data quality in relation to an assumed model. By combining information from two sources, scatterplots are able to highlight correlations and co-variances between the two images that are not apparent simply from observing 1D histograms and images. The scatterplots also enable effects such as field inhomogeneity (within and across slices)

and partial volumes to be seen more easily than from looking at single images or 1D histograms.

The images are assessed on how well they confirm to the expected distributions given our understanding of the image formation process. That is: (a) Gaussian distributed pure tissues accounting for measurement noise and natural tissue density variation, and (b) linear partial volume distributions that link between the means of the pure tissues. The plots are also assessed in terms of the variablity of the distributions both within a single image and across slices from the same volume.

Early investigations illustrated quite obvious problems with parallel imaging and associated image reconstruction, which invalidated the distribution model assumptions. It was therefore found necessary to compare data quality with legacy datasets which were known to have the required behaviour. The results presented include only our best findings, (i.e. excluding multi-coil acquisition, SENSE and CLEAR). This document compares data distributions found from a modern 1.5T MR scanner using two different head coil configurations to the distributions seen in 18 year old 1.5T MR data. Previous work has also demonstrated that, with the noise levels and data distributions found in the legacy data, an improved parameter fitting algorithm is capable of giving volume estimates with around 1% statistical error. We believe that this level of accuracy will now be needed for the current study of heterogeneity, so our target is to at least match the quality of the old dataset.

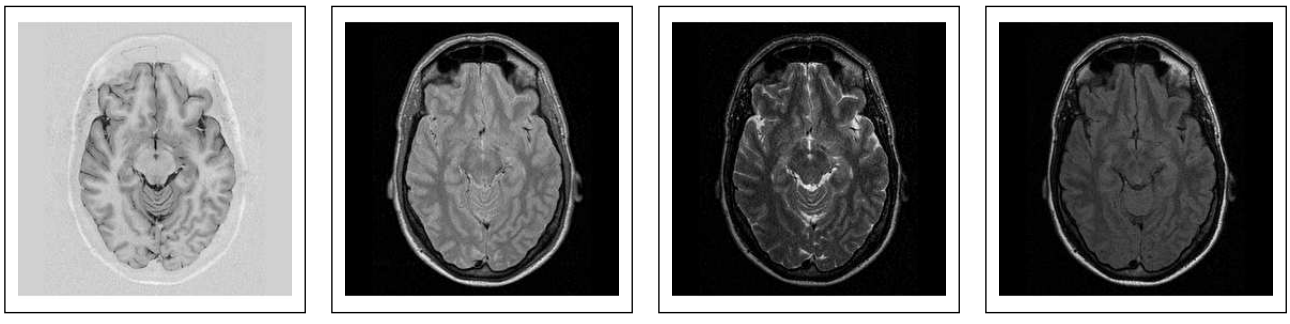
Data Sets

New data was collected using the SENSE coil in quadrature (NEW-Q) and with a birdcage coil (NEW-BC) for comparison with an original data set collected using a birdcage coil (ORIG-BC) . The original data set was collected with 1 x 1 x 5mm voxels giving each image slice a resolution of 256 x 256 voxels. The new data sets were collected with 1.875 x 1.875 x 2.5 mm voxels giving each image slice a resolution of 128 x 128 voxels. The new data was collected with regards to a new study aimed at investigating tumour heterogeneity. The change in the voxel dimensions was determined on the basis of reducing the amount of through plain partial volume voxels while attempting to maintain/minimise the statistical image noise.

The four MR sequences collected in each coil configuration were the same as those used to collect data from the original 1.5T scanner and were: Inversion Recovery Turbo Spin Echo (IRTSE), Variable Echo - Proton Density (VE-PD), Variable Echo - T2 (VE-T2) and Fluid Attenuated Inversion Recovery (FLAIR). The protocol parameters used for each sequence in the data set is given in Table 1 below. Figures 1, 2 and 3 show typical co-registered images, across sequences, for the ORIG-BC, NEW-BC, and NEW-Q datasets respectively.

