

Video Article

Identification of Disease-related Spatial Covariance Patterns using Neuroimaging Data

Phoebe Spetsieris¹, Yilong Ma¹, Shichun Peng¹, Ji Hyun Ko¹, Vijay Dhawan¹, Chris C. Tang¹, David Eidelberg¹

¹Center for Neurosciences, The Feinstein Institute for Medical Research

Correspondence to: David Eidelberg at david1@nshs.edu

URL: <https://www.jove.com/video/50319>

DOI: [doi:10.3791/50319](https://doi.org/10.3791/50319)

Keywords: Medicine, Issue 76, Neurobiology, Neuroscience, Anatomy, Physiology, Molecular Biology, Basal Ganglia Diseases, Parkinsonian Disorders, Parkinson Disease, Movement Disorders, Neurodegenerative Diseases, PCA, SSM, PET, imaging biomarkers, functional brain imaging, multivariate spatial covariance analysis, global normalization, differential diagnosis, PD, brain, imaging, clinical techniques

Date Published: 6/26/2013

Citation: Spetsieris, P., Ma, Y., Peng, S., Ko, J.H., Dhawan, V., Tang, C.C., Eidelberg, D. Identification of Disease-related Spatial Covariance Patterns using Neuroimaging Data. *J. Vis. Exp.* (76), e50319, doi:10.3791/50319 (2013).

Abstract

The scaled subprofile model (SSM)¹⁻⁴ is a multivariate PCA-based algorithm that identifies major sources of variation in patient and control group brain image data while rejecting lesser components (**Figure 1**). Applied directly to voxel-by-voxel covariance data of steady-state multimodality images, an entire group image set can be reduced to a few significant linearly independent covariance patterns and corresponding subject scores. Each pattern, termed a group invariant subprofile (GIS), is an orthogonal principal component that represents a spatially distributed network of functionally interrelated brain regions. Large global mean scalar effects that can obscure smaller network-specific contributions are removed by the inherent logarithmic conversion and mean centering of the data^{2,5,6}. Subjects express each of these patterns to a variable degree represented by a simple scalar score that can correlate with independent clinical or psychometric descriptors^{7,8}. Using logistic regression analysis of subject scores (*i.e.* pattern expression values), linear coefficients can be derived to combine multiple principal components into single disease-related spatial covariance patterns, *i.e.* composite networks with improved discrimination of patients from healthy control subjects^{5,6}. Cross-validation within the derivation set can be performed using bootstrap resampling techniques⁹. Forward validation is easily confirmed by direct score evaluation of the derived patterns in prospective datasets¹⁰. Once validated, disease-related patterns can be used to score individual patients with respect to a fixed reference sample, often the set of healthy subjects that was used (with the disease group) in the original pattern derivation¹¹. These standardized values can in turn be used to assist in differential diagnosis^{12,13} and to assess disease progression and treatment effects at the network level^{7,14-16}. We present an example of the application of this methodology to FDG PET data of Parkinson's Disease patients and normal controls using our in-house software to derive a characteristic covariance pattern biomarker of disease.

Video Link

The video component of this article can be found at <https://www.jove.com/video/50319/>

Introduction

Neurodegenerative disorders have been extensively studied using techniques that localize and quantify abnormalities of brain metabolism as well as non-inferential methods that study regional interactions¹⁷. Data-driven multivariate analytical strategies such as principal component analysis (PCA)^{1,2,4,18} and independent component analysis (ICA)^{19,20}, as well as supervised techniques such as partial least squares (PLS)²¹ and ordinal trends canonical variates analysis (OrT/CVA)²² can reveal characteristic patterns or "networks" of interrelated activity. The basics of multivariate procedures, particularly the scaled subprofile model (SSM)^{1,2,4-6,18} have been previously described in JoVE³. This PCA-based approach was originally developed to examine abnormal functional covariance relationships between brain regions in steady-state single volume images of cerebral blood flow and metabolism acquired in the resting state of modalities such as PET and SPECT that exhibit high signal-to-noise characteristics. Disease-specific SSM patterns are imaging biomarkers that reflect overall differences in the regional topography in patients compared to normal subjects^{7,16} and may reflect a single network process or the assimilation of several complex abnormal functions²³. Metabolic covariance pattern brain networks are associated with expression values (subject scores) that can distinguish between normal and disease groups and provide network-based measures that correlate with clinical ratings of disease severity. Typically, subject scores for such patterns increase with disease progression and may even be expressed before symptom onset^{14,24}. Indeed, disease-related network biomarkers have been characterized for neurodegenerative disorders such as Parkinson's disease¹⁰ (PD), Huntington's disease²⁵ (HD), and Alzheimer's disease⁸ (AD). Importantly, disease-related metabolic topographies have also been identified for atypical parkinsonian movement disorders such as multiple system atrophy (MSA) and progressive supranuclear palsy (PSP). These patterns have been used in concert for the differential diagnosis of individuals with clinically similar "look-alike" syndromes^{12,13,26}.

In contrast, typical fMRI voxel-based univariate methods assess the significance of differences between patients and controls in isolated brain clusters. More recently, methods have been developed to measure functional connectivity between variously defined brain regions²⁷⁻²⁹. This definition of functional connectivity is restricted to subject and region specific interactions and deviates from the original SSM/PCA concept that refers to the cross-sectional interconnectivity of intrinsic spatially distributed brain network regions^{1,2,23,30}. To their advantage, MRI platforms are easily installed, widely available, non-invasive and typically require shorter scanning time than traditional radiotracer imaging modalities such

as PET or SPECT resulting in an upsurge of potential methodologies described in recent literature. However, the resulting time-dependent fMRI signals provide indirect measures of local neural activity^{31,32}. The generally complex analytical algorithms employed have been limited by the large size of datasets, the physiological noise inherent in fMRI signals, as well as the high variability in brain activity that exists between subjects and regions^{19,23}. Although interesting information regarding brain organization can be inferred from the properties of fMRI "networks", they have not been sufficiently stable to be used as reliable disease biomarkers. Furthermore, the resulting network topographies are not necessarily equivalent to those identified using established functional imaging methodologies such as SSM/PCA. For the most part, rigorous cross-validation of the resulting fMRI topographies has been lacking with few examples of successful forward application of derived patterns in prospective scan data from single cases.

An advantage of PCA covariance analysis lies in its capacity to identify the most significant sources of data variation in the first few principal components but it is ineffective if the prominent eigenvectors represent random noise factors rather than actual intrinsic network response. By selecting only the first few eigenvectors and limiting to those that show significant differences in patient versus normal control scores, we greatly reduce the influence of noise elements. However, for the basic approach described here, these measures may not be adequate to generate robust estimators in a typical fMRI dataset with the exception of the modalities described below.

