Consistent intensity inhomogeneity correction in water–fat MRI

Thord Andersson, Thobias Romu, Anette Karlsson, Bengt Norén, Mikael Forsgren, Örjan Smedby, Stergios Kechagias, Sven Almer, Peter Lundberg, Magnus Borga and Olof Dahlqvist Leinhard

Linköping University Post Print

N.B.: When citing this work, cite the original article.

Original Publication:
http://dx.doi.org/10.1002/jmri.24778
Copyright: Wiley
http://eu.wiley.com/WileyCDA/

Postprint available at: Linköping University Electronic Press
http://urn.kb.se/resolve?urn=urn:nbn:se:liu:diva-112129
Consistent Intensity Inhomogeneity Correction in Water-Fat MRI

Thord Andersson¹,²*, MSc, Thobias Romu¹,², MSc, Anette Karlsson¹,², MSc, Bengt Norén²,³, PhD, Mikael F. Forsgren²,⁴, MSc, Örjan Smedby²,³, Prof., Stergios Kechagias⁵, Prof., Sven Almer⁶,⁷, Prof., Peter Lundberg²,⁴, Prof., Magnus Borgia¹,², Prof., Olof Dahlqvist Leinhard²,⁸, PhD

¹Department of Biomedical Engineering (IMT), Linköping University, Linköping, Sweden
²Center for Medical Image Science and Visualization (CMIV), Linköping University, Linköping, Sweden
³Department of Radiology and Department of Medical and Health Sciences, Linköping University, Linköping, Sweden
⁴Department of Radiation Physics and Department of Medical and Health Sciences, Linköping University, Linköping, Sweden
⁵Department of Gastroenterology and Hepatology and Department of Medical and Health Sciences, Linköping University, Linköping, Sweden
⁶Department of Gastroenterology and Department of Clinical and Experimental Medicine, Linköping University, Linköping, Sweden
⁷Div of Gastroenterology, Karolinska Institutet, Karolinska University Hospital, Stockholm, Sweden
⁸Department of Medical and Health Sciences, Linköping University, Linköping, Sweden

*Correspondence to: Thord Andersson, Dept of Biomedical Engineering (IMT), Linköping University, SE-58183, Linköping, Sweden. E-mail: thord.andersson@liu.se

ACKNOWLEDGMENTS

Financial support from the Swedish Research Council (VR/M 2007–2884), the Research Council of Southeast Sweden (FORSS 12621), Linköping University, Lions Research Foundation in Linköping, Linköping University Hospital Research Foundations and the County Council of Östergötland is gratefully acknowledged.

Key words: water-fat imaging, Dixon imaging, inhomogeneity correction, intensity correction, water-fat quantification

Running title: Consistent Intensity Correction in WF-MRI
ABSTRACT

Purpose: To quantitatively and qualitatively evaluate the water-signal performance of the Consistent Intensity Inhomogeneity Correction (CIIC) method to correct for intensity inhomogeneities.

Materials and Methods: Water-Fat volumes were acquired using 1.5 T and 3.0 T symmetrically sampled 2-point Dixon 3D MRI. Two datasets: 1) 10 muscle tissue ROIs from 10 subjects acquired with both 1.5 T and 3.0 T whole-body MRI. 2) 7 liver tissue ROIs from 36 patients imaged using 1.5 T MRI at six time points after Gd-EOB-DTPA injection. The performance of CIIC was evaluated quantitatively by analyzing its impact on the dispersion and bias of the water image ROI intensities, and qualitatively using side-by-side image comparisons.

Results: CIIC significantly ($P^{1.5T} \leq 2.3 \times 10^{-4}, P^{3.0T} \leq 1.0 \times 10^{-6}$) decreased the non-physiological intensity variance while preserving the average intensity levels. The side-by-side comparisons showed improved intensity consistency ($P^{\text{int}} \leq 10^{-6}$) while not introducing artifacts ($P^{\text{art}} = 0.024$) nor changed appearances ($P^{\text{app}} \leq 10^{-6}$).

Conclusion: CIIC improves the spatiotemporal intensity consistency in regions of a homogenous tissue type.

Key words: water-fat imaging, Dixon imaging, inhomogeneity correction, intensity correction, water-fat quantification
INTRODUCTION
When using medical imaging techniques, a highly desirable image quality is tissue intensity homogeneity, i.e. that a certain type of tissue should be represented by the same distribution of intensity values, invariant of its position in the image volume and of time. However, this is rarely the case due to scanning process issues such as inhomogeneity of the static magnetic field, variance in reception coil sensitivity, pulse sequence optimization and patient movement. The resulting intensity inhomogeneities can significantly degrade the performance of common medical imaging operations, e.g. segmentation and registration, which depend on the intensity values or their gradients.
Several intensity inhomogeneity correction (IIC) methods have consequently been proposed, in particular for the single volume case. These methods can be categorized as prospective or retrospective. Prospective correction methods focus on the calibration and optimization of the image volume acquisition process. These include methods using phantoms (1), multi-coils (2) and special sequences (3). If surface coils are used, e.g. phased-array coils, the manufacturers of MRI scanners typically implement a prospective correction method as part of the acquisition protocol (2). These methods use coil sensitivity maps, acquired in a separate reference scan, in order to correct intensity inhomogeneities. The surface coil image is normalized using the sensitivity maps, resulting in an image that has the homogeneity of the integrated body coil image, while preserving the increased sensitivity of the surface coil. These methods perform a B1-receive field correction (4), and work in a similar way across all the manufacturers. Retrospective correction methods use information from the acquired image volume, including histograms, spatial frequencies, intensities and gradients. If the target application is quantitative imaging, it is important for the correction method to have a physically based reference, as it should operate only on the image domain of anatomically and physiologically relevant images (5,6). However, many of the correction methods
are not physically based and do not work optimally in a multi-volume setting, e.g. time-series, as they do not use a common intensity reference for all the involved volumes. This may lead to, for example, an image volume time series that is intensity homogeneous per time point, but not considered over the whole time series. This residual intensity variation over the volume set could decrease the performance of a subsequent analysis or image processing method. Moreover, using an intensity correction method without a common intensity reference also makes direct comparisons between patients or repeated examinations difficult.

