

Multi-Center Correction Functions for Magnetization Transfer Ratios of MRI Scans

Arpad Kelemen¹ and Yulan Liang^{2*}

¹Department of Organizational Systems and Adult Health, University of Maryland, Baltimore, Maryland, USA

²Department of Family and Community Health, University of Maryland, Baltimore, Maryland, USA

Abstract

Persistent differences among images obtained with different MRI scanners or different MRI pulse sequences pose a major challenge worldwide in image studies, particularly for clinical applications and clinical trials. One potential solution would be to obtain correction functions, which can be used to correct voxel based Magnetization Transfer Ratios (MTR) obtained at individual MRI scanners in order to estimate what they would have been if the scans had been taken at different MRI scanners. However, with millions of voxels in each MRI scan one must navigate the challenge of high dimensionality in order to obtain such correction functions. In this article, we propose a novel computational approach for obtaining correction functions to standardize MRI data resulting from different scanners, which includes a two-stage procedure: 1) the construction of spatial voxel co-occurrence matrices for dimension and data reduction of voxel based data; 2) Generalized Linear Models (GLMs) including nonlinear models for estimating correction functions between MRI scanners. A working example is given for illustration. Accurate estimations of important MRI statistics and constructions of MTR correction functions between scanners/pulse sequences were achieved from a small sample size of healthy subjects with large number of voxel measures per MRI scans. Estimation of voxel based MTR values using correction functions can overcome comparison problems between scanner/pulse sequence differences in multicenter studies. This may have important application for future clinical trials using MTR as an endpoint.

Keywords: Magnetization transfer ratio; Threshold method; Spatial voxel co-occurrence matrix; Correction function; Magnetic resonance imaging; Multi-center studies

Introduction

Persistent differences among Magnetic Resonance Imaging (MRI) obtained with different MRI scanners or different pulse sequences is currently a major challenge worldwide, where large scale clinical trials, cross-sectional studies, or longitudinal studies are being conducted [1,2]. Routine analysis of thousands of brain and spinal cord MRI images is being conducted worldwide which were taken with various MRI scanners, pulse sequences, magnetic fields, and time points. Artificial objects, called “phantoms”, are widely being used to periodically calibrate and tune individual MRI scanners in order to minimize variation of MRI scans obtained by the same scanners at different time points. This also reduces the variation of the MRI scan outputs obtained with different scanners. When an MRI scan is ordered in real life, it is possible that there are only a few MRI scanners available at the given area, hence the choice of MRI settings may be limited. Although this is not likely to lead to a misdiagnosis that is based on a single MRI scan, it has a significant quantitative effect that impacts clinical measurement and may impair clinical decision making that is based on longitudinal MRI scans, which often contain scans from more than one scanner or setting. For the purpose of large scale cross-sectional and longitudinal data analysis, clinical trials, and following long term patient progression of specific neurological disorders, it is important to have MRI images which are taken under identical, or at least similar, conditions. Therefore, the existence of persistent, unwanted variation among scanners is a significant problem for which correction functions could be the solution. The key question in this article is “Can we obtain optimal predictive correction functions, which can be used to correct voxel based Magnetization Transfer Ratios (MTR) obtained at individual MRI scanners/pulse sequences in order to estimate what they would have been if the scans had been taken at different MRI scanners/pulse sequences”?

Several approaches have been investigated for reducing inter-scanner variability and bias. Studies were performed for reproducibility issues and the implications of inter-scanner variability for clinical trials [3-10]. For instance, the Functional Imaging Biomedical Informatics Research Network (FBIRN) Consortium [7] investigated the reduction of the inter-scanner variability of activation in a multicenter MRI study by controlling for signal-to-fluctuation-noise-ratio differences. Several studies have identified the sources of MTR variations [11,12] and provided various correction schemes including correction for inter-scanner variations. However, there is no existing standard solution for standardizing multiple scanner MRI data and for obtaining predictive correction functions resulting from multiple MRI scanners. Moreover, there are a number of challenges when one attempts to achieve such goals.

First, it is not easy to recruit human subjects to undergo a series of, say, 5 MRI scans within a single 4 weeks period for pure research purposes. To make the matter worse, it is typically even more difficult to recruit subjects with known brain artifacts (such as lesions, black holes [13]) than to recruit healthy subjects due to the unwillingness of human subjects, the various legal procedures involved, and the expenses. In addition, the use of multiple (e.g. five) MRI scanners is a huge barrier. Therefore, only a small sample size is currently available for such research study.

***Corresponding author:** Yulan Liang, PhD, Department of Family and Community Health, University of Maryland, 655 West Lombard Street, Room 404K, Baltimore, Maryland 21201, USA, Tel: 410-706-4812; Fax: 410-706-0780; E-mail: ylian001@umaryland.edu

Received February 27, 2012; **Accepted** July 07, 2012; **Published** July 12, 2012

Citation: Kelemen A, Liang Y (2012) Multi-Center Correction Functions for Magnetization Transfer Ratios of MRI Scans. J Health Med Informat S3. doi:10.4172/2157-7420.S3-002

Copyright: © 2012 Kelemen A, et al. This is an open-access article distributed under the terms of the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

Second, a typical MRI scan results in about 4.6 million voxels, each of which holding information about a very tiny space which is either part of the brain or not. Each voxel has a large number of readings and some of the readings are aggregated into voxel based statistics. One of the important and commonly used voxel based statistics in clinical applications is the Magnetization Transfer Ratio [14]. In order to address the problem of inter scanner variability and standardize MTR data from different MRI scanners or pulse sequences, dimension reduction becomes an intermediate step.

In this paper, we present a general purpose two-stage methodology and a series of steps to correct the data and find functions to correct when using different scanners, which can be applied to any MRI scanner. In stage I we design an algorithm to construct Spatial Voxel Co-occurrence Matrices (SVCMs) for dimension reduction given millions of voxel based MTRs per scan [14]. In stage II we implement Generalized Linear Regression models including nonlinear models for estimating the correction functions between different scanners and for modeling the functional relationships and variations among scanners. Furthermore, we aggregate the multiple correction functions resulting from stage II to obtain an optimal correction function for better predictive ability and generalization performance. Besides the above approach, we design noise and outlier elimination based on threshold methods for data quality control, image acquisition, and exploratory analysis for MRI data preprocessing. These are important steps toward the success of the proposed approach.