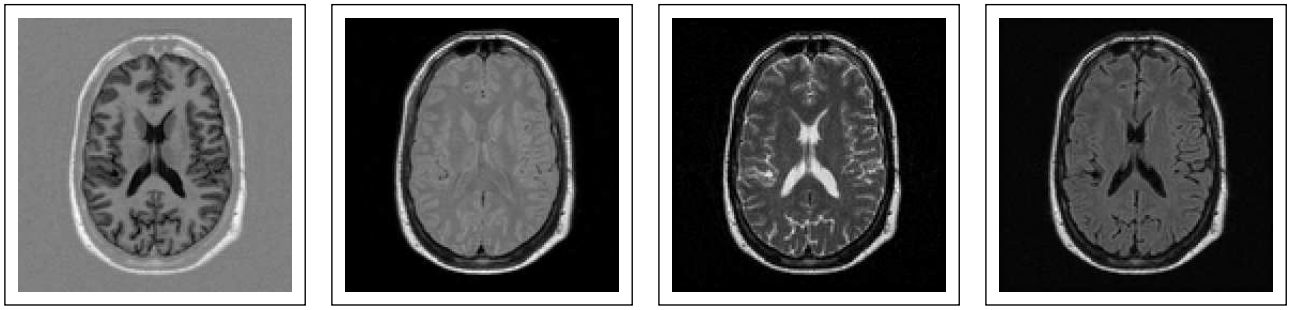
Sequence	TR(ms)	TE(ms)	TI(ms)	TSE factor	echo space (ms)	CSF	Grey	White	error
IRTSE	6850	18	300	9	18	-1800	-650	-200	60
VE (PD)	5500	20	-	8	13.3	1475	1270	970	80
VE (T2)	5500	100	-	8	13.3	1450	435	320	80
Flair	6000	100	2200	19	10	250	750	550	30

Table 1. MR sequence parameters.



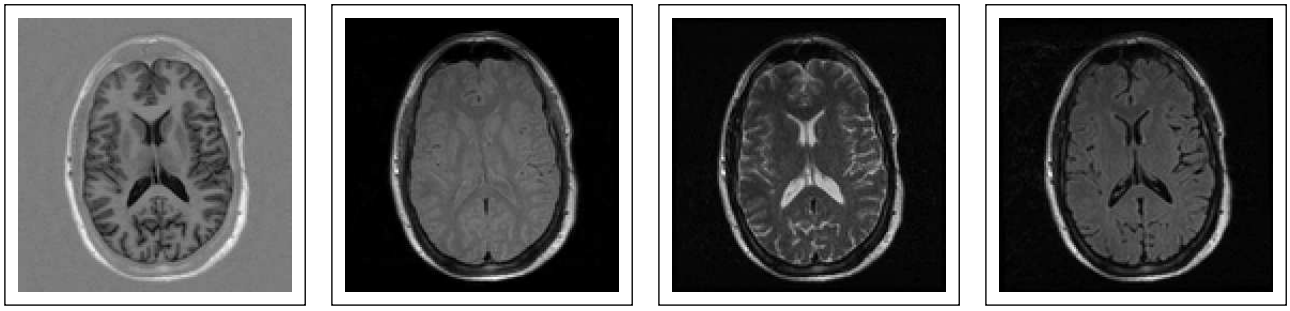
(a) (b) (c) (d)

Figure 1: Original (legacy) data collected with Birdcage coil. (a) Inversion Recovery Turbo Spin Echo (IRTSE), (b) Variable Echo - Proton Density (VE-PD), (c) Variable Echo - T2 (VE-T2), and (d) Fluid-Attenuated Inversion Recovery (FLAIR)



(a) (b) (c) (d)

Figure 2: New data collected with Birdcage coil. (a) Inversion Recovery Turbo Spin Echo (IRTSE), (b) Variable Echo - Proton Density (VE-PD), (c) Variable Echo - T2 (VE-T2), and (d) Fluid-Attenuated Inversion Recovery (FLAIR)



(a) (b) (c) (d)

Figure 3: New data collected with SENSE coil in quadrature. (a) Inversion Recovery Turbo Spin Echo (IRTSE), (b) Variable Echo - Proton Density (VE-PD), (c) Variable Echo - T2 (VE-T2), and (d) Fluid-Attenuated Inversion Recovery (FLAIR)

