Reliability in multi-site structural MRI studies: Effects of gradient non-linearity correction on phantom and human data

Jorge Jovicich,[a,*] Silvester Czanner,[a] Douglas Greve,[a] Elizabeth Haley,[a] Andre van der Kouwe,[a] Randy Gollub,[a] David Kennedy,[a] Franz Schmitt,[b] Gregory Brown,[c] James MacFall,[d] Bruce Fischl,[a,e] and Anders Dale [f]

[a]MGH/MIT/HMS Athinoula A. Martinos Center for Biomedical Imaging, Building 149, 13th Street, Radiology/CNY149-Room 2301, Charlestown, MA 02129, USA
[b]Siemens Medical Solutions, MRPF, Manufacturing Field Generating Units, Allee am Roethelheimpark 2, 91052 Erlangen, Germany
[c]Laboratory of Cognitive Imaging, Department of Psychiatry, University of California, San Diego, La Jolla, CA 92092, USA
[d]Neuropsychiatric Imaging Research Laboratory (NIRL), Duke University, Durham, NC 27708, USA
[e]Computer Science and Artificial Intelligence Lab, Massachusetts Institute of Technology, Cambridge, MA 02142, USA
[f]Departments of Neurosciences and Radiology, University of California San Diego, San Diego, CA 92092, USA

Longitudinal and multi-site clinical studies create the imperative to characterize and correct technological sources of variance that limit image reproducibility in high-resolution structural MRI studies, thus facilitating precise, quantitative, platform-independent, multi-site evaluation. In this work, we investigated the effects that imaging gradient non-linearity have on reproducibility of multi-site human MRI. We applied an image distortion correction method based on spherical harmonics description of the gradients and verified the accuracy of the method using phantom data. The correction method was then applied to the brain image data from a group of subjects scanned twice at multiple sites having different 1.5 T platforms. Within-site and across-site variability of the image data was assessed by evaluating voxel-based image intensity reproducibility. The image intensity reproducibility of the human brain data was significantly improved with distortion correction, suggesting that this method may offer improved reproducibility in morphometry studies. We provide the source code for the gradient distortion algorithm together with the phantom data.

Introduction

Multi-site and longitudinal neuroimaging studies are increasingly becoming a standard element of clinical neuropsychiatric research for diagnosing and evaluating neurological impairments (Ashburner et al., 2003; Grundman et al., 2002; Fox and Schott, 2004). One of the challenges of both longitudinal and multi-site studies is to minimize image variability caused by technological factors (e.g., hardware differences, hardware imperfections), as such variability may be confounded with specific disease-related changes in the images thereby limiting the power to detect and follow the progression of disease biomarkers. Optimization of image reproducibility motivates the calibration of acquisition protocols and the characterization and correction of scanner-specific image variability effects. This is particularly important when data from multiple sites and MRI vendors are to be combined.

An important task in this effort is to correct for site-specific image distortions in order to allow accurate cross-site comparisons of quantitative morphometry results. Image distortions can potentially affect the accuracy of volume (Fischl et al., 2002), shape (Miller, 2004) and boundary (Barnes et al., 2004) measurements. Although distortions in MRI can arise from several factors, one of the most prominent in structural MRI is imaging gradient non-linearity, which degrades both geometric and image intensity accuracy. While in principle gradient distortions may be addressable using manufacturer-supplied software, the currently available correction algorithms work only in two-dimension (2-D) providing an incomplete solution to the problem (Wang et al., 2004a). Three-dimensional (3-D) algorithms to correct gradient non-linearity distortions have been investigated using phantoms. To summarize, two main correction methods have been developed: (a) 3-D measurement of the geometric displacements due to distortions using specially designed phantoms followed by an image transformation to perform the correction (Wang et al., 2004b,c; Langlois et al., 1999) and (b) 3-D calculation of the geometric displacements from the spherical harmonic expansion for the representation of the

* Corresponding author. Fax: +1 617 726 7422.
E-mail address: jovicich@nmr.mgh.harvard.edu (J. Jovicich).
Available online on ScienceDirect (www.sciencedirect.com).

magnetic fields generated by the gradient coils (Schmitt, 1985; Janke et al., 2004; Wald et al., 2001). The second method was used in this work. As yet, there is no quantitative study that systematically compares these correction methods. More importantly, no study investigates the effects of these distortion correction methods on test–retest reproducibility of multi-site human structural MRI data.

The purpose of this work was (i) to quantitatively characterize and correct site-specific image distortions caused by gradient non-linearity in a phantom study, and (ii) to assess if gradient non-linearity distortion correction improves image reproducibility when the same subjects are scanned at multiple sites in multiple sessions. To keep our results independent of brain morphometry analysis tools, here, we focus only on the reproducibility of image intensity for the human data. Parts of these results have been presented at recent meetings (Jovicich et al., 2004; Jovicich et al., 2003).

Materials and methods

Human and phantom image data acquisitions

Four sites with clinical 1.5 T whole body scanners used in regular functional and structural MRI studies participated in this study. These systems included: (a) GE Medical Systems with Cardiac Resonator Module (CRM) gradient coils (maximum strength = 40 mT/m, slew rate = 150 T/m/s) at Duke University Medical Center (Duke) and at Brigham and Women's Hospital (BWH); (b) GE Medical Systems with Brain Resonator Module (BRM) gradient coils at the University of California San Diego (UCSD, 22 mT/m, 120 mT/m/ms); and (c) Siemens Medical Systems with Sonata gradient coils at the Massachusetts General Hospital (MGH, Sonata gradients, 40 mT/m, 200 T/m/s). The GE scanners allowed a default 2-D in-plane distortion correction, which was used for the acquisitions on the GE platforms tested. The version of the Siemens Sonata system used did not enable this, so that data were acquired with no distortion correction at all. Therefore, the 3-D gradient distortion correction effects on image reproducibility can be evaluated against acquisitions with no correction at all (our Siemens data), or acquisitions that had the vendors' 2-D corrections (our GE data).

