

Towards a Power Calculation for ADC Measurement in Clinical Trials.

Neil.A.Thacker, Hossein Ragheb and David Morris.

Last updated
11 / 9 / 2012



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

Towards a Power Calculation for ADC Measurement in Clinical Trials.

N.A.Thacker, Hossein Ragheb and David Morris 11/9/12

Aims

The aim of this document is to make explicit the various factors which would be expected to influence a quantitative assessment of diffusion measurement changes for clinical assessment. In order to do this we must make some assumptions regarding the nature of the study, the data acquisition and the analysis.

We will assume;

- That quantitative measurements will be in the form of a regional averaged ADC, and that a method which allows reproducible identification of the required region (typically of 100 voxels of independent ADC measurements) in biological data has been specified.
- That the acquisition is to be made on a scanner using a standard protocol, with matched phantom data.
- That the expected percentage change in measured ADC following treatment can be predicted from prior understanding of the study.

For convenience we will group the expected level of change into 3 classes, selected according to results found in the literature to be typical of what might be required for future studies ¹.

(Type **A**) A study seeking to establish no observable difference in ADC (for example a fasting study) which requires the most precise estimates available (10% of ADC).

(Type **B**) A study seeking to confirm a typical change in biological behaviour which is expected to be significant (and of known direction) but not complete (30% of ADC).

(Type **C**) A study where the expected response is so strong that all structure in tissue is destroyed and the resulting diffusion coefficient is equivalent to water (50% of ADC).

Changes of ADC value beyond this are expected to be accompanied by such drastic changes in anatomical MR images that using a diffusion measurement to assess change is unnecessary.

The intention now is to be able to claim that for a given class of study (A,B or C), a percentage level change of the required amount will be measurable in an individual subject using a specific MR scanner. We must therefore be able to infer the expected level of accuracy from data obtained from the phantom. We will do this by considering fractional changes in ADC precision which arise from the various forms of MRI degradation.

Introduction

Although we now have a physical design for the phantom, its practical use for calibration of diffusion measurement has yet to be fully specified.

Errors which limit quantitative interpretation are generally grouped into two categories, statistical and systematic. While statistical effects can be reduced to some extent by data averaging, systematic effects persist and place a fundamental limit on how data can be used. Following our preliminary work on phantom analysis and discussions of diffusion data, we can say that the processes leading to systematic errors in the measurement of ADC will have several major components. Not only do we expect 'effective' b values between scanners from different vendors by 10%, we also expect this to be spatially varying 10% (e.g. non-uniform B0). There could also be anisotropic effects in the acquisition which lead to directional scaling of components of the diffusion tensor 10% (i.e. B1 gradient effects). As biological differences are observed to be 20%, any one of these problems (if ignored) could lead to an inability to extract useful information during clinical studies.

Other effects, such as approximating the ADC measurement using the trace of the diffusion tensor, and using mono-exponential rather than bi-exponential curve fits can each contribute another factor of 10% to measurement

¹QBIC have recommended a single reproducibility figure of 15%, which we believe would immediately eliminate some scanners from use on any clinical project. A progressive grading scheme is suggested here as a way of recovering some measure of utility for these machines.

errors in anisotropic regions. However, these can also be considered as systematic effects which will cancel out in repeat scans of the same subject imaged at approximately the same orientation, or can be reduced to acceptable levels (less than a few %) by appropriate selection of b values ².

As a ball-park figure, we assume all systematic effects need to be individually reduced to below 5% in order to generate data of sufficient quality for type A clinical studies. To investigate these effects, and the use of the phantom to measure and remove them, we have constructed the following ‘minimal’ experimental design.

Data acquisition

- A) The experiment should comprise images of the phantom and a number of images of a volunteer acquired on scanners from multiple vendors at 1.5T.
- B) The phantom should be imaged, keeping the separate x,y,z diffusion images.
- C) The phantom should be rotated to 4 separate orientations at 90° separations.
- D) The volunteers should be acquired at two different angles (30° rotation) also keeping the separate x,y,z diffusion images.
- E) All experiments should be repeated 2 weeks later.
- F) If experiments cannot incorporate the fasting study we should adopt the default assumption that fasting is necessary.