In order to correct intensity inhomogeneity in water-fat MRI using Dixon imaging, and to provide a stable common spatiotemporal intensity reference in MRI time series or other multi-volume settings, a method here referred to as ‘Consistent Intensity Inhomogeneity Correction’ (CIIC) has been proposed (7,8). CIIC is a physically based MR scaling method that uses adipose tissue as an internal reference in order to calculate a dense scaling field for each position in the spatiotemporal volume. It is therefore self-calibrating in the sense that it is independent of the effective scaling of the MR scanner at the acquisition time. As such, CIIC accepts original ‘raw’ scanner images as input, as well as pre-processed images using techniques such as sensitivity map normalization, independent of any particular MRI scanner manufacturer or implementation. Since CIIC is applied independently on each acquired volume, temporal inhomogeneities caused by e.g. patient movement can be handled as well as the spatial inhomogeneities. CIIC has successfully been used for fat quantification in previous studies by, for example, Dahlqvist Leinhard et al. (7,8), Lidell et al. (9) and, using a different implementation, by Ludwig et al. (10), and for lean muscle tissue quantification by Karlsson et al. (11).
The purpose of this work is to quantitatively and qualitatively evaluate the water-signal performance of CIIC by comparing it to the output from the Philips 1.5 T Achieva and 3.0 T Ingenia MR-scanners used for data acquisition (Philips Healthcare, Best, The Netherlands).

MATERIALS AND METHODS

The Muscle Dataset
Ten healthy volunteers (6 women and 4 men) participated in this local ethics committee-approved study, after each giving written informed consent. Demographic details of the volunteers are presented in Table 1. Water-fat separated whole body images were acquired using a 1.5 T Philips Achieva and a 3.0 T Philips Ingenia MR-scanner respectively (Philips Health Care, Best, the Netherlands), with a 3D gradient echo sequence with opposite phase and in-phase echo times of 2.3 ms and 4.6 ms respectively for 1.5 T, and 1.15 ms and 2.3 ms respectively for 3.0 T (12). Each image stack was acquired in the axial plane. The 1.5 T scanner used a quadrature body coil (QBC) without any additional inhomogeneity correction processing. The 3.0 T scanner used a sensitivity encoding (SENSE (2)) phased-array body coil with CLEAR inhomogeneity correction, which is Philips’ implementation of coil sensitivity normalization. For 1.5 T, the repetition time was 6.58 ms and the flip angle was 10° with an acquired resolution of 3.50x3.50x3.50 mm³. For 3.0 T, the repetition time was 3.8 ms and the flip angle was 10° with an acquired resolution of 1.75x1.75x1.75 mm³.

The image protocol was applied repeatedly, starting from the head, with a 30 mm and a 28 mm image stack overlap for 1.5 T and 3.0 T respectively, until whole body coverage was achieved. In the abdominal region, expiratory breath-hold acquisition was used minimizing respiratory artifacts. The total scanning time was approximately 10 minutes for 1.5 T and 25 minutes for 3.0
T. Water-fat separation of the in-phase and opposite-phase images was performed using phase sensitive reconstruction (13-15). The time between the 1.5 T and 3.0 T acquisitions was approximately 30 minutes for each subject.

