Global PCA of local moments with application to multi-sequence MRI segmentation

Jacob M. Maronge

maronge@wisc.edu

Department of Statistics,

University of Wisconsin - Madison

John Muschelli

jmuschel@jhsph.edu

Department of Biostatistics,

Johns Hopkins University, Bloomberg School of Public Health

Ciprian M. Crainiceanu

ccraini1@jhu.edu

Department of Biostatistics,

Johns Hopkins University, Bloomberg School of Public Health

August 23, 2021

Abstract

We introduce methods for quantifying the information contained in the moments

of local neighborhoods of complex multi-sequence images. For each voxel and image sequence we construct a vector of the moments (powers) of intensities of the neighborhood voxels, stack these vectors in a matrix, and conduct principal component analysis (PCA) on the resulting matrix. The proposed methods achieve three goals: (1) decompose the observed variability of images at the population level; (2) describe and quantify the main directions of variation; and (3) use these directions of variation to improve segmentation and studies of association with health outcomes. Methods are motivated by and assessed on the 2015 Ischemic Stroke Lesion Segmentation (ISLES) Challenge.

keywords: Magnetic Resonance Imaging (MRI), Principal Component Analysis (PCA), Segmentation.

1 Introduction

1.1 Clinical significance

Stroke is the third most frequent cause of death in industrialized countries (1). Even in people who survive, stroke often results in recurring disability. Currently, stroke is diagnosed via magnetic resonance imaging (MRI) of the brain. This requires manual segmentation of lesion tissue by a radiologist, and is both time-consuming and may be subject to reader variability. There has been a need for methods to automate the process of detecting and segmenting stroke lesions (10).

In this paper we propose a new method for stroke lesion segmentation. For training and testing the method we use the 2015 ISLES challenge data (6). Particularly, we use the Subacute Ischemic Stroke lesion Segmentation (SISS) sub-challenge. These data consists of 28 subjects where each subject has multi-sequence MRIs consisting of fluid attenuated inversion

recovery (FLAIR), diffusion weighted imaging (DWI), T1-weighted (T1w), and T2-weighted (T2w) images. In addition, each subject also has a manual segmentation performed by a radiologist.

1.2 Proposed method

We contend that the meaning, interpretation, and complexity of an image are encapsulated in the collection of all neighborhoods of all locations in the image. More precisely, an image can be considered as a 3-dimensional array, where each element in the array is considered a voxel (volume element) that has a numeric intensity. Every image with V voxels can be represented as an $V \times V$ matrix of voxel intensities, where every row represents a location in the image and every column represents a particular position in the neighborhood of that location. For example, the first column could be the neighbor just above the location, the second column could be the neighbor to the left, and so on. If a particular neighbor does not exist, then the corresponding column contains the mean of any non-missing columns. This matrix contains the first order neighborhood information, but a similar procedure can be applied to the matrix obtained by taking any entry-wise transformation of the original image. Moreover, the same ideas apply to any number of image sequences or modalities that are available.

Consider the case when multi-sequence imaging is available and images are spatially aligned within-subject, but not necessarily across subjects. Such data are routinely available in MRI studies, where FLAIR, T1w, and T2w images are collected for multiple subjects. For such data structures, we are interested in: (1) quantifying the observed variability of local neighborhoods at the population level; (2) obtaining the main directions of variation of local neighborhoods; and (3) using these directions of variation to improve segmentation and studies of association with health outcomes.

To achieve these objectives we propose to decompose the local variability of various moments of the image intensities: the image intensity, squared image intensity, and so on. To build up intuition, consider the case when we only have one image per subject. We will later extend to the case when there the number of image sequences (L), are ≥ 1 images per subject and additional transformations of the neighborhood information is taken into account.

The main idea is to represent this one image as an $X_i = V_i \times K$ matrix, where V_i is the number of voxels in the image for subject i and K is the voxel neighborhood size. This matrix is created by vectorizing the 3D array and column-wise stacking the 3-dimensional adjacent neighbors to the voxel for each row. Here $i = 1, \ldots, I$ denotes the subject i and we allow for images to have a different number of voxels for different subjects. We propose to row-stack these subject-specific X_i matrices across subjects and create the $(\sum_i V_i = V) \times K$ dimensional matrix X. After the matrix X is built we propose to: 1) use principal component decomposition of the X matrix; 2) explore the directions of the main sources of variability; and 3) use the PCA scores to conduct prediction and segmentation.

To illustrate the idea, consider a mock example in 2 dimensions. Figure 1 displays a 3×3 sub-region of a hypothetical image, where each square is a voxel with a particular intensity value. Smaller numbers are colored in darker shades of gray, while larger numbers are colored in lighter shades of gray. The local neighborhood corresponding to the center voxel has the following vectorized representation:

$$x_j = (0.34, 0.58, -0.73, 1.74, -0.69, -1.34, 0.71, -1.87, -1.97).$$

Figure 2 displays the same 3×3 region together with the second, third and fourth order moments. More precisely, the 3×3 region labeled "2nd order" is obtained from the original

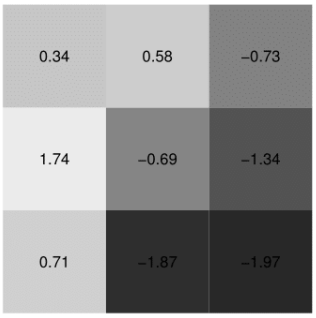

Figure 1: **Example region**: Example region of voxel intensities.

