The Partial Volume Effect in PET / SPECT and Benford's Law

Kevin Wells*, John Chiverton, Mike Partridge, Miriam Barry, Haval Kadhem and Bob Ott

***Abstract*–In this work we consider a probabilistic methodology that models the intensity distributions found in pure and partial volume (PV) voxels. We introduce some methodological developments that enable explicit modeling of the PV voxels prior Probability Density Function (PDF). This new formulation can be applied generically across different imaging modalities including PET and SPECT. In this paper, we establish for the first time, that the prior PDF of voxels that arise from the PV effect in volumetric data can be well described by a simple phenomenological law called Benford's Law, which significantly eases parameter estimation compared to other methods. Results from simulated data are presented, along with a preliminary PET phantom study utilizing registered CT data to determine the quality of the resulting probabilistic voxel classification scheme.**

I. Introduction

Partial Volume (PV) correction techniques in PET or SPECT often rely on the use of high-resolution morphological data extracted from CT or MRI, registered onto the relatively low spatial resolution functional image data. This can then be used to associate structural boundaries with over and under spill from the associated functional distribution, as reviewed in [1].

In this work, we take an alternative approach. We consider a probabilistic methodology that enables this information to be extracted from the functional data directly using voxel intensities. We have previously outlined a methodology in [2], but we have now introduced some developments that enable explicit modeling of the voxel PV *a priori* Probability Density Function (PDF). This new formulation can be applied generically across different imaging modalities such as PET, SPECT as well as high-resolution techniques such as MR and CT.

The major contribution of this paper is to establish for the first time, (to the best of the authors' knowledge), that the *a priori* PDF of voxels that arise from the PV effect in volumetric data can be well described and conveniently represented by a phenomenological observation known as Benford's Law [3]. The *a priori* PV PDF is a fundamental descriptor of the probabilistic behavior of the PV effect. The only alternative for statistical PV analysis is to rely upon heuristic functional descriptions of the underlying *a priori* PV PDF [4]. These heuristic techniques require additional parameters to describe the shape of the PDF, in which, the true shape of the *a priori* PDF is difficult to achieve. The new knowledge that Benford's Law provides excellent governance of the *a priori* PV PDF will be of fundamental interest to those concerned with image quantification in PET/SPECT and other imaging modalities.

II. PV Quantification using Benford's Law

A. Benford's Law

Frank Benford originally observed that the first few pages of books of logarithm tables exhibited greater wear due to higher usage at the beginning of the book, compared to the latter pages. He then went on to observe the same effect in many other seemingly unrelated sources such as the frequency of alpha-numeric characters in newspapers, or the ordering of the atomic weights of molecules. Benford's Law has also found application for the detection of fraudulent IRS/tax returns [5]. This observed distribution can be quantitatively described by a PDF given by:

$$p(\beta) = \frac{1}{K}\log_{10}(1+\frac{1}{\beta}).U(\beta) \qquad (1)$$

where $U(\beta) = 1$ for $0 < \beta < 10$ otherwise $U(\beta) = 0$, and $K = \int_0^1 \log_{10}(1+\frac{1}{\beta}).d\beta$ is a PDF normalization constant. A plot of probabilities drawn from Benford's Law can be seen in fig. 1 below.

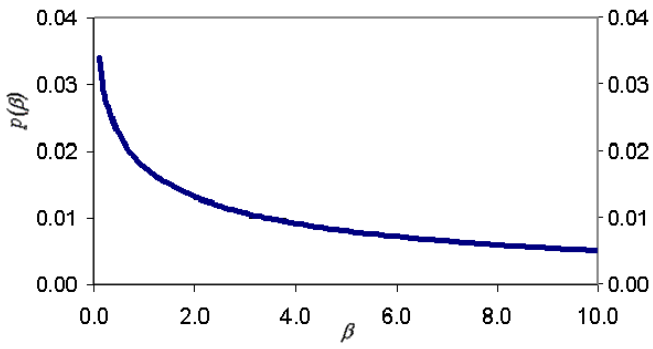

Fig. 1: Example of the PDF described by Benford's Law.

John Chiverton, Kevin Wells and Haval Kadhem are with the School of Electronics & Physical Sciences, University of Surrey, Guildford, Surrey GU2 5XH, UK.

Mirian Barry is now with University Hospital of North Staffordshire, Stoke-on-Trent, Northampton, ST4 7LN, UK.

Mike Partridge and Bob Ott are with the Joint Dept of Physics, Royal Marsden NHS Foundation Trust/Institute of Cancer Research, Sutton, Surrey SM2 5PT, UK.