Thus, because of the stable direct relationship of regional glucose metabolism and synaptic activity³³, this methodology has been applied primarily to the analysis of rest state FDG PET data. However, given that cerebral blood flow (CBF) is closely coupled to metabolic activity in the resting state^{10,11,34}, SPECT^{35,36} and more recently arterial spin labeling (ASL) MRI perfusion imaging methods^{37,38}, have been used to assess abnormal metabolic activity in individual cases. That said, the derivation of reliable spatial covariance patterns with resting state fMRI (rsfMRI) is as previously noted not straightforward^{31,32}. Even so, preliminary SSM/PCA analysis of rsfMRI data from PD patients and control subjects has revealed some topographical homologies between disease-related patterns identified using the two modalities, PET and amplitude of low-frequency fluctuations (ALFF) of BOLD fMRI^{39,40}. Lastly, we also note that this approach has been applied successfully in voxel based morphometry (VBM) structural MRI data^{41,42}, revealing distinctive spatial covariance patterns associated with age-related volume loss and in further comparisons of VBM and ASL patterns in the same subjects⁴³. The relationship between SSM/PCA spatial covariance topographies and analogous brain networks identified using different analytical approaches and imaging platforms is a topic of ongoing investigation.

Protocol

1. Data Collection and Preprocessing

1. The SSM/PCA method can be applied to single volume images obtained from various sources and modalities. Specifically, for on-site PET imaging of metabolism, prepare a suitable radionuclide tracer such as [¹⁸F]-fluorodeoxyglucose (FDG) and administer to each patient. Patients are usually scanned at rest with eyes open, following a fast of at least 12 hr, off medications.
2. Scan each subject for individual or group assessment. For pattern derivation, scan an equal number of gender- and age-matched patients and controls.
3. Transfer data to a workstation and convert to an appropriate format for analysis in your platform. Our windows PC based MATLAB analysis software (scanvp, ssm_pca, www.feinsteinneuroscience.org) requires ANALYZE or *nifti* format images (Mayo Clinic, Rochester, MN). It provides a conversion routine to transform GE Advance (Milwaukee, WI, USA) scanner and other format images to ANALYZE format.
4. Normalize each subject's image to a common stereotaxic space (e.g. Montreal Neurological Institute [MNI]) using a standard neuroimaging software package such as statistical parametric mapping (SPM) (<http://www.fil.ion.ucl.ac.uk/spm>) so that there is a one-to-one correspondence of voxel values between subjects (Figure 2). Masking to limit the analysis to gray matter areas (Figure 3) and log transformation are described in the next steps.

2. Perform Multivariate SSM/PCA

Operations for multivariate SSM/PCA (Figure 4) can be performed by external software. The steps itemized below reflect the procedures performed for the most part automatically by our in-house routines (Figure 5) (scanvp/ssmpca, www.feinsteinneuroscience.org) also available as an SPM toolbox ssm_pca.

1. Mask data using an available 0/1 image to remove unwanted areas of the voxel space such as white matter and ventricles. Using the ssmpca routine, the user is given an option to select the default or other external mask (Figure 3) or to have the program automatically create a mask by removing values lower than a selected fixed percent of each subject's data maximum. Individual subject masks are then multiplied to determine a composite mask. The areas within the mask define a common non-zero voxel space for the analysis.
2. Convert each subject's 3D masked image voxel data to one continuous row vector by appending sequential scan lines from consecutive planes (Figure 4a). Form a group data matrix (**D**) so that each subject's data corresponds to a specific row of the matrix. Each column then represents a particular voxel across subjects. Ideally, for biomarker derivation, the matrix will be composed of an equal number of rows of normal control subject data and disease subject data.
3. Transform each data entry to logarithmic form.
4. Center the data matrix by subtracting each row average or subject mean from the row elements. The average of all centered rows represents a characteristic group mean log image termed a group mean profile (**GMP**, Figure 4a). Subtract the column means that are the elements of the GMP from the corresponding matrix column elements. Each row of the double centered matrix represents a 'residual' image termed a subject residual profile (**SRP**) whose elements represent deviations from both the subject *s* and voxel *v* group means (Figure 4b).

$$SRP_{sv} = \log D_{sv} - \text{mean}_s - \text{GMP}_v$$
5. Construct the subject-by-subject covariance matrix **C** of the composite double centered data matrix by computing the non-normalized covariance between each subject pair *i, j* of **SRP** matrix rows evaluated as a product of corresponding voxel elements summed over all voxels *v* (Figure 4b).

$$C_{ij} = \sum_v (SRP_{iv} \times SRP_{jv})$$

6. Apply PCA to the subject-by-subject covariance matrix **C**. The results will be a set of subject score eigenvectors with associated eigenvalues. Weight each vector by multiplying by the square root of its corresponding eigenvalue. The set of score eigenvectors is represented by the columns of the resulting matrix **S** in **Figure 4c**.
7. Voxel eigenvectors for the same set of eigenvalues can be determined by applying PCA to the column by column covariance matrix or by this alternative computationally less demanding procedure depicted in (**Figure 4c**). Left multiply the previously derived score vector matrix by the transpose **SRP** matrix to derive an array **P** of voxel pattern eigenvectors in descending order of eigenvalues (**Figure 4c**, **Figure 6**, **Figure 7a**).

$$P = SRP^T \times S$$

Each column vector represents a principal component (PC) image pattern of the SSM/PCA analysis attributed to a percent of the total variance accounted for (vaf) corresponding to the relative size of its eigenvalue.