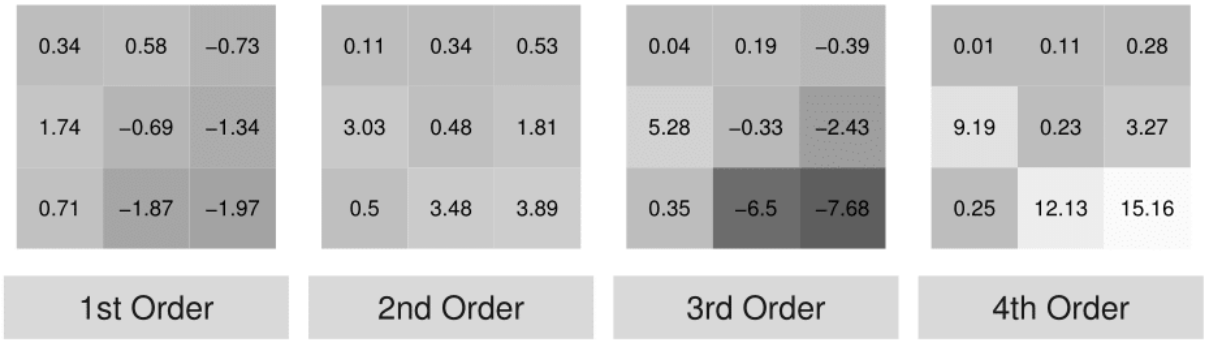

Figure 2: **Higher order moments**: Example region of voxel intensities in 1st through 4th moment images.

image by squaring the entries of the original image and so on. The collection of these images has the following vectorized representation:

$$\begin{aligned} X_j =& (0.34, 0.58, -0.73, 1.74, -0.69, -1.34, 0.71, -1.87, -1.97, \\ & 0.11, 0.34, 0.53, 3.03, 0.48, 1.81, 0.50, 3.48, 3.89, \\ & 0.04, 0.19, -0.39, 5.28, -0.33, -2.43, 0.35, -6.50, -7.68, \\ & 0.01, 0.11, 0.28, 9.19, 0.23, 3.27, 0.25, 12.13, 15.16). \end{aligned}$$

In this work, we will use neighborhoods of size $K = 27$ (directly adjacent neighbors in 3D), but methods can easily be applied to any neighborhood size. We will also transform the neighborhoods using higher order moments of the original image (e.g. square, cube).

Thus, if we consider all moments up to the fourth moment for one imaging sequence, then the matrix representation of the image will be $V \times (4K)$ dimensional. The matrix could be extended to include up to the M^{th} moment for each; in this situation the matrix would be $V_i \times (MK)$ dimensional. The idea is to use the additional information contained in the higher order moments of the image intensities to improve prediction.

For every imaging sequence a matrix can be built. We propose to column-bind these matrices. Thus, if there are L modalities, the subject-specific matrix will be $V_i \times (LMK)$ dimensional. In this paper we will use $L = 4$, $M = 4$, and $K = 27$, which are also the defaults in the associated software. However, these values can be changed depending on the applications and size of the problem. We will redefine the subject-specific matrix X_i as $V_i \times (LMK)$ and perform the aforementioned steps accordingly after aggregating these subject-specific matrices.

In addition to the moment matrices we can also consider the interactions (products) of moment matrices across modalities or other transformations. Because this quickly explodes the dimensionality fo the matrix X, here we focus on the moment matrices only.

2 Materials and methods

2.1 Imaging sequences and preprocessing

The data used is from the 2015 Ischemic Stroke Lesion Segmentation (ISLES) Challenge (6). The ISLES Challenge Data consisted of two sub-challenges, the Sub-acute Ischemic Stroke lesion Segmentation (SISS) Challenge and the acute Stroke outcome/Penumbra Estimation (SPES) Challenge. All data presented here are from the SISS Challenge. Though the method and associated software were developed for and applied to the SISS Challenge, it is general and designed for supervised image segmentation problems. The method could be applied

directly or with small modifications for conducting segmentations of other pathologies as well as healthy brain tissues.

The ISLES SISS Data consisted of twenty-eight training cases and thirty-six testing cases. All MRI sequences were skull-stripped, re-sampled to an isotropic spacing of 1 cubic mm, and spatially aligned to the FLAIR sequence. Stroke regions of interest (ROIs) were hand segmented by an expert for all training cases and by two experts for all cases. As we wish to see the performance of the segmentation in the application, we will use only the training cases as the testing cases have no released ROIs. The training cases are from the same testing center. The MRI sequences for each subject include a T1-weighted, T2-weighted, FLAIR, and DWI sequence.

All images were intensity normalized using a z-scoring approach (11; 8). More precisely, for each subject and each modality is the raw intensity at the voxel v, denoted $I(v)$ were extracted. A mean (μ_{trimmed}) and ($\text{sd}_{\text{trimmed}}$) were estimated after removing the highest and lowest 10% intensities. Each voxel in the brain was then normalized by the following calculation

$$\text{z-score}(v) = \frac{I(v) - \mu_{\text{trimmed}}}{sd_{\text{trimmed}}}.$$

Intensity normalization is crucial in supervised segmentation and the trimming before z-scoring is necessary to reduce the influence of pathology on intensity normalization(11; 3).

2.2 Local neighborhood principal component analysis

To address the problem of stroke ischemia segmentation, we propose a global PCA of local moments to quantify the neighborhood information in a collection of images. The approach generalizes both methods that use voxel intensity only and methods that incorporate neighborhood intensities. While similar ideas have been used in a few papers dedicated to unsu-

pervised clustering (5; 4), this is the first approach that provides a supervised approach that uses higher order moments information. For our analysis, we selected at random 14 subjects out of the 28 subjects in the ISLES data to train our model and PCA. The remaining 14 were used to estimate model performance, after using projecting onto the training-set PC loadings.

The main idea consists of the following easy steps: (1) for each image modality, m, and subject, i, build the image matrix X_{im} with each row corresponding to a voxel and each column corresponding to a position of a neighborhood voxel relative to the neighborhood center voxel; (2) build the k-moment matrices $X_{imk} = X_{im}^{k}.$, where the X^k denotes entry-wise power transformation, and not the kth power of the matrix X; (3) bind the X_{imk} matrices by the corresponding rows (voxels) across all moments and form the matrix $\mathbf{X}_{im}$; (4) bind the matrices $\mathbf{X}_{im}$ by the corresponding rows (voxels) across sequences and form the matrix $\mathbf{X}_i$; and (5) bind the matrices $\mathbf{X}_i$ by the corresponding columns (neighborhood position) across subjects and form the matrix $\mathbf{X}$. Once the matrix $\mathbf{X}$ is built, we center and scale each column to ensure that the observed information is not dominated by the scale of higher moments and conduct PCA. This is equivalent to conducting PCA on the correlation matrix between the columns of $\mathbf{X}$.