Analysis

1. Phantom data should be used to assess scale factors for directional anisotropy using data B.
2. Phantom data should be used to assess spatial variation of effective ‘b’ using C.
3. Results from 1 and 2 should be used to correct data from D, which should be compared for reproducibility between different angles and vendors.
4. Analysis 3 should be repeated first without isotropy and then without spatial correction, to assess their effect on ADC calculation.
5. Results from E should be analysed to estimate drift in calibration parameters measured for the phantom and estimate overall reproducibility of ADC.

Interpretation

We should focus upon answering the following questions;

- (I) Are directional anisotropy effects simply described by 3 directional scale factors?
- (II) Is it sufficient to quantify anisotropy and ignore it (if it falls below a nominal factor) or must we plan to acquire ALL data with separate directional components?
- (III) Is spatial correction needed, and can we calibrate with the phantom at one fixed orientation for clinical trials?
- (IV) Do scanners remain in calibration over a period of at least 2 weeks?
- (V) What is the most appropriate way to estimate SNR and should we set a lower limit?

Only once we know these answers can we define a standard phantom calibration procedure. If we start to acquire this data over the next month (April) we may be able to define the calibration process by the summer (July). Only if this work goes well can any clinical trial seriously be considered (Before this we may simply be wasting scanner time).

²Although this is a significant assumption, the IMI consortium has effectively already assumed that this must be true when adopting a diffusion measurement based on trace imaging with a limited number of b values. As in order to say that a systematic effect can be ignored then it must be at least a factor of 2 smaller than the target precision of the overall system.

Methods

The data are analysed using a combination of techniques which are being developed and evaluated for the purposes of ultimate use as quality control systems in the Keosys product. The first of these is an automatic estimation of ADC in the phantom cylinders. This is based upon object recognition and localisation techniques. An initial rigid (Likelihood) estimate of the phantom is refined locally for each cylinder using patch correlation of local image gradients. An estimate of the spatially dependant multiplicative modification of ADC (caused by effective 'b' changes) is made from the water regions of the phantom. This is done using a technique which estimates a local smooth change in relative ADC for the x and y directions in a given slice. These estimates are then integrated to regenerate the apparent multiplicative correction needed to achieve a uniform ADC estimate throughout the phantom.

Diffusion Measurement Study of the Ice-Water Phantom

Currently the data has been acquired on our local 1.5 T scanner, according to the specifications above (A-F) (Figure 1). Data processing is underway and analyses 2,3 and 4 have been attempted.

Each ADC within a cylinder extracted using the automatic location process (Figure 2) is summarised according to the average and its variance, in order to make an estimate of the accuracy (Table 1). The signal at zero b has been recorded for subsequent use (Table 2).

The ADC estimates have been directly compared for orientations differencing by 180°s (Figure 6(top)). As can be seen, the differences between these values is far greater than the estimated errors, indicative of a systematic error in ADC which we attribute to spatial variation of effective 'b' values. Assuming this is a consistent multiplicative effect we estimate the required correction map from the water regions in the phantom (Figure 3) for each dataset (Figure 4). The use of these corrections on ADC values are demonstrated for a typical profile of ADC data in the direction of greatest change (Figure 5).

Spatial correction coefficients C_i are then estimated at the location of each cylinder (Table 3). Corrected estimates of ADC are then made by multiplying the original measurements by the correction

$$D'_i = C_i D_i^m$$

We assume that a nominal value of true ADC for the phantom calibration cylinders is known (D_i^r), and normalise to the central cylinder values.

$$D_i = D'_i \frac{D_0^r}{D_0'}$$

Adjustments are made to propagate the error from the original ADC measurements through to the newly normalised values and their differences. No estimate of the error on the ADC spatial correction factor is made at this time. Differences in reproducibility beyond the the expected level of error are therefore expected to be attributable to a failure of the correction process. These results are plotted in Figure 6 (bottom), and now illustrate quite good reproducibility within the limits of the error estimates, approximately 3x more consistent than before spatial correction.

tube/img	0°	90°	180°	270°
0	118.50±4.00/26.33	122.43±2.70/25.07	121.30±3.82/25.31	122.61±2.87/23.70
1	75.03±5.24/24.90	77.00±5.04/23.13	78.89±7.31/25.32	82.70±5.61/22.99
2	112.70±1.55/23.94	123.08±1.40/26.47	127.02±0.53/15.00 *	125.51±1.17/19.40
3	106.35±5.03/25.89	106.96±4.12/24.30	103.30±6.61/26.07	100.25±2.81/23.69
4	106.30±2.06/22.84	102.68±2.87/23.79	99.15±2.27/26.08	102.01±1.55/23.41

Table 1: Regional mean of N uncorrected ADC measurements together with the corresponding standard deviation estimates ($\sqrt{\text{var}(ADC)}/\sqrt{2N-1}$); here tube 0 refers to the central tube while tubes numbered 1 to 4 refer to tubes with clockwise order where tube 1 refers to the tube located just after the marker.

