NeuroImage **13,** 931–943 (2001)
doi:10.1006/nimg.2001.0756, available online at http://www.idealibrary.com on IDEAL®

Qualitative and Quantitative Evaluation of Six Algorithms for Correcting Intensity Nonuniformity Effects

James B. Arnold,* Jeih-San Liow,† Kirt A. Schaper,* Joshua J. Stern,‡ John G. Sled,§ David W. Shattuck,¶ Andrew J. Worth,‖ Mark S. Cohen,** Richard M. Leahy,¶ John C. Mazziotta,** and David A. Rottenberg*,†,‡,††

*††Neurology Service and *PET Imaging Center, Minneapolis VA Medical Center, One Veterans Drive, Minneapolis, Minnesota 55417; ‡Department of Neurology and †Department of Radiology, University of Minnesota; §McConnell Brain Imaging Centre, Montreal Neurological Institute; ¶Signal and Image Processing Institute, University of Southern California; ‖Center for Morphometric Analysis, Department of Neurology, Massachusetts General Hospital; **Brain Mapping Center, Neuropsychiatric Institute and Department of Neurology, UCLA School of Medicine*

Received September 22, 2000

The desire to correct intensity nonuniformity in magnetic resonance images has led to the proliferation of nonuniformity-correction (NUC) algorithms with different theoretical underpinnings. In order to provide end users with a rational basis for selecting a given algorithm for a specific neuroscientific application, we evaluated the performance of six NUC algorithms. We used simulated and real MRI data volumes, including six repeat scans of the same subject, in order to rank the accuracy, precision, and stability of the nonuniformity corrections. We also compared algorithms using data volumes from different subjects and different (1.5T and 3.0T) MRI scanners in order to relate differences in algorithmic performance to intersubject variability and/or differences in scanner performance. In phantom studies, the correlation of the extracted with the applied nonuniformity was highest in the transaxial (left-to-right) direction and lowest in the axial (top-to-bottom) direction. Two of the six algorithms demonstrated a high degree of stability, as measured by the iterative application of the algorithm to its corrected output. While none of the algorithms performed ideally under all circumstances, locally adaptive methods generally outperformed nonadaptive methods.

Key Words: **magnetic field inhomogeneity; MRI; tissue segmentation.**

INTRODUCTION

The removal of intensity nonuniformity ("bias") from MRI images is an essential prerequisite for the quantitative analysis of MRI brain volumes. Spatial variation of the image signal unrelated to anatomic information results from inhomogeneity in the B_0 and RF excitation fields and from regional differences in the magnetic properties of the tissues being imaged. While such variation may have little effect on the visual interpretation of brain images, it can significantly reduce the accuracy of computational procedures such as tissue segmentation and brain-surface extraction.

Numerous advantages are associated with the acquisition of magnetic resonance images at high field strengths, including an approximately linear increase in signal strength that can be used to improve spatial resolution; there are also potential advantages in contrast behavior due to the field-strength dependence of tissue relaxation times (Bottomley *et al.*, 1984). However, these advantages come at a cost: intensity nonuniformity is a prominent feature of magnetic resonance images acquired at high field strengths. Although it is possible to create highly uniform radio frequency (RF) fields in a vacuum or in nonconductive materials such as silicon oil (Hayes *et al.*, 1985; Li *et al.*, 1997), biological materials typically have dielectric properties that cause the effective RF field to become nonuniform in a frequency-dependent (i.e., field-strength-dependent) manner (Simmons, 1994; Vaughan *et al.*, 1994; Sled and Pike, 1998). Such RF nonuniformities propagate to image intensity variations, as both the effective transmit power and relative receiver sensitivity become dependent on the location of the sample within the RF field.

Various methods have been devised to detect and correct for intensity nonuniformity in MRI data sets, but there have been relatively few studies comparing the performance of different nonuniformity-correction (NUC) algorithms on standardized sets of brain images. Moreover, the evaluation of NUC algorithms is not a straightforward matter because the true amount of nonuniformity in MRI images is unknown, and it is not possible to directly measure the accuracy of the nonuniformity-correction. Comparisons of NUC methods have been reported using simulated brain volumes

1053-8119/01 $35.00

with known added biases (e.g., Sled *et al.*, 1997; Zaini *et al.*, 1999); however, simulated images may not accurately mimic the signal properties of tissue compartments *in vivo*. Although the impact of NUC on brain-tissue segmentation and tumor response measurements has been studied by DeCarli *et al.* (1996) and Velthuizen *et al.* (1998), intersubject and interscanner variability limit the utility of such data sets for ranking algorithmic performance.

In the present study we evaluated six NUC algorithms representing several different approaches to the estimation and removal of intensity nonuniformity. Quantitative measures were developed to assess the ability of the selected algorithms to remove varying amounts of nonuniformity added to an artificial MRI brain volume and to produce consistent results from a set of repeat brain scans of a single subject on a single scanner. To address the issue of scanner dependence, each algorithm was also applied to high-resolution brain scans acquired on different scanners in different institutions.

METHODS[1]

NUC algorithms were evaluated on T1-weighted MRI volumes because these MRI volumes have become the *de facto* standard of the neuroimaging community for brain segmentation and cortical surface extraction. Whereas dual-echo spin echo (i.e., proton-density- and T2-weighted) volumes can be used for the same purposes, prolonged acquisition times, movement between acquisitions, reduced signal-to-noise, and significantly greater image nonuniformity have led most groups to abandon the dual-echo approach.

The accuracy of nonuniformity correction was assessed qualitatively by visual inspection of the nonuniformity corrected volumes, their intensity histograms, and the corresponding bias volumes (see below); accuracy was also assessed quantitatively by comparing simulated bias volumes applied to an MRI brain phantom with the bias volumes extracted by each algorithm. Precision was evaluated by performing a PCA/CVA analysis on sets of six corrected repeat scans of a single subject (see below) and examining a scatter plot of the first two canonical variables (Strother *et al.*, 1996). All NUC algorithms were run and evaluated on PC/Linux workstations.

Algorithms. Of the six algorithms selected for comparison, four—*n3* (Sled *et al.*, 1998), *hum* (Brinkmann *et al.*, 1998), *eq* (Cohen *et al.*, 2000), and *bfc* (Shattuck *et al.*, 2000)—have been described *in extenso* in the scientific literature, one (*spm*) (Ashburner and Friston, 1998) is part of the SPM99 software package, and one (*cma*) was developed and provided at no cost by the Center for Morphometric Analysis at the Massachusetts General Hospital.

An executable version of the *n3* algorithm was provided by Dr. A. C. Evans at the Montreal Neurological Institute, and program default values were used for all run-time parameters. The *bfc* algorithm was run with a fixed set of three required run-time parameters: (spline spacing = 96, bias estimate spacing = 24 and histogram block radius = 48). Source code for *eq* was provided by Dr. M. Cohen at the Brain Mapping Center at UCLA and compiled at the Minneapolis VA Medical Center. By default, *eq* calculates a "noise" threshold based on a histogram of the raw input volume; however, we observed slightly better results when a masked input volume (i.e., nonbrain voxels set to 0) was used with the noise threshold set to 1; program defaults were used for the other run-time parameters. The *hum* algorithm was implemented based on the published method (Brinkmann *et al.*, 1998). Although we tested two different values for the block-size parameter, all reported results were produced with a blocking window of 13 × 13 × 13 cm. The code for *cma* was extracted from the Nautilos library from the Center for Morphometric Analysis at the Massachusetts General Hospital and recompiled; no run-time parameters were required. The *spm* algorithm, provided as part of the SPM99 software package,[2] ran without required parameter input. Minimal pre- and postprocessing was required in order to reconcile the data formats of our test volumes with those of *n3*, *cma*, and *spm*. Because *spm* did not provide output volumes representing the estimated nonuniformity (bias volumes), postprocessing was required to calculate the bias volume from the uncorrected and nonuniformity-corrected volumes.

Brain masking. For all test volumes, a stripping mask was generated by a consensus technique (Rehm *et al.*, 1999) to remove skull, meninges, and blood vessels. Masked input volumes with nonbrain voxels set to 0 were used for *n3*, *bfc*, *eq*, and *hum;* for *cma* and *spm*, unstripped input volumes were used, and the stripping mask was applied to the output volumes (corrected data and bias) prior to subsequent quantitative analysis. [*cma* required an unstripped input volume, and *spm* produced different bias fields for the brain when stripped and unstripped volumes were input.]