The matrix $\mathbf{X}$ is very large and computation can be sped up considerably by avoiding loading it in the computer memory. Indeed, for the ISLES challenge the size of $\mathbf{X}$ is $(\sum_i V_i = V) \times (LMK)$, where V is approximately 28 million and $LMK = 4 \times 4 \times 27 = 432$. If the size the neighborhood is increased, say to $K = 27$, $K = 64$, or more, then the size of the matrix can become an even more serious problem. Dealing with such a matrix requires careful computational approaches to make the implementation feasible and fast. We now provide the technical details for working with such a matrix.

2.3 Computational problems

Here we show how to conduct the PCA analysis of the matrix $\mathbf{X}$, which has millions of rows and hundreds or thousands of columns. We would like to do this efficiently using a load, compute, and delete strategy, where each subject-specific data is loaded only once, the sufficient statistics are computed and the data are then deleted from the active memory.

To achieve this we calculate the singular matrix decomposition (SVD) of X using the block-wise data access described by Zipunnikov et al. (15). Consider the case when $\mathbf{X}$ is a "skinny long matrix" (e.g. millions of columns and hundreds or thousands of columns) and let $\mathbf{X} = \mathbf{UDV}^T$ be its SVD. Here $\mathbf{U}$ is a $V \times P$ dimensional matrix such that $\mathbf{UU}^T = \mathbf{I}_V$, the identity matrix of dimension $V \times V$, $\mathbf{D}$ is a $P \times P$ dimensional diagonal matrix, and $\mathbf{V}$ is a $P \times P$ dimensional matrix, such that $\mathbf{V}^T\mathbf{V} = \mathbf{I}_P$, the identity matrix of size $P \times P$. The covariance matrix can be written as $\mathbf{X}^T\mathbf{X} = \mathbf{VD}^2\mathbf{V}^T$, which is a $P \times P$ dimensional matrix that can easily be diagonalized. Thus, the matrices $\mathbf{V}$ and $\mathbf{D}$ can be obtained by spectral analysis of $\mathbf{X}^T\mathbf{X}$. The matrix $\mathbf{U}$ can then be obtained by the formula $\mathbf{U} = \mathbf{XVD}^{-1}$ and the sub-blocks of the matrix $\mathbf{U}$ are obtained as $\mathbf{U}_i = \mathbf{X}_i\mathbf{VD}^{-1}$. This requires right multiplication of the subject-specific matrices by $\mathbf{VD}^{-1}$, a simple operation. However, we do not need the matrix $\mathbf{U}$ and focus instead on obtaining $\mathbf{V}$ and $\mathbf{D}$.

Because we want to avoid loading the entire $\mathbf{X}$ matrix in the active memory, computing $\mathbf{X}^T\mathbf{X}$ can be done using sub-blocks of the matrix. More precisely, we consider the case when data are partitioned into blocks on rows as $\mathbf{X} = \left(\mathbf{X_1} \,\middle|\, \mathbf{X_2} \,\middle|\, \ldots \,\middle|\, \mathbf{X_n} \right)^T$. We use $\mathbf{X_i}$ as the block-matrix corresponding to subject i, which is $V_i \times KLM$ dimensional. Denote by j the j^{th} row of the subject-specific blocks and let $\mathbf{X}_{i,j}$ be the j^{th} row of $\mathbf{X}_i$. One potential problem is that the scale of the columns of $\mathbf{X}$ can be quite different and variability can be dominated by columns with a larger scale. This is especially problematic when some of the columns are powers of intensity values. To avoid this problem, we focus on the correlation

instead of the covariance matrix. It can be shown that the correlation matrix is $\mathbf{Z}^T\mathbf{Z}$, where $\mathbf{Z}$ is the matrix of z-scores by columns. More precisely,

$$\mathbf{Z} = \frac{\mathbf{X} - \bar{\mathbf{x}}}{\mathbf{s}}$$

where $\bar{\mathbf{x}}$ is a vector of column means of $\mathbf{X}$ and $\mathbf{s}$ is a vector of column standard deviations of $\mathbf{X}$. The division here is entry-wise, making the columns of $\mathbf{Y}$ vectors have mean 0 and variance 1. If $\mathbf{C} = \mathbf{Z}^T\mathbf{Z}$, the u, v entry of $\mathbf{C}$ is,

$$C_{u,v} = \frac{\sum_{i,j} \mathbf{X}_{i,j}(u)\mathbf{X}_{i,j}(v) - V\bar{x}_u\bar{x}_v}{s_u s_v},$$

where $V = \sum_{i=1}^{I} V_i$ is the number of voxels across subjects, and $\bar{x}_u$ and s_u are the mean and standard deviations of the column u of the matrix $\mathbf{X}$, respectively.

We now show how to obtain the elements necessary for the correlation matrix $\mathbf{C}$ by loading the data for each subject separately and only once. The column mean vector can be written as,

$$\bar{\mathbf{x}} = \frac{\sum_i \bar{\mathbf{x}}_i V_i}{\sum_i V_i},$$

where $\bar{\mathbf{x}}_i$ denotes the column average of the subject i data matrix, $\mathbf{X}_i$, and V_i is the number of voxels for subject i. Denote by $\bar{\mathbf{x}}^k$ the column mean for the data corresponding to the first k subjects and let $V^k = \sum_{i=1}^{k} V_i$ be the number of voxels from the first k subjects. It follows that $\bar{\mathbf{x}}^1 = \bar{\mathbf{x}}_1$, $V^1 = V_1$ and $\bar{\mathbf{x}}^{k+1} = \left(V^k\bar{\mathbf{x}}^k + V_{k+1}\bar{\mathbf{x}}_{k+1}\right)/V^{k+1}$. This recursive formula allows the sequential computation of the column means by loading only subject k data at a time. For the u column of the matrix $\mathbf{X}$ we have $s^2(u) = \sum_{i,j}\{\mathbf{X}_{ij}(u) - \bar{x}_u\}^2/(V-1) = \sum_{ij}\mathbf{X}_{ij}^2(u) - V\bar{x}_u^2/(V-1)$. If $\mathbf{S}^{2,k} = \sum_{ij}\mathbf{X}_{ij}^2(u)$, then $\mathbf{S}^{2,1} = \sum_j \mathbf{X}_{1j}^2(u)$ and $\mathbf{S}^{2,k+1} = \mathbf{S}^{2,k} + \sum_j \mathbf{X}_{k+1j}^2(u)$, indicating that the sum of squares can be calculated