Conclusions

We have succeeded in developing prototype algorithms for the analysis of phantom data. These support automatic estimation of ADC within the calibration cylinders, and estimation of spatial variation in ADC normalisation from

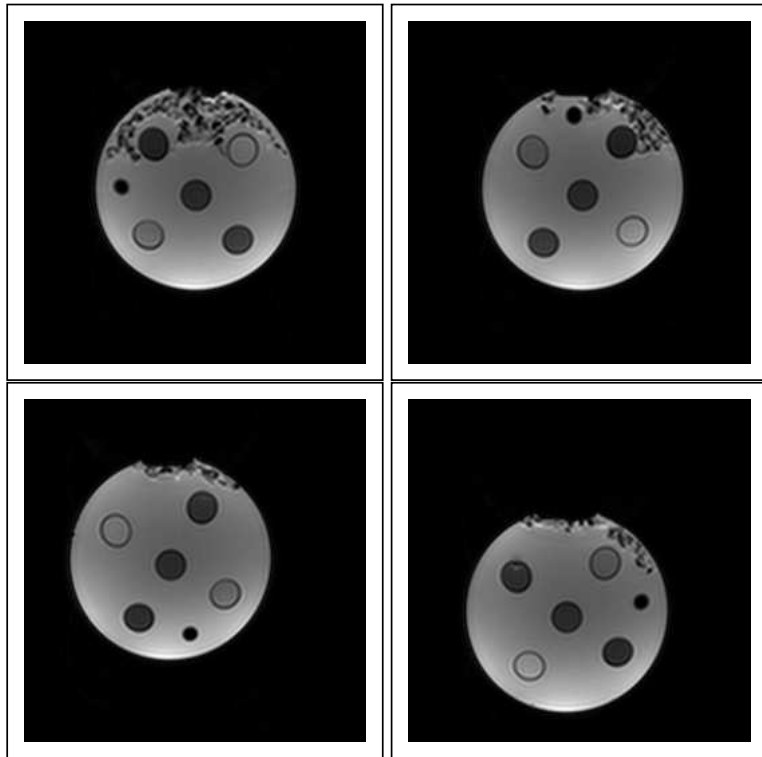


Figure 1: Sample pictures of image slices from the 4 sets of DW-MR image data of the phantom with roughly about 90° orientation difference from each other (MANUNI scanner).

