

Scanner Quality Assurance using the IMI QuIC-ConCePT Ice-Water Phantom (draft #2).

Hossein Ragheb, Neil A. Thacker, David M. Morris and Alan Jackson.

Last updated
29 / 04 / 2014



ISBE, Medical School,
University of Manchester,
Stopford Building, Oxford Road,
Manchester, M13 9PT, UK.

Scanner Quality Assurance using the IMI QuIC-ConCePT Ice-Water Phantom (draft #2).

Hossein Ragheb, Neil A. Thacker, David M. Morris and Alan Jackson

29 / 04 / 2014

Abstract

The aim of this document is to define the quality assurance (QA) methods for scanners using the ice-water phantom and the corresponding software developed within the IMI QuIC-ConCePT project, for the upcoming ‘BOS2’ trials. The outlined methods are what we consider to constitute the best use of currently available data following our recent papers [1] [2] in which we studied ADC (apparent diffusion coefficient) measurements corresponding to five phantom tubes obtained from four scanners located at four different sites.

Scanners are required to gain approval based on their phantom data quality in order to proceed to the clinical trial. The idea is to obtain one phantom data-set from each site before site activation and one week prior to clinical scans. Scanners used in our recent study resulted in about 6% average measurement variability in healthy liver data. These results and those from phantom studies, have been used here to set criteria for QA pass/fail.

To approve the quality of each new scanner, several metrics of image quality will be automatically extracted from phantom data-sets. A simple statistical analysis then provides a three step procedure for the pass or fail decision. The full process is described below with example data.

For the current QA we concentrate on signal to noise and absolute quantitation, ignoring issues of fat suppression and inhomogeneity which will need to be addressed separately (where possible).

Introduction

Each site should upload their phantom DW-MRI data (based on the imaging protocol provided) to the QuIC-ConCePT platform without any delay. An ‘investigator’ should view the corresponding image slices immediately and approve its general quality so that the data can be made accessible to the ‘readers’. The software for automatic localisation of tubes regions and their ADC measurements should be executed automatically upon phantom data upload. This will generate zip files of results ready to be downloaded and analysed. The software should be executed on the 3 central image slices of the phantom. The person(s) in UNIMAN in charge of analysing these results will view the ADC maps as a means for ensuring that the tubes are accurately localised by the software (for rare cases such as severe image distortion). Each text file which provides the mean ADC values corresponding to the five tubes (see Figures 1-2) will then be assessed. Note that before proceeding with the analysis, one must sort the results for the 4 tubes listed in the text file (‘Tube = 1’ to ‘Tube = 4’) to match with the definition of tube order in Figure 1. This is performed based on the reported coordinates of each tube centre (the numbers in parenthesis are x and y respectively) and by viewing the location of the phantom marker (between T0 and T4) on a sample image slice (with the origin being the top-left corner of the image slice, x-axis the horizontal axis and y-axis the vertical axis).

To facilitate the analysis of these text files, UNIMAN has recently developed the phantom QA tool (as part of the phantom software) in order to be implemented in the QuIC-ConCePT platform by KEOSYS. This tool should be executed on the results of phantom ADC tool. The output will be a text file (‘Phantom_QA_Results.txt’) which contains the final 4 numbers needed for the QA (see the section ‘QA Pass Fail Criteria’). Further details are also provided in this file only for inspection by UNIMAN in case of any QA problems (Figure 3). In order for the QA software to work properly, the phantom must be placed in the scanner in a specific orientation based on the marker (as seen in Figure 1) with a maximum of $\pm 30^\circ$ tolerance, as mentioned in the imaging manual.

We have previously obtained 5 average values for ADC measurement, corresponding to the 5 tubes in the phantom from 12 1.5 T data-sets¹ provided by 4 sites A, B, C and E [1] [2], using the automated analysis software. We use these reference ADC values for the QA procedure. We require that the liquid content of each tube remains unchanged and software fitting routines are unmodified. These reference ADC values are $\bar{D}(T0)_{ref} = 112.3$,

¹The results of this report are not guaranteed to be applicable to 3T data.