recursively by loading the subject-specific data sequentially. Thus, the only component that remains to be calculated is the matrix $\mathbf{C}$ is $\sum_{ij} \mathbf{X}_{ij}(u)\mathbf{X}_{ij}(v)$. Note that the (u, v) entry of the matrix $\mathbf{X}_i^T\mathbf{X}_i$ is equal to $\sum_j \mathbf{X}_{ij}(u)\mathbf{X}_{ij}(v)$. Thus, we need to calculate only the matrix $\mathbf{X}^T\mathbf{X} = \sum_{i=1}^{I} \mathbf{X}_i^T\mathbf{X}_i$, which provides an expression of $\mathbf{X}^T\mathbf{X}$ in terms of the subject-specific blocks $\mathbf{X}_i^T\mathbf{X}_i$. If $\mathbf{U}^k = \sum_{i=1}^{k} \mathbf{X}_i^T\mathbf{X}_i$, then $\mathbf{U}^1 = \mathbf{X}_1^T\mathbf{X}_1$ and $\mathbf{U}^{k+1} = \mathbf{U}^k + \mathbf{X}_{k+1}^T\mathbf{X}_{k+1}$. Thus, the correlation matrix can be calculated as $\{\mathbf{U}^I - V\bar{\mathbf{x}}^T\bar{\mathbf{x}}\}/\mathbf{s}^T\mathbf{s}$, where the division is entry-wise division of matrix entries and not matrix division. These results indicate that if we read in each $\mathbf{X}_i$ and retain only $\mathbf{X}_i^T\mathbf{X}_i$, $\bar{\mathbf{x}}_i$, V_i, and $\sum_j \mathbf{X}_{i,j}^2$, then we can calculate $\mathbf{C}$, the matrix of interest.

2.4 Algorithm for calculation of the correlation matrix

We now summarize the formulas from Section 2.3 into a simple algorithm for the calculation of the correlation matrix.

1. Initialize values for $\mathbf{U}_k = \mathbf{X}_k^T\mathbf{X}_k$, $\bar{\mathbf{x}}_k$, V_k, and $\sum_j \mathbf{X}_{k,j}^2$. If $k = 0$, initialize to 0.

2. Read in $k + 1$ subject's sequence of images.

3. Create the local neighborhood moments matrix ($\mathbf{X}_{k+1}$).

4. Calculate and store the number of voxels in $(k + 1)th$ subject's brain mask (V_{k+1}), the column means of $\mathbf{X}_{k+1}$, $\bar{\mathbf{x}}_{k+1}$, the sum of squared entries for each column of $\mathbf{X}_{k+1}$, $\mathbf{S}_{k+1}^2$, and $\mathbf{U}_{k+1}$.

5. Update $k + 1$ subjects with $V^{k+1} = V^k + V_{k+1}$.

6. Update $\mathbf{S}^{2,k+1} = \mathbf{S}^{2,k} + \mathbf{S}_{k+1}^2$

7. Update $\bar{\mathbf{x}}^{k+1} = \left(V^k\bar{\mathbf{x}}^k + V_{k+1}\bar{\mathbf{x}}_{k+1}\right)/V^{k+1}$.

8. Delete all information pertaining to only the $(k+1)th$ subject.

9. Repeat up to Ith subject.

10. Obtain $\mathbf{s} = \sqrt{\frac{\mathbf{S}^{2,I} - V^I(\bar{\mathbf{x}}^I)^2}{V^I - 1}}$.

11. Obtain $\mathbf{C} = \frac{\mathbf{U}^I - V\bar{\mathbf{x}}^T\bar{\mathbf{x}}}{\mathbf{s}^T\mathbf{s}}$.

2.5 Model performance

The principal components can be visualized by mapping them back onto the original image. The scores for each image is then obtained by projection on these principal components. To assess how these scores perform in terms of prediction of stroke ischemia, we employ principal component regression. Principal Component Regression works by conducting logistic regression on the first Q principal components scores. Specifically, we fit the model

$$\text{logit}\{\pi(x)\} = \beta_0 + \beta_1\xi_{1x} + \ldots + \beta_Q\xi_{Qx}, \tag{1}$$

where $\pi(x)$ is the probability that the voxel x is affected by stroke and $\xi_{1x}, \ldots, \xi_{Qx}$ are scores for the first Q principal components corresponding to the voxel x. Recall that the voxel x is represented as the vector of moments of intensities of the image in a neighborhood of x and the scores are obtained by projecting this vector on the individual PCs. To calculate the scores we first obtain the spectral decomposition $\mathbf{Z}^T\mathbf{Z} = VD^2V^T$, and take the first Q vectors of $\mathbf{V}$, denoted $\mathbf{V}_Q$. The scores are obtained by the operation $\mathbf{Z}\mathbf{V}_Q$, where the scores for the qth eigencomponent are on the q column of $\mathbf{Z}\mathbf{V}_Q$, for $q = 1, \ldots, Q$.

We divide the data into a training and testing set. On the training set we implement our method, and choose the number of principal components used in the regression based on a threshold on the cumulative variance explained. After we estimate the coefficients

from the regression model, we construct partial ROC (pROC) curves to determine a cut-off probability for deciding if voxel $v_{i,j}$ is part of a lesion. We use pROC curves instead of ROC curves because lesions are often a small fraction of the overall brain volume. Using ROC curves in these cases would result in very large AUCs simply by wrongly declaring that none of the voxels are in a lesion. Instead, pROCs focus on the high specificity area of the ROC curve, which provides for meaningful comparisons of stroke segmentation algorithm performance. To the best of our knowledge, this was first proposed in the context of brain image segmentation (12; 13). In this paper the pROC curve refers to the ROC curve up to a false positive rate (FPR) of 0.05. The maximum AUC for this pROC would be at 0.05 (perfect prediction) and chance prediction would correspond to AUC $= 0.05 \times 0.05 = 0.0025$. To normalize the pAUC, we provide the proportion of the AUC out of the maximum AUC. For example, if an AUC up to a False Positive Rate of 0.05 is 0.04 we will say that the pAUC (out of the total possible) is $0.04/0.05 = 0.80$. This normalization is done so that the pAUC is scaled between 0 and 1, much like traditional AUC (14). In addition to pROC curves, we also consider the Dice Coefficient. The Dice for two sets A and B is defined as Dice $= \frac{2|A \cap B|}{|A|+|B|}$. In our calculations A is the gold standard segmentation and B the proposed automatic segmentation. The Dice was first proposed in (2).