3. Pattern Biomarker Derivation

1. Examine the results of the preceding analysis to determine PC patterns that are associated with high vaf values. Within our routine, voxel pattern vectors are Z-transformed so that their values represent positive and negative standard deviations from their mean value. They are then reshaped into image format prior to display (**Figure 7a**).
2. In some cases patterns with known regional deviations associated with the disease can be visually identified. The mathematical formulation considers positive and negative forms of the derived PCs as equivalent solutions in which case both the PCs and their associated scores may be multiplied by minus one to conform to a physically correct interpretation.
3. Scores corresponding to each PC pattern are displayed as bar graphs and scatter plots (**Figure 7**). An optional ROC plot can be generated (**Figure 7d**). To identify a disease-specific pattern, notice the differentiation of subject scores between patients and controls reflected by the p-values of the corresponding two-sample t-tests and AUC values. Limit the analysis to those PCs associated with a high vaf and high differentiation for some fixed cutoff values (e.g., $p < 0.05$ and $vaf > 5\%$). Typically, only one or two PCs satisfy this criterion for PET data.
4. There are various other ways for PC selection that can be considered⁴⁴. The scree plot of sequential eigenvalues (**Figure 6**) may give a sharp cutoff value represented by a bend in the curve and a sharp reduction in the slope of the curve where eigenvalues begin to degenerate. Another approach is to include all PCs that account for the top 50% of the variance. A widely accepted procedure is to compute the Akaike Information Criterion (AIC)⁴⁵ to determine which combination of PCs define the model with the lowest AIC value that can distinguish between patients and controls.
5. The selected PC(s) can be vector normalized and linearly combined to yield a single disease-related pattern. Our software optionally uses the MATLAB function *glmfit* to determine coefficients based on logistic or other regression models applied to subject scores. Although differentiation of patient and control groups usually improves with the additional PCs considered in the derivation group, the resultant patterns are a composite representation that may not correspond to a single physical network or may incorporate outlier deviations (**Figures 7a** and **7c**).
6. Further validation is required for reliability and prospective significance. Bootstrap resampling can be performed as discussed below⁸ to identify the most reliable voxels within the original derivation dataset associated with the least standard deviation in repeated pattern derivation. Forward validation is performed to test for the sensitivity and specificity of independent group discrimination by deriving scores for prospective groups of patients and controls using the single-subject score evaluation method (**Figure 4d**) described in the next segment of the protocol.

4. Single-subject Prospective Score Evaluation using a Predetermined Biomarker

1. Once a significant SSM-GIS biomarker pattern has been identified, a score for its expression in a prospective subject can be evaluated from that individual's scan using a simple computation of the internal vector product of the subject's SRP vector and the GIS pattern vector (**Figure 4d**, **Figure 7d**).

$$SCORE = SRPs \cdot Pattern$$
2. The previous operation is automated by our TPR routine. However, to independently derive the subject **SRPs** vector use the associated intrinsic GIS mask on the log transformed data and subtract both the subject mean and the corresponding voxel values of the prederived reference group **GMP** as in step 2.4. This insures that scores can be compared to the predetermined reference range. Scores for a new group can be similarly evaluated as a set of prospective single subject scores. For use with a different reference group or setting, **GMP** can be recalibrated while the pattern is unchanged.

Representative Results

A simple application of multivariate SSM/PCA analysis to derive a neuroimaging biomarker pattern for PD is illustrated below. PET FDG images of ten clinically diagnosed PD patients (6M/4F, 59y $\pm$ 7y sd) of variable diseased duration (9y $\pm$ 5y sd) and ten age and gender matched normal controls (6M/4F, 58y $\pm$ 7y sd) were analyzed using our *ssmpca* routine. The twenty corresponding spatially pre-normalized images were initially selected under the categories disease subjects or controls along with the default mask (**Figure 3**). The program output the first 16 principal component image files and a file listing of their associated scores along with a scree plot of sequential eigenvalues (**Figure 6**). Sequential PC images and associated bar and scatter graphs were displayed on the screen for initial review as depicted in **Figures 7a** and **7b**. Note that the first PC component accounted for the largest vaf (24%) and its associated subject scores significantly discriminate patients from controls ($p=0.0002$) based on two-sample t-test comparison and ROC characteristics (AUC=0.99, sensitivity=0.93 for specificity=0.95). For these reasons it could be considered a biomarker by itself and is in fact highly spatially correlated with a previously derived PD biomarker PDRP^{46,47} ($R^2=84\%$, $p < 0.001$) validated in numerous prospective datasets⁵. The second pattern ($p=0.81$, AUC=0.54) and third pattern ($p=0.38$, AUC=0.63) did not discriminate significantly and may originate from normal processes that are not highly affected by disease. The fourth pattern (vaf=7 %) discriminates to a far lesser degree than the first PC ($p=0.13$, AUC=0.71) and may be associated with less prominent disease factors. Further PCs were ignored because of decreasing vaf values ($< 7\%$) combined with non-significant discrimination of normal and patient scores ($p > 0.4$). The degeneration of eigenvalue differences is illustrated in **Figure 6**.

The first and fourth patterns were vector normalized and linearly combined using our software to determine associated coefficients (.78, .63) for optimum patient/control score discrimination using a logistic discrimination model. The composite GIS results in improved discrimination as indicated by the p-value ($p=10^{-5}$, **Figure 7b**) and a perfect AUC value of 1.0 for the derivation group. The composite vaf (17.3%) associated with this pattern is evaluated as the sum of the individual vafs (24%, 7%) modified by the squares of the linear coefficients. However, note that this model achieved higher discrimination by attributing a disproportionately high coefficient (in comparison to vaf values) to the less significant PC4. Even higher discrimination within the derivation group may be achieved using additional components with the same reservation. In our case there was only slight improvement by adding PC3 ($p=3 \times 10^{-6}$) (**Figure 7c**) and no improvement by including all four PCs ($p=10^{-5}$) because of the non-discriminant capacity of PC2. All of the combined patterns had perfect ROC characteristics (AUC=1 and sensitivity=1, specificity=1 for a low Z-score threshold). However, these values are specific to the derivation group. The validity of the final biomarkers has to be verified in prospective evaluation as performed here for an independent group of 22 patients (15M/7F, $57y \pm 9y$ sd; disease duration $10y \pm 4y$ sd) and 22 controls (4M/18F, $55y \pm 15y$ sd) (**Figure 7d**). Although significantly higher differentiation was achieved in the derivation group by combining more than one PC, the same relative advantage was not maintained in the test group although all four combinations performed well. The first PC demonstrated higher mean differentiation in the prospective group data ($p=5 \times 10^{-8}$, AUC=0.95) than the combined patterns PC1_3_4 ($p=2 \times 10^{-7}$, AUC=0.95) and PC1_2_3_4 ($p=6 \times 10^{-7}$, AUC=0.92) attributed to its innate validity. A minor advantage was achieved by the additive PC1_4 ($p=3.2 \times 10^{-8}$, AUC=0.96). Although the prospective AUC value was slightly higher for PC1_4, the sensitivity for a specificity of 0.95 appears to decrease with additive PCs from a value of approximately 0.8 for PC1 to 0.5 for PC1_2_3_4. Clearly more extensive sampling would be necessary to more accurately predict these values judging from the increasing irregularity of the ROC plot. In addition, higher accuracy could have been achieved by using a larger derivation group as demonstrated in previous publications. However, it is apparent from these displays that the same discriminant accuracy attained in the derivation group does not always generalize to independent group evaluation, necessitating prospective validation of the derivation patterns.

Thus, our original dataset of 20 subject images was reduced to one essential PC image pattern PC1 that performed well as a diagnostic biomarker of disease in 44 prospective subjects.