$\bar{D}(T1)_{ref} = 74.0$, $\bar{D}(T2)_{ref} = 113.3$, $\bar{D}(T3)_{ref} = 94.8$ and $\bar{D}(T4)_{ref} = 95.5$ ($10^{-5} \text{ mm}^2/\text{s}$) corresponding to the tubes T0, T1, T2, T3 and T4 (see Figure 1 and Table 1).

An example ADC data analysis is provided in Table 1 and Figures 1-5.

Timeline

For the initial QA (before site activation), the phantom scan must be performed one week before the patient scan. The phantom data must then be uploaded (by the corresponding site) and processed automatically (by the phantom software ADC and QA tools) upon data upload. The results of the analysis should only be accessible to the corresponding site and UNIMAN. UNIMAN should be informed and respond within 5 days.

Once a scanner passes the QA, for further regular QAs, phantoms must continue to be acquired within one week of clinical scans. The results of the analysis should be accessible to the corresponding site and UNIMAN. The person who has uploaded the data should confirm that new measurements are consistent (within limits defined below) with previous scanner measurements. If the scanner is outside the specified limits, the site should immediately inform UNIMAN to investigate whether the scanner actually failed or there has been a technical problem

Methods

Phantom Ice Packing: One possible source of ADC inconsistency is expected to be lack of sufficient temperature control [3]. Consequently, the amount of ice present in the ice-water regions of the phantom may have an impact on the reliability of measurements. Hence, we visually estimate the quantity of ice in each phantom data-set as a percentage (see Figure 4) to the nearest 5%.

Phantom Normalised Goodness of ADC Fit: For each individual tube in the phantom, an average goodness of fit value $\chi^2(T)$ (chi-squared), an average ADC value $D(T)$, and an average signal value $S0(T)$ (corresponding to b-value=0) is stored in the output text file. Also for the corresponding image slice an estimate of the standard deviation of the image noise σ_I is stored (see Figures 1-2 and Table 1). Note that in order to get an estimate of image noise σ_I in the output text file, the Ricean correction parameter should be set to 1.0 when the software is executed in the KEOSYS platform. The normalised goodness of fit (the quantity used for the vertical axis of Figure 5) is then computed using

$$\eta(P) = \kappa \sqrt{\chi^2(P) \bar{\sigma}_I(P) / \bar{S}0(P)} \quad (1)$$

where $\kappa = 84.5$ is a fixed coefficient (based on our previous study [2]) which normalises these values to a level comparable to $\sqrt{\chi^2(P)}$ and the corresponding error bars (see Figure 5). As explained in [2], κ is the ratio of the mean of all 12 mean signals to the mean of all 12 noise standard deviations from all the 4 sites studied.

In Table 1, we tabulate the S0 signal $S0(T)$, the goodness of fit $\chi^2(T)$ and standard deviation of image noise σ_I for three adjacent slices, where (c) refers to the central slice while ($c - 1$) and ($c + 1$) refer to its adjacent slices. For each of these three quantities, we find a mean value which is then used in Equation (1). These are referred to as $\bar{S}0(P)$, $\bar{\chi}^2(P)$ and $\bar{\sigma}_I(P)$. Note that in our notations (T) refers to a quantity corresponding to each individual tube, while (P) refers to the mean quantity computed from these tube quantities to represent the corresponding phantom.