The pROC curves are created based on the gold standard, which is a manual segmentation done by an expert. Once a cut-off probability is set, we calculate pAUC to assess the model performance in the training set. The estimates for the coefficients and the cut-off probability obtained on the training data are then used on the testing data and pROCs curves and pAUCs estimates are calculated on the test data.

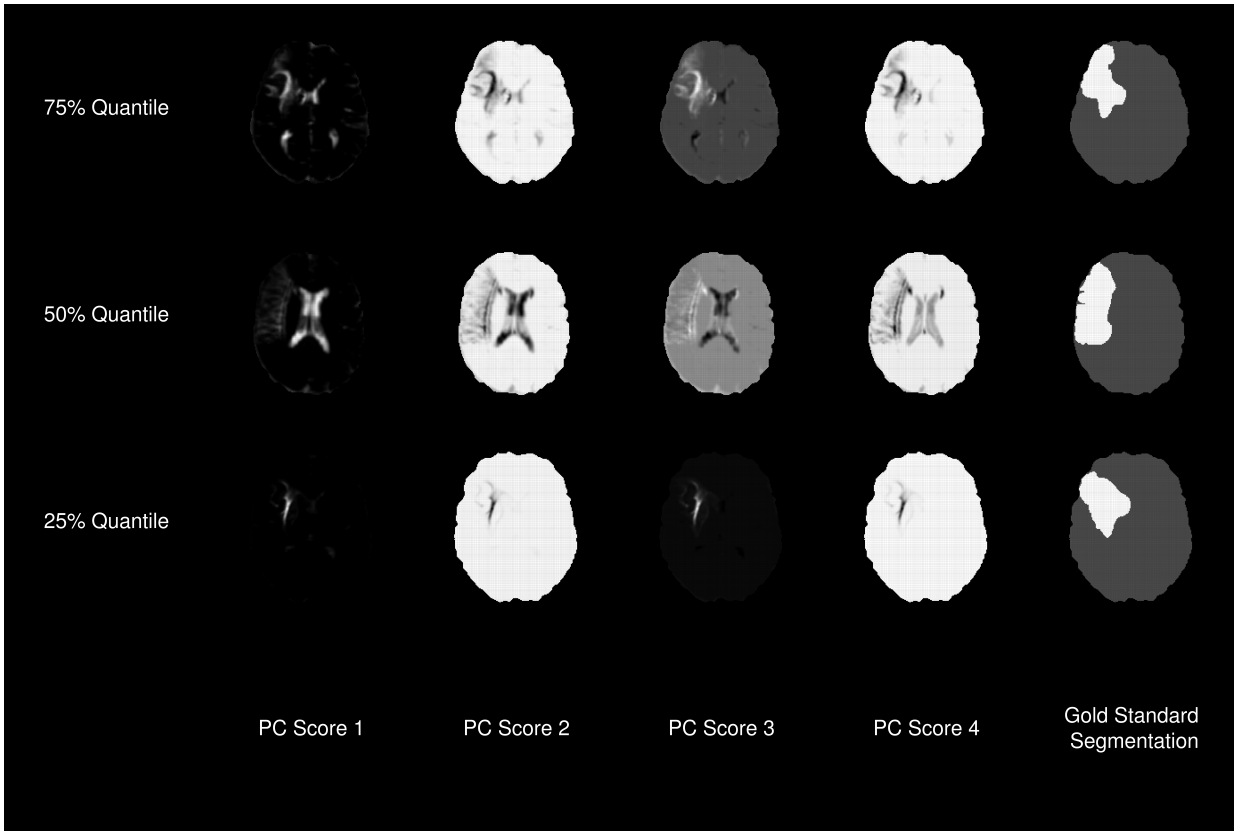

Figure 3: **Resulting PC scores**: Principal component scores for first four PCs and gold standard segmentation for three subjects. Principal component scores from our local neighborhood PCA method generally detect variation due to lesion and non-lesion tissue.

3 Results

3.1 Score images

For our analysis we use all four imaging sequences included in the ISLES Data (FLAIR, DWI, T1w, and T2w). We also included up to the 4th order moment of local neighborhoods.

Figure 3 displays the score images of three subjects in the testing group (on rows) on the first four PCs (first four columns) and the gold standard segmentations (last column). These three subjects were selected based on their pAUC for the testing data. Specifically, these subjects were the 25%, 50%, and 75% quantiles in pAUC. These quantiles correspond to below-median, median, and above-median segmentation in the testing data. The first PC