*Corresponding author (email k.wells@eim.surrey.ac.uk)

B. PV Quantification

We have found that this PDF can be utilized within probabilistic descriptions of 3-D medical imaging data that possess voxels affected by the PV effect. One possible probabilistic description is given by [6]:

$$E[\boldsymbol{\alpha} \mid p(\boldsymbol{\alpha}|\mathbf{x})] = \int_{\alpha} \frac{\boldsymbol{\alpha}}{p(\mathbf{x})}\, p(\mathbf{x}|\boldsymbol{\alpha})\, p(\boldsymbol{\alpha}).d\boldsymbol{\alpha} \qquad (2)$$

where $p(\boldsymbol{\alpha}|\mathbf{x})$ is the probability of a voxel with a feature measurement vector $\mathbf{x}$, where $\mathbf{x}$ can be a multidimensional feature vector [2], but in this work limited to the 1-D signal intensity, g. $\boldsymbol{\alpha} = (\alpha_1\ \alpha_2\ ....\alpha_N)^T$ is a vector of scalar signal mixing values, where N is the number of signal components in the feature space. E.g. one might consider $N = 3$ for anatomical neurological imaging data consisting of components representative of White Matter (WM), Gray Matter (GM) and Cerebrospinal Fluid (CSF) for CT data or $N = 2$ for tumor or cardiac uptake in a uniform background in a PET or SPECT Region of Interest (ROI). $p(\mathbf{x}|\boldsymbol{\alpha})$ is the probability of obtaining a particular feature measurement vector given a particular vector of component mixture values. $p(\boldsymbol{\alpha})$ is the *a priori* probability of obtaining a particular vector of component mixture values. In this work, this latter PDF is modeled on equation (1), i.e. the individual inter-voxel mixing is governed by Benford's Law.

To investigate this, we propose a simple 2-class formulation where the PDF is representative of the probability of obtaining a particular mixture of the two classes, *a* and *b*. As has been stated previously, we propose that the *a priori* PDFs are governed by Benford's Law, so combining two scaled distributions (one is reflected about $\alpha = 0.5$), given by equation (1), we obtain:

$$p_{ab}(\boldsymbol{\alpha}) = \frac{P_{PV}(a,b)}{K} U(\boldsymbol{\alpha}) \left[\log_{10}\left(1+\frac{0.1}{\alpha_a}\right) + \log_{10}\left(1+\frac{0.1}{\alpha_b}\right) \right] \qquad (3)$$

The factors of 0.1 appear in this formulation, as compared to equation 1, because we choose to normalize our distribution in the interval (0,1) rather than the (0,10) interval that Benford originally proposed. Note that this demonstrates there is only a single scaling parameter to fit to the PV component in the intensity histogram. Within the context of PET or SPECT imaging, these two distributions within equation (3), represent the intrinsic under or over spill from a high activity area with mean μ_b, surrounded by a lower activity area with mean intensity μ_a.

III. MATERIALS & METHODS

A. Partial Volume Data Simulation

We can illustrate the above effect with the use of simulated data, wherein the availability of ground truth data enables the true *a priori* PV distribution to be determined. This is described below for a two-class problem. Initially high-resolution 3-D data are generated where each high-resolution voxel takes an intensity value equal to either one of two user-defined means, μ_a or μ_b. This initial dataset represents the continuous case found in the actual object to be imaged.

Gaussian distributed noise is then added to each voxel intensity, where the standard deviation, σ_a or σ_b, of the Gaussian distributed noise is dependent upon the original mean value the voxel was assigned, representative of either the hot insert, μ_b or background, μ_a**.** The data are then filtered with a kernel equivalent to the PSF of the imaging system. The filtering process introduces a blurring of the edges containing high-frequency components, simulating the action of the point spread function of the imaging system or scanner, thereby introducing boundary mixing artifacts that can be considered to simulate the PV effect. The filtered data are then down sampled onto a relatively coarse voxel grid representative of that found in PET/SPECT image data.

By first ignoring any form of additive noise contributing to σ_a or σ_b,, the resultant noise-less filtered, down sampled simulated PV data sets can be considered as a ground-truth or idealized representation of the image data in the absence of any corruption artifacts. In our studies we find that, as far as we can ascertain, any low-pass PSF filter can be selected for Benford's Law to remain a plausible fit to the resulting noise-free voxel intensity histogram.