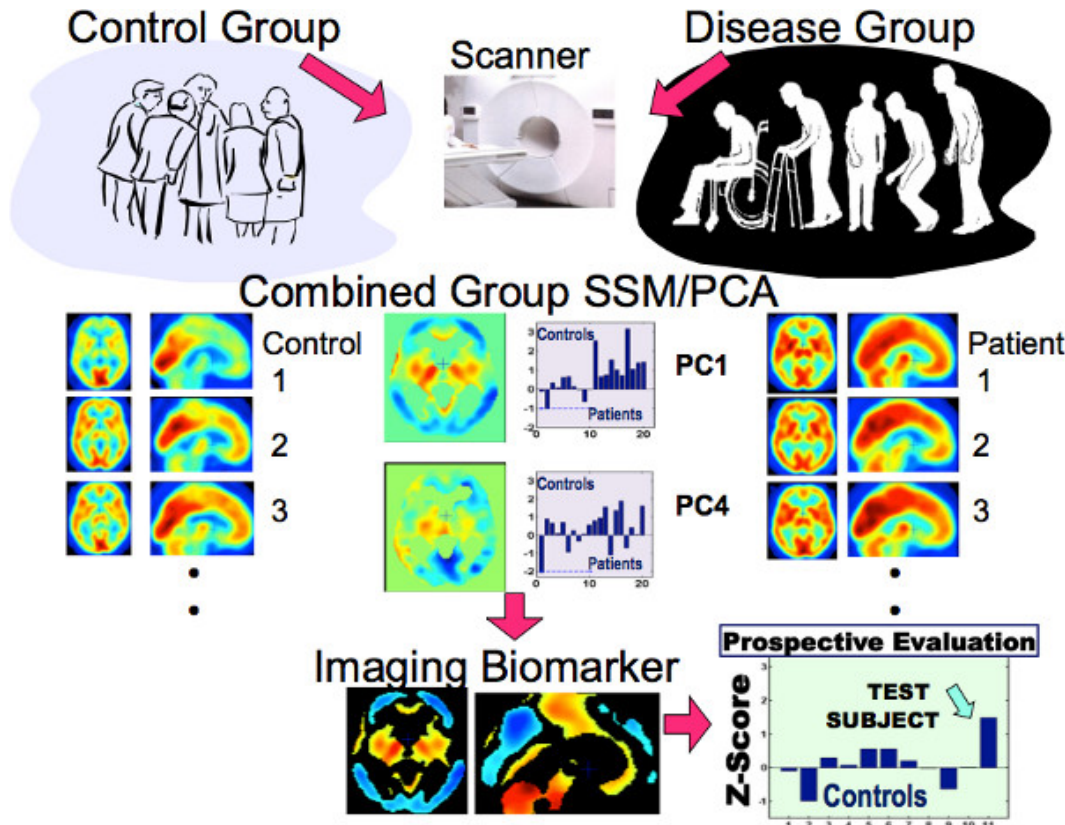


Figure 1. Schematic of SSM/PCA Modeling Strategy. Normal and patient scanner data is processed by the SSM/PCA algorithms to derive a neuroimaging biomarker and associated subject scores. The mean difference between patient and normal subject scores is significant. [Click here to view larger figure.](#)

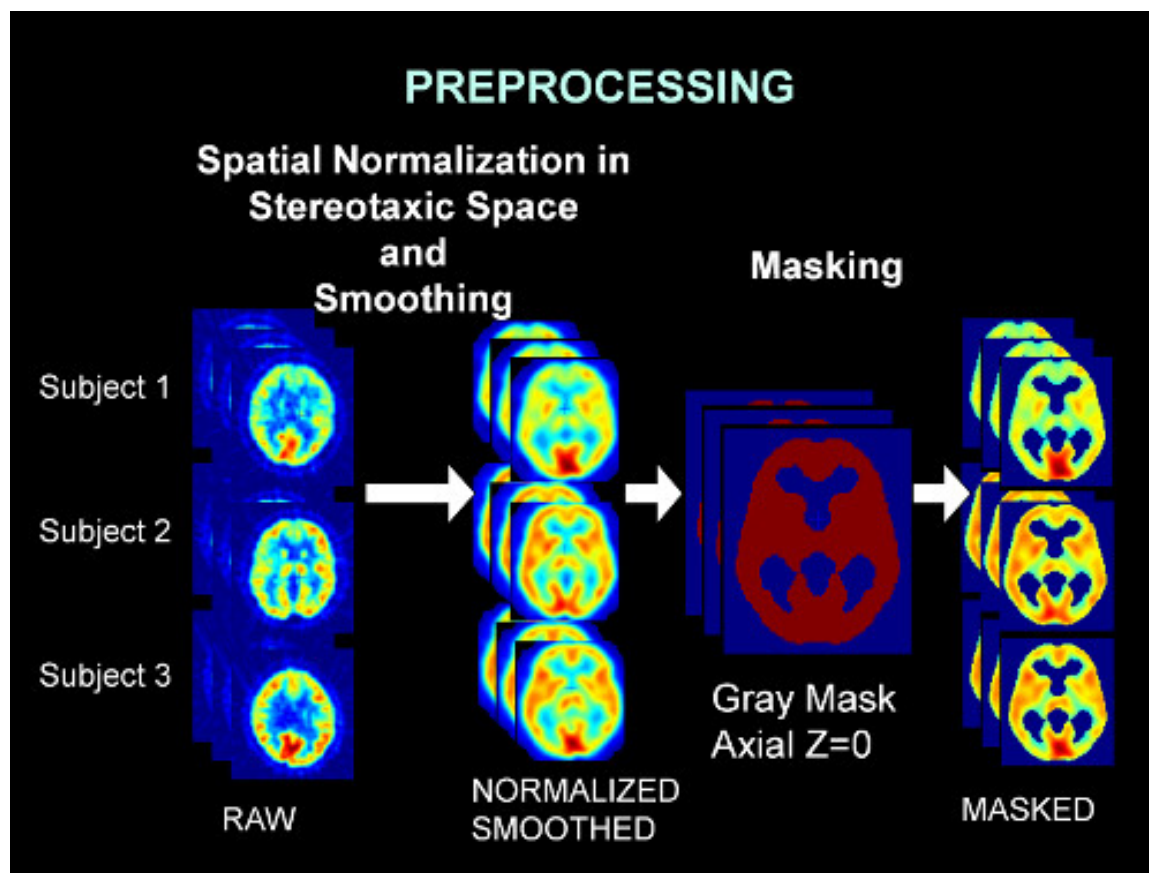


Figure 2. Preprocessing and Masking of Data. Raw scanner images are spatially normalized to map voxels to a common stereotaxic space for all subjects and smoothed using a Gaussian kernel. The masking operation removes ventricular space and noise.