Methodology

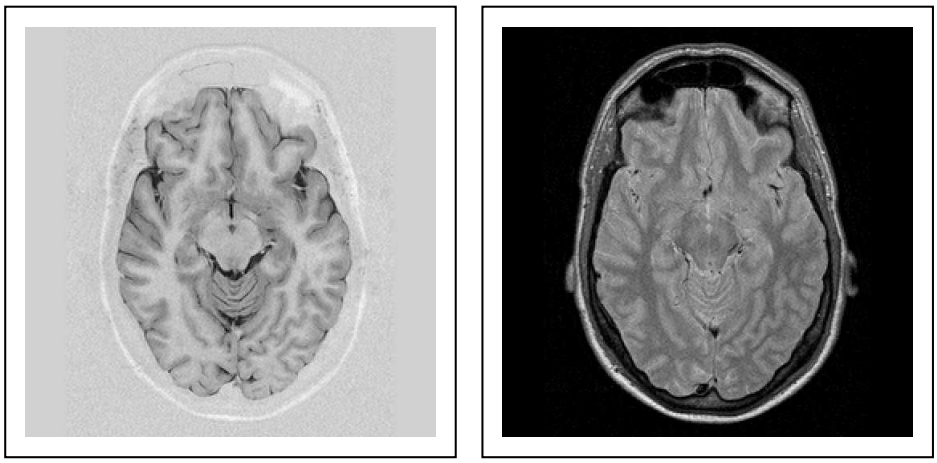
The intensity data for each pixel location in the two images is displayed as a collection of points on a Cartesian co-ordinate system i.e. a scatterplot. Each point is located at co-ordinates using the pixel intensities from the first image on the horizontal axis and the corresponding pixel intensities in the second image on the vertical axis. The image intensity of each image is scaled such that the intensity range is between 0 and 255. As an example, figure 4 shows two images (ORIG-BC dataset) of the same brain slice using (a) IRTSE and (b) VE (PD) imaging sequences. The associated histograms (c) and scatterplot for these two images (d) are also shown. The scatterplot has been labelled with the approximate location of the pure tissue means (GM - grey matter, WM - white matter).

From the histograms it is difficult to identify the distribution means due to the overlapping pure distributions and the partial volume distributions. In the two dimensional scatterplots the distributions are generally more clearly separated. It is also clear that the overall distribution is not suitably modelled as a pure Gaussian mixture model. The plots clearly show partial volume distributions as broad linear structures between the pure tissue distributions.

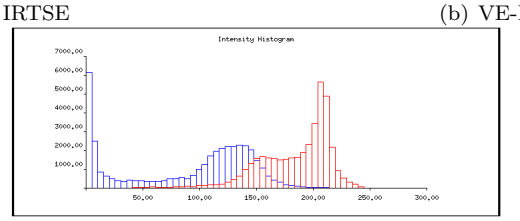
Co-Variance and Noise Estimates

For a noiseless image with no natural pure tissue variation, all the points in the scatterplot related to single pure tissue will be at the same co-ordinate. Uncorrelated Gaussian additive noise will move the points away from a single point and will form an ellipsoid. In the case of no natural tissue variation the principal axis of the ellipsoid will be aligned with the scatterplot axis. If there is any correlation in the noise across the images then the principal axis of the ellipsoid will be rotated with respect to the co-ordinate system. If there is variation in the pure tissue density then we would expect some degree of correlation between the images and the ellipsoid due to noise will be convolved with the line of correlation.

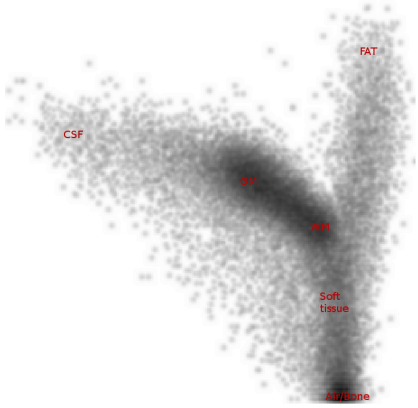
As an example, the airspace/bone distribution in figure 4d is semi-circular, showing that in the scaled intensity spaces both images have similar noise levels. Since there is no 'stretching' of the the distribution in any particular direction we may infer that any correlation between the images in the air/bone space is negligible in comparison



(a) IRTSE (b) VE-PD



(c) Scaled Histograms



(d) Scaled Scatterplot

Figure 4: Two MR images of the same brain slice with associated histograms and scatterplot; (a) Inversion Recovery image - IRTSE, (b) Variable Echo (Proton Density) - VE-PD, (c) Scaled histograms (Red- IRTSE, Blue - VE-PD), and (d) Scatterplot.

to the image noise. With knowledge of the scalings factor it is possible to obtain estimates of the image noise in the air/bone space for both images.