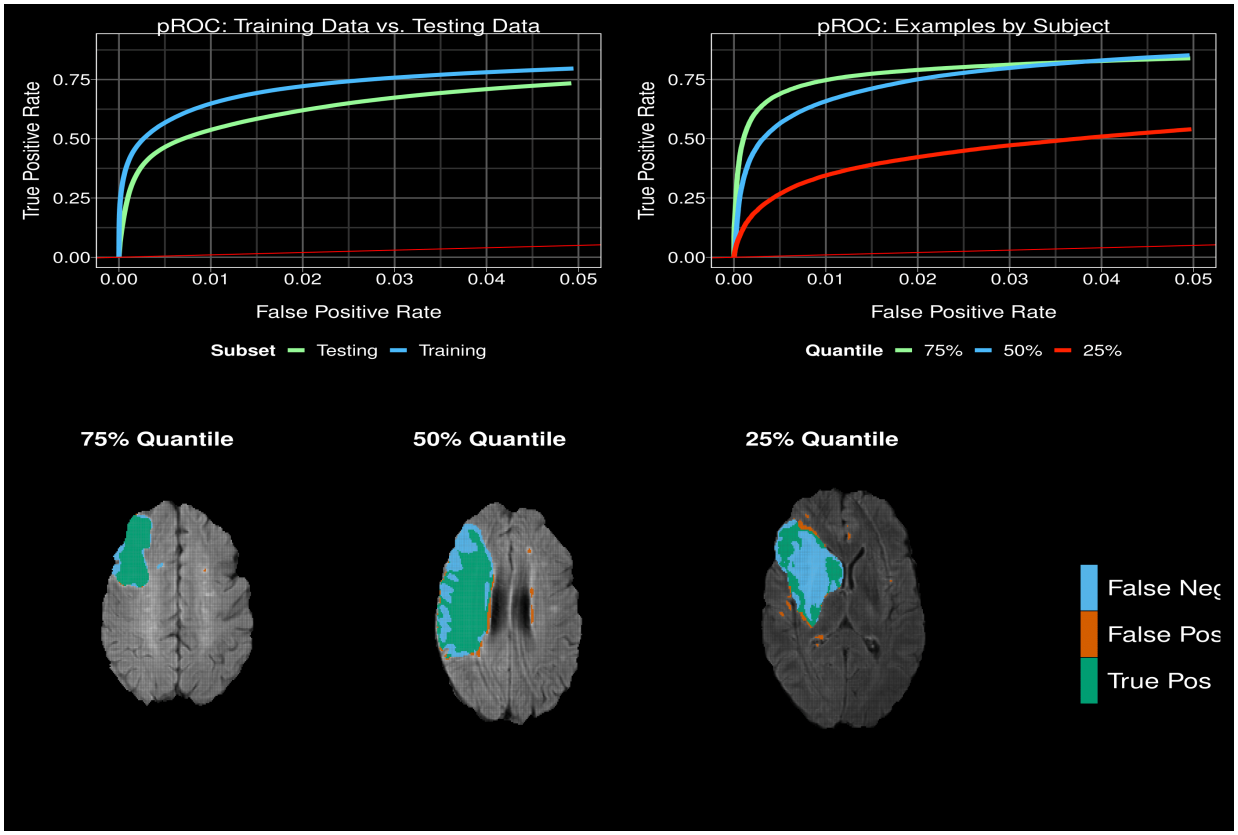

Figure 4: **Segmentation results**: Segmentation results in the form of pROC curves for training and testing data (top-left), pROC curves for three subjects with varying degrees of performance (top-right), and examples of segmentations for those subjects (bottom).

score (column one) seems to capture the overall mean intensity with some contrast between the ventricles and the rest of the brain (especially visible in the second row, first column image). The second, third, and fourth PC scores (columns two, three, and four) seems to capture complementary information about the lesion; note, for example, the second row, where the lesion appears clearly in the score images.

3.2 Segmentation

In this section we assess the segmentation performance for stroke lesions. We used a sufficient amount of PCs to explain 80% of the variability of the neighborhood information in the training data. These PCs were plugged into a logistic regression, where the outcome is

binary at every voxel (voxel is or not in the manual segmentation of the lesion) and the predictors are the voxel-specific PC scores. The resulting output from the PCR model is a probability image where the value at each voxel is the probability that that particular voxel is in stroke lesion. To provide a hard segmentation we must establish a probability threshold for calling lesions; this, together with the number of PCs are the two tuning parameters of our procedure. We do this by creating pROC curves for all subjects in the training data and by selecting a global threshold based on the overall pROC curves in testing data collapsed over subject. This results in a global threshold, that is used for every subject and is based on the false positive rate. More precisely, for our application we selected a threshold corresponding to 5% of the voxels in the population level training data being classified as false positives, which corresponds to an estimated probability of a voxel to be in a lesion of 3.4%.

Figure 4 shows the overall pROC curves for the training (blue lines) and testing (green lines) data; not all pROC curves are shown because there are too many lines and the plot becomes difficult to read. Instead, we display the pROC curves for the three subjects introduced in section 3.1. The pROCs of these subjects correspond to the 25%, 50%, and 75% quantiles in pAUC. The values of pAUC for these subjects corresponds to 0.418, 0.726, and 0.770, respectively. Finally, we show the resulting segmentations for these subjects and compare them with the gold standard segmentation. In Figure 4, green corresponds to true positive, blue corresponds to false negative, and red corresponds false positive voxels. Overall, many voxels are correctly classified but more misclassifications occur with false negatives than false positives.

3.3 Choice of moments

In this section we consider: (1) how changing the number of moments affects the resulting score images; and (2) how these score images change segmentation performance. Figure 5

Table 1: Mean and standard deviation of pAUC, and Dice for segmentations using local neighborhood PCA including up to specified moment with enough PCs to account for 80% of total variation.

	pAUC				Dice			
	Training		Testing		Training		Testing	
	Mean	SD	Mean	SD	Mean	SD	Mean	SD
1st Order	0.617	0.182	0.681	0.191	0.259	0.263	0.301	0.265
2nd Order	0.621	0.181	0.687	0.180	0.258	0.262	0.299	0.266
3rd Order	0.632	0.179	0.641	0.181	0.258	0.259	0.281	0.253
4th Order	0.634	0.180	0.631	0.206	0.263	0.262	0.279	0.252

displays an example of how the score images change when more or less moments are included in the analysis. We show one slice for one subject, with data shown on the first row together with the manual segmentation.

In Table 1 we show how the pAUC and Dice changes in the training and testing data when we include up to a given number of moments. The automatic segmentation was obtained by using a global threshold that set the percent of false positive voxels at 5% of the voxels in the training set. Once this threshold is established, we calculate pAUC and Dice for each subject then report the mean and standard deviation from these calculations.