Materials and Methods

Multiple MRI image acquisition

Five healthy subjects (4 males, 1 female, and age range 20-40 years) were recruited for the multiple MRI scanner correction study. All subjects provided informed consent for this study and underwent a series of 5 MRI scans made with 4 available MRI scanners (A, B, C, D; D with 2 different pulse sequences: 600 and 1200) at the given location using proton density and T1 gradient-echo sequence within a 4 weeks period. Technical specifications of the MRI scanners varied, and were as follows: Single/Multi Slice: 0.05sec/image; TR: SE: 30-10,000 msec GE: 20-10,000 msec IR: 50-16,700 msec FSE: 200-16,7000 msec; TE: SE: 10-250 msec IR: 10 -250 msec GE: 5-50 msec FSE: 15-2,000; FOV: 300; BW: 10 & 40 KHz; flip angle 45°; pulse sequences: 600 and 1200; Parallel Imaging Techniques: Sensitivity Encoding (SENSE), Simultaneous Acquisition of Spatial Harmonics (SMASH), Array Spatial Sensitivity Encoding Technique (ASSET).

During such a short time, it is believed that neither the subjects nor the scanners underwent significant changes. All images with and without MT contrast were registered (transformed different sets of data into one coordinate system) prior to computation of the MTR [15,16]. Therefore, the resulting scans can be compared to one another without introducing additional biases. Each scan was obtained under 1.5 T magnetic field and the scans were all Spoiled-Gradient Recalled (SPGR) type. One subject dropped out after providing three scans.

Therefore, a total of 23 scans were obtained. They were independently evaluated by 2 neurologists to assure scan quality and to make sure the scans did not contain unexpected artifacts, unwanted anomalies, and to confirm that the scans were taken under “similar” conditions, such as scanner settings, coil, voxel sizes, slices, alignments. All of the 23 scans passed these evaluations and they were used for model building

and analysis. Standard protocols were performed and various MRI measures and statistics were obtained in order to summarize the various properties of each voxel into voxel based Magnetization Transfer Ratio (MTR) values [14]. Magnetization Transfer (MT) refers to the transfer of longitudinal magnetization from free water protons to hydration water protons in MRI. MTR is used to highlight abnormalities in brain structures and is defined as $MTR = (Mo - Mt) / Mo$, where Mo , Mt represent the 2 different MT measures at the same region of a brain MRI at 2 different time points. According to existing imaging protocol, MTR values were rounded to the nearest integers. MTR is one of the commonly used MRI statistics in clinical applications and bears high significance for various neurological disorders.

MTR values obtained through a single scan form an “MTR map”, about which we want to find out how they change from scanner to scanner for individual subjects. Each image file contained a sequence of 70 slices, centered on the brain (from the top to the bottom of the head) after application of the “Brain Extraction Tool” deskulling process, which replaced skull related voxel values by 0’s [17]. Each slice contained 256 voxels per row and 256 voxels per column. Therefore, there were a total of 4,587,520 voxels per MTR map. Individual voxels correspond to a uniform 0.9375 mm × 0.9375 mm × 2.5 mm 3D space. Each scanner uniformly met these specifications. There were a total of 23 scans and therefore we had 23 MTR maps. All pairs of MTR maps were organized and co-registered by using the JIM 3.0 software [18] in such a way that the orders of the voxels were the same in the pairs and they each corresponded to the same physical space of the same brain. Note that during image acquisition there are small variations in the orientation of the head and the brain. The JIM 3.0 software corrected these orientation variations by relying on anatomical features, such as the shape of the skull and rotated images to a common orientation. These “co-registered” images were then used by our method reported in this paper. Figure 1 shows the 5 MTR maps of one of the subjects

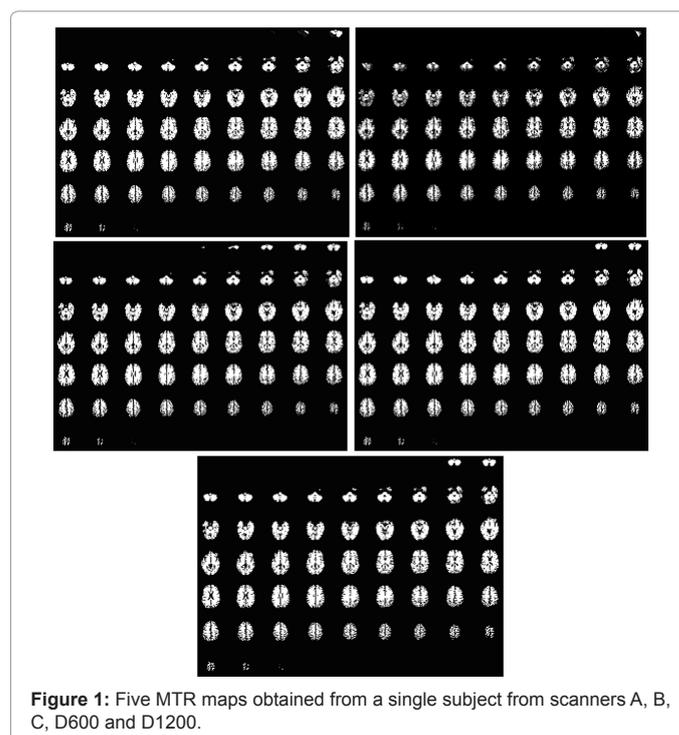
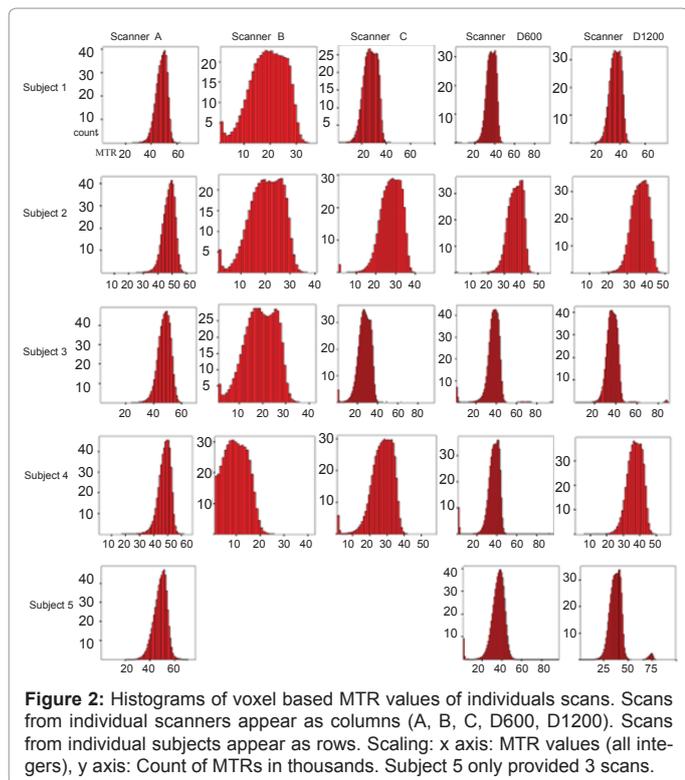