Phantom ADC Scale and Scaled ADC Standard Deviation: In Table 1, we also tabulate the ADC measurements on five tubes from three image slices corresponding to an example phantom data-set. The last column on the table provides the mean ADC reference value $\bar{D}(T)_{ref}$ corresponding to each tube (from our previous study). Hence, we can compute the mean and standard deviation for the phantom (new data-set) using the ADC measurements $D(T)$ from the five tubes in each image slice. These are referred to as $\bar{D}(P)$ and $\sigma_D(P)$. The standard deviation $\sigma_D(P)$ is computed using all ADC values from tubes $D(T)$ together with their corresponding reference mean value $\bar{D}(T)_{ref}$. Specifically $\sigma_D(P)$ is given by

$$\sigma_D(P) = \sqrt{\frac{1}{M} \sum_{m=1}^M (D(Tm) - \bar{D}(Tn)_{ref})^2} \quad (2)$$

where $M = 15$ (for 3 adjacent slices, each having 5 tube measurements) and m refers to the rows 1 to 15 in Table 1. Also while $D(Tm)$ includes ADC measurements from the adjacent slices, $\bar{D}(Tn)_{ref}$ refers to the single mean reference ADC for tube n , where $0 \leq n \leq 4$. For example, $\bar{D}(T0)_{ref} = 112.3$.

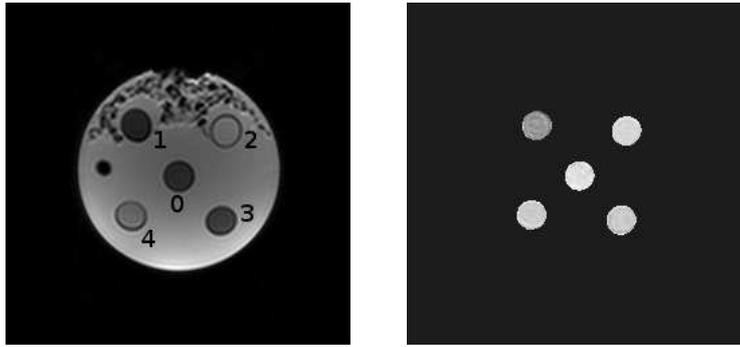


Figure 1: Left: a sample image slice ($b=0$) from the phantom data used in the study with tube numbers superimposed (tubes 0 to 4 are referred to as T0 to T4); right: the corresponding ADC map generated using the project software.

```

Tube = 0 ( 129.39 , 112.93 )
ADC = 114.56 +- 3.14 / 26.80
S0 = 288.34 +- 20.63 / 23.32

Chisq mean = 2.64

Tube = 1 ( 159.36 , 75.56 )
ADC = 115.72 +- 2.36 / 25.76
S0 = 756.80 +- 34.21 / 21.85

Chisq mean = 3.54

Tube = 2 ( 165.49 , 143.66 )
ADC = 95.55 +- 4.97 / 24.02
S0 = 377.41 +- 21.32 / 20.89

Chisq mean = 4.03

Tube = 3 ( 98.44 , 148.59 )
ADC = 96.94 +- 2.11 / 22.00
S0 = 738.12 +- 44.52 / 22.05

Chisq mean = 3.59

Tube = 4 ( 93.21 , 81.20 )
ADC = 75.66 +- 7.92 / 25.66
S0 = 259.90 +- 16.78 / 22.36

Chisq mean = 2.69

image noise used for Ricean correction (factor=1.00) = 1.70

```

Figure 2: The output text file (corresponding to an example phantom data-set) generated using the project phantom ADC software (available through QuIC-ConCePT platform); three of these text files (corresponding to three adjacent image slices) are used by the phantom QA tool to compute 4 final numbers needed for QA (as shown in Figure 3).

```

Results to be used for QA:

phantom scaling factor {rho(p)} = 1.021698

phantom standard deviation on scaled ADCs {sigmaD(p)} = 0.937578

phantom normalised goodness of ADC fits {eta(p)} = 0.551178

phantom criteria parameter {A*sigma(p) + B*eta_p} = 1.987371

Results below are only printed for UNIMAN to help resolving possible QA
problems.