Phantom data (from the 4 sites) and test–retest human data (from MGH, UCSD and Duke) were collected using an acquisition protocol that included a 3-D-spoiled gradient echo volume (TR = 20 ms, TE = 6 ms, flip angle 30°, 256 × 192, 1.3 mm thick 124 sagittal slabs, FOV 25 cm, 8 min 12 s acquisition) and used the vendors' standard head RF coil. Five healthy volunteers gave written informed consent to participate in this multi-site study, which was approved by the institutional review boards at each participating site. Each subject (1 female, 4 males, and average age 39) was scanned twice on each site, in different sessions using this acquisition protocol. For within-site repetitions the average time interval for rest–retest scans was 19 days (minimum time 1 day, maximum time 8 months), and for across-site repetitions the average time interval was 8 months (minimum, 2 months; maximum, 15 months).

A special cylindrical phantom (250 mm length × 220 mm diameter) consisting of 25 plastic plates, each 10 mm thick was specifically built for assessing geometric distortions from gradient non-linearities (Franz Schmitt, personal communication). Each plate had a pattern of holes 3 mm in diameter, going through the thickness of the plate and perpendicular to the faces of the plate. These holes formed a 2-D rectangular grid with 10 ± 0.05 mm

spacing. In addition, on the two sides of each plate, each hole was enlarged with a half spherical depression of 7 mm in diameter. In this way, with the plates glued to each other and the 3 mm holes aligned, the phantom formed a 3-D 10 ± 0.05 mm grid of 7 mm diameter spheres that could be filled with a fluid (water doped with salt) through the passing 3 mm holes. The phantom acquisition protocol scan was as for the humans, but with 2 mm slices and a 45-cm FOV. In addition to the 3-D geometric phantom, which was scanned at all participating sites and is not suitable for measuring image intensity uniformity, a standard saline filled uniform cylindrical phantom (250 mm, length; 150 mm, diameter) was scanned at one of the sites (MGH) to evaluate image intensity uniformity improvements from gradient distortion correction. This uniform phantom was scanned three times within the same session, at slightly different positions within the field of view, using the standard 3-D T1-weighted sequence described above.

Gradient distortion correction

The goal of the distortion correction is to transform (interpolate) the original distorted image into a corrected one by displacing each voxel into an estimate of its correct 3-D location and by scaling each voxel's intensity to account for voxel-size distortions. A laboratory-based coordinate system was used to calculate the displacements in Euclidean coordinates (x, y, z). This coordinate axis system had its origin at the iso-center of the scanner, with the z direction aligned with the main field B_0, the x direction the left/right and the y direction superior/inferior. The 3-D displacements along each of the three directions can be calculated from the non-linear terms of the magnetic field generated by each of the gradient coils (Janke et al., 2004). These fields are usually provided by the vendor in the form of a truncated series of spherical harmonic coefficients and instructions for how these coefficients are normalized to calculate field magnitudes (5 terms were available for the GE CRM and BRM gradients and 11 for Siemens Sonata gradients). The intensity correction is the Jacobian determinant calculated from the non-linear magnetic field terms. The displacements and intensity correction tables were pre computed (using trilinear 3-D interpolation) in a 3-D cylindrical grid (300 mm in diameter, 300 mm long) that is large enough to contain any imaging volume acquired within the head RF coil.

Evaluation of unwarping effects on test–retest reproducibility

The quantitative characterization of the distortion corrections (unwarping) on the phantom data was evaluated in two ways: (a) measurement of the phantom diameter at several positions along the z axis in the raw and corrected images and comparison of these values with the true phantom diameter (220 mm) to quantify how the % errors were reduced with distortion correction, and (b) estimation of the range of image intensity errors and magnitude displacements within a spherical region of interest centered at the magnet's iso-center and large enough to encompass the full brain volume in a conventional scan.