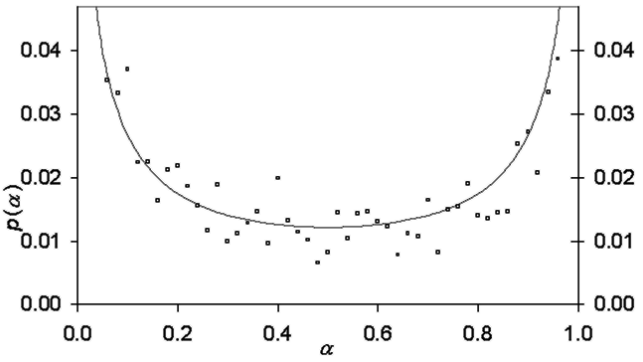

Fig. 2: Histogram produced from a simulated noise-free PV data set. The result of an idealized noiseless situation composed of two peaks (off scale) corresponding to high frequency occurrences of the means of the pure voxel regions in the ground truth data and a function linking the two peaks together, corresponding to the simulated PV voxels. The continuous line demonstrates excellent agreement between the Benford model (equation 3) and the synthetic data.

Equation (3) results in a function that has an excellent fit to the noiseless data. Fig. 3 was plotted without any fitting, as there was no requirement for any parameter estimation. $P_{pv}(a,b)$, the prior probability of the PV voxels was calculated from prior knowledge of the pure and partial voxel membership. In a real imaging situation, it could easily be estimated by least squares or ML fitting.

To examine the effect of noise on the intrinsic performance of a statistical classifier using Benford's Law, a variety of simulated datasets were generated. Parameters representing the means and variances of the two-voxel classes were extracted using ROI analysis from actual PET phantom data (see below). For simplicity, the simulation ignores modality-specific effects such as attenuation and scatter in PET/SPECT, (or beam hardening in CT etc), in order to

demonstrate the intrinsic performance of the approach. However we return to these issues in the Discussion.

B. Experimental Data Acquisition

In order to investigate the practical imaging performance of the Bayes'/Benford Classifier, a preliminary PET phantom study was undertaken. Experimental work was carried out using a Phillips Gemini PET/CT scanner and a cylindrical phantom filled with Ga-68. Registered CT data from the same scanner were acquired, and processed to represent an idealized PET image. This was to be used as a 'ground truth' dataset upon which to compare the experimental classification based on the Benford model.

The phantom used comprised a main cylindrical compartment of 19cm x 20cm diameter, containing three axially parallel PTFE, air and user-fillable inserts each having external diameters of 49mm. For the purpose of the experiment the PTFE insert was removed to provide a large background region for subsequent analysis. The air insert could not be removed.

The fillable 'hot' insert and the main phantom compartment were to be used to generate partial volume measurements at their interface by filling with different activity concentrations. Although the 1mm (cold) insert wall thickness could potentially result in PV underestimates in the classification process, initial PV data simulations using a hot insert/warm background/cold wall thickness geometry suggested that for the scanner PSF and phantom geometry/activities used here, there was a maximum ~3% estimated classification error introduced due to finite wall thickness. (In fact we found wall thickness only becomes significant when it approaches the scanner PSF FWHM).

Initially Gallium-68 was extracted from a Germanium-68 generator in the form of Gallium Chloride. The vial containing the radioactivity was then placed in a calibrator to determine the activity (227 MBq in 5.13ml). This was used to produce a ~5:1 activity concentration ratio between the hot insert and the background, considered to be similar to the uptake found in many hot spot imaging applications. After injection of radioactivity into the water-filled phantom compartments, the phantom was shaken to ensure uniform mixing of the activity in each compartment.

The phantom was wedged vertically at an angle of ~30^{o} on the scanner bed, in order to improve the occurrence and distribution of Partial Volume Effect voxels around the hot insert boundary. X-ray CT and PET image data were acquired using a modified Whole Body PET/CT acquisition protocol, and a 15% energy acceptance window centered on 511keV for the PET data. The PET data were acquired for 15 minutes in a single bed position to give good count statistics. As the PET scanner field of view only covers 18cm, the phantom was positioned to avoid obtaining PET image slices of the tilted, and thereby incompletely sampled end faces.

The PET sinogram data were reconstructed on a 4.00 x 4.00 x 4.00 mm^3 grid using various reconstruction methods and applied corrections as summarized in Table 1 below. Those datasets using RAMLA [8,9] were reconstructed using two iterations of the reconstruction scheme. Each was reconstructed onto the same size grid using Philips supplied software. The corrections applied to the data were implemented using Philips software, which uses registered CT data to determine attenuation correction factors at 511keV. These attenuation correction factors are then utilized in a single scatter simulation to generate a scatter correction image. Random coincidences were acquired using a delayed anti-coincidence channel during acquisition.