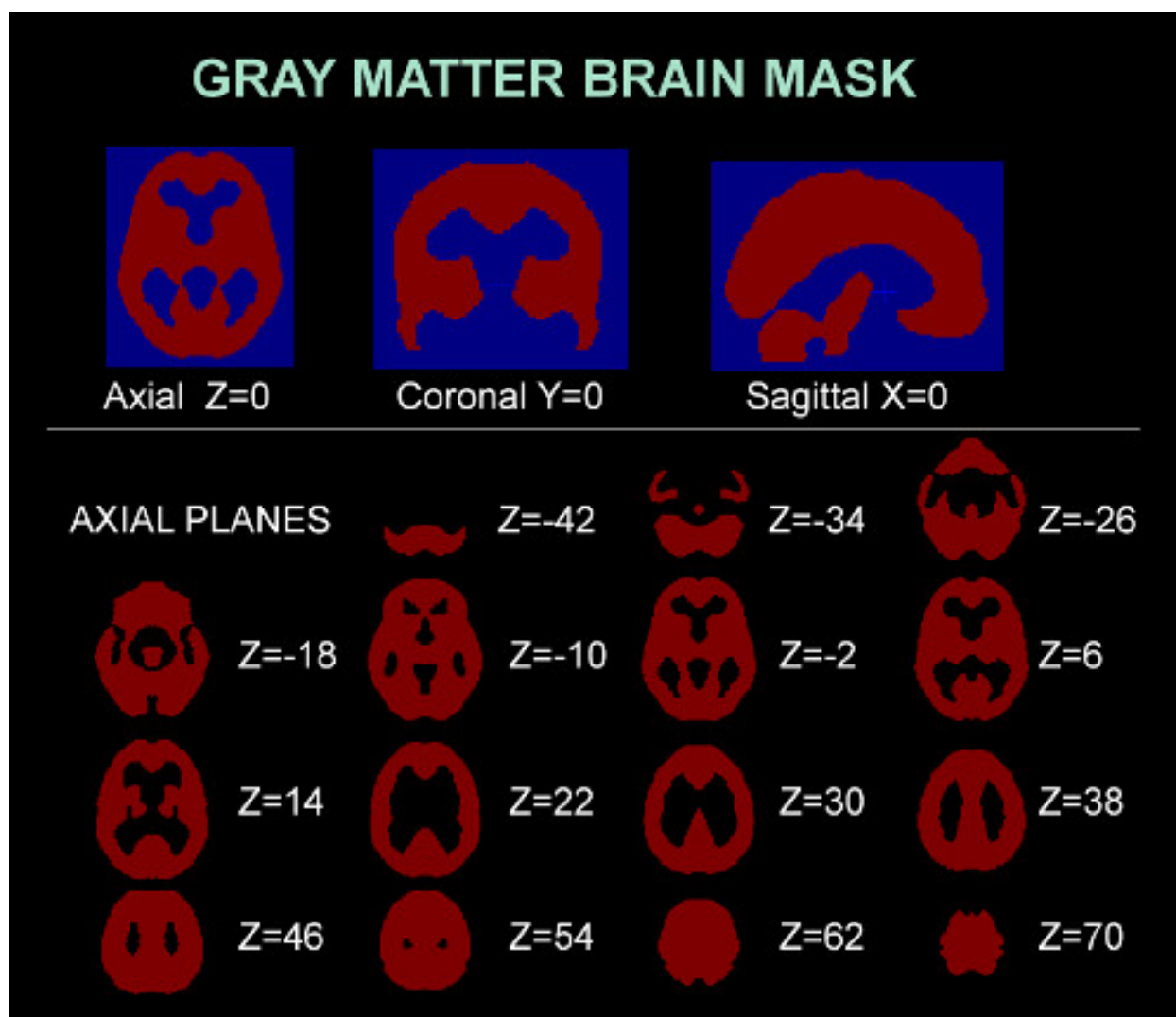


Figure 3. This default 0/1 gray matter mask was created from the processing of 72 normal subjects. Individual spatially normalized data was thresholded at 38% of their maximum value to create individual masks that were multiplied together. The multiplicative mask was flipped left to right through the origin and multiplied with the original to create the final symmetrized mask.