The unwarping effects on the human data were evaluated by assessing if voxel-based image intensity variability was reduced with distortion correction, both within and across sites. For each subject, the T1-weighted volume was skull stripped, co-registered with the other scans of the same subject, and intensity normalized (brain mean 100). The images were then grouped in the following ways to compute variability: within-site test–retest (for each site
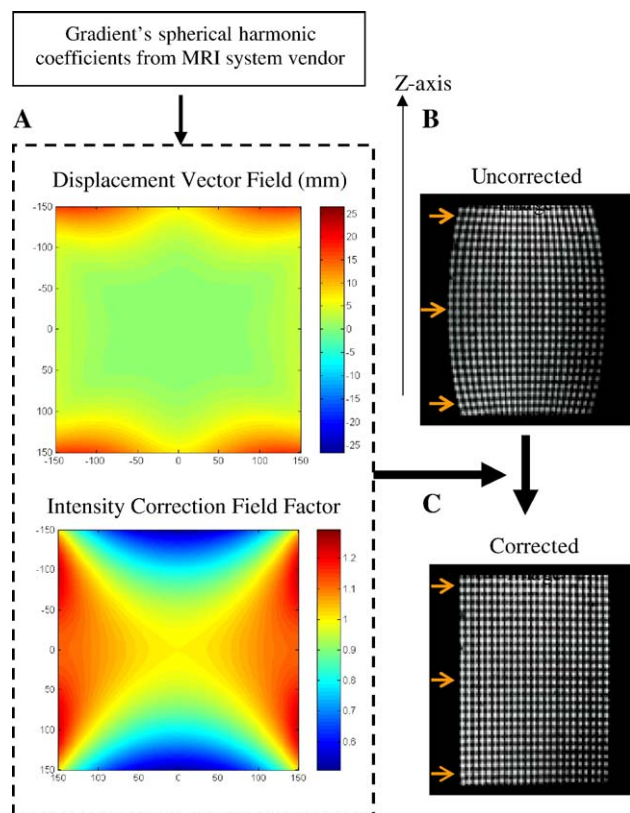
Fig. 1. Schematic representation of the distortion correction process. The spherical harmonic coefficients specific to the MR system are used to generate 3-D x, y, z displacement fields (here, we show the magnitude) and intensity correction fields arising from voxel distortions (A). The correction fields correspond to a 300 mm × 300 mm coronal slice through and centered on the scanner iso-center. The distortions on the uncorrected phantom image appear as diameter variations along the z axis (B). These geometric distortions are significantly reduced in the corrected image (C).

and subject the image intensity variability across the two sessions, with and without distortion correction, was computed), across-site test–retest (for the first session image data for each subject the image intensity variability across the three sites, with and without distortion correction, was computed). To quantify image intensity variability voxel-based relative errors were calculated for the cases described above (standard deviation divided by mean intensity, for each voxel). The relative errors were then assessed by inspecting their spatial distribution using maps (thresholded relative errors

overlaid on structural scans) and also by inspecting their magnitude distribution, using histograms. The maps allowed us to see if there was visible structure to the spatial distribution of the image intensity fluctuations, and also to see how these distributions would change with distortion correction. The histograms provided a more quantitative measure for comparing the magnitude of the overall brain intensity variability with and without distortion correction. Statistical Kolmogorov–Smirnov tests were used to quantify the significance of the differences between histograms. In addition, the corrected and uncorrected histogram means were compared using 2-sample t tests to assess if distortion correction was producing the expected change: a reduction in the overall intensity variability, reflected in a smaller mean error.

Results and discussion

(1) Gradient distortion correction: phantom validation

To validate the distortion correction we measured the diameter of the phantom images obtained at four sites, and compared the measures with the true diameter, with and without image distortion correction. Fig. 1 gives a schematic summary of the gradient distortion correction process, with sample phantom images from the MGH site. The spherical harmonic coefficients from the MR vendor's gradient were used to calculate 3-D displacement and intensity correction fields within the scanner's field of view (Fig. 1A). Intensity correction is necessary to compensate for the change in effective voxel size with distortion. The correction fields, shown as color maps in Fig. 1, are specific to the gradients' hardware and thus need to be calculated only once for each site for each gradient upgrade. Three displacement fields are calculated for each of the three x, y, z axes, Fig. 1 shows only the magnitude displacement as a summary. The correction fields correspond to a 300 mm × 300 mm coronal slice through and centered on iso-center. The vertical axis corresponds to the scanner's z axis (head–foot direction) and the horizontal axis to the scanner's x axis (left–right direction). The red areas in the maps correspond to the areas where the distortions are strongest (Fig. 1). Note how the phantom diameter clearly varies along the longitudinal z axis in the uncorrected image (Fig. 1B) and is uniform in the corrected one (Fig. 1C). Similar results were obtained for each of the four MR systems.

Table 1 summarizes the group results for the phantom scans across the 4 sites (MGH, BWH, Duke, and UCSD), showing how both the mean and standard deviations of the multi-site diameter errors (% difference between the measured and true diameters

Table 1
Multi-site gradient distortion correction evaluation using a phantom

Site	Uncorrected images				3-D corrected images			
	% Diameter errors along z axis				% Diameter errors along z axis			
	0	40	80	124	0	40	80	124
GE-BWH	3.6	1.8	−2.3	−4.5	0.4	0	0.4	0
GE-Duke	4.1	2.3	−3.2	−3.6	0.4	0.4	0.4	0.4
GE-UCSD	1.8	0.9	−1.8	−5.9	0	0	0	0
Siemens-MGH	2.7	2.3	−0.4	−7.7	0	0	0.4	0
Mean errors	3.1	1.8	−1.9	−5.4	0.2	0.1	0.3	0.1
Std	1.0	0.7	1.2	1.8	0.2	0.2	0.2	0.2

Variability of phantom diameter relative errors (% difference between the measured and true diameters relative to the true diameter) across the MGH, BWH, Duke, and UCSD sites, with and without 3-D gradient distortion correction, as a function of the distance from magnet's iso-center along the z axis (mm).

relative to the true diameter, averaged across the four sites) are significantly reduced after distortion correction ($P < 0.001$) at various points along the z axis. At the phantom edges the uncorrected diameter errors were about $5 \pm 2\%$, whereas for the corrected images the deviation from the true diameter was $0.5 \pm 0.08\%$.