Visual analysis. All six NUC algorithms were applied to a single T1-weighted MRI volume (see below), and the estimated bias fields were extracted—or computed in the case of *spm*. Intensity histograms of the original, and corrected volumes and voxel-wise scatter plots of the original vs each of the six corrected volumes

[1] The data sets described in this study, as well as BIAS_COMPARE.PRO, the IDL procedure used to generate Figs. 1–4, are available for download at http://pet.med.va.gov:8080/papers/tech_reports/bias_correction.html.

[2] SPM99 is available for download at http//www.fil.ion.ucl.ac.uk/spm/spm99.html.

were computed. Images of an axial slice through the basal ganglia of the original volume and the corresponding slices from the six nonuniformity corrected volumes were visually compared.

Phantom studies. We evaluated the performance of the six NUC algorithms on a set of six brain images created by adding known biases to the Montreal Brain Phantom (MBP) with no added noise (http://www.bic.mni.mcgill.ca/brainweb/selection_normal.html; Collins *et al.,* 1998). Bias volumes with different spatial distributions were created as the product of (i) three orthogonal parabolic functions and (ii) three orthogonal sinusoidal functions. Each simulated bias volume was computed at three different magnitudes of bias (± 2, ± 4, and $\pm 8\%$) centered about a value of 1.0; the periods of the sinusoids varied from 0.8 to 1.2 times the cardinal dimensions of the MBP. The six resulting bias volumes were applied multiplicatively to the MBP followed by the addition of Gaussian random noise (mean = 0, SD = 4.0; mean signal value of white matter ~115) to produce the test set of phantom brain volumes. After the NUC algorithms had been run on the phantom test set, the corrected volumes were normalized to have the same mean as the original, unbiased volume. To account for differences between the correction algorithms with regard to their representation of the estimated nonuniformity and, thereby, to allow for meaningful comparisons, all calculated bias volumes were appropriately scaled and offset.

Performance of the NUC algorithms was also evaluated by comparing the corrected volumes to the unbiased phantom volume and by comparing the extracted bias volumes to the bias volumes that had been applied to the MBP. The root-mean-squared (RMS) error between the corrected and unbiased phantom volumes and between the extracted and applied bias volumes was calculated on a slice-by-slice basis and plotted for consecutive axial, coronal, and sagittal slices to facilitate the discrimination of spatial variations in the nonuniformity correction. The correlation coefficient between extracted and applied bias volumes was also used for quantitative comparisons. All measures were computed from the masked brain subvolume, and slices with fewer than 2000 brain voxels were excluded from the comparisons.

Repeat scans of a single subject. Six T1-weighted brain volumes of a normal volunteer were acquired over a 6-month period on a Siemens Vision 1.5T MRI scanner using the 3DFLASH protocol (TR 35 ms, TE 6 ms, flip angle 45°, one excitation, 165 × 220 mm FOV, 192 × 256 matrix, 100 slices, voxel dimensions 0.86 × 0.86 × 2.0 mm). Stripping (cortical isolation) masks were created for each volume as described above, and the raw volumes and their corresponding brain masks were aligned as follows: all possible six-parameter rigid-body transformations between pairs of scans were computed using AIR 3.0 (Woods *et al.,* 1998). The six-parameter rigid-body transformation matrices between any two scans ($\mathbf{T}_{ij}$) were used to obtain a consensus transformation by averaging the 4×4 homogeneous coordinate products, $\mathbf{T}_{ij}^{k} = \mathbf{T}_{ik}\mathbf{T}_{kj}$, over all values of k to form the average matrix $\langle \mathbf{T}_{ij}^{k} \rangle_k$. The average transformation matrix was then converted to a six-parameter rigid body transformation by using $\langle \mathbf{T}_{ij}^{k} \rangle_k$ to transform an evenly spaced $20 \times 20 \times 20$ point grid covering the average brain mask and then calculating the six-parameter Procrustes transformation of the original to the transformed grid. After the NUC algorithms were run on the six aligned brain volumes, each corrected data volume was mean-normalized to the corresponding input volume to facilitate quantitative comparisons between algorithms.

PCA/CVA. To determine the within-algorithm similarity and between-algorithm differences for the six-repeat-scan evaluation, we employed a voxel-based principal component analysis followed by a canonical variables analysis (PCA/CVA) to the nonuniformity-corrected volumes (Strother *et al.,* 1996). PCA, performed on 42 [six uncorrected and $6 \times 6 = 36$ nonuniformity-corrected] MRI volumes, was followed by a seven-group CVA [no correction and six NUC methods]. A scatter plot of canonical variates (CVs) from the significant dimensions was then generated, and performance was evaluated under the assumption that the "best" algorithm would produce the tightest cluster of nonuniformity-corrected volumes. The canonical eigenimages associated with the significant CVs were displayed for assessment of spatial differences between NUC algorithms.

Stability. In order to assess the stability of the six NUC algorithms, we recursively applied each of algorithms to its corrected output volume for five iterations following the initial correction. Results were evaluated under the assumption that an ideal algorithm would completely remove any bias on the first pass and that subsequent applications of the algorithm would have no further effect. Therefore, we assumed that the "goodness" of a correction algorithm is related to the stability of its output, i.e., successively extracted bias volumes rapidly approach uniformity—recognizing, of course, that a "bad" method could converge to a bad solution.

High-resolution MRI scans. The performance of the NUC algorithms on two University of Minnesota (UMN) MRI scans (1.5T Siemens Vision scanner, axial 3DFLASH, TR 35 ms, TE 6 ms, TI 300 ms, flip angle 45°, one excitation, voxel size 0.82 × 0.82 × 2.0 mm, acquired 10 months apart) and two UCLA MRI scans (3.0T GE Signa scanner, sagittal 3DGRASS, TR 24 ms, TE 4 ms, flip angle 35°, one excitation, voxel size 1.2 × 0.98 × 0.98 mm, acquired 3 months apart), all of different normal subjects, was compared. For each of the