Figure 1: Five MTR maps obtained from a single subject from scanners A, B, C, D600 and D1200.



obtained from 5 MRI scanners/pulse sequences after co-registration and deskulling with JIM. Although we may expect the images to be nearly identical, there are visible differences among the images, which indicate that correction functions are necessary to standardize MRI data for clinical applications.

Data preprocessing, threshold method for noise and outlier elimination

Using the applied standard image processing protocol, individual voxel based MTR values were captured on a 0-100 scale (integers only), where value 0 (black color) is considered to be non-brain area and higher values signaling higher intensity. The 23 image files (MTR maps) were converted to text files that contained only numeric values, then were all merged into a single file, which contained 23 columns (scans), and 4,587,520 rows (voxel vectors). All scans had their voxels in the same order. Next, columns were merged one after another for creating one column per scanner/pulse sequence by listing all scanner A MTR values first, then all scanner B values, followed by all scanner C, D600, and D120 values respectively. This was done in order to determine scanner/pulse sequence specific distributions and the total amount of noise and outliers in the data for elimination. We had 4 or 5 scans for each scanner/pulse sequence and we created 5 such merged columns, one for each scanner/pulse sequence. These columns contained all MTR values for individual scanners/pulse sequences and the order of the MTR values was preserved.

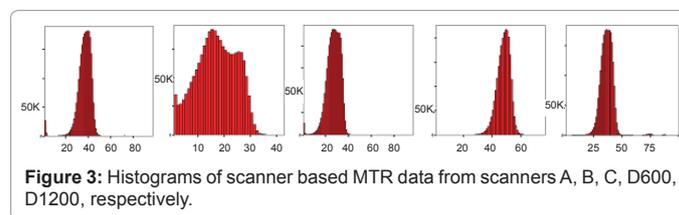
Individual histograms were created for each of the 23 columns (scans) and also for the 5 merged columns (scanners/pulse sequences) containing the MTR values. Scan based histograms are given in Figure 2. Histograms obtained from individual scanners/pulse sequences appear as columns. Each histogram's x axis was scaled to display the actual observed range of MTRs, but do not display MTR = 0 values,

due to those are being considered non-brain areas. The following can be seen in these histograms: scans of different individuals (all taken from healthy subjects) taken on the same scanner and pulse sequence are roughly similar to one another, with respect to their shapes, distributions, and ranges. On the other hand, scans taken with different scanners or pulse sequences are remarkably different from one another, even for the same subjects. This indicates that inter-individual variability for measurements on a single scanner is lower than inter-scanner variability. This further suggests that correction functions among scanners are important. Moreover, there are some outliers and noise, especially near 0 on several scans. These near 0 intensities are likely to be biological or technical noise and probably correspond to non-brain volumes or brain boundaries, which should be eliminated before further analysis. Also, there are some small "spikes" on the right side of some histograms. These might be due to technical error or perhaps the subject moved during the scan. However, all the 23 scans were admissible at this point and we kept them for further analysis.

Figure 3 shows the 5 histograms that were created for the scanner/pulse sequence based (merged) MTR values and their overall MTR value distributions for studied scanners/pulse sequences. As it can be seen on these histograms, for scanners A, B, C there are "unwanted spikes" near zero. Again, these are likely to be biological or technical noise, and are likely to show non-brain material or background, which should be eliminated. For scanner (with pulse sequence) D1200 there are long (but low) tails both on the left and right.

After empirical evaluation of the histograms in Figure 3 and the corresponding frequency statistics, a decision was made together with an image expert that a uniform, symmetric 6.5% of MTR values are to be eliminated on both sides. This decision is usually straightforward, and the goal is to eliminate spikes and long tails on both the left and right sides of the histograms. If the decision is not easy to make then a larger percentage is to be chosen, as eliminating some good data is better than keeping some "noisy" data. Note, that for comparison we have also chosen 3.5% and 9.5% for our presented data, and the impact of this choice was low at the end. Different image sets may result in different percentages, but based on the hundreds of images which we have worked with (taken over 10 years from about 20 scanners) we recommend that 6.5% can be used as a default value for amount to eliminate.

Once this single percentage is determined (6.5%), the "valid ranges



Scanner	Valid MTR interval	% trimmed on left	% trimmed on right
A	25-44	6.6%	5%
B	4-28	6.7%	6.2%
C	17-35	6.4%	6.5%
D600	42-54	7.5%	6.1%
D1200	30-44	5.7%	7.1%

Table 1: Valid ranges of voxel based MTRs and actual amounts of trims for scanners. Percentages were chosen to be as close to 6.5% as possible subject to the discrete nature of MTRs.