The phantom diameter distortions represent the maximal potential errors because they are in the periphery of the field-of-view, which is where the gradient non-linearity effects are strongest. We also looked at the predicted effects of image intensity errors (voxel distortions) and displacements within a conservative region of interest (ROI) of 100 mm in diameter centered at the magnet's iso-center. Such an ROI is likely to contain the entire brain volume in a standard scan. Fig. 2 shows an example of the results for a coronal view of the Sonata system (the

others systems gave similar results), showing how the intensity (Fig. 2A) and magnitude displacement (Fig. 2B) errors grow as one moves away from the iso-center but within this ROI. These error ranges are shown on the right side of Fig. 2 as gray-shaded areas that give the magnitude of the intensity correction factors (top) and magnitude of 3-D displacements in mm (bottom) along the x axis within this ROI. The gray shaded areas are delineated by the intersection (black dots) of the various color-coded contours defined by $z = 0, 20, 40, 60, 80, 100$ and 120 mm and the corresponding vertical reference lines that define the ROI boundaries. The results show that if the same voxel-sized tissue sample would be placed at various positions within the 100 mm diameter ROI in this system, then the image intensity (or voxel volume) could fluctuate up to 20% and the location error could be up to 2 mm. The magnitude of these errors represent a problem for
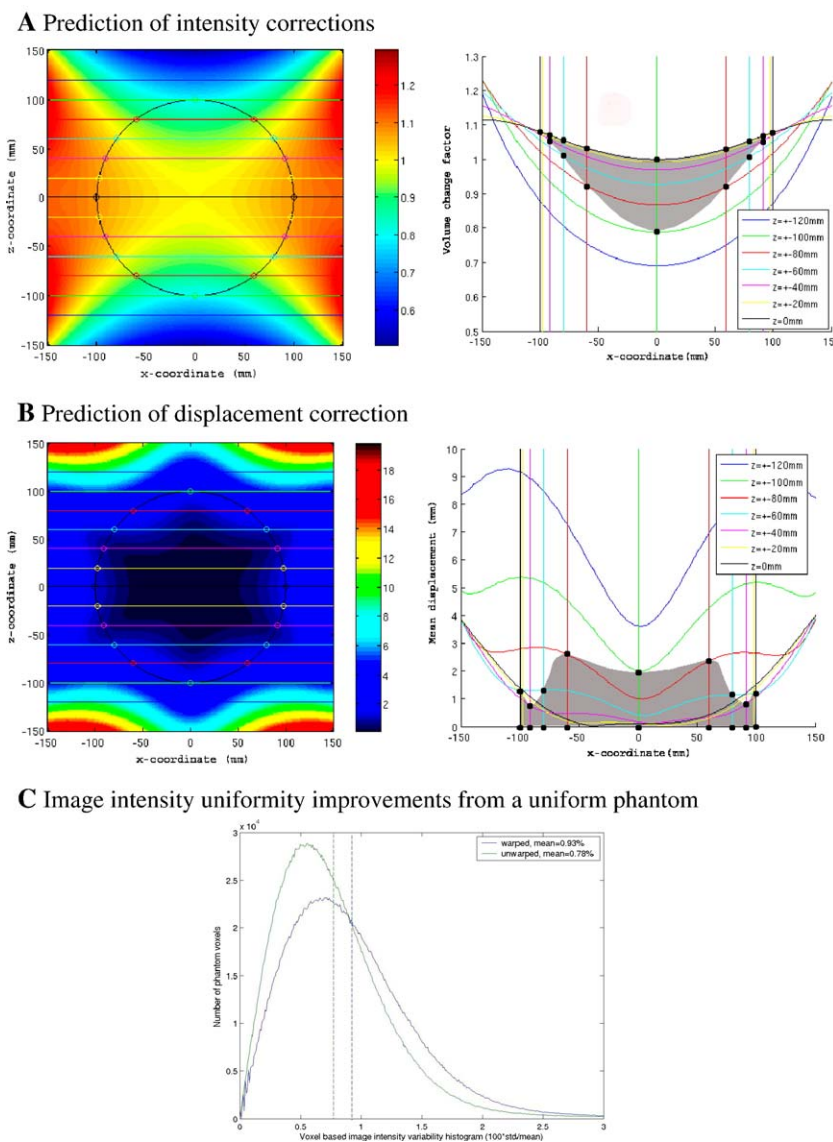

Fig. 2. Prediction of intensity and displacement errors within a region of interest (ROI) where the brain volume is likely to be positioned (A and B). The left shows field maps for volume change factor (A) and magnitude of displacement (B). The entire FOV corresponds to a 300 mm × 300 mm coronal slice through and centered on iso-center. In these maps, a conservative spherical ROI is marked where brain is likely to occur using a radius of 100 mm. The plots in the right column show in gray shaded areas the range of intensity corrections (A) and displacement errors (B) that are expected within the ROI as a function of the x coordinate. Fig. 3C shows the histograms of image intensity variability from a uniform phantom scanned three times (MGH site, Siemens Sonata), with (0.78% mean intensity variability) and without (0.93% mean variability) 3-D distortion correction.

quantitative brain morphometry given that cross-sectional or longitudinal studies typically report size effects that are fractions of mm (Barnes et al., 2004). Finally, Fig. 2C shows that the image intensity uniformity of a uniform phantom scanned multiple times within a same session at slightly different positions is improved following gradient distortion correction. The mean intensity variability of the phantom signal is reduced from 0.93% to 0.78%.

Overall, the phantom results validate the correction method by showing a significant reduction of site-specific geometric distortion effects due to gradient non-linearity, as well as showing image uniformity improvements.