In the example just given the pure tissue distribution was well separated from the other pure tissue distributions and with enough data points to be distinct from partial volume distributions. For other tissue types, such as the fat and CSF distributions, there are not enough data points available to clearly view the pure distributions despite good separation from the other tissues. In this case the partial volume distributions may be used to make an assessment of the noise characteristics.

In theory, the grey level of a partial volume voxel is a linear combination or pure tissue grey levels, in proportion to the fraction of each tissue within the voxel. From this we may deduce that, in two dimensions and with no noise, the partial volume distributions lie along the lines joining the pure tissue centres. In noise, the partial volume line is convolved with the ellipsoid due to noise and natural tissue variation, as shown in figure 4d. This allows an estimate of the image noise to be made by measuring the width of the partial volume distributions orthogonal the the joining centre lines.

Imhomogeneity

In MR imaging, inhomogeneities in the magnetic field are unavoidable at some level. This leads to geometrical distortions and systematic variations in the grey level intensities across both the image plain (within a single slice) and in the through plain (across slices). Problems associated with geometrical distortions are discussed in the next section as part of image alignment. The effects of field inhomogeneities are often subtle and difficult to detect with a cursory visual inspection.

In the scatter plots field inhomogeneity will show as a shift in the mean position of the pure distributions when considering different areas of the images. This shift would also be contained with the 1D histograms but is difficult to observe with any clarity. To assess inhomogeneity we may take data from different areas of the images and overlay the scatter plots.

Results

The following results show overlaid scatter plots of the IRTSE vs DE-PD images for the all of the data sets. In all cases the scatter plots show: (a) plots of the data taken only from the top half of the images overlaid onto the plots using only the data from the bottom half of the images and (b) plots of the data taken only taken from the left half of the images overlaid onto the plots using only the data from the right half of the images. Systematic colour changes across partial volume or pure tissue regions indicate systematic differences in signal behaviour. In this way the overlaid plots can be used to assess how the image intensity distributions vary across the images.

Original Data (ORIG-BC)

Figure 5 shows the overlaid scatter plots for the IRTSE vs DE-PD images from the original (legacy) data sets. For the top-bottom overlays there is a generally good agreement between the plots with some possible discrepancy in the mean position of the grey matter pure distribution (green enhancement of the lower edge of the GM-WM pure and partial volume distributions). In the left-right overlay there is a similar shift in the fat pure (and partial) distributions. Such variations are relatively minor and can probably be addressed by suitable pre-processing.

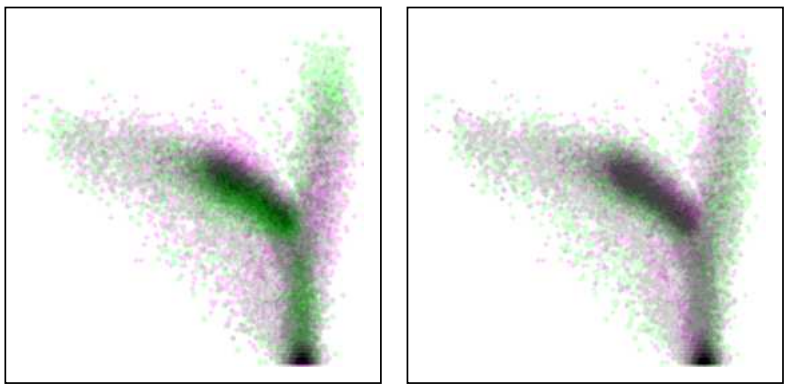


Figure 5: Overlaid scatterplots for the ORIG-BC data (a) Top-Bottom Overlays & (b) Left-Right Overlays