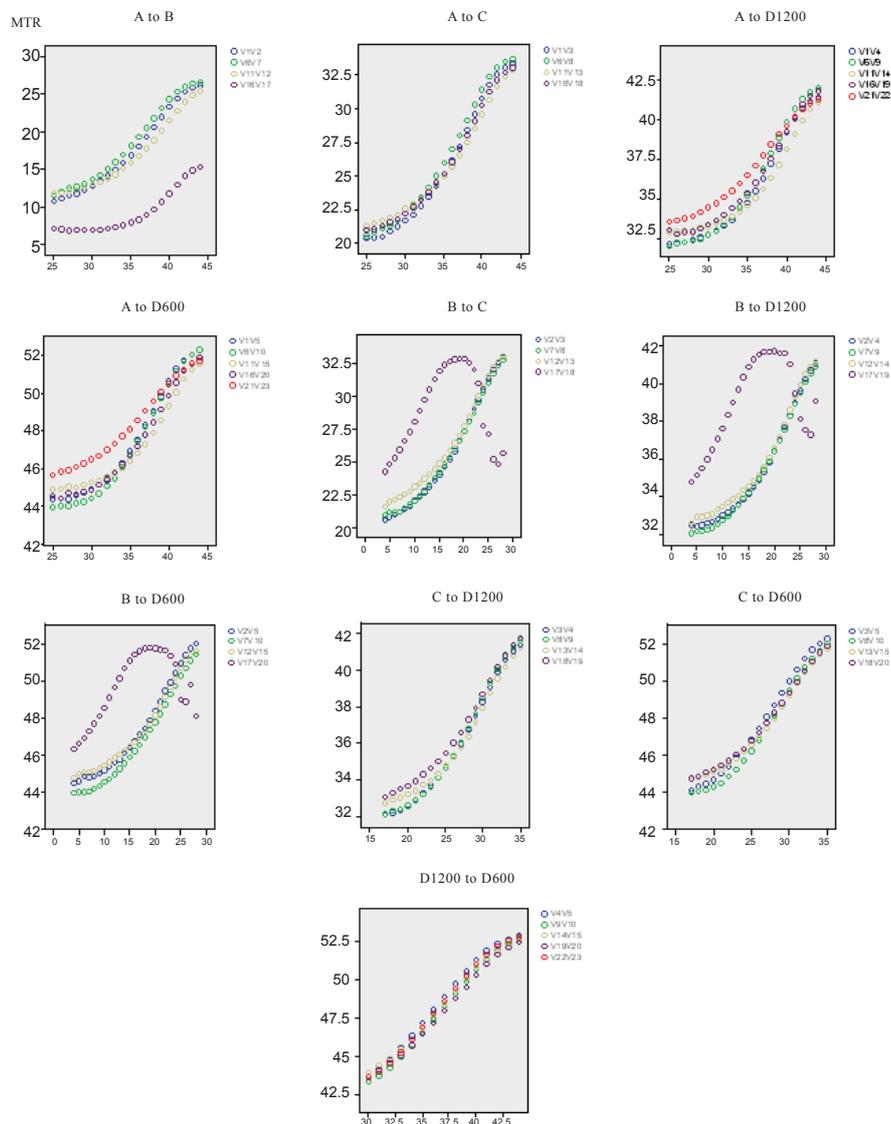


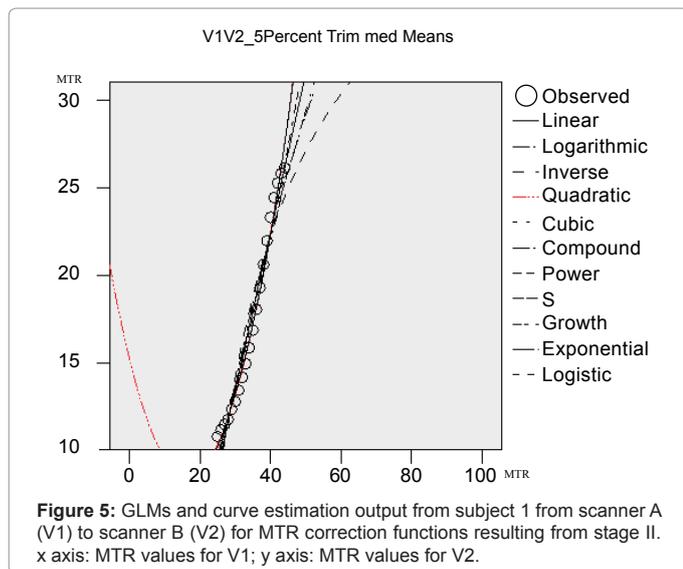
Figure 4: Overlay scatterplots to compare pairs of scanners. They are the following in line sequential order: A to B, A to C, A to D1200, A to D600, B to C, B to D1200, B to D600, C to D1200, C to D600, D1200 to D600. Scan pairs are shown on the right. x axis: MTR values for scan 1; y axis: MTR values for scan 2 (in the pair).

of MTR values” were determined for each scanner/pulse sequence based on the pooled data. Since the values in each scan were on a discrete scale of (0, 1000 (0 was already considered non-brain area), it was not possible to cut exactly 6.5%. We selected thresholds to be as close to the desired 6.5% cutoffs as possible subject to the discrete nature of MTRs: any value below the selected lower threshold or above the upper threshold was eliminated. Table 1 shows these thresholds, the resulting valid intervals of MTR values for the 5 scanners/pulse sequences and the actual percentages of lower and upper elimination. For example, choosing 44 as the upper limits of the valid range for scanner A eliminated 5%, but if we would have chosen to be 43, then we would have eliminated 8.1% or more. Therefore, the 5% was chosen, since it was closer to 6.5% than 8.1%.

Next, these thresholds were applied to each of the 23 individual scans, and the eliminated percentages were calculated for each scan. See Table 2 for the eliminated percentages for individual scans. Most

strikingly, the scan of subject 5 on D1200 yielded 20.9% of its MTR values higher than the upper threshold.

To further validate the above threshold method, the scans and particularly this outlying scan were reevaluated visually on the image files with the JIM 3.0 software [18]. The questionable low and high intensity voxels were not randomly distributed across the space, but rather they clustered on the first 10 and the last 10 slices out of the 70 slices on most images, which may be due to the B1-effect [19]. In order to obtain robust correction functions, additional outlier elimination was performed. For all images we eliminated slices 1-10 and 61-70 and used the middle 50 slices for further analysis. In this way, entire images did not have to be eliminated. Note that eliminating 20 slices out of 70 did not eliminate 28.6% of the valid MTR values, but much less (<4%), since the first several and last several slices typically don’t show much brain matter according to scanning protocol.



Computational approach

Given 4,587,520 MTRs per MTR map for a total of 23 MTR maps, a key statistical challenge is how to reduce such high dimensions in order to obtain correction functions. We develop the following two stage approach: Stage I: Spatial voxel co-occurrence matrix constructions for dimension reduction based on whole brain voxel-by-voxel data; Stage II: Obtaining predictive correction functions through curve estimations and aggregations.