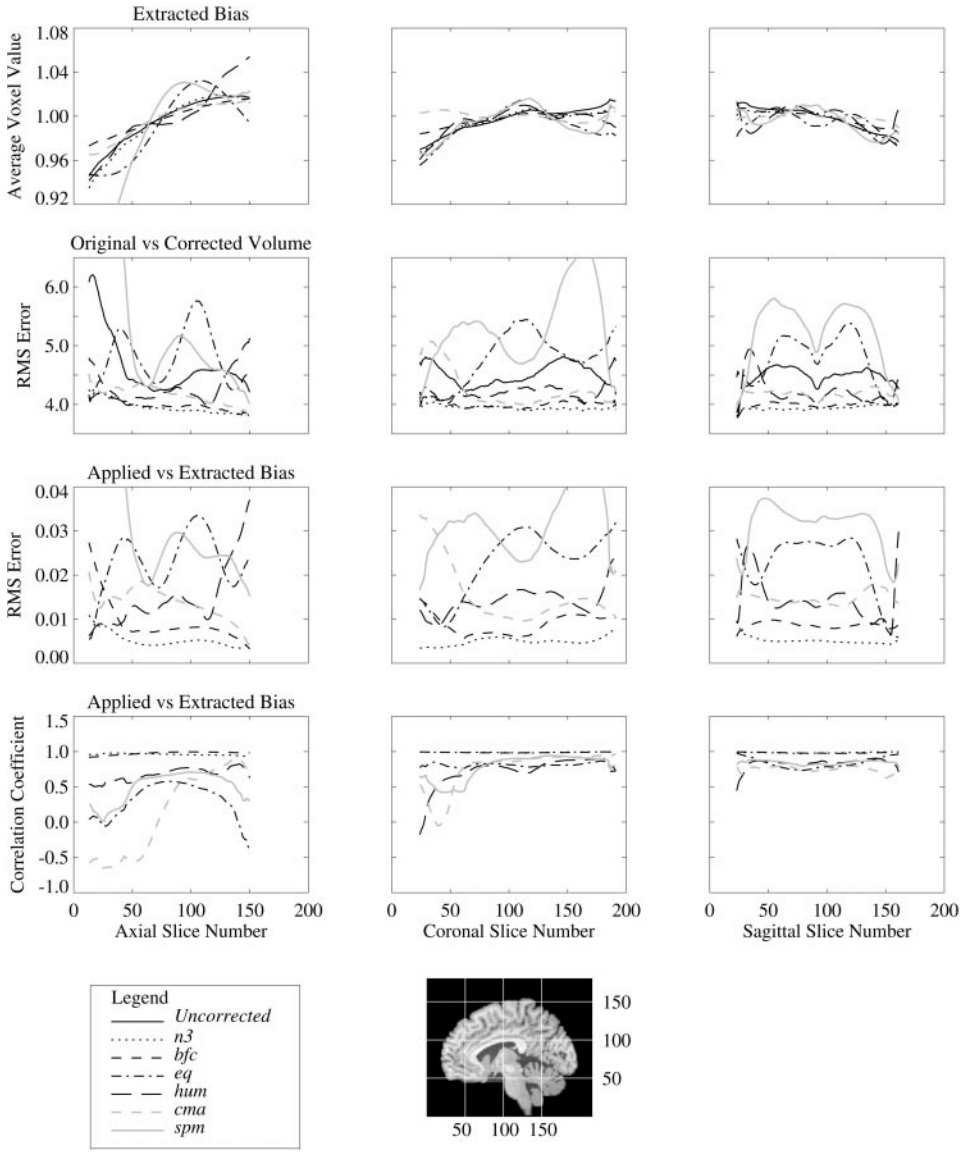

FIG. 1. Comparison of six NUC algorithms using the Montreal Brain Phantom with ±8% paraboloidal bias. Columns (left to right) represent measurements derived from consecutive axial, coronal, and sagittal slices. Top row, Average voxel values for the applied (solid black line) and extracted bias volumes. Second row, RMS error between nonuniformity-corrected and unbiased volumes. The solid black line represents the RMS "error" between the biased and unbiased volumes. Gaussian random noise was added after biasing the phantom volume, so that all measures include an RMS error of ~4.0 based on noise alone. Third row, RMS error between the extracted and applied bias volumes. Fourth row, Voxel-wise correlation coefficient for the extracted vs applied bias volumes.

four scans, the set of nonuniformity-corrected volumes plus the uncorrected volume were subjected to a principal component analysis. Scatter plots of the first two scaled eigenvectors from each PCA were compared to evaluate pattern similarity among the five algorithms across scanning platforms. Accuracy was assessed by visual inspection of the corrected volumes and the extracted bias fields.

RESULTS

Phantom studies. Comparisons of the unbiased and nonuniformity-corrected MBP, and the applied and extracted bias volumes for the ±8% paraboloidal bias are summarized in Fig. 1. Comparison measures were computed for consecutive axial, coronal, and sagittal slices. The *n3* and *bfc* bias estimates most closely approximate the applied bias, whereas the *eq*, *hum*, *cma*, and *spm* estimates differ noticeably from the applied bias (Fig. 1, top row). In general, deviation from the applied bias is greatest along the axial dimension.

The RMS error between the nonuniformity-corrected volumes and the unbiased MBP is plotted in Fig. 1 (second row), together with the RMS "error" (RMS bias) between the biased and unbiased MBP; because Gaussian random noise with a standard deviation of

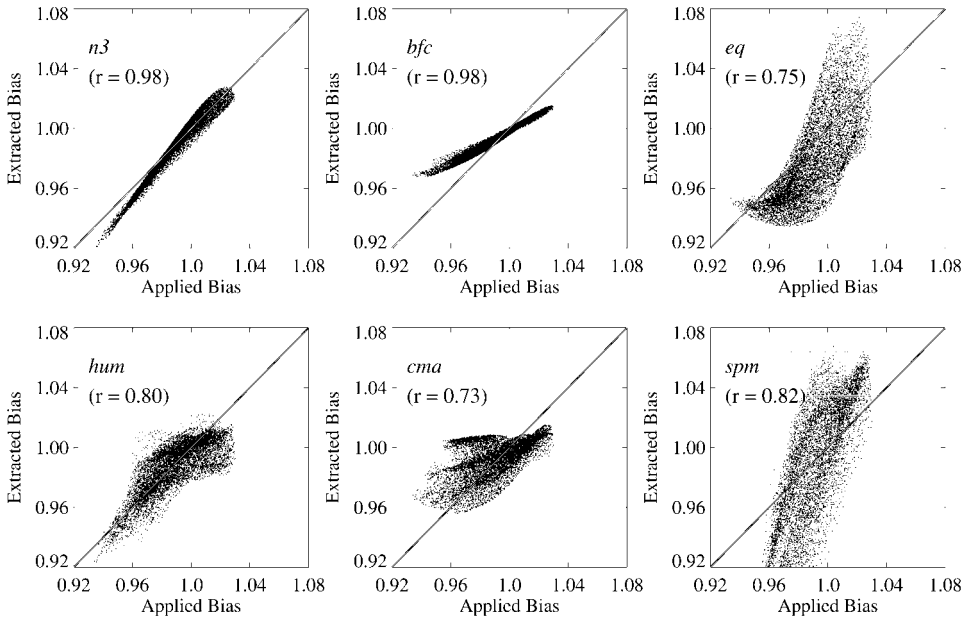

FIG. 2. Voxel-wise scatter plots of the extracted vs applied (±8% paraboloidal) bias volumes for the six NUC algorithms; *r* values are linear Pearson correlation coefficients.

4.0 was added after the bias was applied, these RMS errors have an expected minimum value of approximately 4.0. For the *n3*- and *bfc*-corrected volumes, RMS errors in the nonuniformity-corrected volumes are close to 4.0, implying that the applied bias has been nearly completely removed. For the *hum*- and *cma*-corrected volumes, RMS errors generally range between 4.0 and those for the biased phantom, suggesting that the applied bias has been only partially removed. RMS errors for the *eq*- and *spm*-corrected volumes range above those for the biased phantom, indicating that *eq* and *spm* have introduced nonuniformity.

Comparing the applied and extracted bias volumes (Fig. 1, third row), the low RMS errors for *n3* and *bfc* indicate that the estimated bias is similar to the applied bias, although the error for *bfc* increases in the low axial slices; RMS errors for *eq*, *hum*, *cma*, and *spm* are higher and more variable across the volume. Comparing the slice-by-slice correlation coefficients for the estimated vs applied bias (Fig. 1, fourth row), *n3* and *bfc* estimates are consistently highly correlated with the applied bias across all slices. Correlations for *hum*, *eq*, *cma*, and *spm* vary widely and are generally low across the axial slices and higher and more consistent across the coronal and sagittal slices. Scatter plots of the applied ±8% paraboloidal bias vs the extracted bias for the six algorithms (Fig. 2) reinforce the conclusions derived from slice-wise correlations plotted in Fig. 1—that *n3* and *bfc* most accurately capture the shape of the applied bias.

The plots in Fig. 3, organized in the same manner as those in Fig. 1, summarize comparisons for the ±8% sinusoidal bias. The top-row graphs, illustrating average voxel values from the applied and extracted bias volumes, are similar to those in Fig. 1. The *n3* bias estimate is most similar to the applied bias, while the *bfc* estimate is marginally poorer for the sinusoidal bias than it was for the paraboloidal bias (cf. Fig. 1, top row). The profiles for the sinusoidal and paraboloidal biases extracted by the other four algorithms are similar, suggesting that they are influenced by the underlying anatomy. As illustrated in Fig. 3 (second row), the *n3*-corrected volume has the smallest RMS errors relative to the original phantom, barely above the 4.0 noise level, while RMS errors for *bfc*, *hum*, and *cma* are slightly higher and more variable than those produced by the same algorithm for the paraboloidal bias corrections (cf. Fig. 1, second row). RMS errors for the *eq*- and *spm*-corrected volumes are similar in magnitude to those in Fig. 1. As regards the RMS error in the extracted bias volumes, *n3* estimates vary least from the applied bias except in the lower axial slices, followed by *bfc* and *hum* estimates and then by *cma*, *eq*, and *spm* estimates (Fig. 3, third row).