New Data with Birdcage coil (NEW-BC)

Figure 6 shows the overlaid scatter plots for the IRTSE vs DE-PD images from the new data sets with the birdcage coil. This data was collected with broader voxels but with less depth than the original data. From the scatter plots it may be seen that in comparison to the original data there are, in general, less partial volume voxels especially especially between the CSF and Skin/Muscle Distributions. There is, however, still a significant amount of partial volume voxels between the fat and air/bone distributions. This may be expected since the fat tissue is only a thin layer in the in-plane direction, thus through plane partial volume effects may be reduced but there will always remain the in-plane partial volumes.

As with the original data there is a good agreement between the overlays in both the top-bottom plots and the left-right plots. The width of the partial volume distribution between the gray and white matter gives some indication

of the image noise. It appears that the noise may be less than in the original dataset but this is difficult to assess because there is no absolute scaling of the different tissue types. However, we may confirm that the partial volume distributions appear to relatively straight lines between the pure tissue distributions, have a comparable widths, and thus conform well to the model assumptions.

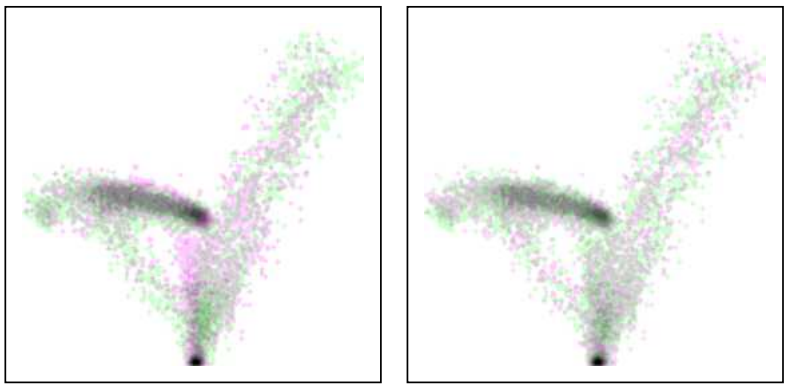


Figure 6: Overlaid scatterplots for the NEW-BC data (a) Top-Bottom Overlays & (b) Left-Right Overlays

New Data with SENSE Coil in Quadrature (NEW-Q)

Figure 7 shows the overlaid scatter plots for the IRTSE vs DE-PD images from the new data sets with the SENSE coil in quadrature mode. In comparison with previous data, we see that the expected linear behaviour of partial volume distributions is not well preserved, and pure tissue distributions are no longer as compact (or Gaussian). From the figure 7a it is also clear there is a problem with the images in that there is a poor agreement in between the overlays. There is a distinct shift in the means of the pure and partial volume distributions for the grey matter, white matter, and CSF tissue classes. The left-right overlays do not show the shift but do show two clear distributions where there should be a single white matter pure distribution. The shift in the distributions tends to be mostly in the vertical direction on the plots indicating the problem is mostly confined to the DE-PD image although there may also be a slight shift in the IRTSE image.

A closer inspection of the raw image data (figure 7) shows that there does appear to be a step shift between the top third of the DE-PD image and the bottom two thirds of the image. This step change is perhaps indicative that the problem lies with the coil and is not due to a slowly varying field inhomogeneity. Consequently, attempts to rectify this problem using a multiplicative (classic) correction would be unlikely to succeed. We discuss this further below.

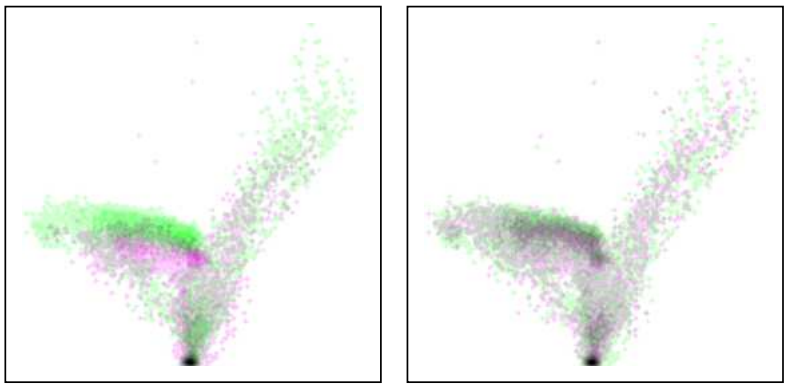


Figure 7: Overlaid scatterplots for the NEW-Q data (a) Top-Bottom Overlays & (b) Left-Right Overlays