```

Figure 3: The top of the output text file (corresponding to an example phantom data-set) generated using the project phantom QA software (available through QuIC-ConCePT platform); the first number in the file to be used for QA is $\rho(P)$ ($\rho(p)$), the second one is $\sigma'_D(P)$ ($\sigma_D(p)$), the third one is $\eta(P)$ ($\eta(p)$), and the fourth one is $A\sigma'_D(P) + B\eta(P)$ ($A*\sigma_D(p) + B*\eta(p)$).

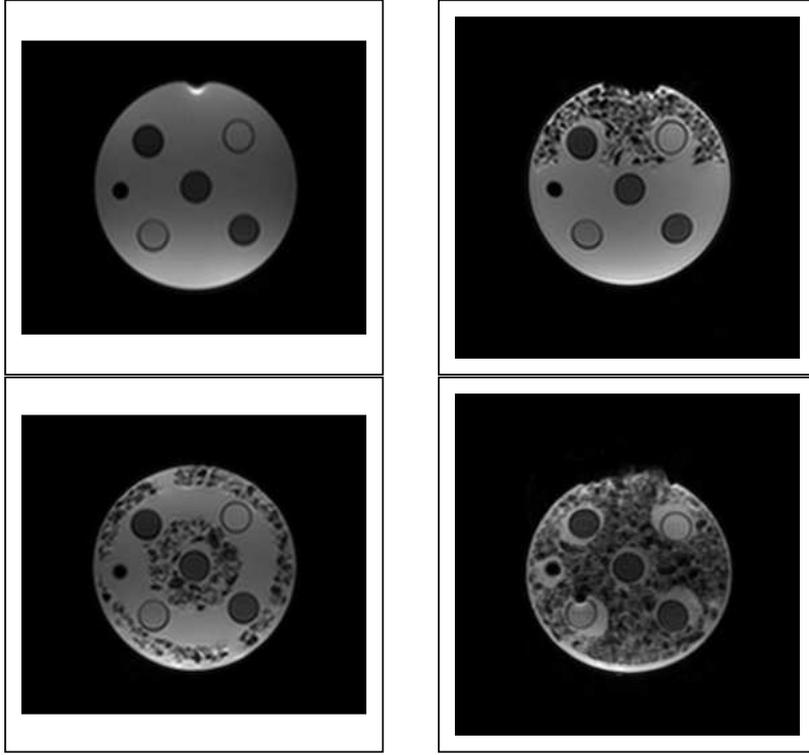


Figure 4: Pictures of phantom central image slice from four different scanners with different amounts of ice; no ice (top-left), 35% ice (top-right), 60% ice (bottom-left) and 95% ice (bottom-right).

Further, the overall mean reference ADC value $\bar{D}(P)_{ref} = 98.02$ for the mean values $\bar{D}(Tn)_{ref}$ is used to compute a scale value $\rho(P)$ for the phantom.

$$\rho(P) = \bar{D}(P) / \bar{D}(P)_{ref} \quad (3)$$

The scale value provides important information about scaling between each scanner and the reference data. It also provides the factor with which a new standard deviation can be computed for each data-set making the comparison between scanners more tractable. Hence $\sigma'_D(P)$ (the quantity used for the horizontal axis of Figure 5), which refers to the standard deviation after scaling the ADC values by the scale value, is given by

$$\sigma'_D(P) = \sqrt{\frac{1}{M-1} \sum_{m=1}^M \left(\frac{D(Tm)}{\rho(P)} - \bar{D}(Tn)_{ref} \right)^2} \quad (4)$$

Finally in Figure 5, we show the plot of the normalised goodness of ADC fit $\eta(P)$ against the standard deviation of the scaled ADCs $\sigma'_D(P)$ corresponding to phantom data from four different sites A, B, C and E (three scans per site). This plot has appeared in our recent publication [1] (and the corresponding report [2]). We have modified the plot only by drawing a polygon using three solid lines for QA (see below).

QA Pass Fail Criteria

The initial QA decision can be summarised in three steps as follows.