Stage I: Spatial voxel co-occurrence matrix constructions: Given two MTR maps (scans) of the same subject, but from two different scanners/pulse sequences, we look for the relationship between the two. We first create a spatial voxel co-occurrence matrix in the following way: calculate the number of times the MTR value = 0 in scan 1 is also a value 0 in scan 2 and place this number as the first element of the SVCVM (matrix index 0, 0). Next, calculate the number of times the MTR value = 0 in scan 1 corresponds to value = 1 in scan 2 and place this number as the second element of the SVCVM (matrix index 0, 1). Since there are a total of 101 possible MTR values, we end up with a 101 × 101 sized matrix for the given pair of scans. Note that our MTR data was recorded on the integer scale as per existing clinical imaging protocol. Floating point data acquisition is planned for the future. There were 23 scans from 5 subjects and 5 scanners/pulse sequences; a total of 43 SVCVMs were constructed (four subjects gave 5 scans each, which gives 40 SVCVMs: 5 choose 2 times 4; one subject gave 3 scans, which gives 3 SVCVMs).

Applying the thresholds of valid ranges, which were discussed in section 2.2, reduces the size of the matrices. The parts of the matrices that show exactly the areas of valid MTRs for both scanners/pulse sequences are maintained, while the other parts are eliminated. This results in matrices of average size of 15 × 20. Finally, some robust statistics including median, 5% trimmed mean, and mode are calculated for each row of these matrices [20,21]. These calculations are made based on the reduced matrices, but row indexes are not changed during the matrix reduction.

Let Z_k represent the number of scans for scanner/pulse sequence k ($k=1, 2, \dots, K$) (in our case, $Z_k=5$ for $k=1, 4, 5$; $Z_k=4$ for $k=2, 3$; $K=5$). Let S_m represent the number of scans for subject m ($m=1, 2, \dots, M$) (in

our case $S_m=5$ for $m=1, 2, 3, 4$; $S_m=3$ for $m=5$; $M=5$); N denote the total number of scans for all subjects (in our case, $N=23$), therefore, $N=\sum_k(Z_k) = \sum_m(S_m)$. Let V_i, V_j ($i, j=1, \dots, N$) denote scan based vectors for MTRs, which take values from 0, 1, 2, ..., 100. The details of the full algorithm are given as follows:

Step 0: Repeat Step 1 through Step 4 for each index pair i, j ($i, j=1, \dots, N$) for which $i \neq j$; i and j are from same subject m ($m=1, 2, \dots, M$).

Step 1: Construct a spatial voxel co-occurrence matrix by a cross tabulation for V_i, V_j that displays the numbers of co-occurrences of individual MTR values.

Step 2: Apply the threshold method to eliminate the outliers and noise from the matrix obtained in Step 1.

Step 3: Calculate means, standard deviations, frequencies, and sums for individual valid MTR values in V_i (rows in the matrices).

Step 4: Calculate 5% trimmed means for individual valid MTR values in V_i .

The decision to use the 5% trimmed means in the algorithm was made after careful evaluation of the resulting matrices and central tendencies. They displayed the most desirable properties and they were the most robust statistics available at this stage, largely free from noise and outliers.

The above spatial voxel co-occurrence matrix constructions and calculated statistics reduce the high dimensional voxel-based data to low dimensional matrices (from millions of dimensions to hundreds). More importantly, the algorithm allows us to incorporate inter-regional

Subject ID	Scanner	Valid MTR interval	% trimmed on left	% trimmed on right
1	A	25-44	4.9%	2.1%
1	B	4-28	5.2%	7.2%
1	C	17-35	5.4%	5.3%
1	D600	42-54	6.8%	4.2%
1	D1200	30-44	6.0%	1.2%
2	A	25-44	2.2%	0.8%
2	B	4-28	4.2%	10.9%
2	C	17-35	5.2%	4.4%
2	D600	42-54	6.2%	3.3%
2	D1200	30-44	4.6%	1.9%
3	A	25-44	6.3%	6.8%
3	B	4-28	4.8%	7.0%
3	C	17-35	6.0%	8.3%
3	D600	42-54	7.5%	5.4%
3	D1200	30-44	6.4%	3.9%
4	A	25-44	9.6%	5.8%
4	B	4-28	12.8%	0.0%
4	C	17-35	8.6%	7.1%
4	D600	42-54	7.1%	3.4%
4	D1200	30-44	4.6%	3.7%
5	A	25-44	8.7%	7.8%
5	D600	42-54	9.1%	12.3%
5	D1200	30-44	6.6%	20.9%

Table 2: Actual percentages of trims for individual scans.

MTR	42	43	44	45	46	47	48	49	50	51	52	53	54
30	2138	2211	1884	1231	643	322	158	57	29	19	5	3	3
31	2484	2933	3061	2525	1573	869	372	180	83	27	12	8	5
32	2010	3183	3899	3974	3045	2019	975	495	191	53	39	15	9
33	1361	2502	3937	4806	4703	3747	2302	1093	457	172	75	32	18
34	765	1613	2984	4548	5687	5408	4036	2328	1183	438	156	69	30
35	340	814	1832	3301	5247	6136	5621	4121	2323	1022	373	133	66
36	133	366	848	1895	3734	5659	6484	5888	3988	2085	832	262	108
37	58	145	327	906	2008	3937	5891	6641	5643	3694	1727	628	206
38	22	52	138	350	903	2201	4088	6084	6735	5502	3343	1553	477
39	9	23	43	117	374	1006	2417	4480	6534	7007	5338	2883	1154
40	5	11	26	44	143	401	1202	2755	4874	7146	7295	4962	2412
41	3	3	7	16	54	123	445	1336	3171	5656	7286	6425	3768
42	1	4	5	9	22	47	158	544	1647	3497	5568	6070	4376
43	1	0	3	6	12	28	55	191	615	1596	3073	4117	3578
44	3	0	1	3	6	5	13	56	178	499	1169	1878	1926

Table 3: Example of a constructed spatial voxel co-occurrence matrix from Stage I for subject 1 at D1200 scanner (V4) vs. subject 1 on D600 scanner (V5) from MTR maps. Left column shows valid MTR range for V4, top row for V5. Matrix elements show co-occurrences of MTR values between 2 scans.

(in brain) correlations in whole MTR maps and further measure the variations of MTR values both within scanners/pulse sequences from different subjects and between scanners/pulse sequences from the same subjects. The calculated statistics from stage I will be further utilized in stage II, in which correction functions, curve estimations and fitting procedures will be developed.

Stage II: Correction functions: The aim of this study is to find correction functions between individual MRI scanners/pulse sequences. To do so, an intuitive idea is to build linear and/or nonlinear regressions between two scanners/pulse sequences' statistics calculated in Stage I and use these for correction functions. Therefore, Generalized Linear Models (GLMs) were considered to obtain optimal correction functions.