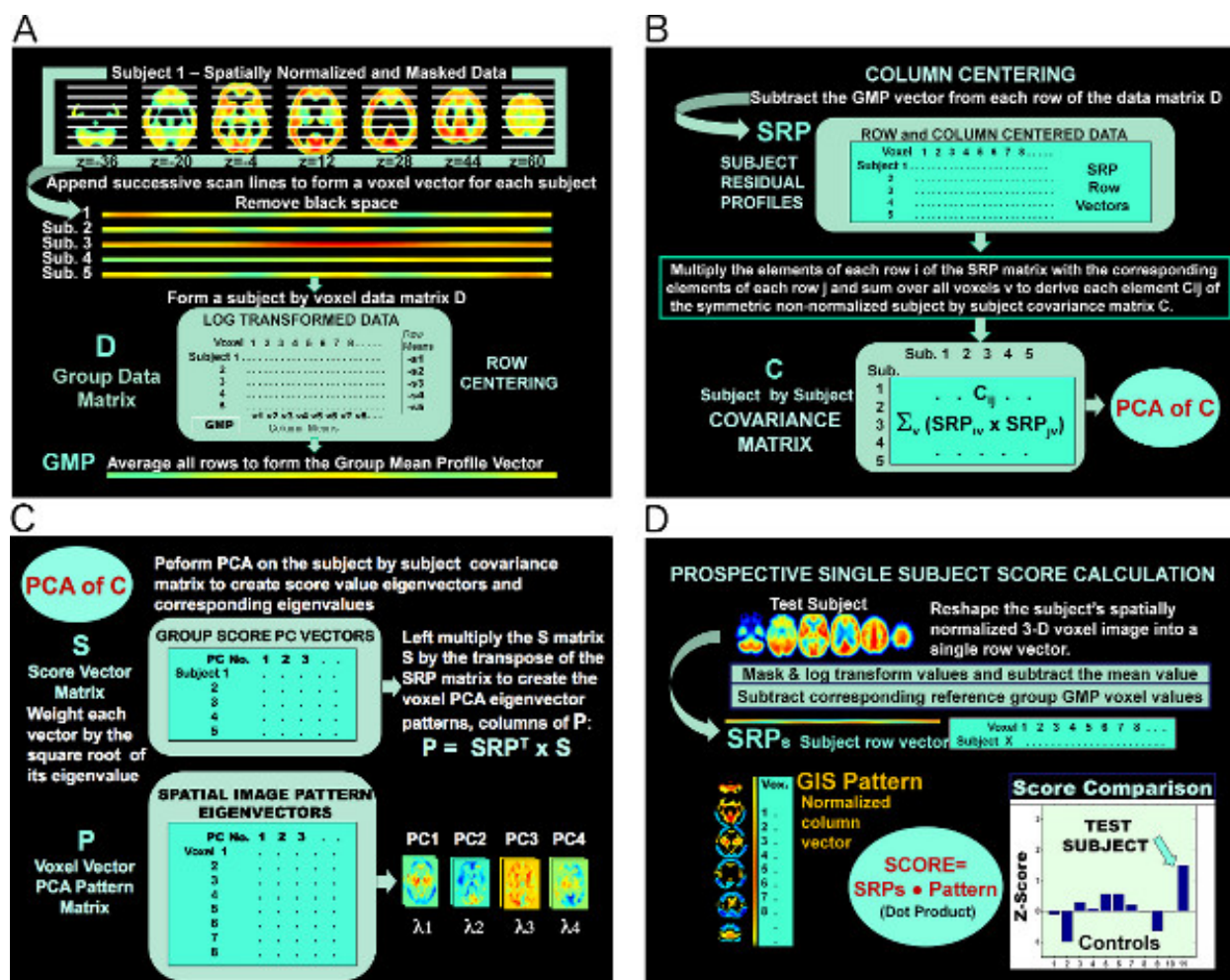


Figure 4. Flow Path of SSM/PCA Processing. a) Subject vector formulation of data matrix. Log transformation. Row centering. Evaluation of the GMP vector. b) Column centering. Derivation of SRP. Construction of the covariance matrix C. c) Principal Component Analysis. Derivation of PC score vectors and voxel patterns. d) Prospective single subject score calculation. [Click here to view larger figure.](#)

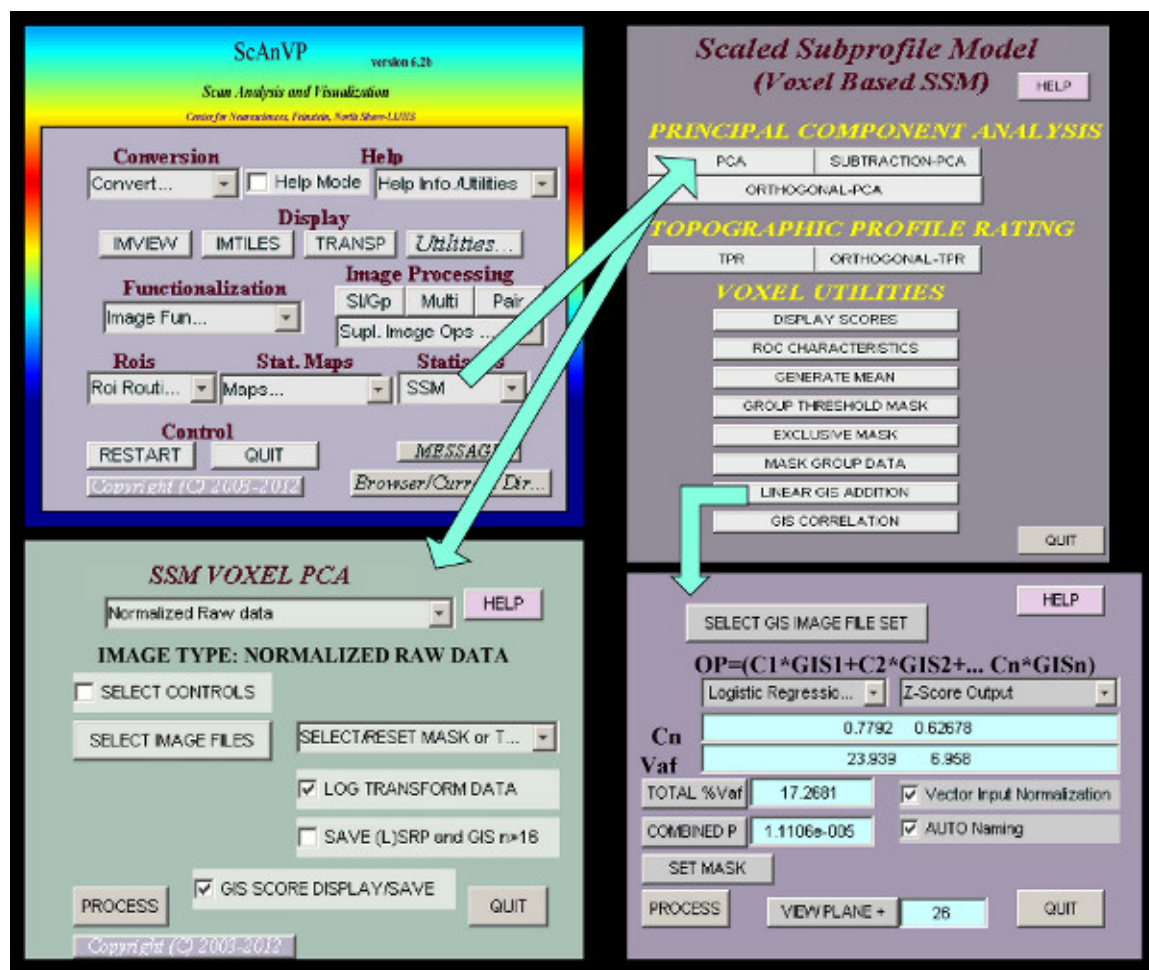


Figure 5. Software. To perform SSM/PCA the user can select the voxel based SSM/PCA option from the ScAnVP menu, select patient image files and normal controls, a preferred mask and then process. Log transformation is performed by default. After reviewing output and screen displayed plots and images, the user can choose to linearly combine selected PCs using logistic regression or other models to create a biomarker with higher patient/normal score discrimination. [Click here to view larger figure.](#)