Water image data $D$ were extracted from ten different muscle ROIs, defined by a trained operator as distinct groups with homogeneous muscle parenchyma. The muscle ROIs were from the {Abdomen, Arm, Lower-Leg, Upper-Leg-Back, Upper-Leg-Front} from the subjects’ left and right side respectively. The intensity values are denoted $D^m(p, x, y, z)$, where $m$={1.5 T, 3.0 T} indicates the originating scanner, $p$ is the subject index and $x, y, z$ are the spatial coordinates. The dataset after the CIIC post-processing is denoted $D_{\text{CIIC}}^m$. In order to be able to compare $D^m$ and $D_{\text{CIIC}}^m$ statistically, they must be described using the same global scale. Each one of the four subsets $D_{1.5T}, D_{3.0T}, D_{\text{CIIC}}^{1.5T}$ and $D_{\text{CIIC}}^{3.0T}$ was therefore scaled so that their estimated probability density functions (pdfs) all share the same mode (maximum peak): $\text{Mo}[D^m]=\text{Mo}[D_{\text{CIIC}}^m]=1$ where $\text{Mo}[\cdot]$ denotes the first mode of the operand’s pdf. See Figure 1 for the rescaled pdfs of $D^m$ and $D_{\text{CIIC}}^m$ for $m$={1.5 T, 3.0 T} respectively. A muscle ROI, “Lower-Leg”, with significantly increased intensity levels was manually identified as a large inhomogeneity in the $D_{3.0T}$ dataset. This can also be seen in the corresponding, almost bimodal, pdf for $D_{3.0T}$ in Figure 1, as well as in the example in Figure 2. The rescaled datasets were finally reduced by representing each ROI with its average $\mu$ and standard deviation $\sigma$, resulting in the 10x10 matrices $\mu^m(p, r), \mu_{\text{CIIC}}^m(p, r)$ and $\sigma^m(p, r), \sigma_{\text{CIIC}}^m(p, r)$, where $r$ is the ROI index. Let $\varepsilon_{\text{CIIC}}$ and $\varepsilon$ denote the squared difference between 1.5 T and 3.0 T ROI-average intensity data, with and without CIIC respectively.
\begin{equation}
\varepsilon_{\text{CIC}}(p, r) = \left( \mu_{\text{CIC}}^{1.5T}(p, r) - \mu_{\text{CIC}}^{3.0T}(p, r) \right)^2
\end{equation}

\begin{equation}
\varepsilon(p, r) = \left( \mu^{1.5T}(p, r) - \mu^{3.0T}(p, r) \right)^2
\end{equation}

for all subjects \( p \) and ROI indices \( r \). Let \( s''(p) \), \( s''_{\text{CIC}}(p) \) be the standard deviations of \( \mu''(p, r), \mu''_{\text{CIC}}(p, r) \) respectively for each subject \( p \). Since \( \mu'' \) and \( \mu''_{\text{CIC}} \) are dependent, we let

\[ \mu''_{\text{diff}} = \mu'' - \mu''_{\text{CIC}} \]

to simplify the statistical analysis. For the same reason, we also define \( \sigma''_{\text{diff}}, \varepsilon_{\text{diff}} \) and \( s''_{\text{diff}} \) in an analogous way.

**The Liver Dataset**

We used a subset of 36 patients from a non-invasive liver biopsy study, described fully by Norén et al. (16). The study was approved by the local ethics committee and written informed consent was obtained from all subjects (Reference No. M72-07T5-08). A 1.5 T Philips Achieva MR-scanner (Best, The Netherlands) with sensitivity encoding and CLEAR inhomogeneity correction was used together with a phased-array body coil for image acquisition.

The patients received a bolus injection intravenously containing the liver specific contrast agent Gd-EOB-DTPA (0.025 mmol/kg, 1 mL/s). This was followed by a 30 mL saline flush. Dynamic Contrast Enhanced (DCE) MRI time series were acquired using single-breath-hold symmetrically sampled 2-point Dixon 3D imaging (12) (TE = 2.3 and 4.6 ms, TR = 6.5 ms, flip angle = 13°, breath hold \( \approx 20 \) s). Each image stack was acquired in the axial plane. Water and fat images were obtained by reconstructing the acquired in- and opposite-phase images using the inverse gradient method (13-15). The measurements were performed before (non-enhanced), during arterial and venous portal phase, and after the injection (at 3, 10, 20 and 30 min). The non-enhanced measurements were repeated 2 or 3 times. The measurements from the arterial phase time point
did not have complete liver coverage and were not used further in this work. Water image data $D$ were extracted at seven different liver ROIs, placed per time point in homogeneous liver parenchyma by a radiologist (co-author B.N.) with 27 years experience. Figure 4d shows an example of the positions and the sizes of the seven ROIs. See Norén et al. (16) for more examples of ROI placements.

The intensity values are denoted $D(p, x, y, z, t)$ where $p$ is the patient index, $x, y, z$ are the spatial coordinates and $t$ is the time index. The dataset was reduced by representing each ROI with its average, resulting in $D(p, r, t)$ instead, where $r$ is the ROI index. In summation, for each patient there is a time series, or curve, for each liver ROI. Let $X$ and $Y$ be matrices with all curves from all patients and ROIs, where $X$ and $Y$ denote the data from the standard IIC (CLEAR) and CIIC method respectively. The matrices consequently have six columns and $36 \times 7 = 252$ rows, where each row constitutes one ROI time series. CIIC scales the dataset so that a pure adipose voxel has the intensity of 1.00 in the fat-image. In order to be able to compare $X$ and $Y$ statistically, they must be described using the same global scale. All of the data $X$ were therefore scaled using the single global scalar $s = 2.606 \times 10^{-6}$ which was chosen so that the difference between the mean time series $\mathbf{m}_X$ and $\mathbf{m}_Y$ were minimized in a least square sense:

$$
\min_s \sum_{i=1}^{6} (s \mathbf{m}_X^i - \mathbf{m}_Y^i)^2,
$$

where $i$ denotes the time index, $\mathbf{m}_X = E[X]$ and $\mathbf{m}_Y = E[Y]$ respectively, and $E[\cdot]$ denotes the expectation operator. The expectation values were estimated using the sample means throughout this work. Finally, let $z_{\text{lever}}(p,t), z_{\text{CIIC}}^\text{liver}(p,t)$ be the standard deviations over all liver ROIs, where the ROI-levels have been normalized such that the ROI-median always is 1.00 for each individual patient $p$ and time index $t$. Let also $s_{\text{lever}} = \text{vec}(z_{\text{lever}}), s_{\text{CIIC}}^\text{liver} = \text{vec}(z_{\text{CIIC}}^\text{liver})$ be the collection, or
vectorization, of all these standard deviations. Since $s_{\text{liver}}^\text{liver}$ and $s_{\text{CIIC}}^\text{liver}$ are dependent, we let

$$s_{\text{diff}}^\text{liver} = s_{\text{liver}}^\text{liver} - s_{\text{CIIC}}^\text{liver}$$

to simplify the statistical analysis.

**Statistical Analysis**

The performance evaluation of CIIC was made using the muscle and liver datasets in three steps.

The tested hypotheses were:

1. **CIIC improves the spatial intensity homogeneity within a single image volume:**
   Homogeneous parenchyma should have equal signal intensity distribution within the organ. Hence, the hypothesis was that the spatial dispersion of intensity values a) within muscles and b) inside the liver should decrease after the application of CIIC. In addition, the dispersion of intensity values c) between muscles within the same subject should decrease. For the muscle tests, CIIC was compared to uncorrected data (1.5 T) and to CLEAR corrected data (3.0 T). For the liver test, CIIC was compared to CLEAR corrected data.

2. **CIIC improves the inter-volume intensity homogeneity:** Homogeneous parenchyma should have equal signal intensity distribution between closely spaced repeated examinations. The inter-volume dispersion of intensity values between the same set of muscles should decrease after the application of CIIC. Uncorrected muscle data (1.5 T) as well as CLEAR corrected muscle data (3.0 T) were used for this test. The liver dataset could not be used for testing of this hypothesis because of the intensity changes induced by the contrast agent between the time points.

3. **In the absence of intensity inhomogeneity errors, CIIC will maintain the relative intensity levels on average, except for a global scale constant.** Local differences are due to anticipated improvement of the intensity homogeneity of the data. Thus, the hypothesis
was that these expected differences are generally small compared to the inherent total variation in the data. However, the intensity levels may still be different in regions with large intensity errors, as CIIC will reduce these. Uncorrected data as well as CLEAR corrected data were used for this test.

In order to estimate the actual impact of CIIC, we calculated the Cohen’s $d$ effect size measure (17,18) in parallel to the hypotheses tests where appropriate. The effect size tells us how large the average difference is between two groups in relation to the standard deviation of the groups. It is therefore a measure of the effective magnitude of a change. For multivariate data, we used the Mahalanobis distance $d_m$ (19), which is a non-negative multivariate extension of the univariate Cohen’s $d$ effect size. All the image and signal analysis in this work were done with Matlab R2011b (MathWorks Inc., MA, USA).

*Methods for Hypotheses Tests*

It should be noted that the 1.5 T muscle dataset is ‘raw’ and is not a result from the CLEAR inhomogeneity correction method. The comparisons consequently include both CIIC versus original data, and CIIC versus CLEAR-processed data.

The hypotheses were tested according to Table 2. The significance level was $\alpha=0.01$ for all tests. We performed the Jarque-Bera (20,21) normality test on univariate data and the Henze-Zirkler's test (22,23) on multivariate data. The paired right-tailed t-test was performed for the normal univariate cases, while the nonparametric resampling techniques Bootstrap and Permutation were used otherwise (24-28). $N_{\text{samp}} = 10^6$ samples were used when calculating the bootstrap and null-distribution respectively. For hypothesis 3 for the muscle data, we used Cohen’s $d$ effect size measure for paired data, see (17,18):


\[ d_{\text{diff}}^m = \frac{E\left[\mu_{\text{diff}}^m\right]}{SD\left[\mu_{\text{diff}}^m\right]} = \frac{E\left[\mu^m - \mu_{\text{CIIC}}^m\right]}{SD\left[\mu^m - \mu_{\text{CIIC}}^m\right]} \]  

Let \( d_{\text{diff*}}^{3.0T} \) denote \( d_{\text{diff}}^{3.0T} \) but with the “Lower-Leg” ROI excluded, as this ROI was identified as a large intensity inhomogeneity for the \( D^{3.0T} \) dataset.

For hypothesis 3 for the multivariate liver data, we used the Mahalanobis distance \( d_m(X_{t:2}^{\text{PCA}}, Y_{t:2}^{\text{PCA}}) \) (19). A principal component analysis (PCA) (29) was performed on the signals \( X \) and \( Y \) in order to avoid differences due to noise in the Mahalanobis effect size calculation. The first two principal components contributed to more than 96% of all signal variance, as shown in Table 3. The noise level in the non-enhanced images was estimated to 0.96% of all signal variance by comparing this time point with additional, independent measurements of these images. Assuming that all time points are affected by approximately the same level of noise, the estimated total noise level for all time points was 5.76%. For this reason, since the noise was assumed to be uncorrelated with the main physiological signal, only the first two principal components \( X_{t:2}^{\text{PCA}} \) and \( Y_{t:2}^{\text{PCA}} \) were included in the subsequent analysis.