More specifically, we intend to model functional relationship between scanners/pulse sequences' V_i 's ($i=1, \dots, N$; $N=23$ in our case) at the valid ranges obtained from the threshold method in section 2.2 (Table 2). For this we use SVCMs and the calculated statistics obtained in Stage I and formulate the problem of finding correction functions with GLMs, the linear or nonlinear relationships between scans are modeled through link functions [22,23]. Several link functions between two scanners/pulse sequences including linear, logarithmic, inverse, quadratic, cubic, power, compound, logistic, growth, exponential and S function are applied. Maximum Likelihood (ML) estimation with Fisher-Scoring method is used to obtain the parameter estimation for the correction functions. Model comparisons are based on goodness of fit of models, adjusted R^2 , and standard error of the estimates, which are obtained for each of these GLMs. We select the best model for each pair of scans (which were taken from the same subject; in our example: 23 scans, 43 scan pairs) based on the highest adjusted R^2 . If the second best model is less than 0.001 away (adjusted R^2) from the best model, then break these "ties" by choosing the model that has the lowest standard error of estimate.

We end up with up to Z_k (four or five in our example) correction functions for each pair of scanners/pulse sequences (if correcting in one direction). It becomes necessary to find an optimal correction function for each scanner/pulse sequence pair for achieving better predictive and generalization performance [24,25], which will be used for future voxel based corrections. This is achieved by aggregating correction

functions, which were obtained from the same scanner/pulse sequence pair by weighting them equally.

To implement the above two stage approach for constructing SVCMs and calculate statistics based on voxel-based data and to obtain correction functions, Visual C++ 9.0 was used. Using an Intel T7200 CPU with 2GHz clock speed and 2GB RAM the overall program execution took approximately 5.5 hours.

Results

Table 3 displays one of the constructed SVCMs. Figure 4 shows the overlay scatter plots for the 5% trimmed mean statistics, which were obtained in Stage I for all the 10 one-way corrections for the 5 scanners/pulse sequences (the figure doesn't show the other 10 scatter plots for the opposite directions). Each subfigure shows 4-5 "curves" (one for each ordered pair of variables). Each curve shows the relationship between MTR values for the same subject from one scanner/pulse sequence to another. As it can be seen in the figure, most of the curves are very similar to one another for each (ordered) pair, which indicates that obtaining correction functions from a small number of subjects with large number of voxels is possible. Also, since the curves are not identical, aggregating these functions is needed in order to obtain the optimal predictive correction function.

Most of these curves appear to approximate some nonlinear, strictly monotonic functions, apparently some sigmoid functions. However, there are a few curves, which do not agree with the others as follows: V16V17, V17V18, V17V19, and V17V20. This can be confirmed both visually and by applying similarity metrics, such as SVCM's. Since V17 is the only variable that appears in all of these outlying curves as well as all of V17's curves are outliers, there must be something wrong with V17, which was derived from the image taken on subject 4 with scanner B. Further visual examination of the corresponding questionable image by an expert confirmed that it was an outlier missed scan and was eliminated at this point. All the other 22 variables were found admissible at this point and their corresponding 78 curves showed very similar patterns to one another for each scanner/pulse sequence pair, which validates the appropriateness of Stage II.

Figure 5 shows one of the results of curve estimation outputs from Stage II. The goodness of fit of models for each of the above

	Best model	Adj R ²	Std Err of Est	Function
Scanner A to B				
V1V2	Quadratic	0.989	0.577	0.025*x*x-0.821*x+15.263
V6V7	Compound	0.986	0.035	3.193*(1.050**x)
V11V12	Quadratic	0.995	0.329	0.037*x*x-1.784*x+33.158
Aggregate	Model Averaging			0.020*x*x-0.868*x+16.140+1.064*(1.050**x)
Scanner A to C				
V1V3	Quadratic	0.988	0.522	0.025*x*x-0.955*x+28.032
V6V8	Quadratic	0.987	0.541	0.019*x*x-0.571*x+22.036
V11V13	Quadratic	0.996	0.246	0.028*x*x-1.342*x+36.825
V16V18	Quadratic	0.992	0.374	0.025*x*x-1.091*x+31.752
Aggregate	Quadratic			0.024*x*x-0.990*x+29.661
Scanner A to D1200				
V1V4				
V6V9	Quadratic	0.987	0.389	0.027*x*x-1.333*x+48.314
V11V14	Quadratic	0.984	0.456	0.025*x*x-1.179*x+45.197
V16V19	Quadratic	0.994	0.213	0.030*x*x-1.675*x+55.794
V21V22	Quadratic	0.993	0.271	0.026*x*x-1.306*x+48.897
Aggregate	Quadratic	0.990	0.266	0.013*x*x-0.464*x+36.589 0.024*x*x-1.191*x+46.958
Scanner A to D600				
V1V5	Quadratic	0.984	0.371	0.019*x*x-0.890*x+53.962
V6V10	Quadratic	0.983	0.397	0.020*x*x-0.886*x+53.164
V11V15	Quadratic	0.994	0.181	0.023*x*x-1.269*x+61.897
V16V20	Quadratic	0.992	0.235	0.021*x*x-1.029*x+56.914
V21V23	Quadratic	0.991	0.203	0.009*x*x-0.287*x+46.876
Aggregate	Quadratic			0.018*x*x-0.872*x+54.563
Scanner B to C				
V2V3	Quadratic	0.998	0.166	0.043*x*x-0.349*x+21.570
V7V8	Quadratic	0.999	0.100	0.043*x*x-0.354*x+21.814
V12V13	Quadratic	0.997	0.196	0.030*x*x-0.191*x+22.162
Aggregate	Quadratic			0.039*x*x-0.298*x+21.849
Scanner B to D1200				
V2V4	Quadratic	0.997	0.148	0.027*x*x-0.305*x+33.390
V7V9	Quadratic	0.999	0.112	0.027*x*x-0.242*x+32.702
V12V14	Quadratic	0.995	0.207	0.015*x*x-0.133*x+33.141
Aggregate	Quadratic			0.023*x*x-0.226*x+33.078
Scanner B to D600				
V2V5	Quadratic	0.998	0.118	0.027*x*x-0.267*x+45.311
V7V10	Quadratic	0.999	0.062	0.028*x*x-0.259*x+44.587
V12V15	Quadratic	0.995	0.155	0.019*x*x-0.178*x+45.368
Aggregate	Quadratic			0.025*x*x-0.235*x+45.089
Scanner C to D1200				
V3V4	Quadratic	0.993	0.271	0.018*x*x-0.394*x+33.212
V8V9	Quadratic	0.992	0.304	0.020*x*x-0.468*x+34.022
V13V14	Quadratic	0.996	0.190	0.022*x*x-0.669*x+37.601
V18V19	Quadratic	0.995	0.201	0.018*x*x-0.428*x+35.062
Aggregate	Quadratic			0.019*x*x-0.490*x+34.974
Scanner C to D600				
V3V5	Quadratic	0.991	0.267	0.009*x*x+40.911
V8V10	Quadratic	0.989	0.299	0.014*x*x-0.267*x+43.841
V13V15	Quadratic	0.994	0.189	0.015*x*x-0.405*x+46.810
V18V20	Quadratic	0.996	0.156	0.013*x*x-0.297*x+45.628
Aggregate	Quadratic			0.013*x*x-0.242*x+44.298
Scanner D1200 to D600				
V4V5	Linear	0.993	0.278	1.237*x+9.425
V9V10	Linear	0.993	0.274	0.725*x+21.302
V14V15	Linear	0.991	0.284	0.668*x+23.668
V19V20	Linear	0.994	0.229	0.683*x+22.771
V22V23	Linear	0.992	0.292	1.074*x+13.201
Aggregate	Linear			0.877*x+18.074