1) The procedures outlined for loading the phantom with ice should be followed. This should result in significant quantities of ice in the phantom at the point of image acquisition. However, if there is less than about 50% ice appearing in the phantom image data, the corresponding scanner fails to pass the QA.

2) In order to believe that a new scanner is delivering data of a similar quality to the reference data, we demand that the scaling factor $\rho(P)$ should be between symmetric minimum (0.96) and maximum (1.04) values (based on our recent study).

$0.96 \leq \rho(P) \leq 1.04$ (for regular QAs by sites, $\rho(P)$ should be within %2 of the initial QA measurement)

m	n	tube (slice no)	$S0(T)$	$\chi^2(T)$	σ_I	$D(T)$	$\bar{D}(T)_{ref}$
1	0	$T0(c-1)$	294.3	3.22	1.79	113.39	112.3
2		$T0(c)$	288.3	2.64	1.70	114.56	
3		$T0(c+1)$	295.8	3.12	1.78	115.23	
4	1	$T1(c-1)$	260.3	2.56		75.71	74.0
5		$T1(c)$	259.9	2.69		75.66	
6		$T1(c+1)$	261.0	2.46		78.16	
7	2	$T2(c-1)$	762.0	2.97		115.79	113.3
8		$T2(c)$	756.8	3.54		115.72	
9		$T2(c+1)$	747.9	2.92		116.41	
10	3	$T3(c-1)$	379.2	3.96		96.15	94.8
11		$T3(c)$	377.4	4.03		95.55	
12		$T3(c+1)$	385.1	3.71		97.05	
13	4	$T4(c-1)$	729.6	3.66		96.88	95.5
14		$T4(c)$	738.1	3.59		96.94	
15		$T4(c+1)$	744.2	3.65		98.39	
derived quantities (below this line)							
		$\bar{S}0(P)$	458.7				
		$\bar{\chi}^2(P)$		3.248			
		$\bar{\sigma}_I(P)$			1.75		
Eq. (1)		$\eta(P)$			0.581		
		$\bar{D}(P)_{ref}$					98.02
		$\bar{D}(P)$				100.1	
Eq. (2)		$\sigma_D(P)$				2.07	
Eq. (3)		$\rho(P)$				1.0212	
Eq. (4)		$\sigma'_D(P)$				0.85	

Table 1: ADC values measured in each tube $D(T)$ on each of the three central adjacent slices ($c-1$, c and $c+1$) used; the mean and standard deviation of these values across the tubes ($\bar{D}(P)$ and $\sigma_D(P)$); the expected (reference) mean ADC values for each tube $\bar{D}(T)_{ref}$ from our recent study [2]; $\sigma'_D(P)$ refers to the standard deviation after scaling the ADC values by the scale value $\rho(P)$; ADCs are measured in units of $10^{-5} \text{ mm}^2/\text{s}$; signal values $S0(T)$ and goodness of fit values $\chi^2(T)$ computed for each tube on each of the three slices used; the standard deviation of image noise σ_I for each image slice; and the overall mean values for each of these columns, i.e. $\bar{S}0(P)$, $\bar{\chi}^2(P)$ and $\bar{\sigma}_I(P)$; c refers to central, (T) refers to tube and (P) refers to phantom.

3) Based on the polygon drawn in Figure 5, in order for a scanner to pass the QA, the following conditions must hold on the corresponding data.

$\sigma'_D(P) < 6.0$ (for regular QAs by sites, $\sigma'_D(P)$ should be within %40 of the initial QA measurement)

$\eta(P) < 1.55$ (for regular QAs by sites, $\eta(P)$ should be within %40 of the initial QA measurement)

$A\sigma'_D(P) + B\eta(P) < C$, where $A = 0.65$, $B = 2.5$ and $C = 6.15$ (for regular QAs by sites, $A\sigma'_D(P) + B\eta(P)$ should be within %40 of the initial QA measurement).