A qualitative inspection of the CIIC data was also performed by two radiologists (co-authors Ö.S, B.N) and one image processing researcher (author T.A.) with 30, 27 and 14 years of experience respectively. Side-by-side image comparisons with/without CIIC (random order) were made for the muscle data sets (1.5 T, 3.0 T) and for two time points (pre, 20 min) for the liver data set. The quality of the right image compared to the left was judged in the three categories Number of Artifacts (Q1), General Appearance (Q2) and Intensity Homogeneity (Q3). Artifacts were in this context defined as missing data, image discontinuities or distorted data that clearly do not represent the tissue structures. The available quality scores were +1 (right better), 0 (equal) and -1 (left better).
The collected quality data were decoded before the statistical analysis so that +1 imply better quality with CIIC. Let $q^m$ be the collection of all the quality data where $m \in \{Q1, Q2, Q3\}$.

RESULTS

The Muscle Dataset

Figures 2 and 3 show two representative examples of the muscle dataset, including the positions of the different ROIs and the dispersion of intensity values before and after CIIC.

Hypothesis 1a, Decreased Intra-Muscle Dispersion

The 99% lower confidence bounds for the average differences in intra-muscle dispersion $E[\sigma_{\text{diff}}^m]$ were $E[\sigma_{\text{diff}}^{1.5T}] > 5.3 \times 10^{-3}$ and $E[\sigma_{\text{diff}}^{3.0T}] > 9.3 \times 10^{-2}$ for 1.5 T and 3.0 T respectively. The observed sample estimates were $\sigma_{\text{Obs \ diff}}^{1.5T} = 1.6 \times 10^{-2}$, $\sigma_{\text{Obs \ diff}}^{3.0T} = 1.1 \times 10^{-1}$ and the corresponding effect sizes were $d_{\text{Obs \ diff}}^{1.5T} = 0.49$ and $d_{\text{Obs \ diff}}^{3.0T} = 1.3$ standard deviations. The significances were calculated to $P_{\text{diff}}^{1.5T} = 2.3 \times 10^{-4}$ and $P_{\text{diff}}^{3.0T} = 1.0 \times 10^{-6}$. We can therefore reject the null hypothesis $H0$ in favor of $H1$ for both 1.5 T and 3.0 T at the significance level $\alpha=0.01$; CIIC decreased the intra-muscle dispersion.

Hypothesis 1c, Decreased Inter-Muscle Dispersion

The test for the inter-muscle dispersion hypothesis $H0$: $E[s_{\text{diff}}^m] = 0$ resulted in the significances $P_{\text{diff}}^{1.5T} = 6.9 \times 10^{-6}$ and $P_{\text{diff}}^{3.0T} = 1.1 \times 10^{-7}$ respectively. The corresponding 99% lower confidence bounds were $E[s_{\text{diff}}^{1.5T}] > 3.5 \times 10^{-2}$ and $E[s_{\text{diff}}^{3.0T}] > 2.6 \times 10^{-1}$. We can therefore reject the null hypothesis $H0$ in favor of $H1$ for both 1.5 T and 3.0 T at the significance level $\alpha=0.01$; CIIC
decreased the inter-muscle dispersion. The effect sizes were $d^{1.5T}_{\text{obs}} = 3.0$ and $d^{3.0T}_{\text{obs}} = 4.7$ standard deviations.

**Hypothesis 2, Decreased Inter-Volume Dispersion**

The 99% lower confidence bound for the average difference in inter-volume dispersion $E[\varepsilon_{\text{diff}}]$ was $E[\varepsilon_{\text{diff}}] > 7.9 \times 10^{-2}$. The permutation test resulted in a calculated significance of $P_{\varepsilon_{\text{diff}}} = 1.0 \times 10^{-6}$, based on the observed sample estimate of $\varepsilon_{\text{obs, diff}} = 1.5 \times 10^{-1}$. We can therefore reject the null hypothesis $H_0$ in favor of $H_1$ at the significance level $\alpha=0.01$; CIIC decreased the inter-volume dispersion. The effect size was $d_{\text{obs}} = 0.44$ standard deviations. $d^{\text{CIIC}}_{\text{obs}} = 5.0$ standard deviations if we instead used the standard deviation of CIIC in the effect size calculations.

**Hypothesis 3, Preserved Intensity on Average**

The bootstrap 99% confidence intervals for the Cohen effect sizes were $-0.19 \leq d^{1.5T}_{\text{diff}} \leq 0.21$ and $-0.11 \leq d^{3.0T}_{\text{diff}} \leq 0.35$. The permutation test resulted in a calculated significance of $P^{1.5T}_{d-\text{diff}} = 0.91$ and $P^{3.0T}_{d-\text{diff}} = 0.28$, based on the observed effect sizes of $d^{1.5T}_{\text{obs, diff}} = 0.01$ and $d^{3.0T}_{\text{obs, diff}} = 0.12$ standard deviations. We can therefore not reject the null hypothesis $H_0$: $d^{m}_{\text{diff}} = 0$ at the significance level $\alpha=0.01$, for either 1.5 T or 3.0 T.