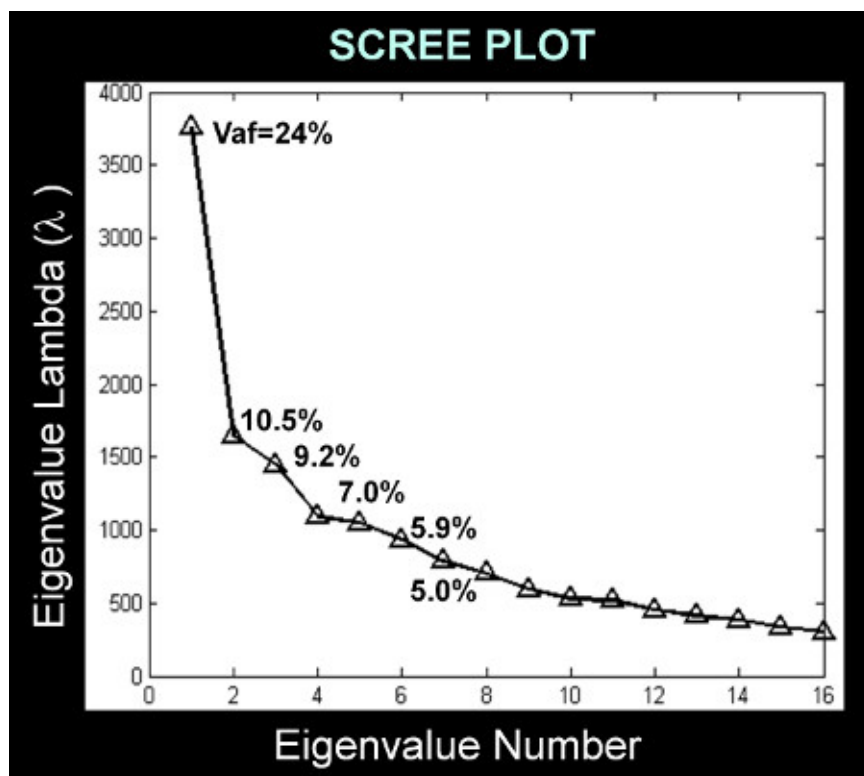


Figure 6. Scree Plot. This Scree plot of sequential eigenvalues of the non-normalized subject by subject covariance matrix resulted from the SSM/PCA analysis of 10 PD patients and 10 normal controls. Notice the sharp drop in the eigenvalues after the first PC. Differences between eigenvalues are small after the 4th PC.

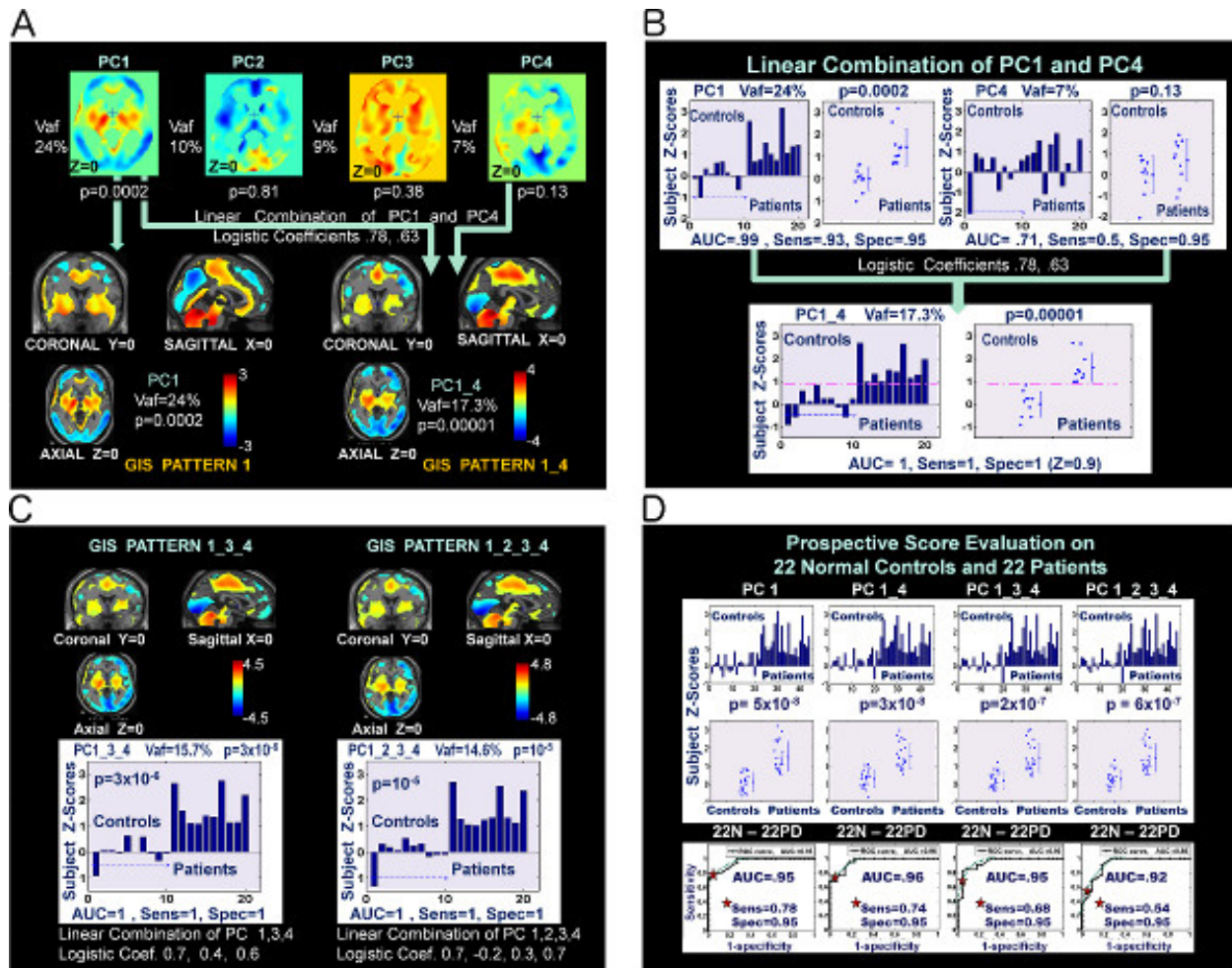


Figure 7. SSM/PCA program output. **a)** Top: Axial display ($Z=0$) of the first four principal component images (PC1 to 4). Bottom: Orthogonal views through origin ($X,Y,Z=0,0,0$) of PC1 and combined PCs 1 and 4 (PC1_4) displayed over a structural MRI image. PC voxel values represent positive and negative Z-score deviations relative to the mean voxel loading. Hot colors indicate relative increases in metabolic activity within the PC's contribution to the overall SRP whereas cold colors indicate relative metabolic decreases. For PC's that discriminate patients from controls these deviations can be attributed to disease. **b)** Bar graphs and scatterplots of the derivation subject Z-scores of PC1 and PC4. Only the first principal component significantly discriminates patients from controls ($p=0.0002$) while the fourth demonstrates a trend ($p=0.13$). The linear logistic combination of PC1 and PC4 (coefficients .78, .63) improves discrimination ($p=0.00001$). The ROC AUC value and the sensitivity for a specificity of 0.95 are also displayed for each pattern. The combined pattern PC1_4 demonstrates perfect separation at a Z-score threshold of 0.9 ($AUC=1$). **c)** Orthogonal views of combined PC1_3_4 and PC1_2_3_4 and corresponding subject bar graphs for derivation subject Z-scores. Discrimination increased ($p=3 \times 10^{-6}$ and 10^{-5} , respectively) compared to that of PC1 for both of these different combinations. AUC characteristics indicated perfect separation in both cases ($AUC=1$, sensitivity=1, specificity=1). See note in legend 7a regarding the color scale. **d)** Prospective score evaluation of patterns PC1, PC1_4, PC1_3_4 and PC1_2_3_4 on 22 normal controls and 22 PD patients. Bar and scatter-plot graphs as well as the ROC plot for subject Z-scores are displayed and AUC values as well as the sensitivity at a specificity of 0.95 is indicated. The AUC values and sensitivity appear to decrease for more than two combined principal components while p-values tend to become less robust. An insignificant improvement was noted for PC1_4 in AUC and p values invalidating the significant difference predicted in the derivation sample. [Click here to view larger figure.](#)