Slice-wise correlations between the applied and extracted bias volumes are illustrated in Fig. 3 (fourth row). Correlations for the *n3* and *bfc* bias volumes were slightly lower in the axial direction than in the paraboloidal-bias comparison (cf. Fig. 1, fourth row). For coronal and sagittal slices the correlations for *eq*, *hum*, *cma*, and *spm* were comparable to those illustrated in Fig. 1; however, the patterns of correlation were very different for the four algorithms along the axial direction—for *hum*, the correlation was higher for the pa-

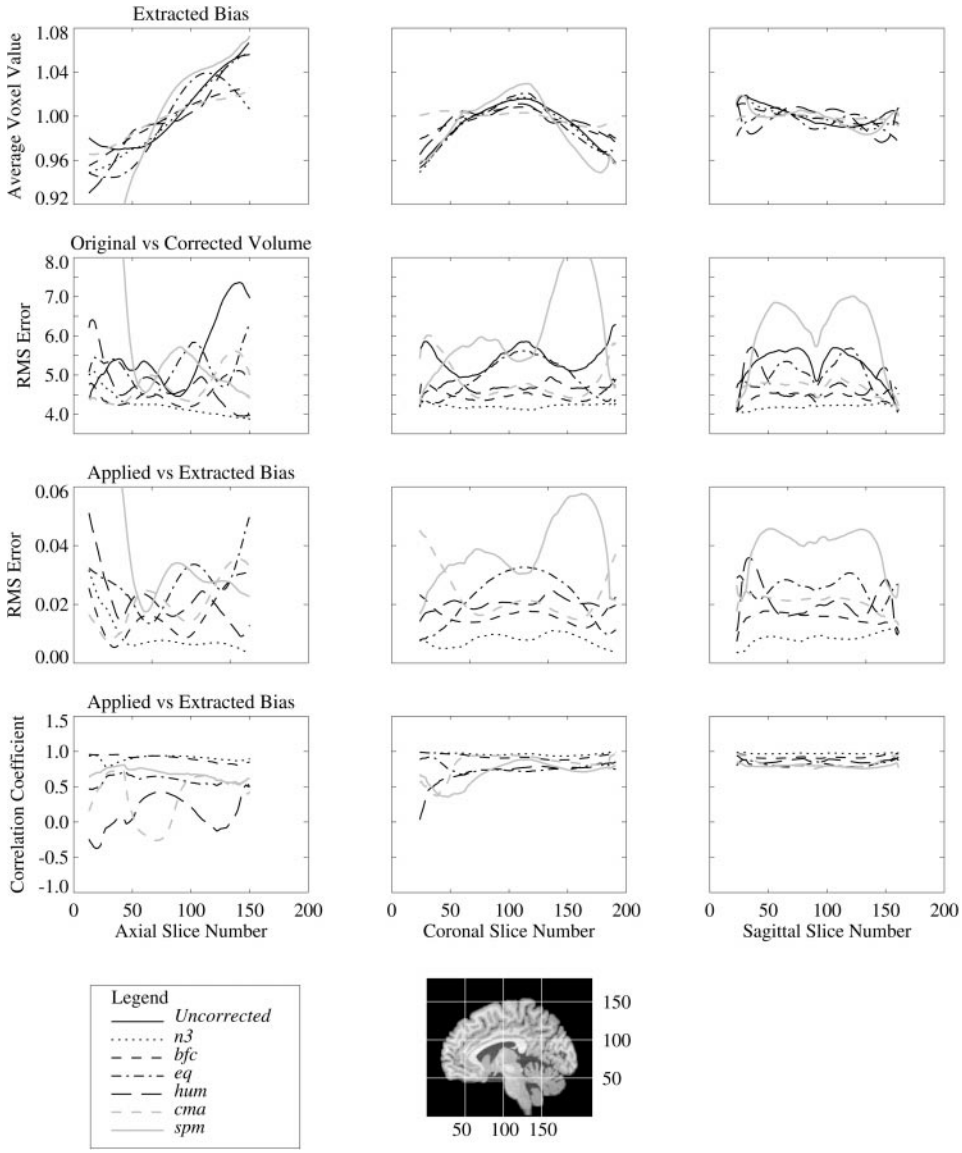

FIG. 3. Comparison of six NUC algorithms using the Montreal Brain Phantom with ±8% sinusoidal bias. Columns (left to right) represent measurements derived from consecutive axial, coronal, and sagittal slices. Top row, Average voxel values for the applied (solid black line) and extracted bias volumes. Second row, RMS error between nonuniformity-corrected and unbiased volumes. The solid black line represents the RMS "error" between the biased and unbiased volumes. Gaussian random noise was added after biasing the phantom volume, so that all measures include an RMS error of ~4.0 based on noise alone. Third row, RMS error between the extracted and applied bias volumes. Fourth row, Voxel-wise correlation coefficient for the extracted vs applied bias volumes.

raboloidal bias, whereas the reverse was true for *eq*, *cma*, and *spm*.

Scatter plots of the applied ±8% sinusoidal bias vs the extracted bias for the six algorithms are illustrated in Fig. 4. Comparison of the plots in Fig. 4 with those in Fig. 2 demonstrates that the *n3* and *bfc* algorithms performed better with the paraboloidal bias than with the sinusoidal bias, and, in both cases, outperformed the other four algorithms.

Comparisons identical to those summarized in Figs. 1–4 were also made for corrections of ±2% and ±4% for both paraboloidal and sinusoidal biases. In general, the relative ordering of the six algorithms was comparable at the three levels of bias for each of the two bias shapes; however, for *bfc* the estimated bias was as close or closer to the applied bias than the *n3* estimate at the lower bias levels.

Repeat scans of a single subject. Figure 5 illustrates a representative slice from the index (first) MRI brain volume before and after nonuniformity correction. There is little difference in the appearance of the *n3*-,

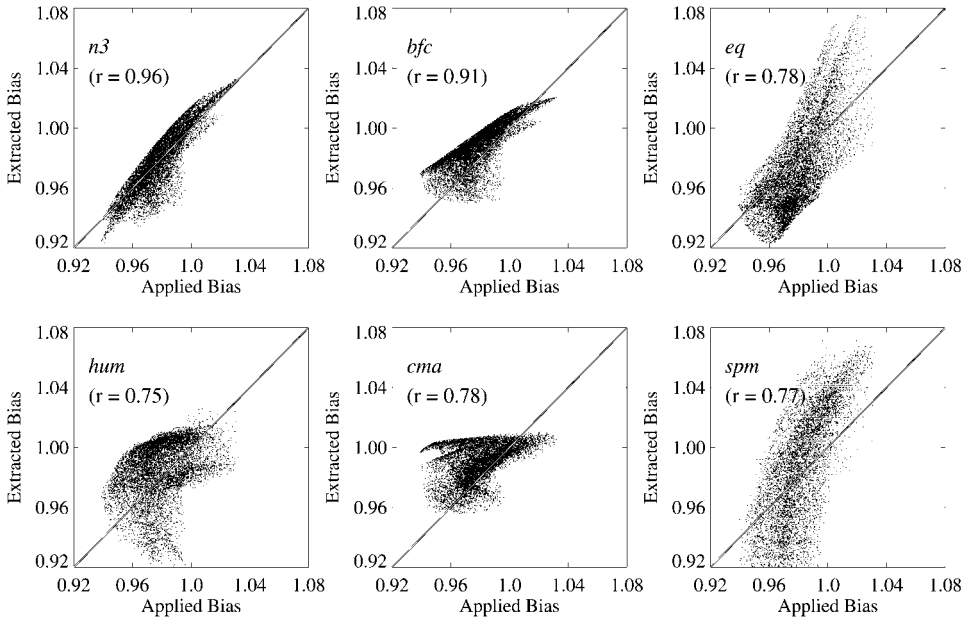

FIG. 4. Voxel-wise scatter plots of the extracted vs applied (±8% sinusoidal) bias volumes for the six NUC algorithms; *r* values are linear Pearson correlation coefficients.

bfc-, *eq*-, *hum*-, and *cma*-corrected slices, whereas white-matter nonuniformity is slightly increased in the *spm*-corrected image. Intensity histograms of the masked brain data from the uncorrected index volume and the six nonuniformity-corrected volumes are illustrated in Fig. 6. A composite graph (Fig. 6A) demonstrates that the corrected histograms are similar to each other and to the histogram of the uncorrected volume, with distinct GM and WM peaks; however, the *spm*-corrected volume demonstrates a reduction in the height of the gray-matter peak and a global shift of the white-matter peak to lower intensity values (Fig. 6B). Voxel-wise scatter plots of the uncorrected index volume vs each of the six nonuniformity-corrected volumes are illustrated in Fig. 7. The *n3*-, *bfc*-, *eq*-, *hum*-, and *cma*-corrected volumes are more highly correlated with the original biased volume than the *spm*-corrected volume.