**The Liver Dataset**

The correlation matrix $R_X$ showed high correlations between the time indices. Let $r_X$ be the vector with all the cross-correlations from $R_X$. The median cross-correlation was then $\tilde{r}_X = 0.86$ and the mean cross-correlation was $\bar{r}_X = 0.86$ with a standard deviation of $\sigma_{r_x} = 0.10$. The results of the PCA of $X$ are shown in Table 3: $X_{1:2}^{\text{PCA}}$ represents more than 96% of the signal variation in $X$. 
**Hypothesis 1b, Decreased Intra-Liver Dispersion**

The test for the intra-liver dispersion hypothesis $H_0$: $E[\hat{s}_{\text{liver}}^{\text{diff}}] = 0$ resulted in the significance $P_{\text{diff}} = 1.1 \times 10^{-15}$. The corresponding 99% lower confidence bound was $E[\hat{s}_{\text{liver}}^{\text{diff}}] > 1.3 \times 10^{-2}$. We can therefore reject the null hypothesis $H_0$ in favor of $H_1$ at the significance level $\alpha=0.01$; the liver ROI intensity homogeneity was better with CIIC compared to standard scaling. The effect size of the difference of the estimated standard deviations was $d_{\text{liver}}^{\text{obs}} = 0.44$ standard deviations.

**Hypothesis 3, Preserved Intensity on Average**

The observed Mahalanobis distance was:

$$d_{M}^{\text{obs}}(X_{t:2}^{\text{PCA}}, Y_{t:2}^{\text{PCA}}) = 0.15 \text{ standard deviations} \quad [4]$$

The probability of measuring at least the observed effect size was calculated to $P_M = 0.035$ under the null hypothesis, and the hypothesis that $d_{M}(X_{t:2}^{\text{PCA}}, Y_{t:2}^{\text{PCA}}) = 0$ could consequently not be rejected at the significance level $\alpha=0.01$.

**Qualitative Inspection**

The bootstrap 95% lower bounds for the mean of the quality data were $\bar{q}^{Q1} \geq 0.0036$, $\bar{q}^{Q2} \geq 0.11$ and $\bar{q}^{Q1} \geq 0.41$. The significances of the corresponding tests were $P^{Q1} = 0.024$, $P^{Q2} \leq 10^{-6}$ and $P^{Q1} \leq 10^{-6}$. The visual comparison of image slices before and after CIIC therefore showed that CIIC did not introduce any artifacts, nor did it degrade the appearance. The intensity variations over regions with homogeneous tissue were also smaller after CIIC. Two representative example image pairs from the muscle dataset are shown in Figure 2 (3.0 T) and in Figure 3 (1.5 T), together with the muscle ROI masks. One liver image pair is shown in Figure 4 together with intensity profiles, illustrating reduced intensity variation across the liver after the application of CIIC.
Moreover, Figure 4 also shows the positions of the seven liver ROIs, orthogonally projected into the shown image plane. In Figure 5 the mean liver time series with both standard and CIIC scalings are shown together with their standard deviations. The mean differences were qualitatively small and the univariate standard deviation was smaller for CIIC.

**DISCUSSION**

It has previously been shown that CIIC successfully produces a bias field suitable for fat quantification applications (7-10). This implies that CIIC, in contrast to many other retrospective inhomogeneity correction methods, produces a physically relevant bias field. The CIIC method uses adipose tissue as an internal reference in order to correct the intensity inhomogeneities. This makes it self-contained and also provides the physical relevancy. In fact, if T2* signal effects are properly corrected and T1 signal saturation effects are avoided by using low flip angles, long repetition times, or are corrected, the bias field represents the unsaturated proton density signal in pure adipose tissue. This is valid not only for fat quantification (7,30), but also for quantification of water concentration and total proton density.