Table 4: Examples of fitted correction functions from Stage II and Stage III. Bold functions are aggregates of individual functions and these are the final voxel-based correction functions. Functions are only applicable for the valid ranges of voxel based MTRs given in Table 1.

curve estimations resulting from GLM models are reported including adjusted R², and standard error of the estimate. Table 4 shows the

selected models based on the above model comparison criteria. All the reported models and correction functions fit the data well, with

the lowest adjusted $R^2 = 0.949$, and highest standard error of estimate = 1.33. Also, the aggregate functions by model averaging are shown in bold. The aggregate models can be used for voxel based MTR correction/prediction from any scanner/pulse sequence (A, B, C, D600, D1200) to any other scanner/pulse sequence. The reported functions are only applicable for the valid ranges of voxel based MTRs given in Table 1. Invalid ranges are considered noise and outliers and correction for those is not necessary for healthy subjects.

Due to the small sample size, we also calculated the average error for all the correction functions using the leave one out cross validation method. More specifically, each correction function, which was based on a scan pair and reported in Table 4 was left out once and compared to the average of the other correction functions for the same correction direction. The highest error was 0.0277 and the average error was 0.0075. This indicates that our obtained correction functions agree with one another and the resulting correction functions are very stable and robust. This also indicates that a small sample of 4-5 volunteers was sufficient for obtaining high quality correction functions and larger sample is not needed.

Discussion and Conclusions

In this paper we propose a novel approach that includes 1) spatial voxel co-occurrence matrix constructions and calculation of related statistics for voxel-level dimension reduction; 2) correction function with GLMs and smoothing curve fitting for estimating functional relationships among scanners/pulse sequences; and aggregate correction functions for robust function approximations and better generalization performance. Results demonstrate that between scanner/pulse sequence correction functions can reliably and accurately be obtained even from small sample size (e.g. 5) of healthy subjects with large number of available voxel measures (about 4,600,000) per MRI scans. Although it may be surprising to some that the obtained relationships between MRI scanners/pulse sequences were nonlinear, it is almost certain that the true relationships between MTR values in MRI scanners/pulse sequences are at least monotonic increasing. Accurate estimation of voxel based MTR values using correction functions overcomes comparison problems between scanner/pulse sequence differences in multicenter studies. This may have important application for future clinical trials using MTR as an endpoint. Moreover, our method has the potential to be applicable to other image acquisition techniques including X-Ray, CT, PET for which multi-scanner corrections are desired.

The proposed nonparametric spatial voxel co-occurrence matrix constructions help to avoid the multiple testing issues for high dimension reduction and are more robust than parametric approaches. Our approach has other key merits besides dimension reductions and modeling the functional relationships between scanners/pulse sequences. For instance, the resulting voxelbased correction functions can be potentially used for MTR map reconstructions. We could apply them to entire MTR maps and change MTR intensity values, voxel by voxel, effectively reconstructing entire MTR maps. Ultimately, we could be able to create “artificial MRI images” which could be evaluated by neurologists as if they were taken on an MRI scanner/pulse sequence of their choice. This way, disease progression could be better evaluated, documented, and predicted without having to ask the patient to retake MRI scans at different scanners/pulse sequences.

It is also possible to calculate a “difference MTR map” between a

known MTR map and a corresponding projected map by using our voxel-based correction procedure. If the correction procedure results in a map (and image), which show very low amount of difference (mostly black image) then the procedure can be considered to be a good procedure for MTR map correction. With this method, even multiple correction procedures can be evaluated simultaneously and compared to one another, providing a higher quality, visualized validation study.

A potential limitation is that based on this study, we expect for future MTR maps an approximate 13% of the non-zero MTR values to be outside of the valid range of MTR values we reported in Table 2. It is desirable for multicenter studies to have functions that can correct any MTR value, but we currently don't have multi-scanner data available on patients with neurological anomalies, such as black hole areas, white matter hyperintensities and lesions. Therefore, in order to estimate such MTR values for patients we made the following three attempts: first, extrapolate the correction functions; second, extrapolate with linear regression functions by substituting correction functions with linear functions for ranges below the lower threshold and above the upper threshold; third, extrapolate with constant functions through converting values to the minimum value given by the correction functions for the range below the lower threshold and to the maximum value for the range above the upper threshold. We have found that the third method worked best. It was not surprising that extrapolating the correction functions to outside of the “valid ranges of MTRs” did not work well, since the functions were estimated based on the valid ranges only. Some of the correction functions even stopped being monotonic when we extrapolated to outside the thresholds. Invalid ranges of MTR values (at least for healthy subjects) tend to contain mostly noise and outliers and are not likely to match between different scans, even on two consecutive scan of the same subject on the same scanner and pulse sequence.