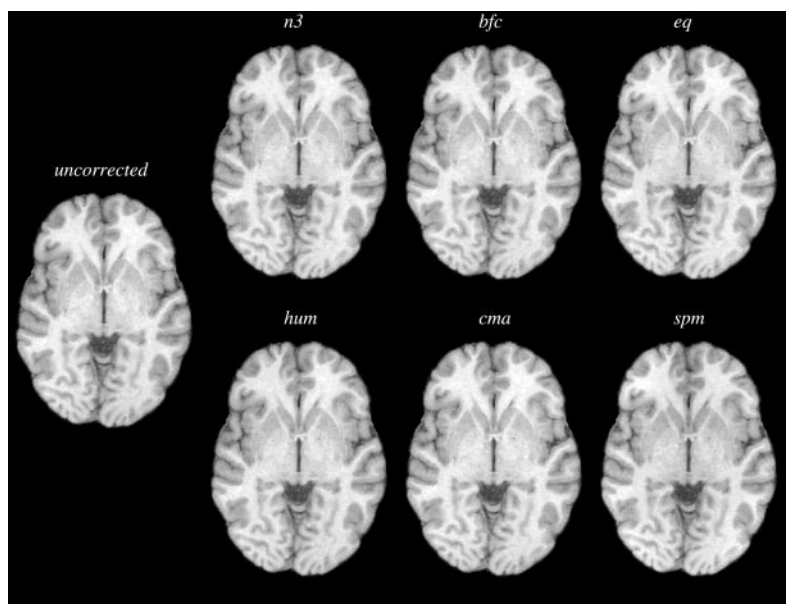

FIG. 5. Brain slice from a single MRI brain volume before and after nonuniformity correction. There is little difference in the appearance of the *n3*-, *bfc*-, *eq*-, *hum*-, and *cma*-corrected slices. White-matter nonuniformity is slightly increased in the *spm*-corrected image.

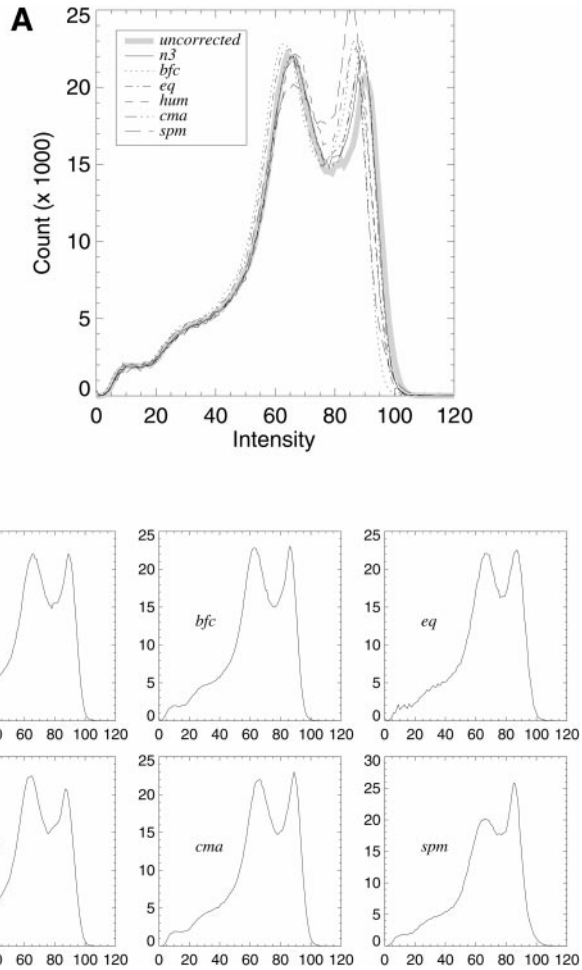

FIG. 6. (A) Intensity histograms of an MRI volume before and after nonuniformity correction. (B) Individual intensity histograms of the six nonuniformity-corrected volumes. Histograms of the *n3*-, *bfc*-, *eq*-, *hum*-, and *cma*-corrected volumes are similar to each other and to that of the uncorrected volume, whereas the histogram of the *spm*-corrected volume demonstrates a reduction in the height of the gray-matter peak and a global shift of the white-matter peak to lower intensity values.

Figure 8 illustrates the bias fields extracted by the six correction algorithms for the brain slice in Fig. 5 for each of the six repeat scans. Bias fields extracted by *n3*, *bfc*, and *spm* have a low-spatial-frequency pattern, whereas those extracted by *eq*, *hum*, and *cma* incorporate higher spatial frequencies and appear to be significantly influenced by the underlying brain anatomy.

PCA/CVA. A scatter plot of the first two canonical variates (CV1, CV2) derived from a voxel-based PCA/CVA of the six repeat scans before and after nonuniformity correction is illustrated in Fig. 9A. [*spm* was excluded because 98% of the variance in the data decomposition was accounted for by the difference between *spm* and the other five algorithms along the CV1* dimension (inset).] The first dimension (CV1) is driven by *eq*, whereas the second dimension (CV2) is driven by *cma* and *hum*, although in opposite directions. Together, the first two dimensions account for approximately 84% of the total variance in the CVA decomposition. The clustering of CV values demonstrates that the between-group (algorithmic) differences are, in all cases, greater than the within-group differences (repeat scans). The first canonical eigenimage (corresponding to CV1, Fig. 9B) contains a central hypodensity and prominent structural detail, which reflect differences between the *eq*-corrected volumes and those corrected by the other algorithms. The second canonical eigenimage (corresponding to CV2, Fig. 9C) illustrates a predominantly axial effect.

Stability. Figure 10 illustrates the results of applying the six algorithms recursively to their respective corrected output for five iterations after the initial correction. The mean (dot) and 5th-to-95th percentiles of the data range (bar) are plotted for each successive application. Only *n3* and *bfc* behaved in accordance with the assumption that the mean signal value of successively extracted bias volumes should rapidly approach a constant value, i.e., the variation of successively extracted bias volumes should rapidly approach zero. For *eq* and *cma*, the mean signal value appeared to be slowly converging. For *hum*, the magnitude of the correction did not change appreciably after five iterations; *spm* behaved erratically.

High-resolution MRI scans. Scatter plots of the first two scaled eigenvectors from separate PCAs of the two UMN high-resolution scans are illustrated in Fig. 11; *spm* was excluded once again in order to appreciate differences in the behavior of the other five algorithms. In both scans the first two principal components accounted for approximately 85% of the total variance of the decomposition. The *bfc*-, *cma*-, and *hum*-corrected volumes lie closest to the uncorrected volume; the *n3*- and *eq*-corrected volumes lie farthest away. Scatter plots of the first two scaled eigenvectors from the two UCLA PCAs, which account for approximately 87% of the total variance, are also illustrated in Fig. 11. The patterns of the two UCLA plots are remarkably similar, though they differ from the UMN pattern. In the UCLA plots, the *cma*-corrected volume lies closest to the uncorrected volume, followed by *bfc* and *hum;* once again, the *n3*- and *eq*-corrected volumes lie farthest away.

For each of the four high-resolution brain volumes, a single brain slice taken from the uncorrected MRI volume and each nonuniformity-corrected volume, and the corresponding slice from each of the extracted bias volumes are illustrated in Fig. 12. Although the general shape of the bias differs between the two scanners, the relative behavior of the five algorithms is similar. Both *n3* and *bfc* detected a low-spatial-frequency inhomogeneity, although the magnitude of the bias detected by *n3* was always greater. The performance of