(2) Gradient distortion correction: reproducibility effects on human data

Fig. 3 shows qualitatively the type of tissue boundary shifts due to distortion that are corrected by the distortion correction unwrapping method. Sample results from one of the Sonata scans are shown, other sites and subjects show similar effects. Fig. 3 shows that boundary shifts of several mm can take place when the cortical surfaces of the distortion corrected volume are overlaid on the original distorted volume (Fig. 3C). These displacements may vary across scan sessions as a function of the exact location of the head in the gradient field.

The effects of gradient distortion correction on image intensity reproducibility were evaluated to determine if reproducibility errors could be reduced by the corrections for both within-site test–retest scans and also across site comparisons for each of the subjects. Gradient distortions are expected to affect image intensity reproducibility because of voxel size distortions.

Image intensity variability was interrogated in two ways: (1) by calculating voxel-based relative error maps (standard deviation across multiple measures divided by the mean intensity) inside the brain to assess the spatial distribution of the intensity fluctuations, and (2) by calculating the corresponding variability histogram to assess the intensity distribution of the errors. Fig. 4 shows sample results from a single subject. The intensity variability maps corresponding to two within-site test–retest
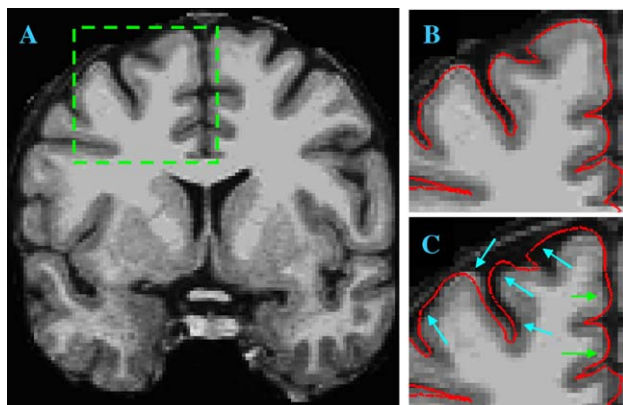

Fig. 3. Qualitative demonstration of tissue boundary shifts due to unwrapping (subject 2, Siemens site). A shows a representative coronal slice of the original structural T1 volume. The delineated area in A is enlarged on the right to show the pial surface (in red) of the distortion-corrected volume overlaid on the corresponding distortion-corrected volume (B), and the same surface overlaid on the original distorted volume (C). The blue arrows on C indicate areas with visible displacements, and the green arrows areas where there was much less displacement.
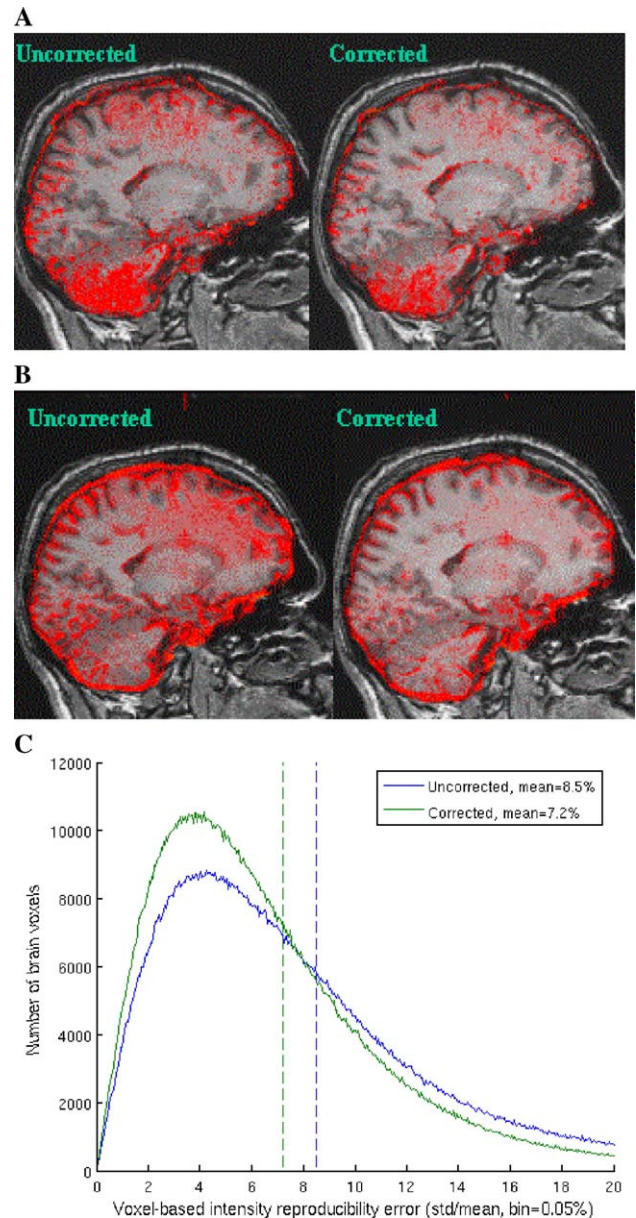

Fig. 4. Distortion correction effects on image intensity reproducibility of human structural data. The color maps show the voxel-based variability maps (standard deviation divided by the voxel mean), for a single subject (Subject #3 from Table 1) test–retest within site (A, two sessions at MGH) and the same subject test–retest across 3 sites (B, UCSD, Duke and MGH). Intensity variations larger than 8% are shown in red overlaid on one of the subject's structural scans. C shows the effects of the distortion correction on the histograms of the variability maps corresponding to B.