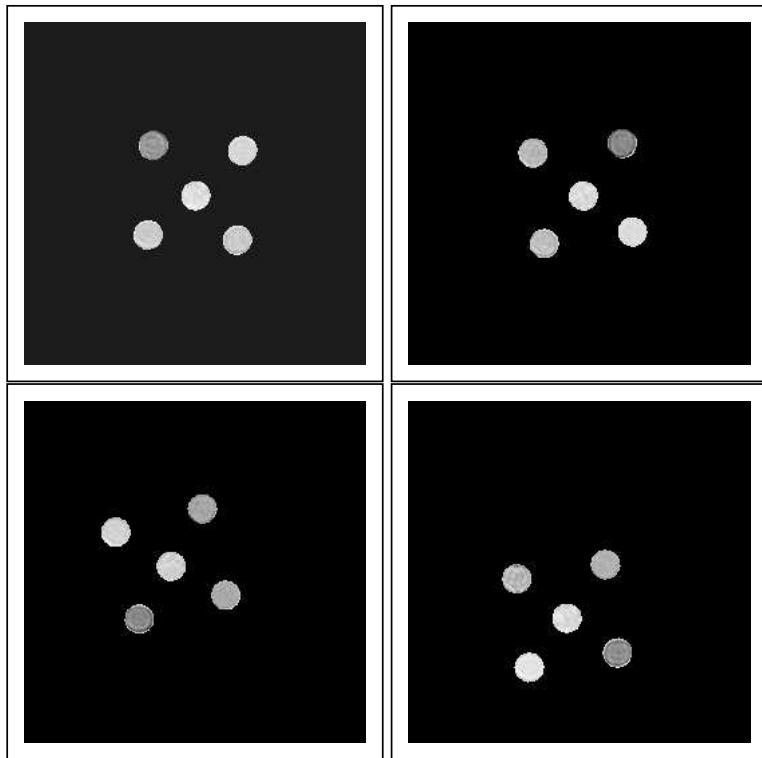


Figure 2: ADC maps of all 5 tubes (which are automatically localised) for the corresponding images shown in Fig. 1.

the water filled region. Previous experience of our correction algorithm suggests that such spatial normalisation processes are generally correct to no better than approximately 25% of the required correction. This level of performance is consistent with the results presented here. It is therefore better to avoid the need for correction

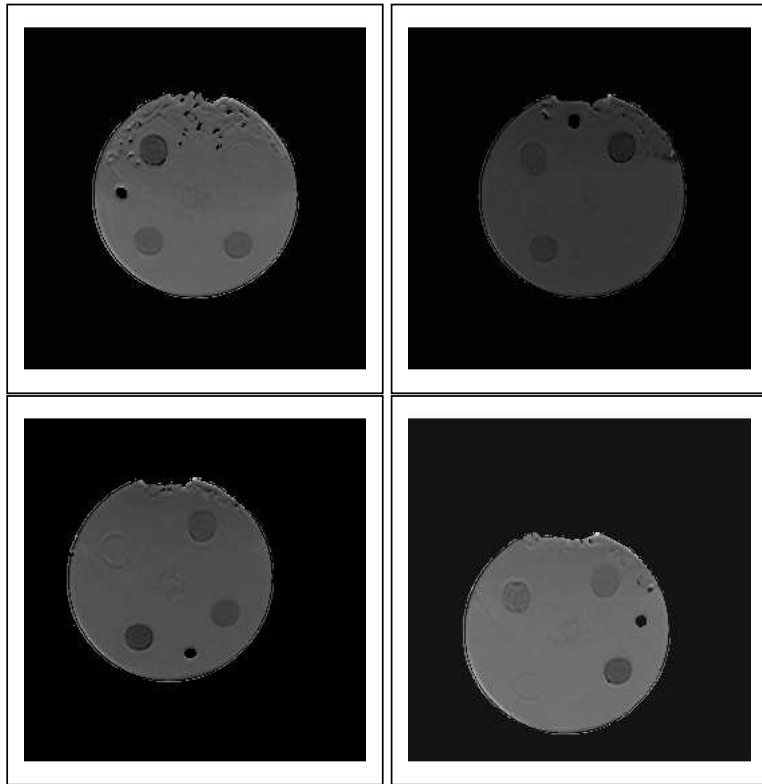


Figure 3: ADC maps of the whole phantom (without localisation) for the corresponding images shown in Fig. 1.