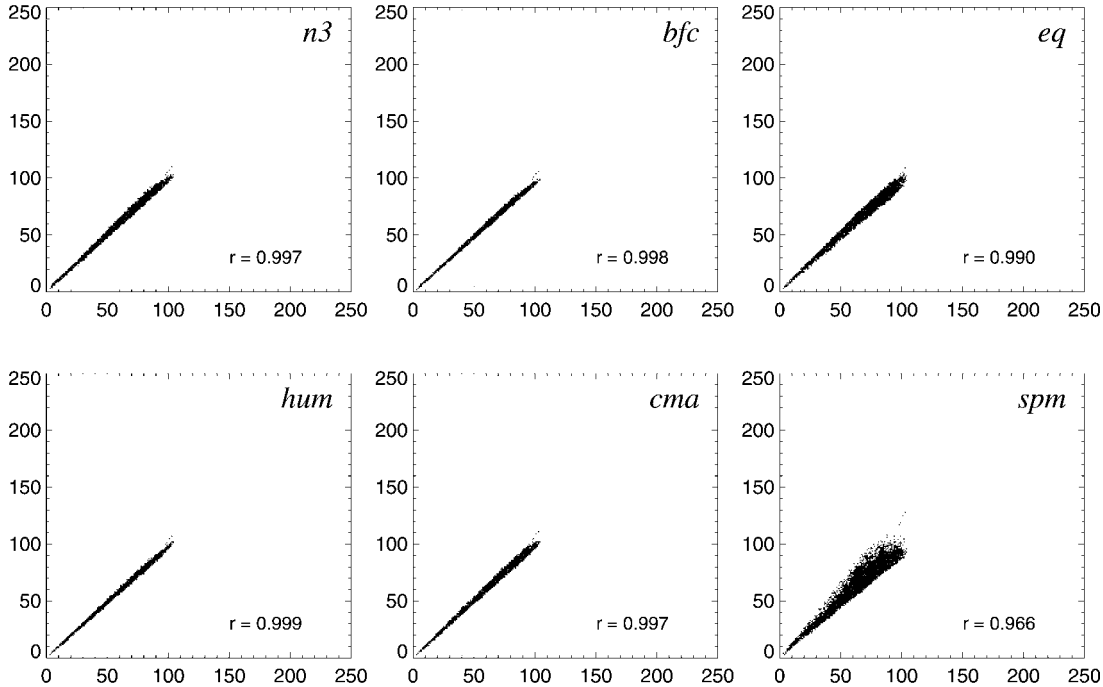

FIG. 7. Voxel-wise scatter plots of an uncorrected MRI volume (*x*-axis) vs each of six nonuniformity-corrected volumes (*y*-axis). The *n3*-, *bfc*-, *eq*-, *hum*-, and *cma*-corrected volumes are more highly correlated with the original uncorrected volume than is the *spm*-corrected volume.

eq, *hum*, and *cma* appears to be influenced to some degree by the underlying anatomic structure, as evidenced by their extracted bias volumes; this effect was greater for *eq* than for *hum* and *cma*.

DISCUSSION

All six of the NUC algorithms that we compared estimate intensity nonuniformity based on a *post hoc*

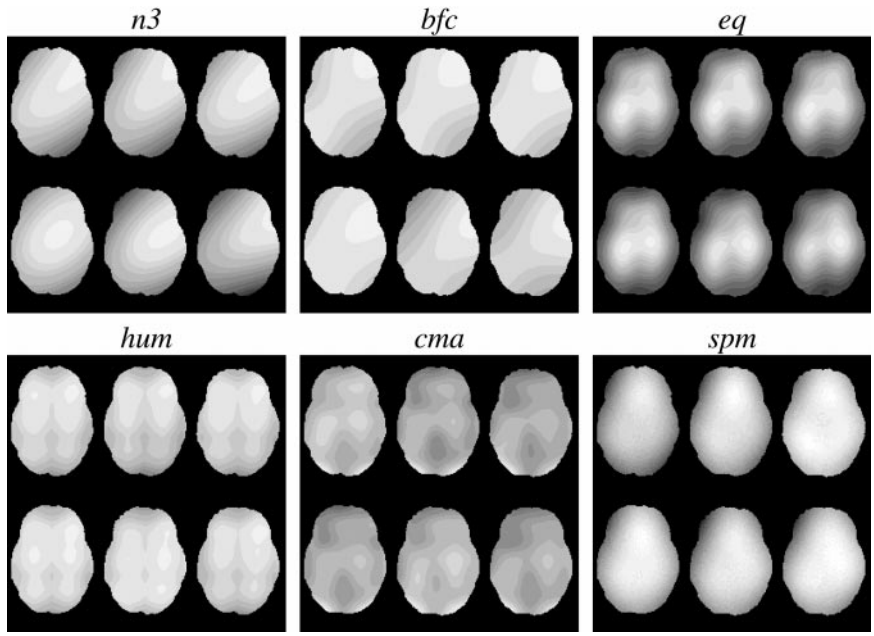

FIG. 8. Images of nonuniformity "extracted" by the six different correction algorithms (one slice from each of six repeat scans). The *n3*, *bfc*, and *spm* images exhibit a low-spatial-frequency pattern; however, the *spm* pattern appears to reflect the underlying brain anatomy. The *eq*, *hum*, and *cma* patterns include higher spatial frequencies and appear to be significantly influenced by brain anatomy.

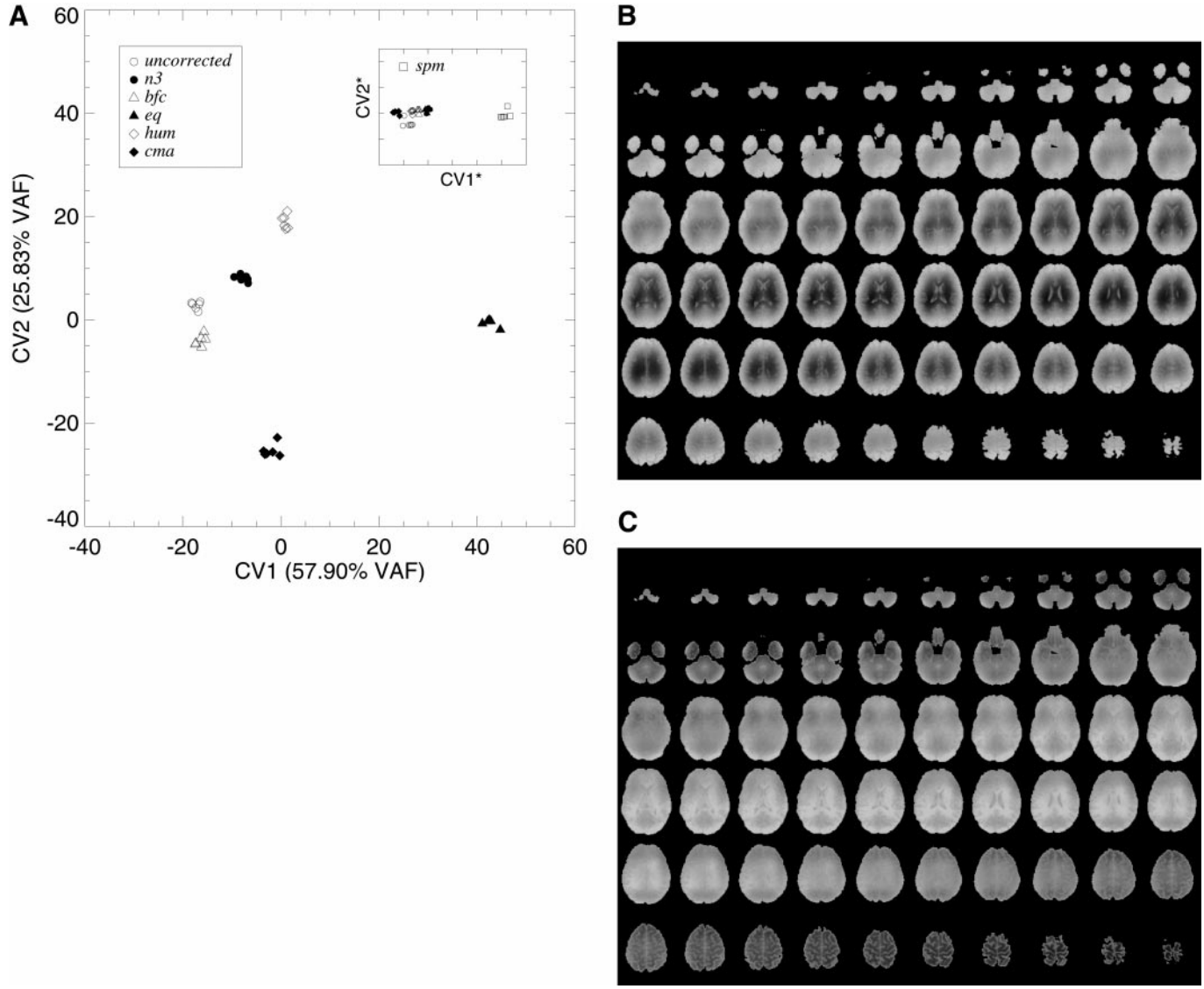

FIG. 9. (A) Scatter plot of the first two canonical variates (CV1, CV2) from a canonical variables analysis of six repeat scans before and after nonuniformity correction; *spm* was excluded from the analysis, because 98% of the variance in the data decomposition was accounted for by the difference between *spm* and the other five methods along the CV1* dimension (inset). Note that the *eq* cluster (closed triangles) is separated from the other four corrected volume clusters and the uncorrected volumes along the CV1 axis, while the CV2 axis separates the *hum* and *cma* clusters (closed and open diamonds) from the other three corrected volume sets and the uncorrected volumes. (B) The canonical eigenimage corresponding to CV1. Note the central hypodensity and the prominence of structural detail. (C) The canonical eigenimage corresponding to CV2. Nonuniformity is most marked along the axial direction. Note that (B) and (C) are scaled differently.

analysis of individual T1-weighted MRI image volumes. In general, *post hoc* techniques depend on the assumption that image inhomogeneity is a low-frequency spatial variation that can be distinguished from higher-frequency components representing anatomic information; these techniques can be broadly divided between methods based on low-pass filtering and those which evaluate spatial variations in tissue intensity parameters.