CT data were acquired by translating the phantom through the CT part of the Gemini scanner. Following fan-beam reconstruction**,** the resultant CT image data set contained voxels of 1.17 x.1.17 x 5.00 mm^3.

TABLE I
PET DATASETS GENERATED FOR PHANTOM IMAGING.

Dataset Abbreviation	Reconstruction Method	Correction Applied
FBP-AC	Filtered Backprojection	attenuation only
RAMLA	RAMLA OS-EM	none
RAMLA-AC	as above	attenuation only
RAMLA-FC	as above	attenuation scatter randoms

C. Experimental Data Analysis

Initial analysis of the PET data involved considering the mean and variance of user-defined ROIs in contiguous transverse slices. This was because the probabilistic classifier assumes an ergodic voxel response, which was examined by considering the mean and variance in each axial slice. As a result, only the central 14 PET slices were used for subsequent analysis, where the ROI mean and variance appeared almost constant regardless of whether RAMLA or FBP was used for reconstruction. However, the FBP-AC data exhibited a variance almost twice that of the RAMLA-AC data.

The PET datasets were then resampled and registered onto the CT grid using a rigid body transformation defined by the scanner manufacturer. Although the resampled voxel size of the PET data sets were the same as that of the CT, due to the differing fields of view, a further translation was required to register the data to the same origin in the analysis software. This was achieved by using 12 user-defined landmarks from which a mean displacement vector was calculated. The precision of the calculated transformation matrix was then verified by inspection of profiles in the *x, y* and *z* directions.

The registered CT data were then processed to produce 'ground truth' or idealized noiseless PET datasets onto which the classified real PET data would be compared. As the CT voxel intensity values of the phantom hot insert and its surroundings were similar (as both contained water) 3D seeded region growing was performed with a variety of thresholds on voxels representing the thin (1mm) Perspex

insert wall thickness. This produced an initial binary wall template for the insert wall.

To determine a resultant 'hot' insert volume, this 3D insert wall template was then subjected to a further 3D region growing process, initiated inside the blank volume enclosed by the segmented wall. Once thresholds had been optimized, the segmented CT insert volume was found to be within 3% of the calculated volume.

The voxels in the resulting 3D binary template (insert and background) were then assigned the corresponding mean values extracted from the datasets listed in Table I to generate unique ground truth datasets for each reconstruction/correction combination investigated. Each dataset was then filtered using a Gaussian fitted to a Ga-68 point spread function, experimentally acquired from the same scanner. This produced idealized noiseless 'ground truth' representations for each dataset listed in Table I.

To generate simulated data using the same geometry, the ground truth data were noise-corrupted, prior to PSF filtering, by applying Gaussian noise with variances for insert and background determined from the aforementioned ROI analysis (see fig. 3 a,b) for each dataset. Thus, simulated data for each experimentally acquired dataset was generated to examine intrinsic classifier performance.

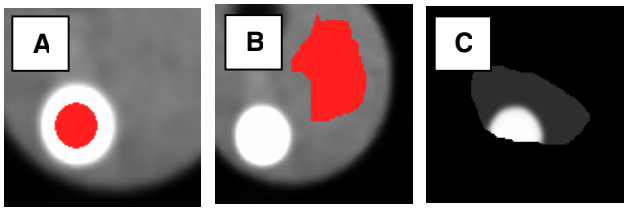

Fig. 3. ROI placement in red for the hot insert (A) and the warn background (B) used for parameter estimation illustrated using RAMLA data. The ROI are kept clear of the edges to avoid PV bias on the extracted parameter values. The ROI template used to define the PV analysis area in each slice (C) ensures air/outer wall mixing effects are avoided. The software described in [7] was used to create these ROI templates.

The various simulated and experimental PET datasets were input to the Benford Classifier, along with the relevant ground truth of the hot insert derived from the CT data. In order to avoid unwanted multiple partial volume effects at the outer edge of the hot insert close to the phantom's outer wall, the PV analysis was limited in each slice to an area defined by set of binary templates as illustrated in fig. 3 c.

IV. Results & Discussion

A. Partial Volume Data Simulation

Two exemplar histograms from the ROI analysis assumed by the classifier are shown in fig. 4. The resultant RMS error for each dataset is presented in Table II.