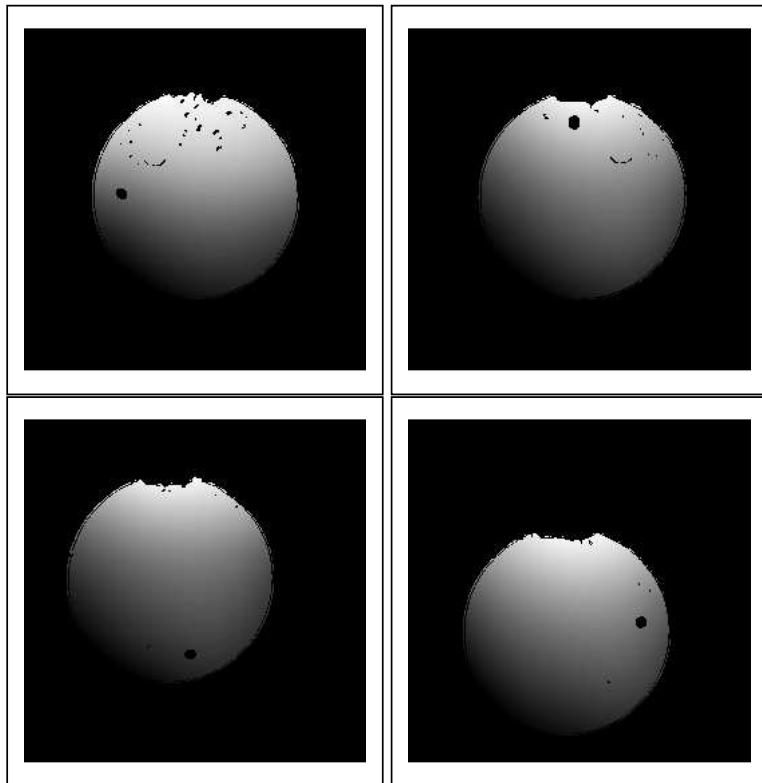


Figure 4: ADC correction maps of the whole phantom for the corresponding images shown in Fig. 1., displayed to make the variation visible. Maximum corrections (at the edges of the phantom) are typically less than 10%.

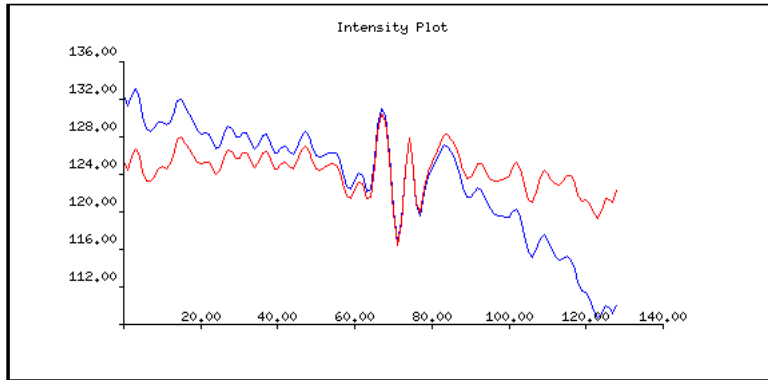


Figure 5: Intensity profile for the uncorrected and corrected ADC maps across a direction in the phantom with relatively large changes in the ADC measures of the ice water (using the 270° image).

tube/img	0°	90°	180°	270°
0	362.36±16.50/20.76	365.66±16.45/20.70	373.41±28.20/23.13	388.06±17.90/20.38
1	307.29±23.54/23.92	310.24±24.71/23.38	378.95±34.11/21.94	419.53±40.00/22.25
2	912.87±35.04/21.63	1209.83±66.26/20.70	1275.79±65.45/21.29	1010.33±33.52/21.23
3	525.64±40.25/22.04	519.28±33.66/21.73	426.79±33.33/22.04	428.99±34.07/22.92
4	903.29±42.55/20.53	685.53±25.01/20.70	745.19±32.69/22.32	903.09±36.03/18.82

Table 2: Regional signal at $b = 0$ (S_0) measurements together with the corresponding standard deviation estimates; tube numbering are the same as those in Table 1.

tub/img	0°	90°	180°	270°
0	1.00	1.00	0.96	1.00
1	1.07	1.01	0.97	1.01
2	1.05	1.05	1.01	1.05
3	0.94	0.98	0.94	0.99
4	0.92	0.94	0.91	0.96

Table 3: Average coil correction coefficients for all 5 tubes in images of phantom with different orientations; here tubes numbered 1 to 4 refer to tubes located on the top-left, top-right, bottom-right and bottom-left quarters of the phantom, no matter what the orientation of the phantom is (tube 0 still refers to the central tube).

by better sequence design if at all possible. Preliminary results indicate that without a correction process it will probably be difficult to achieve a level of reproducibility of better than 10% (our highest category of clinical study) on any current MRI scanner however, accuracies of 15 % may be achievable.

With regard to our list of questions (I-V), we have partial answers to III and V. Spatial correction is already seen to be possible to correct for effective b changes, and it also seems possible to perform a correction based upon one acquisition (orientation) of the phantom. On the issue of signal to noise (V); if we are intending to always take regional averages of ADC which incorporate at least 100 samples in relatively homogenous regions ($\sqrt{\text{var}(ADC)} < 20\%$), then the contribution to the statistical error on ADC from image noise is predicted to be negligible in comparison to the systematic effects from effective b changes, even following our best attempt at correction. A value of $\sqrt{\text{var}(ADC)} = 20\%$ is comparable with the largest variation we could ever measure in all but the most heterogenous of tissues. Put bluntly, when it comes to quantitative measurement of regional ADC, random image noise does not seem to be an issue.