Filter-based correction methods derive from well-established image processing theory (Brinkmann *et al.*, 1998). Three of the methods that we compared—*hum*, *eq*, and *cma*—represent variations of low-pass filtering. While *hum* and *cma* filter image data in the spatial domain, *eq* uses Fourier methods to filter in the frequency domain. The *cma* algorithm further modifies the process by filtering a thresholded subvolume representing white matter and then extrapolating the estimated inhomogeneity to the entire brain volume.

In contrast to filtering methods, the *bfc* algorithm utilizes an approach based on normalization of regional tissue intensity histograms to global values, under the assumption that tissue parameters are consistent throughout the MRI volume. The *n3* algorithm repre-

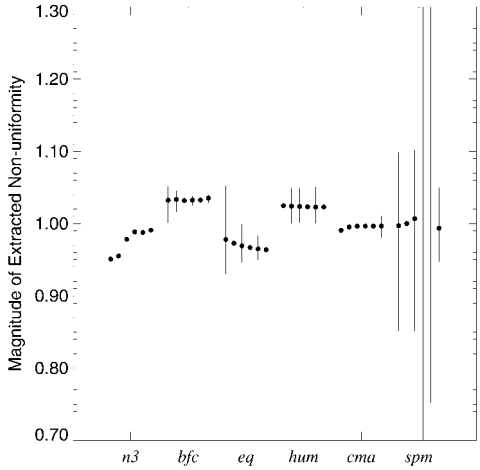

FIG. 10. Summary plot of the mean and range of extracted nonuniformity for the initial and five iterations of *n3*, *bfc*, *eq*, *hum*, *cma*, and *spm*.

sents an elaboration of tissue signal analysis, using iterative modeling of low-frequency spatial variations in the data to maximize high-frequency information in the intensity histogram of the corrected volume. The *spm* algorithm estimates intensity nonuniformity as a low-frequency component in an iteratively optimized tissue mixture model initialized with *a priori* information about anatomic tissue distributions.

The relative amount of nonuniformity in our T1-weighted MRI volumes was generally small; thus, one might expect that the differences between uncorrected and corrected volumes should also be small. This was indeed the case for normal-subject scans, as indicated by the high correlations of the scatter plots for five of the six algorithms illustrated in Fig. 7. Visual inspection of these scatter plots suggests that *spm* and, to some extent, *eq* may have introduced additional nonuniformity, producing relatively greater scatter about

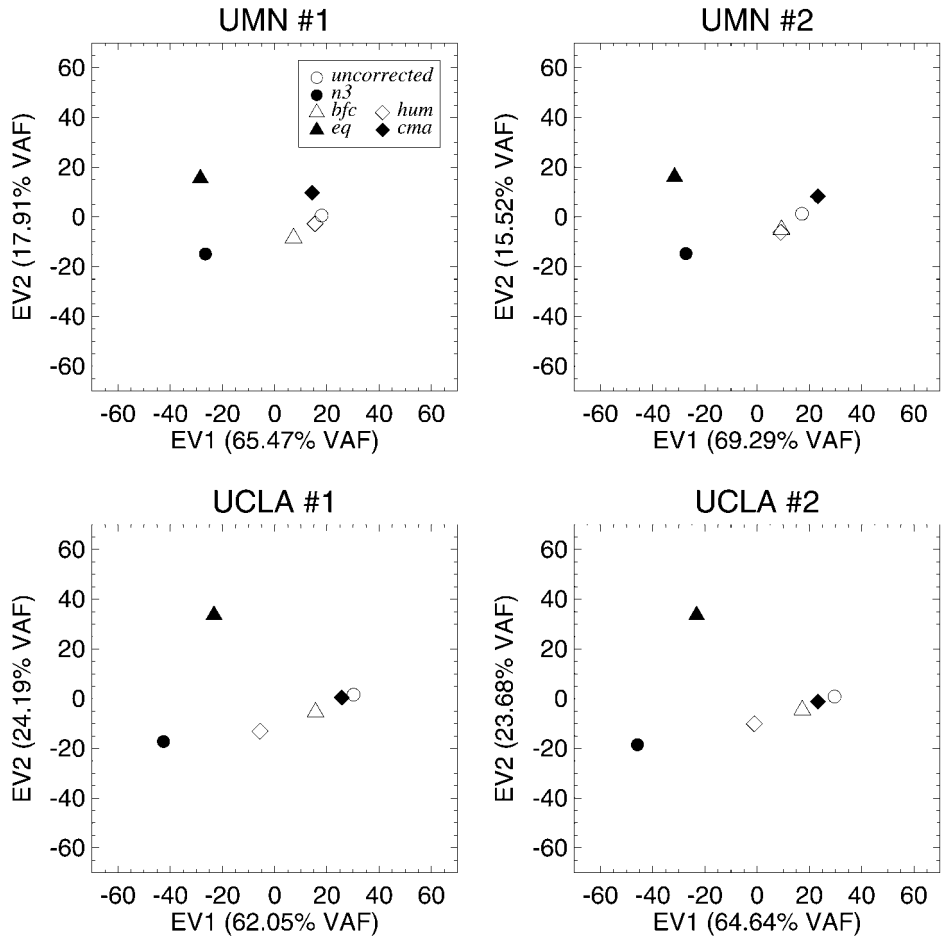

FIG. 11. Top row, Scatter plots of the first two scaled eigenvectors (EV1 and EV2) from principal component analyses of uncorrected and five corrected high-resolution UMN MRI volumes acquired ten months apart. Bottom row, Scatter plots of EV1 and EV2 for a similar analysis of two high-resolution UCLA MRI scans acquired three months apart. The similarity of the 2-D patterns can be quantified as follows: in each group (UMN1, UMN2, UCLA1, UCLA2), each method is represented as a 2-D vector originating at the center of gravity of the cluster of *x*,*y* coordinates. Pairwise similarities between the positions of each method in the plot are then summarized by computing the 6-choose-2 dot products of the vectors for each pair of methods. From these data one can compute the average within-site and between-site correlations, 0.879 and 0.708, respectively.

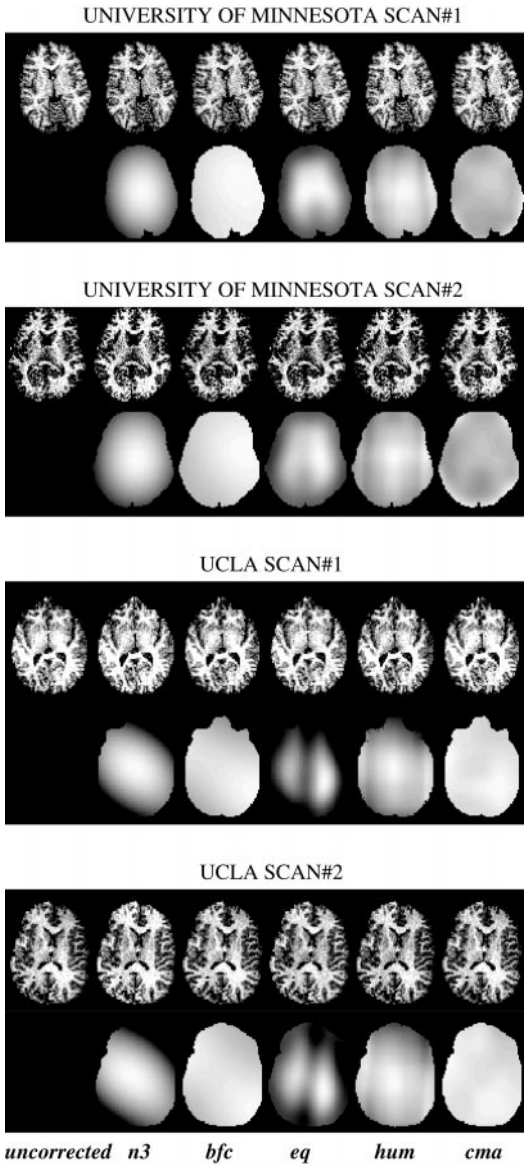

FIG. 12. A representative slice from the uncorrected and corrected volumes (upper rows) and from the extracted bias volumes (lower rows) for each of the UMN and UCLA high-resolution MRI brain scans.

the diagonal. This is consistent with the performance of *eq* and *spm* in the phantom studies and with the large RMS error in *eq*- and *spm*-corrected phantom volumes.