As can be seen from comparison of fig.s 4 and Table II, there is relatively little difference between the RAMLA-AC and RAMLA-FC histograms, and this yields very similar overall classification results. However the (uncorrected) RAMLA and FBP datasets exhibit the worst classification results, attributed to deviations from the expected Gaussian behavior for pure voxels.

Comparing each of the three RAMLA datasets, we see this effect verified, as the RMS error is reduced as the histograms produce better fits to the assumed statistical behavior.

B. PET/CT Phantom Study

Using registered CT-derived ground truth data for each corresponding dataset listed in Table I, RMS error values were calculated following classification using the Benford classifier as shown in Table III. The RAMLA-FC data provides the overall best classification result, although this represents a relatively marginal improvement compared to RAMLA-AC, and in fact a decrease in pure voxel classification error, attributed to the deviations from Gaussian variance observed once the scatter and randoms corrections had been applied.

TABLE II
RMS DIFFERENCES IN CLASSIFIFCATION PERFORMANCE BETWEEN SIMULATED DATA SETS USING VOXEL CLASS PARAMETERS TAKEN FROM PET DATASETS LISTED IN TABLE I.

Dataset Abbreviation	overall rms error	rms mean PV error	rms mean pure error
FBP-AC	0.017	0.026	0.012
RAMLA	0.017	0.028	0.011
RAMLA-AC	0.009	0.016	0.005
RAMLA-FC	0.006	0.011	0.003

The small overall difference between the RAMLA–FC and –AC datasets may be due to the reduced magnitude and structure associated with randoms and scatter corrections as compared to the changes on the data produced by attenuation correction, at least for the geometry used here.

The RAMLA data (i.e. no correction applied) has the worst result. This may be attributed to photon attenuation effects, which produced significant image artifact in the experimental data. This is also demonstrated in fig. 4 by the arrowed deviation from the assumed Gaussian behavior for pure voxels from the warm background.

The FBP-AC data also provided similar relatively poor performance as the (uncorrected) RAMLA dataset. As the RAMLA-AC and FBP-AC datasets have the same level of correction, one contributing factor may be the method of reconstruction that limits the performance of the classifier. As noted prior to classification, the variance in the FBP data was higher than the RAMLA data.

Comparing the results of Table II and III, there appears to be consistency across different datasets in that RAMLA-FC produces the best result and the ordering of each dataset in terms of overall RMS error seems consistent between simulated and experimental data. However the absolute differences between the real and simulated results are significant (~×8).

The results in Table II represent an intrinsic 'best-case' classification result for the subsequent PET study, and assume

perfect registration and other corrections. We attribute this difference at this stage to the need for better sub-voxel registration, which is key for observing low PV errors in methodological assessment using processed CT data as ground truth. Any deviation from the assumed statistical behavior will affect the classification process.

TABLE III
THE RMS CLASSIFICATION ERROR BETWEEN GROUND TRUTH (PROCESED CT DATA) AND CLASSIFIED EXPERIMENTAL PET DATA SETS

Dataset Abbreviation	overall rms error	rms mean PV error	rms mean pure error
FBP-AC	0.190	0.369	0.063
RAMLA	0.221	0.412	0.100
RAMLA-AC	0.063	0.126	0.004
RAMLA-FC	0.061	0.122	0.011

V. Conclusions

We have outlined a method of quantifying PV effects in volumetric data, with particular reference to PET and SPECT imaging modalities. This may be used to accurately determine metabolically active volumes without *a-priori* knowledge of morphological boundaries. The advantages of this approach are that (1) we make no prior assumptions about the physical boundaries of hot or cold objects; (2), this new formulation is easy to use as Benford's Law requires no parameter estimation other than a scaling term, to determine the shape of the *a priori* PV PDF, and (3), the PDF arising from Benford's Law can be considered to provide a good model of the intrinsic under or over spill from high/low activity boundaries.

A preliminary phantom study using co-registered CT data has been completed in order to experimentally validate the quality of PV correction available with this approach. Out of the 4 methods used to reconstruct the PET images, the RAMLA-FC gave the smallest errors when compared to ground truth, with RMS errors of just over 12% for PV voxels and just over 1% for pure voxels.

It is worth noting that the quality of the registration step in aligning experimental object boundaries and the accurate determination of the PSF filter is critical for realizing low observed PV classification error. It may be that further registration improvements may produce slightly better PV classification errors, alongside a more accurate representation of the PSF filter in each case, thereby providing closer agreement with the intrinsic ~1% RMS error limit suggested by PV simulation for this level of contrast.