sessions, uncorrected (left) and distortion corrected (right) are shown in Fig. 4A, and those corresponding to three across-site sessions are shown in Fig. 4B. The color maps in Fig. 4 display voxels with image intensity variability greater than 8% (in red) overlaid on a representative anatomical gray scale image. Figs. 4A and B show that the largest image intensity variability effects appear along areas where gray or white matter tissue borders cerebral spinal fluid (CSF) spaces, creating sharp image intensity edges due to the very low MR signal that CSF has in our T1-weighted images. Such intensity variability effects appear as local image co-registration errors, but are most likely due to variability

Table 2
Gradient distortion effects on the mean brain intensity measures of within-site and across-site test–retest structural MRI data

Within site image intensity variability

Sites	Subject	Uncorrected histogram error mean (a)	3-D Corrected histogram error mean (b)	% reduction of the mean errors (100*(a−b)/a)
UCSD	1*	2.97%	3.03%	−2.11
	2	3.02%	2.88%	4.73
	3	2.97%	2.86%	3.79
	4*	2.79%	2.78%	0.4
	5	3.23%	3.14%	2.97
	Average	3%	2.94%	1.92
Duke	1	4.04%	3.66%	9.42
	2	4.14%	3.39%	17.99
	3*	4.00%	3.98%	0.37
	4	4.46%	4.14%	7.22
	5	5.40%	4.73%	12.32
	Average	4.41%	3.98%	9.46
MGH	1	8.44%	4.88%	42.16
	2	5.53%	4.99%	9.71
	3	4.83%	3.98%	17.71
	4	6.32%	5.79%	8.43
	Average	5.9%	4.72%	19.50

Across site (UCSD, Duke, MGH) image intensity variability

Subject	Uncorrected histogram error mean	Corrected histogram error mean	% reduction of the mean errors
1	8.82%	7.31%	17.16
2	10.09%	7.61%	24.61
3	8.37%	7.07%	15.52
4	8.09%	7.20%	11.02
5	9.55%	8.36%	12.48
Average	8.98%	7.51%	16.16

In all cases, except those marked with an asterisk, the 3-D distortion correction gave image intensity variability histograms with lower mean intensity errors ($P < 0.001$, see text) compared to the raw acquired data. Improvements can be seen regardless of whether data had been acquired with the default 2D correction offered by the vendors (GE sites, Duke and UCSD), or whether it had been acquired with no correction at all (Siemens site, MGH).

from partial volume and B_0 inhomogeneity effects, which are not corrected by the distortion correction from imaging gradient non-linearities. A qualitative assessment of the error maps from all subjects indicated that there was no consistent structure to the spatial distribution of the variability maps and that distortion correction errors reduced but did not fully eliminate the image intensity variability. This is consistent with the fact that other sources of image intensity variability, such as inhomogeneities of the transmit and receive RF excitations, are not being considered in this study. The histograms of the uncorrected (blue) and corrected (green) across-site intensity error maps from Fig. 4B are shown in Fig. 4C. The image intensity variability distribution shifts towards a lower mean when distortion correction is applied, indicating that variability errors are reduced. Similar results were obtained for other subjects. The statistical tests showed that overall for both within- and across-site comparisons (3 sites, 5 subjects), the distortion corrected histograms were significantly different from the uncorrected ones and that the reproducibility error means were reduced by distortion correction (Table 2). One

of the subjects (subject 5 in Table 2) could not complete the retest session at one of the sites (MGH), so his within-site variability test is missing. The results in Table 2 show that, in the group of subjects tested, distortion correction improves image intensity reproducibility within the brain, and that the effects of the gradient distortion correction are strongest in across-site comparisons. Furthermore, Table 2 also shows that the improved image intensity reproducibility given by the 3-D distortion correction was observed regardless of whether data had been acquired with the default 2-D correction offered by the vendors (GE sites, Duke and UCSD), or whether it had been acquired with no correction at all (Siemens site, MGH).

The brain intensity variability in the within-site test–retest comparisons shows a strong site effect for the uncorrected data: the intensity variability is approximately 3.2% for UCSD, 4.4% for Duke and 5.9% for MGH (Table 2). To further investigate the sources of these site-specific effects, we looked at the reproducibility of the brain center coordinates for all subjects at each of the three sites. For each skull-stripped volume, the brain center-of-mass scanner coordinates were calculated. The means and standard deviations of these coordinates were then obtained from all the sessions at each site (Table 2). The results show that the variability in the brain coordinates is strongest along the z axis (head–foot direction), i.e., positioning along the main axis of the magnet, and lowest along the x axis (left–right direction). It was also found that there is a very strong positive correlation (0.99) between intensity and positioning reproducibility. To visually inspect how these positioning fluctuations mapped onto the distortion fields, we overlaid the brain center coordinates for all sessions on the intensity correction fields for each of the sites (Fig. 5). To facilitate viewing, a brain "radius" was estimated for each session, as the
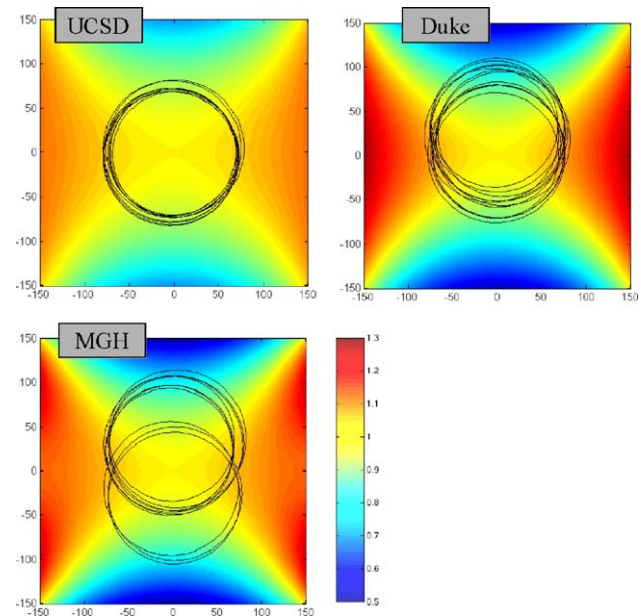