Note that in the preprocessing part of our presented method, the upper and lower 10 slices were removed, which is only relevant for brain images. Our method could be slightly modified in order to better fit other types of image studies and different organs. Moreover, for better generalization ability of our proposed correction function approach, especially for patients with neurological anomalies, such as black holes, “Region of Interest” (ROI) based MTR maps should be obtained and correction functions estimated separately for each ROI, based on MTR values in those areas only, which will be our future work. This will also include separate analysis of MTR for white matter and gray matter. Also, we plan to extend our approach by extending and replacing GLM with a generalized additive model in a hierarchical mixture regression model framework, which allows MTR distribution to be a mixture of normal distributions, each representing white matter and gray matter [26,27]. Finally, we plan to use beta-kernel density, which has potential to further improve the results [28].

References

1. Miglioretti DL, McCulloch C, Zeger SL (2002) Combining images across multiple subjects: a study of direct cortical electrical interference. *J Am Stat Assoc* 97: 125-135.
2. Magland JF, Jones CE, Leonard MB, Wehrli FW (2009) Retrospective 3D registration of trabecular bone MR images for longitudinal studies. *J Magn Reson Imaging* 29: 118-126.
3. Filippi M, van Waesberghe JH, Horsfield MA, Bressi S, Gasperini C, et al. (1997) Interscanner variation in brain MRI lesion load measurements in MS: implications for clinical trials. *Neurology* 49: 371-377.
4. Cercignani M, Bammer R, Sormani MP, Fazekas F, Filippi M (2003) Inter-

- sequence and inter-imaging unit variability of diffusion tensor MR imaging histogram-derived metrics of the brain in healthy volunteers. *AJNR Am J Neuroradiol* 24: 638-643.
5. Friedman L, Glover GH, Krenz D, Magnotta V, FIRST BIRN (2006) Reducing inter-scanner variability of activation in a multicenter fMRI study: Role of smoothness equalization. *Neuroimage* 32: 1656-1668.
 6. Benedetti B, Rigotti DJ, Liu S, Filippi M, Grossman RI, et al. (2007) Reproducibility of the whole-brain N-acetylaspartate level across institutions, MR scanners, and field strengths. *AJNR Am J Neuroradiol* 28: 72-75.
 7. Friedman L, Glover GH; Fbirn Consortium (2006) Reducing interscanner variability of activation in a multicenter fMRI study: Controlling for signal-to-fluctuation-noise-ratio (SFNR) differences. *Neuroimage* 33: 471-481.
 8. Sormani MP, Iannucci G, Rocca MA, Mastronardo G, Cercignani M, et al. (2000) Reproducibility of magnetization transfer ratio histogram-derived measures of the brain in healthy volunteers. *AJNR Am J Neuroradiol* 21: 133-136.
 9. Stonnington CM, Tan G, Kloppel S, Chu C, Draganski B, et al. (2008) Interpreting scan data acquired from multiple scanners: a study with Alzheimer's disease. *Neuroimage* 39: 1180-1185.
 10. Wang L, Lai HM, Barker GJ, Miller DH, Tofts PS (1998) Correction for variations in MRI scanner sensitivity in brain studies with histogram matching. *Magn Reson Med* 39: 322-327.
 11. Madabhushi A, Udupa JK (2006) New methods of MR image intensity standardization via generalized scale. *Med Phys* 33: 3426-3434.
 12. Madabhushi A, Udupa JK, Moonis G (2006) Comparing MR image intensity standardization against tissue characterizability of magnetization transfer ratio imaging. *J Magn Reson Imaging* 24: 667-675.
 13. Cadavid D, Cheriyan J, Skurnick J, Lincoln JA, Wolansky LJ, et al. (2009) New acute and chronic black holes in patients with multiple sclerosis randomised to interferon beta-1b or glatiramer acetate. *J Neurol Neurosurg Psychiatry* 80: 1337-1343.
 14. Samson RS, Wheeler-Kingshott CA, Symms MR, Tozer DJ, Tofts PS (2006) A simple correction for B1 field errors in magnetization transfer ratio measurements. *Magn Reson Imaging* 24: 255-263.
 15. Brown LG (1992) A survey of image registration techniques. *ACM Computing Surveys* 24: 325-376.
 16. Goshtasby AA (2005) 2-D and 3-D Image Registration for Medical, Remote Sensing, and Industrial Applications. Wiley Press.
 17. Smith SM (2002) Fast robust automated brain extraction. *Hum Brain Mapp* 17: 143-155.
 18. JIM 3.0 software (Xinapse Medical Systems, Thorpe Waterville, UK).
 19. Cunningham CH, Pauly JM, Nayak KS (2006) Saturated double-angle method for rapid B1+ mapping. *Magn Reson Med* 55: 1326-1333.
 20. Hettmansperger TP, McKean JW (1998) Robust nonparametric statistical methods. Kendall's Library of Statistics. 5. London: Edward Arnold.
 21. Maronna RA, Martin RD, Yohai VJ (2006) Robust Statistics - Theory and Methods, Wiley.
 22. Beckmann CF, Tracey I, Noble JA, Smith SM (2000) Combining ICA and GLM: a hybrid approach to fMRI analysis. Sixth Annual Meeting of the Organization for Human Brain Mapping, San Diego, California S643.
 23. Calhoun VD, Adali T, McGinty VB, Pekar JJ, Watson TD, et al. (2001) fMRI activation in a visual-perception task: network of areas detected using the general linear model and independent components analysis. *Neuroimage* 14: 1080-1088.
 24. Hoeting JA, Madigan D, Raftery AE, Volinsky CT (1999) Bayesian Model Averaging: A Tutorial. *Stat Sci* 14: 382-401.
 25. Raftery AE, Madigan D, Hoeting JA (1997) Bayesian Model Averaging for Linear Regression Models. *J Am Stat Assoc* 92: 179-191.
 26. Grun B, Leisch F (2006) Fitting Finite Mixtures of Linear Regression Models with Varying & Fixed Effects in R*. In A. Rizzi, M. Vichi (eds.), Physica Verlag, Heidelberg, Germany 853-860.
 27. Leisch F (2004) FlexMix: A general framework for finite mixture models and latent class regression in R. *J Stat Softw* 11.
 28. Chen SX (1999) Beta kernel estimators for density functions. *Comput Stat Data Anal* 31: 131-145.

This article was originally published in a special issue, **Biomedical Informatics** handled by Editor(s), Dr. Emekalam Antony, Elizabeth City State University, USA