Our next step will be to investigate the individual directional diffusion images in order to assess the possible effects of differing effective b changes. If significant effects are found an additional correction based upon phantom data can be attempted. If it is found that there is very little directional variation in the effective ‘ b ’ spatial correction maps (i.e. $< 2\%$ on ADC), it may be possible to use the method described above to determine a calibration process for clinical studies, sufficient for use in the area of a subject which overlaps the region calibrated by the phantom. However, the loading on the scanner coil changes when a real subject replaces the phantom in the scanner, this is expected to render any spatial correction less effective than observed here. This will need to be quantified.

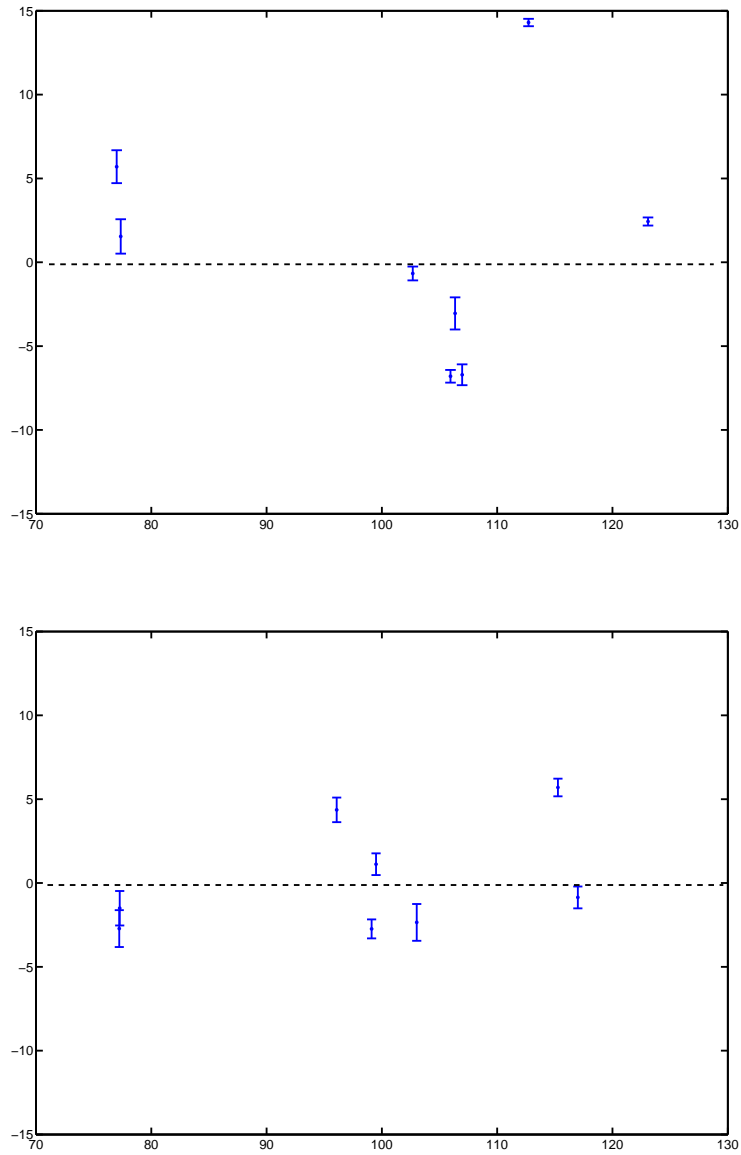


Figure 6: The difference in regional ADC measures for each pair of the phantom images with 180° orientation difference versus the regional ADC measures; uncorrected regional ADCs (top) and corrected regional ADCs (bottom). Error bars are for 3 S.D. of the predicted error on the difference. Spatial correction of ADC helps to recover some reproducibility in measurement, the remaining differences are attributed to imprecise correction.

In vivo data and machines with 3T fields will be likely to display additional problems. The current point of greatest concern is that the fat suppression used in 3T scanners often generates significant levels of artefact. Not only are these effects likely to be difficult to correct but the magnitude of the errors generated are inestimable from phantom data.