Fig. 5. Variability of subject's brain center positioning at the different sites from all test–retest scan sessions overlaid on the image intensity correction fields for each of the sites (coronal slice). The horizontal axis corresponds to the right–left direction, and the vertical axis corresponds to the head–foot direction if the field of view (distance in mm). The brain centers were obtained by calculating the centroid of each skull stripped brain volume. The radius for each location was calculated as the radius of a sphere with the volume of that brain.

Table 3
Variability subject's brain center positioning at the different sites

Site	Group scanner coordinates of the brain volume centers		
	X (Left–Right) (mean ± std, mm)	Y (Anterior–Posterior) (mean ± std, mm)	Z (Head–Foot) (mean ± std, mm)
UCSD	−1.8 ± 1.8	−6.1 ± 3.9	−1.1 ± 3.9
Duke	0.5 ± 3.4	−9.7 ± 5.6	17.6 ± 12.5
MGH	0.0 ± 3.0	11.1 ± 10.1	12.9 ± 27.0

The table shows the positioning variability in terms of means and standard deviations (std) of the brain center scanner coordinates, for each of the sites, across all sessions for all subjects. The biggest variability was measured along the Anterior–Posterior and Head–Foot directions, particularly at the MGH site.

radius that a sphere of that brain volume would have. We then overlaid the spheres for each imaging session on their corresponding site-specific intensity correction maps for the coronal sections, which showed the strongest variability along the z axis (Fig. 5). The strongest subject-positioning variability at the MGH site is likely to be the cause for that site's higher within-site test–retest intensity variability for the uncorrected data (Table 2). The size effect of the distortion correction improvements results show also a correlation with subject-positioning error: the stronger the positioning variability the stronger the correction effects. Therefore, reproducible head positioning along the z axis can have a significant impact in reducing inter-session image intensity variability.

How important is the contribution to image intensity reproducibility brought by the proposed distortion correction? To address this, we considered the following approximations: (i) we assume that the total image voxel-based intensity test–retest variability has a systematic component due to the MRI electronics noise, a component due to gradient non-linearity distortions, and a component that includes a combination of physiological, B_1, B_0, and other effects. (ii) The results from Table 2 show that the percent reduction of the mean intensity reproducibility errors thanks to the distortion correction is approximately 16% for across-site comparisons. (iii) The uniform phantom test–retest image intensity variability of ~1% can be used as a rough estimate of the MRI system's electronics (since the phantom was uniform and since B_0 and B_1 inhomogeneity are likely to remain systematic and not add variability through the multiple acquisitions). When compared to the total ~9% of uncorrected across-site variability (Table 2), the electronics noise figure (~1%) represents approximately 11% of the total test–retest noise. (iv) Therefore, if 11% of the noise is due to electronics and 16% is due to gradient non-linearity distortions, then about 72% of the total intensity reproducibility noise is most likely due to a combination of variability factors including physiological and B_0 and B_1 inhomogeneities. Overall, our results show that distortion correction produce significant intensity reproducibility improvements but that these effects are not the most important cause leading to the variability (Table 3).

The 3-D gradient unwarping software with documentation and our phantom data are publicly available for distribution on http://www.nbirn.net/Resources/Downloads/GradientNonLinearity/index.htm. Proprietary information about the vendors' gradient fields spherical harmonics expansion is not distributed. Users should request this information by contacting their local vendor representatives. Scripts are provided that read in the proprietary coefficients files (Siemens Sonata, GE CRM and BRM) for the unwarping algorithm.

Conclusions

Structural MRI studies offer the potential for quantifying subtle brain structural differences between patient populations or changes over time, such as during development or atrophy in neuro-degenerative diseases. Further, the ability to combine calibrated data acquired across multiple sites offers the possibility for increased statistical power by analyzing large collections of datasets. However, the effectiveness of such approaches is limited by image reproducibility errors, which can translate into accuracy errors for the derived morphometric data used to characterize brain structures. It is therefore important to understand and try to correct for the various sources of variance that can affect test–retest image variability, within- and across-sites. Sources of image intensity variability include: noise of the electronics of the MRI system, subject's physiological noise, partial volume effects, imaging gradient non-linearities, spatial inhomogeneities in the main field B_0 and in the RF field B_1. In this work, we only focused on image intensity reproducibility errors introduced by site-specific non-linearities in the imaging gradients. These errors introduce geometric distortions and as a result image intensity errors due to voxel-size distortions.