Table 1 indicates that, in general, pAUC is slightly larger in the testing data than in the training data, except for in the 4th order analysis where the pAUC is slightly larger. This seems to suggest there might be an overfitting issue in this dataset when more than three moments are used. It may seem odd that pAUC and Dice are generally larger in the testing data, which may be an artifact of more lesion voxels in the testing data that training data (672476 vs. 553783 by gold standard). It also appears that including information only up to the second moment provides the largest pAUC in the testing data. This suggests that, for this problem, increasing the number of moments increases the complexity of the problem without adding anything in terms of prediction performance. Note that for each analysis we included enough principal components to account for 80% of the total variation seen in the

data. For this reason, results from this analysis shown in Figure 5 have fewer or more images depending on the analysis because of the different number of principal components required to account for 80% of the variability. It is also important to note that though Dice is fairly modest in Table 1, this is because we did not attempt to optimize the probability threshold for Dice, we only optimized to control the FPR. If one wants to optimize with respect to Dice, this is possible. As an example, if we optimize the 4th order analysis with respect to Dice in the training data the probability threshold for classifying lesions becomes 44.2% vs. 3.4% if we only control FPR at 5%. This results in a Dice of 0.402 in the training data and 0.413 in the testing data.

4 Discussion

We have provided a method for exploring the variation of local neighborhoods in multisequence imaging data. Our proposed approach is computationally efficient and can be implemented with limited computing resources. The associated R (9) package *****(removed for anonymous review) provides an efficient implementation of the methods described here and is deployed on Neuroconductor (7). We have shown that the method produces reasonable results using the 2015 ISLES challenge data. It is worth noting that this method could be used in other types of 3-dimensional imaging structures, not only medical images.

Our method provides a new way to explore the structure of higher order moments of local neighborhoods of multi-sequence structural MRI data. We believe that exploring multisequential MRI data in this way can provide additional insight. It also provides an alternative way of conducting segmentation using a novel combination of basic tools for information representation. The method is generally applicable to image segmentation problems and could be used as a first line approach, benchmark approach for future methods.

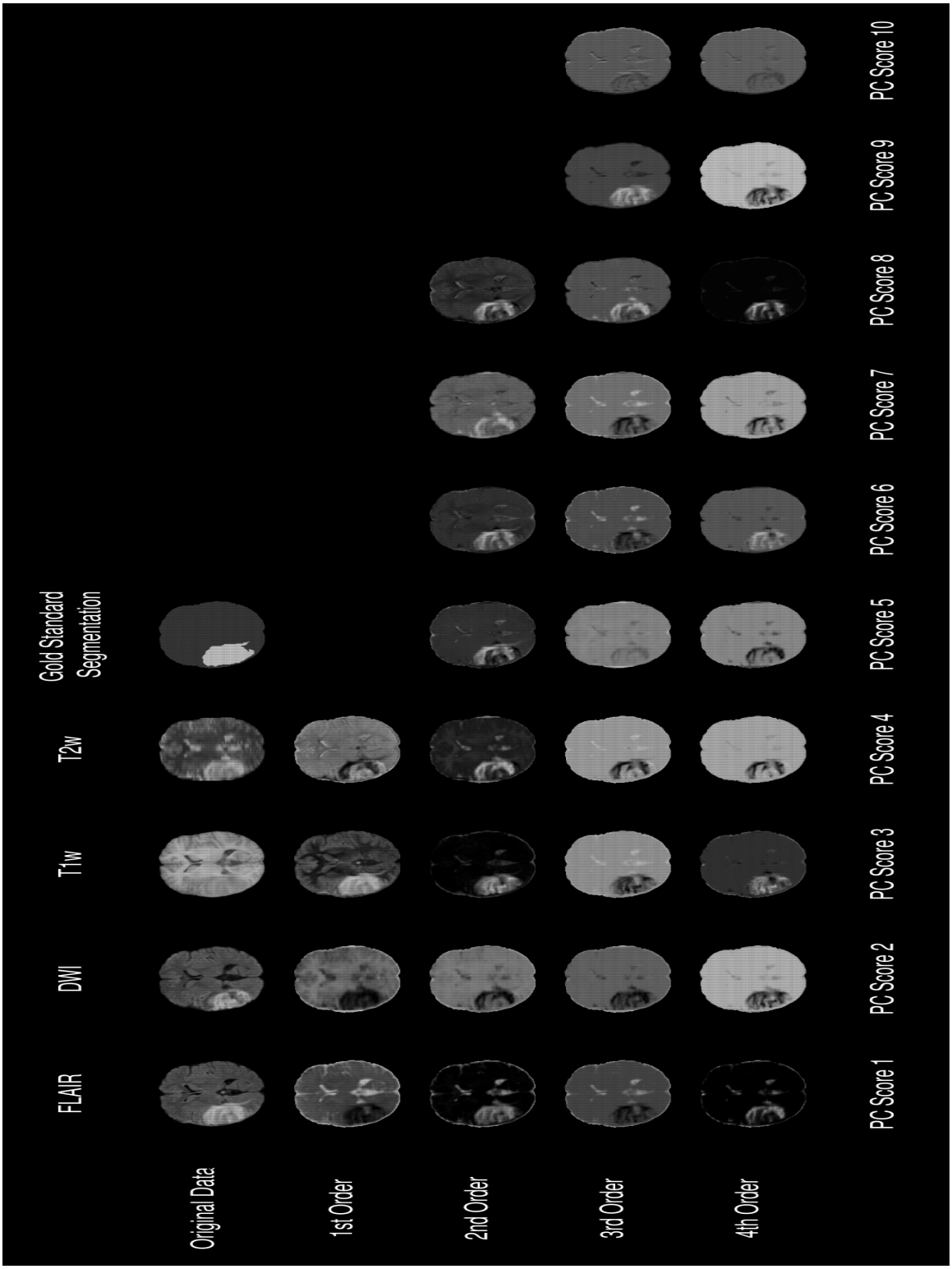

Figure 5: **Moments and their impact on PC scores**: Original data (top row) and score images from analyses that include up the the 1st, 2nd, 3rd, and 4th local neighborhood moments (bottom four rows) for one subject. Score images change depending on the number of moments included in analysis.

Acknowledgements

This work was supported by NIH grants R01HL123407, R01HL094786, and R01NS060910. This work was also supported by a University of Wisconsin - Madison Morse Fellowship.

References

[1] Ruth Bonita. Epidemiology of stroke. *The Lancet*, 339(8789):342 – 344, 1992. Originally published as Volume 1, Issue 8789.

[2] Lee R. Dice. Measures of the amount of ecologic association between species. *Ecology*, 26(3):297–302, 1945.