Summary

When collecting data for the detection/characterisation of subtle features of the data, like investigating the heterogeneity of tumours, we require the highest level of accuracy which may be obtained. A cursory visual inspection of the individual images does not easily show if there are any problems with the data in terms of noise, signal linearity, image alignment or inhomogeneity. Consequently, images with such artefacts can even be considered good quality for clinical practice. However, for systematic bias to be reduced to a minimum in quantitative analysis, it is necessary that the data be a good match to the model assumptions. Specifically, pure tissue distributions must be compact Gaussians and partial volume data must lie on a line between pure tissue means. If this is not true for normal tissues then it will be impossible to meaningfully quantify differences seen in pathology (heterogeneity). As any difference between two pathological regions could be interpreted either as heterogeneity or imaging artefact. Although constructing an automated method to quantify these effects would be difficult, these characteristics can be visually assessed using overlaid scatterplots.

This paper has described the collection of new sets of data, taken from a 1.5T scanner with a birdcage head coil and a SENSE coil in quadrature. The data sets have been assessed by means of overlaid scatterplots and compared to an original dataset collected 18 years ago on 1.5T scanner with a birdcage head coil.

The overlaid scatterplot for the SENSE coil in quadrature immediately show there to be problems with the data in terms of inhomogeneity (parallel imaging using SENSE and CLEAR are not shown here but were no better). The nature of the shift in the means of the tissue distributions (and from a post hoc inspection of the image data) indicate that the problem lies with the response of the coil and not an effect of "classic" field inhomogeneity.

In this work, we conclude that the SENSE coil does not produce images of sufficient quality to be immediately useful for investigation of tumour heterogeneity. Pure tissue distributions are not compact and partial volume distributions are non-linear. Although we can use knowledge of normal tissues to try to correct these effects in normal regions, the observed behaviour does not seem to be easy to quantify. It would in any case be wrong to assume that any data driven approach could be applied reliably in regions of (unknown) pathology. In addition to signal changes, the use of parallel imaging is also expected to introduce spatially varying noise characteristics, which again contradict the assumptions of any practical segmentation model. These variations are not something which seem amenable to any realisable form of automated correction in the context of multi-spectral segmentation. We certainly cannot expect to address this issue and also make headway with our initial hypothesis regarding characterisation of heterogeneity.

Our requirements for quantitative analysis seem to be at odds with recent advances in MR design, which (for the last decade) appear to have focused instead on generating faster imaging sequences. Currently the only viable options for data acquisition must avoid parallel imaging. We are aware that this conclusion has unfortunate consequences for the integration of multi-spectral segmentation with other MR studies. However, analysis of the new data collected with the birdcage coil, using overlaid scatterplots, shows that there appears to be minimal effects due to inhomogeneity and that the tissue distributions conform well with the expected distributions of the model. In comparison to the older original data the quality of the new data may even show an improvement in image noise, but absolute noise levels are difficult to assess due to the lack of absolute scale.

It is now our intention to use multi-spectral segmentation to assess "normal" tissue in regions around tumours and to use the parameters estimated from normal tissue to standardise data measured between different subjects. As poorly behaved data would make interpretation of pathology impossible, we suggest using overlaid scatterplots as a form of quality assessment in this work. Given the challenges posed, we expect it will also be necessary to use every method at our disposal to control and improve analysis reliability. In this way we hope to establish a method for quantification of tumour heterogeneity and investigate correlations with histology and other bio-markers.

At the time of writing, we have no idea as to the long term practicalities of using sophisticated multi-spectral approaches in clinical practice. Neither do we know if bird cage head coils are likely to continue to be supported by manufacturers. However, in the short term we may be able to determine something of biological significance regarding the structure and development of heterogeneity in tumours.