On the basis of previous data, we estimate that scanners generating data of similar quality to those used for reference [2] will fail these criteria at a level less than 1%.

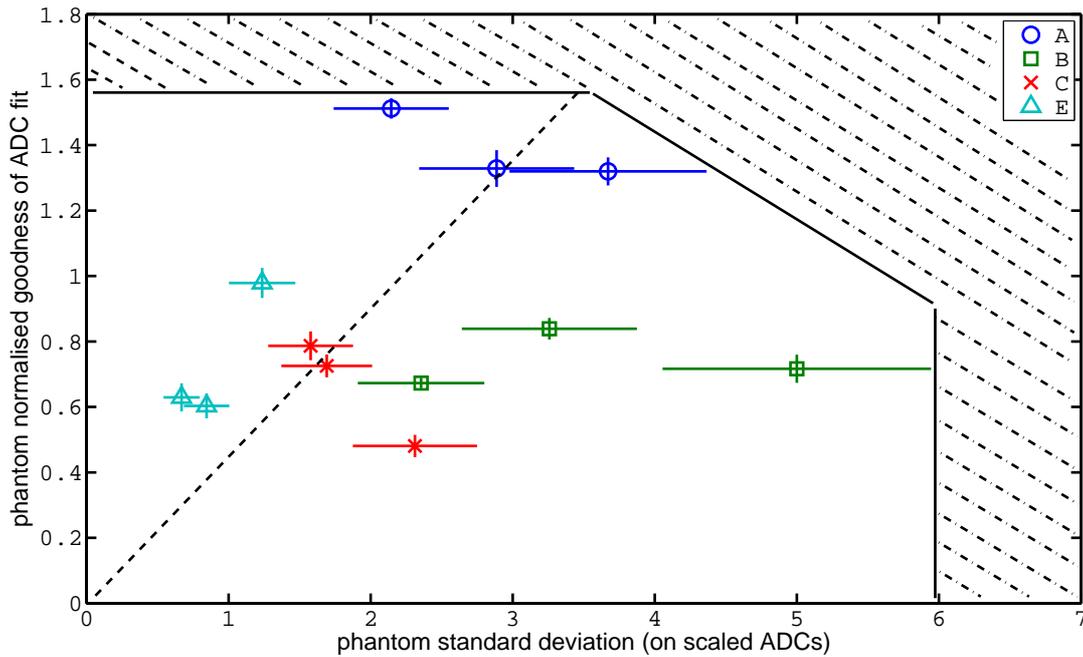


Figure 5: Phantom normalised goodness of fit against the standard deviation of the scaled ADCs for different sites A, B, C and E (three scans per site) from our recent study [2]; the dashed line shows the best fit correlation between the two variables on horizontal and vertical axes; the three solid lines form a polygon with the horizontal and vertical axes which defines the area for passing the QA; the points falling outside this area fail to pass the QA (the dash-dotted area).

References

- [1] H. Ragheb, N.A. Thacker, D.M. Morris and A. Jackson, Interpreting Ice-Water Phantom Data for Prediction of Clinical ADC Measurement, *Tina memo* 2013-005, 2013, <http://www.tina-vision.net/docs/memos/2013-005.pdf>.
- [2] H. Ragheb, N.A. Thacker, D.M. Morris, N.H.M. Douglas and A. Jackson, Predicting the Quantitative Accuracy of In-Vivo ADC using an Ice-Water Phantom, *International Society for Magnetic Resonance in Medicine (ISMRM)*, Milan, Italy, May 10-16, 2014, to appear.
- [3] T.L. Chenevert, C.J. Galban, M.K. Ivancevic, S.E. Rohrer, F.J. Londy, T.C. Kwee, C.R. Meyer, T.D. Johnson, A. Rehemtulla and B.D. Ross, Diffusion Coefficient Measurement Using a Temperature-Controlled Fluid for Quality Control in Multicenter Studies, *J. Magnetic Resonance Imaging*, 34:983-987, 2011.