The multi-site phantom results validated the gradient distortion correction method, showing that the reproducibility and geometric accuracy of phantom image data can be significantly increased. The test–retest human data, within- and across-site results showed that image intensity reproducibility is significantly improved with distortion correction. As expected, the correction effects are strongest in across-site comparisons. This is consistent with the fact that in multi-site scanning the variability in subject's positioning is added to the variability in distortion fields from the different sites. We also found that the site-specific differences in within-site reproducibility errors could be explained by the positioning of the subjects. Overall, the proposed 3-D distortion correction significantly improves image intensity reproducibility. However, our estimates suggest that these effects explain approximately 17% of the image intensity test–retest reproducibility of structural images at 1.5 T, and that the approximately 72% of the remaining uncorrected variability is most likely due to a combination of effects including physiological, B_0 and B_1 inhomogeneities. We provide the source code for the 3-D distortion correction and the phantom data.

In conclusion, correction for gradient non-linearity errors has the potential for improving the accuracy of morphometric analysis in longitudinal and multi-site imaging studies, by improving both geometric accuracy and image intensity reproducibility. These corrections, however, do not account for all the sources of image intensity variability. Future studies will investigate the effects of distortion correction on the reproducibility of brain morphometric measures.

the gradient's spherical harmonics information. This research was supported by a grant (#U24 RR021382) to the Morphometry Biomedical Informatics Research Network (BIRN, http://www.nbirn.net), that is funded by the National Center for Research Resources (NCRR) at the National Institutes of Health (NIH).

References

Ashburner, J., Csernansky, J.G., Davatzikos, C., Fox, N.C., Frisoni, G.B., Thompson, P.M., 2003. Computer-assisted imaging to assess brain structure in healthy and diseased brains. Lancet Neurol., 79–88 (Feb 2).

Barnes, J., Scahill, R.I., Boyes, R.G., Frost, C., Lewis, E.B., Rossor, C.L., Rossor, M.N., Fox, N.C., 2004. Differentiating AD from aging using semiautomated measurement of hippocampal atrophy rates. Neuro-Image 23 (2), 574–581 (Oct).

Fischl, B., Salat, D.H., Busa, E., Albert, M., Dieterich, M., Haselgrove, C., van der Kouwe, A., Killiany, R., Kennedy, D., Klaveness, S., Montillo, A., Makris, N., Rosen, B., Dale, A.M., 2002. Whole brain segmentation: automated labeling of neuroanatomical structures in the human brain. Neuron 33 (3), 341–355 (Jan 31).

Fox, N.C., Schott, J.M., 2004. Imaging cerebral atrophy: normal ageing to Alzheimer's disease. Lancet 363, 392–394 (Jan 31).

Grundman, M., Sencakova, D., Jack Jr., C.R., Petersen, R.C., Kim, H.T., Schultz, A., Weiner, M.F., DeCarli, C., DeKosky, S.T., van Dyck, C., Thomas, R.G., Thal, L.J., 2002. Alzheimer's disease cooperative study brain MRI hippocampal volume and prediction of clinical status in a mild cognitive impairment trial. J. Mol. Neurosci. 19 (1–2), 23–27 (Aug–Oct).

Janke, A., Zhao, H., Cowin, G.J., Galloway, G.J., Doddrell, D.M., 2004. Use of spherical harmonic deconvolution methods to compensate for nonlinear gradient effects on MRI images. Magn. Reson. Med. 52 (1), 115–122 (Jul).

Jovicich, J., Greve, D., Haley, E., Fischl, B., Dale, A., 2003. Brain Morphology BIRN Biomedical Informatics Research Network. Characterization and Correction of Image Distortions in Multi-Site Structural MRI. NeuroImage. 19; Suppl. 1 S45, p. 863.

Jovicich, J., Greve, D., Haley, E., Kennedy, D., Tosa, Y., Gollub, R.L., Fischl, B., Dale, A., 2004. B Morphometry BIRN. Multi-site structural MRI studies: an evaluation of image distortions and image intensity reproducibility. Proceedings of the 12th International Society of Magnetic Resonance in Medicine, Kyoto, Japan. International Society for Magnetic Resonance in Medicine, Inc., pp. 136. (ISSN 1545-4436).

Langlois, S., Desvignes, M., Constans, J.M., Revenu, M., 1999. MRI geometric distortion: a simple approach to correcting the effects of non-linear gradient fields. J. Magn. Reson. Imaging 9 (6), 821–831 (Jun).

Miller, M.I., 2004. Computational anatomy: shape, growth, and atrophy comparison via diffeomorphisms. NeuroImage 23 (Suppl. 1), S19–S33.

Schmitt, F., 1985. Correction of geometric distortion in MR Images. Proceedings of the computer assisted radiology (CAR). Springer, Berlin, pp. 15–25.

Wald, L.L., Schmitt, F., Dale, A., 2001. Systematic spatial distortion in MRI due to gradient nonlinearities. Proceedings of the 7th Annual Human Brain Mapping Meeting, Brighton, UK, pp. 203.

Wang, D., Strugnell, W., Cowin, G., Doddrell, D.M., Slaughter, R., 2004. Geometric distortion in clinical MRI systems Part I: evaluation using a 3 D phantom. Magn. Reson. Imaging 22 (9), 1211–1221 (Nov).

Wang, D., Doddrell, D.M., Cowin, G., 2004. A novel phantom and method for comprehensive 3-dimensional measurement and correction of geometric distortion in magnetic resonance imaging. Magn. Reson. Imaging 22 (4), 529–542 (May).

Wang, D., Strugnell, W., Cowin, G., Doddrell, D.M., Slaughter, R., 2004. Geometric distortion in clinical MRI systems Part II: correction using a 3D phantom. Magn. Reson. Imaging 22 (9), 1223–1232 (Nov).