In the PCA/CVA analysis, corrected volumes produced by the four non-locally adaptive methods (*hum*, *eq*, *cma*, and *spm*) differed most from the uncorrected volumes; inspection of the canonical eigenimages (Figs. 9B and 9C) suggests that *eq* contaminates the extracted bias with anatomical information and that *cma* and *hum* incorrectly estimate the relative magnitude of the bias along the axial dimension (although in opposite directions).

In the principal component analyses of four subjects from two scanners (Fig. 11), the only source of variation in each of the PCA data sets was the effect of the nonuniformity corrections. The fact that the pattern of scaled eigenvectors in the scatter plots was similar for the two subjects scanned at each location, *but different for the two scanners*, demonstrates that differences between scanners may also influence the relative performances of NUC methods.

An alternative to evaluating the relative performance of different algorithms on real data sets by means of a quantitative benchmark is provided by the iterative application of the correction algorithms (Fig. 10). The rapid convergence of the output of *n3* and *bfc* to a static image volume suggests that the majority of identifiable bias has been removed during the first application of the correction.[3] The apparent lack of convergence in the iterated *hum* output suggests that something other than nonuniformity is being progressively removed from the image—or that nonuniformity is being introduced, and the erratic behavior of *spm* suggests that it is unstable when operating on relatively uniform image volumes.

Differences in algorithmic performance may be related to specific design considerations, e.g., to require input images that are masked to exclude nonbrain voxels, or to expect specific biases or highly nonuniform image volumes. Thus, as provided, *cma* requires unmasked volumes, whereas *eq* was optimized to operate on unmasked volumes[4] with a central hyperintensity. A related issue, which we did not address in this paper, is whether nonuniformity correction or brain masking should be performed first.

The use of multiple methods for evaluating the performance of NUC algorithms and the demonstration of similar rankings across different comparisons using real image data provide a basis for extending the results obtained with simulated or phantom data. While none of the algorithms that we evaluated performed ideally under all circumstances, locally adaptive methods (*n3* and *bfc*) generally outperformed nonadaptive methods (*eq*, *cma*, *hum*, and *spm*). The fact that the highest and lowest correlations of the extracted with the applied bias occurred in the transaxial (left-to-right) and axial (top-to-bottom) directions, respectively (Figs. 1 and 3), presumably reflects the relative degree of anatomic symmetry in the corresponding directions, and it seems reasonable to assume that locally adaptive techniques may be able to better accommodate anatomic asymmetry than nonadaptive techniques.

[3] For *n3*, the rate of convergence is somewhat arbitrary, since the algorithm is implemented as an iterative procedure with two stopping conditions; for *cma* and *eq*, the step-size parameter was, by default, set to 1.0.

[4] However, as noted above, *eq* performed better when the input volumes were masked.

ACKNOWLEDGMENTS

This study was supported by NIH Grants NS33718, MH57180, and MH52176.

REFERENCES

Ashburner, J., and Friston, K. 1998. MRI sensitivity correction and tissue classification. *NeuroImage* **7:** S706.

Bottomley, P., Foster, T., Argersinger, R., and Pfeiffer, L. 1984. A review of normal tissue hydrogen NMR relaxation times and relaxation mechanisms from 1-100 MHz: Dependence on tissue type, NMR frequency, temperature, species, excision and age. *Med. Phys.* **11:** 425.

Brinkmann, B. H., Manduca, A., and Robb, R. A. 1998. Optimized homomorphic unsharp masking for MR grayscale inhomogeneity correction. *IEEE Trans. Med. Imag.* **17:** 161–171.

Cohen, M. S., DuBois, R. M., and Zeineh, M. M. 2000. Rapid and effective correction of RF inhomogeneity for high field magnetic resonance imaging. *Hum. Brain Mapp.* **10:** 204–211.

Collins, D. L., Zijdenbos, A. P., Kollokian, V., Sled, J. G., Kabani, N. J., Holmes, C. J., and Evans, A. C. Design and construction of a realistic digital brain phantom. 1998. *IEEE Trans. Med. Imag.* **17:** 463–468.

Dale, A. M., Fischl, B., and Sereno, M. I. 1999. Cortical surface-based analysis I: Segmentation and surface reconstruction. *NeuroImage* **9:** 179–194.

DeCarli, C., Murphy, D. G., Teichberg, D., Campbell, G., and Sobering, G. S. 1996. Local histogram correction of MRI spatially dependent image pixel intensity nonuniformity. *J. Magn. Reson. Imag.* **6:** 519–528.

Hayes, C., Edelstein, W., Schenck, J., Mueller, O., and Eash, M. 1985. An efficient, highly homogeneous radiofrequency coil for whole-body NMR imaging at 1.5 T. *J. Magn. Reson.* **63:** 622–628.

Li, S., Collins, C. M., Dardzinski, B. J., Chin, C. L., and Smith, M. B. 1997. A method to create an optimum current distribution and homogeneous B1 field for elliptical birdcage coils. *Magn. Reson. Med.* **37:** 600–608.

Rehm, K., Shattuck, D., Leahy, R., Schaper, K., and Rottenberg, D. 1999. Semi-automated stripping of T1 MRI volumes: I. Consensus of intensity and edge-based methods. *NeuroImage* **9:** S86.

Shattuck, D. W., Sandor-Leahy, S. R., Schaper, K. A., Rottenberg, D. A., and Leahy, R. M. 2000. Magnetic resonance image tissue classification using a partial volume model. *NeuroImage,* in press.

Simmons, A. 1994. Sources of intensity nonuniformity in spin echo images at 1.5T. *Magn. Res. Med.* **32:** 121–128.

Sled, J. G., and Pike, G. B. 1998. Standing-wave and RF penetration artifacts caused by elliptic geometry: An electrodynamic analysis of MRI. IEEE *Trans. Med. Imag.* **17:** 653–662.

Sled, J. G., Zijdenbos, A. P., and Evans, A. C. 1997. A comparison of retrospective intensity nonuniformity correction methods for MRI. In *Lecture Notes in Computer Science: Information Processing in Medical Imaging* (J. S. Duncan and G. R. Gingi, Eds.), pp. 459–464. Springer, New York.

Sled, J. G., Zijdenbos, A. P., and Evans, A. C. 1998. A nonparametric method for automatic correction of intensity nonuniformity in MRI data. *IEEE Trans. Med. Imag.* **17:** 87–97.

Strother, S. C., Lange, N., Savoy, R. L., Anderson, J. R., Sidtis, J. J., Hansen, L. K., Bandettini, P. A., O'Craven, K., Rezza, M., Rosen, B. R., and Rottenberg, D. A. 1996. Multidimensional state-spaces for fMRI and PET activation studies. *NeuroImage* **3:** S98.

Vaughan, J. T., Hetherington, H. P., Otu, J. O., Pan, J. W., and Pohost, G. M. 1994. High frequency volume coils for clinical NMR imaging and spectroscopy. *Magn. Reson. Med.* **32:** 206–218.

Velthuizen R. P., Heine, J. J., Cantor, A. B., Lin H., Fletcher L. M., and Clarke, L. P. 1998. Review and evaluation of MRI nonuniformity corrections for brain tumor response measurements. *Med. Phys.* **25:** 1655–1666.

Woods, R. P., Grafton, S. T., Holmes, C. J., Cherry, S. R., and Mazziotta, J. C. 1998. Automated image registration: I. General methods and intrasubject, intramodality validation. *J. Comput. Assist. Tomogr.* **22:** 141–154.

Zaini, M. R., Schaper, K. A., Anderson, J. R., Shattuck, D. W., Liow, J-S., Sled, J. G., and Rottenberg, D. A. 1999. Method for evaluating the performance of nonuniformity correction algorithms. *NeuroImage* **9:** S99.