These first results using Benford's Law with a statistical partial volume classifier demonstrate that this approach has high potential for accurate PV quantification in PET/SPECT imaging.

VI. Acknowledgment

This work has been supported by the Engineering and Physical Sciences Research Council (EPSRC). John Chiverton would like to thank the NIH for a conference student award in 2004 and the EPSRC for funding his PhD.

VII. References

[1] M. Quarantelli, K. Berkouk, A. Prinster, B. Landeau, C. Svarer, L. Balkay, B. Alfano, A. Brunetti, J.-C. Baron, and M. Salvatore, "Integrated Software for the Analysis of Brain PET/SPECT Studies with Partial-Volume Effect Correction"; *J. Nucl. Med.*; Vol. 45; 2004.

[2] J.P. Chiverton and K. Wells, "Volumetric Partial Volume Quantification via a Statistical Model of 3-D Voxel Gradient Magnitude", In Conf. Proc. *IEEE NSS & MIC 2004, Rome, Italy*, October 2004.

[3] F. Benford, "The Law of Anomalous Numbers", *Proc. Amer. Phil. Soc.*, 78, 551-572, 1938.

[4] P. Santago and H. Gage, "Statistical Models of Partial Volume Effect", *IEEE Trans. Image Proc.*, Vol. 4(11), pp. 1531-1540, 1995.

[5] C. Durtschi, W. Hillison and C. Pacini, "The Effective Use of Benford's Law to Assist in Detecting Fraud in Accounting Data", *J. Forensic. Acc.*, pp. 17-34, 2004.

[6] J.P. Chiverton and K. Wells, "Estimation of Partial Volume Mixtures with a Confidence Measure Combining Intensity and Gradient Magnitude Information", In Conf. Proc. *Med. Image Und. Anal.*, Bristol, UK, 2005.

[7] C. Rorden, and M. Brett, "Stereotaxic display of brain lesions", *Behav. Neurol.*, Vol. 12, 191-200, 2000.

[8] J. Browne and A.R. De Pierro, "A row-action alternative to the EM algorithm for maximizing likelihoods in emission tomography", *IEEE Trans. Med. Imag.*, 15(5), 1996.

[9] M.E. Daube-Witherspoon, S. Matej and J.S. Karp, "Assessment of image quality with a fast fully 3D reconstruction algorithm", In Conf. Proc. *IEEE NSS & MIC 2001, San Diego, USA*, October 2001.

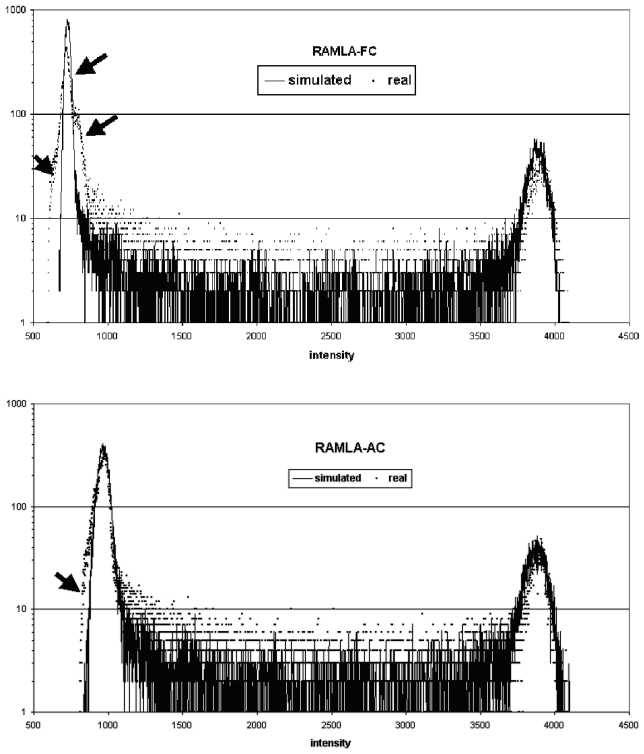

Fig. 4. Comparison of the intensity histograms for the RAMLA-AC and RAMLA–FC datasets obtained using the ROIs shown in Fig.3. Arrows denote significant differences between the simulated and real datasets, demonstrating that the application of the randoms and scatter correction used in RAMLA-FC has produced an apparent deviation from the assumed Gaussian-like noise model for the pure background voxels. See Table III for corresponding RMS error classification results.