[3] Jean Philippe Fortin, Elizabeth M. Sweeney, John Muschelli, Ciprian M. Crainiceanu, and Russell T. Shinohara. Removing inter-subject technical variability in magnetic resonance imaging studies. *NeuroImage*, 132:198–212, May 2016.

[4] D.J. Kroon, E.S.B. van Oort, and C.H. Slump. Multiple sclerosis detection in multispectral magnetic resonance images with principal components analysis. In *3D Segmentation in the Clinic: A Grand Challenge II: MS lesion segmentation*, Website, September 2008. Kitware.

[5] Lihong Li, Dongqing Chen, Hongbing Lu, and Zhengrong Liang. Segmentation of brain mr images: a self-adaptive online vector quantization approach. In *Medical Imaging 2001: Image Processing*, pages 1431–1438. International Society for Optics and Photonics, July 2001.

[6] Oskar Maier, Bjoern H Menze, Janina von der Gablentz, Levin Ḧani, Mattias P Heinrich, Matthias Liebrand, Stefan Winzeck, Abdul Basit, Paul Bentley, Liang Chen, Daan

Christiaens, Francis Dutil, Karl Egger, Chaolu Feng, Ben Glocker, Michael Götz, Tom Haeck, Hanna-Leena Halme, Mohammad Havaei, Khan M Iftekharuddin, Pierre-Marc Jodoin, Konstantinos Kamnitsas, Elias Kellner, Antti Korvenoja, Hugo Larochelle, Christian Ledig, Jia-Hong Lee, Frederik Maes, Qaiser Mahmood, Klaus H Maier-Hein, Richard McKinley, John Muschelli, Chris Pal, Linmin Pei, Janaki Raman Rangarajan, Syed M S Reza, David Robben, Daniel Rueckert, Eero Salli, Paul Suetens, Ching-Wei Wang, Matthias Wilms, Jan S Kirschke, Ulrike M Kr Amer, Thomas F Münte, Peter Schramm, Roland Wiest, Heinz Handels, and Mauricio Reyes. Isles 2015 - a public evaluation benchmark for ischemic stroke lesion segmentation from multispectral mri. *Medical image analysis*, 35:250–269, 01 2017.

[7] J. Muschelli, A. Gherman, J. P. Fortin, B. Avants, B. Whitcher, J. D. Clayden, B. S. Caffo, and C. M. Crainiceanu. Neuroconductor: an R platform for medical imaging analysis. *Biostatistics*, Jan 2018.

[8] John Muschelli, Elizabeth M Sweeney, Natalie L Ullman, Paul Vespa, Daniel F Hanley, and Ciprian M Crainiceanu. Pitchperfect: Primary intracranial hemorrhage probability estimation using random forests on ct. *NeuroImage : Clinical*, 14:379–390, 2017.

[9] R Core Team. *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing, Vienna, Austria, 2018.

[10] Islem Rekik, Stéphanie Allassonnière, Trevor K. Carpenter, and Joanna M. Wardlaw. Medical image analysis methods in mr/ct-imaged acute-subacute ischemic stroke lesion: Segmentation, prediction and insights into dynamic evolution simulation models. a critical appraisal. *NeuroImage: Clinical*, 1(1):164 – 178, 2012.

[11] Russell T. Shinohara, Elizabeth M. Sweeney, Jeff Goldsmith, Navid Shiee, Farrah J.

Mateen, Peter A. Calabresi, Samson Jarso, Dzung L. Pham, Daniel S. Reich, and Ciprian M. Crainiceanu. Statistical normalization techniques for magnetic resonance imaging. *NeuroImage: Clinical*, 6:9–19, 2014.

[12] E. M. Sweeney, R. T. Shinohara, C. D. Shea, D. S. Reich, and C. M. Crainiceanu. Automatic lesion incidence estimation and detection in multiple sclerosis using multisequence longitudinal MRI. *AJNR Am J Neuroradiol*, 34(1):68–73, Jan 2013.

[13] E. M. Sweeney, R. T. Shinohara, N. Shiee, F. J. Mateen, A. A. Chudgar, J. L. Cuzzocreo, P. A. Calabresi, D. L. Pham, D. S. Reich, and C. M. Crainiceanu. OASIS is Automated Statistical Inference for Segmentation, with applications to multiple sclerosis lesion segmentation in MRI. *Neuroimage Clin*, 2:402–413, 2013.

[14] S. D. Walter. The partial area under the summary roc curve. *Statistics in Medicine*, 24(13):2025–2040, 2005.

[15] Vadim Zipunnikov, Brian Caffo, David M. Yousem, Christos Davatzikos, Brian S. Schwartz, and Ciprian Crainiceanu. Multilevel functional principal component analysis for high-dimensional data. *J Comput Graph Stat.*, 20(4):852–873, 2011.

Appendix

A Choice of number of principal components

In Section 3, we explored how the choice of order of local moments incorporated effected segmentation. For the analysis we included the number of principal components to account for 80% of the variation in the data. Here we inspect this analysis by increasing the number of PCs to account for 90% of the variation in the data.

In Table 2 we show pAUC and Dice for the resulting segmentations when we include enough PCs to account for 90% of the variation in the original data. This is done when we include only the 1st order moments, 1st and 2nd order moments, and so on up to the 4th order moments. This is similar to the analysis done in Section 3.3, but there we only used enough PCs to account for 80% of the variation in the original data. The results of the analysis including 90% of the variation in the data seem very similar to the analysis including 80% of the variation. This seems to suggest that the analysis in Section 3 contains a sufficient number of PCs.

Table 2: Mean and standard deviation of pAUC and Dice for segmentations using local neighborhood PCA including up to specified moment with enough PCs to account for 90% of total variation.

	pAUC				Dice			
	Training		Testing		Training		Testing	
	Mean	SD	Mean	SD	Mean	SD	Mean	SD
1st Order	0.620	0.184	0.679	0.191	0.260	0.263	0.301	0.265
2nd Order	0.629	0.183	0.696	0.180	0.260	0.263	0.302	0.266
3rd Order	0.642	0.189	0.666	0.186	0.261	0.261	0.286	0.254
4th Order	0.662	0.191	0.634	0.213	0.262	0.261	0.276	0.251