The most important result in this work is that CIIC also produces reasonable bias fields in *water* image regions where the reference adipose signal is weak or absent. This can be seen in the results as CIIC, in support of Hypothesis 1 and 2, significantly improves the intensity consistency in regions of a homogeneous tissue type. The statistical analysis shows this both locally (Hypothesis 1ab), globally (Hypothesis 1c) and for repeated examinations with a different scanner (Hypothesis 2). The inter-volume and spatial dispersions were decreased by CIIC with effect sizes ranging from 0.44 up to 4.7 standard deviations in the different tests. Commonly used definitions of the effect size magnitudes are “small” as \( d \approx 0.2 \), “medium” as \( d \approx 0.5 \) and “large” as \( d \approx 0.8 \) (17). The decreased dispersions are helpful for any method that depends directly or indirectly on the intensity
values or their gradients, e.g. lipid and water signal quantification, registration and segmentation. The decreased dispersions may also be very useful in applications such as DCE, where a certain contrast agent concentration ideally should be represented by the same intensity value throughout the DCE time series, as well as in all positions in the image. For example, an important target application of DCE MRI is the construction of parametric, functional maps of e.g. blood flow. The enforcement of intensity homogeneity in the underlying images would potentially reduce artificial variations in these maps. The inter-volume, or temporal, intensity inhomogeneity correction is an inherent feature of CIIC as it independently recreates the scaling information for each time point. The results also support Hypothesis 3 that states that ROI intensity variance reduction is performed while preserving the average intensity levels and appearances of the data and time series, excluding large intensity errors that should be corrected. This suggests that CIIC behaves well and also illustrates an important generalization property, as the evaluation in this work was performed on water volumes in regions where the reference adipose-signal, from the corresponding fat volumes, is weak or non-existent. The quantitative results show that we cannot reject the hypothesis of equal ROI-means on average for the muscle dataset, excluding a manually identified large intensity inhomogeneity in the lower leg for the 3.0 T data that was corrected. In addition, the hypothesis of equal ROI time series on average for the liver dataset could also not be rejected. The observed effect size between muscle ROIs with raw/CLEAR and CIIC scaling were 0.01 and 0.12 standard deviations for 1.5 T and 3.0 T respectively. For the liver ROI time series, the observed effect size between standard and CIIC scaling was 0.15 standard deviations. Altogether, this means that the average intensity level changes caused by CIIC are not statistically significant and, in relation to the normal variations within the datasets, indeed small in magnitude. This is important, as CIIC should improve scaling consistency while keeping the average intensity level, and shape of the
average time series, intact in the absence of large intensity errors. The statistical comparisons in this work include both CIIC versus original data and CIIC versus CLEAR-processed data, giving similar results and showing the robustness of CIIC to scale changes in its input data. CIIC should give comparable results for the corresponding implementations of other manufacturers, e.g. PURE (GE Healthcare) and Prescan Normalize (Siemens), as they are all using coil sensitivity maps and CIIC is independent of any specific implementation. While this study shows that CIIC reduces the inhomogeneities both with and without prospective calibration based on field sensitivity maps, the exact gain compared to different manufacturers’ implementations of such methods has not been investigated.

The physical relevancy of the bias field leads to intensity values that are consequently physically meaningful, not only relatively, but also in their absolute magnitudes. Previous studies have, as we have seen, successfully used the intensity values for quantitative imaging applications. Dahlqvist Leinhard et al. use CIIC for quantification of abdominal fat (7), Lidell et al. for quantification and identification of brown adipose tissue (9), Ludwig et al. for whole-body fat quantification (10), Karlsson et al. for muscle volume quantification (11), and Norén et al. for quantification of the hepatobiliary uptake of the Gd-EOB-DTPA contrast agent (16). Progress towards successful automatic quantification of body functions and morphology is clinically important, as it would decrease the need for invasive procedures. In order to assure accurate quantifications however, the intensity level interpretations need to acknowledge factors such as $R_2^*$-relaxation, $T_1$-saturation and fatty infiltration. It should be pointed out that the present study is limited to evaluate the necessary first step towards this – the intensity inhomogeneity correction – and has not evaluated the physical interpretations of the resulting intensity values. A general problem in chemical shift based fat and water separation techniques is the occurrence of fat/water swaps. Such swaps will
cause a local non-physical effect on the bias field in the affected region. However, no global effects of swaps affecting the bias field in the regions of interest were observed in this study. It should also be noted that there must exist observable pure adipose tissue in order for CIIC to detect an inhomogeneity, and to track fast inhomogeneity changes. Proper requirements on image volume resolution and/or field of view are therefore necessary to facilitate good coverage of pure adipose voxels, and the limits of these requirements have not been investigated in this study. One important question is how the performance of CIIC changes in presence of variations in MR-parameters, e.g. the flip angle. More experiments are needed in order to quantify and model this dependence. Future work should also include experiments on a larger population, including older patients, in order to better evaluate the performance of CIIC on different demographics.

In conclusion, the results show that CIIC improves the spatiotemporal intensity consistency while preserving the intensity levels on average. The large significant decrease in non-physiological variance of water image intensities is particularly useful, considering that the CIIC method only uses image information from the adipose tissue surrounding the corresponding region. These results show that CIIC is self-calibrating in the sense that it can recreate standardized, physically meaningful and homogeneous global scaling information directly from the data. This implies that CIIC can successfully be used as a regular scaling method in chemical shift based water and fat separated MRI, with a significant potential for providing both robust and consistent results in MR image analysis and in quantitative imaging applications.
ACKNOWLEDGMENTS

Financial support from the Swedish Research Council (VR/M 2007–2884), the Research Council of Southeast Sweden (FORSS 12621), Linköping University, Lions Research Foundation in Linköping, Linköping University Hospital Research Foundations and the County Council of Östergötland is gratefully acknowledged.
REFERENCES


### TABLES

**Table 1:** Demographic details of subjects for the muscle dataset.

<table>
<thead>
<tr>
<th></th>
<th>Mean ± SD</th>
<th>Range</th>
</tr>
</thead>
<tbody>
<tr>
<td>Age (years)</td>
<td>24.9 ± 2.4</td>
<td>21-29</td>
</tr>
<tr>
<td>Weight (kg)</td>
<td>67.9 ± 11.1</td>
<td>57-90</td>
</tr>
<tr>
<td>BMI (kg/m²)</td>
<td>23.6 ± 3.8</td>
<td>20.1-32.3</td>
</tr>
</tbody>
</table>
**Table 2:** The statistical analysis for each hypothesis

<table>
<thead>
<tr>
<th>Dataset, Case</th>
<th>H0</th>
<th>H1</th>
<th>Normality Assumption</th>
<th>Hypothesis test</th>
</tr>
</thead>
<tbody>
<tr>
<td>Muscle, 1a</td>
<td>$E\left[\sigma_{diff}^m\right] = 0$</td>
<td>$E\left[\sigma_{diff}^m\right] &gt; 0$</td>
<td>Reject $P^{1.5T} = 0.0013$, $P^{3.0T} = 0.0027$</td>
<td>Resampling</td>
</tr>
<tr>
<td>Muscle, 1c</td>
<td>$E\left[s_{diff}^m\right] = 0$</td>
<td>$E\left[s_{diff}^m\right] &gt; 0$</td>
<td>No Reject $P^{1.5T} = 0.084$, $P^{3.0T} = 0.112$</td>
<td>t-test (paired right-tailed)</td>
</tr>
<tr>
<td>Muscle, 2</td>
<td>$E\left[q_{diff}\right] = 0$</td>
<td>$E\left[q_{diff}\right] &gt; 0$</td>
<td>Reject ($P = 2.4 \times 10^{-7}$)</td>
<td>Resampling</td>
</tr>
<tr>
<td>Muscle, 3</td>
<td>$d_{diff}^m = 0$</td>
<td>$d_{diff}^m \neq 0$</td>
<td>Reject $P^{1.5T} = 2.9 \times 10^{-4}$, $P^{3.0T} = 8.1 \times 10^{-5}$</td>
<td>Resampling</td>
</tr>
<tr>
<td>Liver, 1b</td>
<td>$E\left[s_{diff}^{liver}\right] = 0$</td>
<td>$E\left[s_{diff}^{liver}\right] &gt; 0$</td>
<td>No Reject ($P = 0.41$)</td>
<td>t-test (paired right-tailed)</td>
</tr>
<tr>
<td>Liver, 3</td>
<td>$d_M = 0$</td>
<td>$d_M &gt; 0$</td>
<td>Reject $P_X = 7.1 \times 10^{-4}$, $P_Y = 2.3 \times 10^{-4}$</td>
<td>Resampling</td>
</tr>
<tr>
<td>Qualitative</td>
<td>$\bar{q}^m = 0$</td>
<td>$\bar{q}^m &gt; 0$</td>
<td>Ordinal data</td>
<td>Resampling</td>
</tr>
</tbody>
</table>
Table 3: Results of the principal components analysis (PCA) of X

<table>
<thead>
<tr>
<th>Principal Component #</th>
<th>Variance (Eigenvalue)</th>
<th>Proportion of Total Variance (%)</th>
<th>Cumulative Sum of Proportions of Total Variance (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>0.19</td>
<td>91.4</td>
<td>91.4</td>
</tr>
<tr>
<td>2</td>
<td>0.011</td>
<td>5.1</td>
<td>96.5</td>
</tr>
<tr>
<td>3</td>
<td>0.0031</td>
<td>1.5</td>
<td>97.9</td>
</tr>
<tr>
<td>4</td>
<td>0.0019</td>
<td>0.9</td>
<td>98.8</td>
</tr>
<tr>
<td>5</td>
<td>0.0015</td>
<td>0.7</td>
<td>99.5</td>
</tr>
<tr>
<td>6</td>
<td>0.0010</td>
<td>0.5</td>
<td>100.0</td>
</tr>
</tbody>
</table>
Figures

**Figure 1**: The probability density functions (pdfs) of the muscle datasets \( D^{1.5T} \), \( D^{3.0T} \), \( D^{1.5T}_{\text{CIIC}} \) and \( D^{3.0T}_{\text{CIIC}} \). They are scaled so all share the same mode (maximum peak): \( \text{Mo} \left[ D^m \right] = \text{Mo} \left[ D^{m\text{CIIC}} \right] = 1 \).
Figure 2: Example of a qualitative comparison between water image slices from the whole-body 3.0 T (muscle) dataset, with standard IIC (CLEAR) (b) and CIIC scaling (c) respectively. The arrows mark transition areas between different overlapping image stacks. Pseudo-colors have been used in order to increase the visibility of the differences. (a): The ROI positions in the current image slice. Colors here indicate different ROI indices.
Figure 3: Example of a qualitative comparison between water image slices from the whole-body 1.5 T (muscle) dataset, without intensity inhomogeneity correction (b) and CIIC scaling (c) respectively. The arrows mark an intensity inhomogeneity caused by differences in T1-saturation due to flip angle variations. Pseudo-colors have been used in order to increase the visibility of the differences. (a): The ROI positions in the current image slice. Colors here indicate different ROI indices.
Figure 4: Example of a qualitative comparison between water image slices from the 1.5 T liver dataset, with standard IIC (CLEAR) (a) and CIIC scaling (b) respectively. (c): Intensity profiles along the marked lines. (d): The ROI positions, projected on the current image slice.
Figure 5: A qualitative comparison between the mean time series of the 1.5 T liver (water) dataset, with standard ICC (CLEAR) and CIIC scaling respectively. The standard deviations are also shown.