Discussion

The SSM/PCA model originally presented by Moeller *et al.*⁴ has evolved¹⁻³ into a straightforward and robust technique for the analysis of neuroimaging data. However, there have been ambiguities in the application of this methodology that we have attempted to clarify here and in previous publications^{5-7,10}. Some of these issues have been addressed in the text but are reemphasized here because of their importance. As detailed in the Introduction, SSM/PCA is primarily effective in resting state FDG PET data, but has also been successful using other single volume imaging techniques and scanning platforms including H_2^{150} PET and 99mTc-ECD SPECT, as well as perfusion ASL and volumetric VBM MRI. Its successful application to time-series rsfMRI necessitates the future development of greatly enhanced techniques⁴⁸.

The main strength of SSM lies in its ability to identify sources of variation in the data and to separate them into spatially orthogonal components. In most studies, only a few significant components are attributable to disease effects, thus dramatically reducing the relevant data set. To identify disease-specific components, we ideally include equal numbers of disease subjects with gender and age-matched healthy control subjects in a combined group analysis^{2,5}. In this situation, deviations from normal values due to disease are generally small in absolute terms and obscured by larger global factors that do not discriminate patients from controls^{2,5,6}. PCA is an effective analytical tool when the distribution of data is near-

Gaussian⁴⁹. The SSM procedure performs log transformation which basically converts the data into a log normal distribution⁵⁰ and separates multiplicative factors into additive terms. Subsequent subject mean centering (row mean subtraction) is a necessary step that removes any subject global scaling factors from the log data. This implies that preliminary subject global mean normalization has no effect in SSM even though it is generally a necessary preprocessing step in analytical neuroimaging procedures. Similarly, pre-normalization by any regional mean has no effect.

The mean centered log images are averaged over subjects to generate the GMP image that can be considered an invariant characteristic image of similar cohorts. The subsequent column centering over subjects removes regional means (GMP) to create subject residual SRP images. The SRP vectors are entered into the PCA analysis to generate a set of orthogonal PC patterns and reference group scores for each PC. In prospective single subject score evaluation, group analysis is not necessary; the subject's SRP is evaluated by the subtraction of the subject mean and reference group GMP and the subject's score is then determined as the scalar inner product of the SRP vector and the prederived PC pattern vector. This score usually falls within the range of the reference scores provided that the new data is obtained in a similar manner as in the original derivation. In a different setting, the PC vector may still be a valid pattern even for a different modality imager but to obtain an interpretable score, a recalibration procedure must be performed to evaluate new reference scores. The recalibration procedure does not require PCA analysis but is a simple determination of single subject scores for a complete new cohort of patients and controls using the predetermined pattern and a newly determined GMP.

It should be noted that even though a specific modality may reflect certain physiological parameters, derived SSM/PCA patterns cannot be interpreted as absolute measures of such quantities. Firstly, they are derived from the interregional covariance of voxel differences from mean values. These differences may result from various network functions and may represent composite effects. Hence, the PCA pattern is a mathematical construct that may reflect covariance within a single network or more complex processes. By performing correlation analysis of pattern scores with clinical measures of disease severity or normal psychosomatic measures, and by anatomically identifying highly weighted pattern areas, we can indirectly interpret the significance of these patterns. Furthermore, composite PC patterns are constructed as linear combinations of PCs with coefficients derived using logistic regression, with Akaike or other information theoretic criteria imposed on the corresponding score values. Although these may result in better discrimination, they may incorporate less significant components with disproportionately higher coefficients. Thus, covariance topographies designed for accurate differential diagnosis of clinically similar "look-alike" conditions ("diagnostic biomarkers") may be composites of complex physio-chemical processes, not necessarily corresponding to easily interpretable physical networks¹³. By contrast, other patterns may be optimal for capturing the topographical hallmarks of specific disease manifestations such as parkinsonian tremor or cognitive dysfunction^{15,51,52}, or for monitoring disease progression^{7,16}. The specific objective of the biomarker's function should be considered in making a final selection.

It is essential to validate a potential biomarker by evaluating prospective subject scores in independent cohorts¹⁰. Preferably, cross-validation should be initially performed in the derivation sample to determine the reliability of the pattern topography on a voxel-by-voxel basis and to exclude potential outlier effects. To this end, bootstrapping procedures^{8,9} are routinely implemented. This approach requires iterative pattern derivation in resampled derivation data to identify the most robust voxels, i.e. those with the lowest standard deviation in voxel weight. By optionally masking the original data to remove less reliable voxels, higher discrimination can be achieved for a rederived pattern that must then be validated in prospective data sets.

As previously noted, once an SSM/PCA biomarker has been established it can be applied to prospective individual or group data. Derived scores can be used to monitor a patient's status by gauging network progression in longitudinal imaging data, or to evaluate treatment effects. In complex cases of undetermined diagnosis, sets of subject scores can be entered in logistic regression analysis (or in other discriminant function models) to distinguish between different alternatives^{12,13}. This approach can be vitally important given the high rate of clinical misdiagnosis in individuals with early disease, and the varying prognosis and treatment outcome associated with the different underlying pathologies. As a final point, SSM pattern-based group score evaluation can be a valuable tool in clinical trials of experimental therapies for brain disorders^{15,53,54}.

Disclosures

Dr. Eidelberg is listed as co-inventor of two U.S. patents on the use of imaging markers to screen patients for nervous system dysfunction (without financial gain). There are no further disclosures.

Acknowledgements

This work was supported by Grant No. P50NS071675 (Morris K. Udall Center of Excellence in Parkinson's Disease Research at The Feinstein Institute for Medical Research) to D.E. from the National Institute of Neurological Disorders and Stroke. The content is solely the responsibility of the authors and does not necessarily represent the official views of the National Institute of Neurological Disorders and Stroke or the National Institutes of Health. The sponsor did not play a role in study design, collection, analysis and interpretation of data, writing of the report or in the decision to submit the paper for publication.

References

1. Moeller, J.R. & Strother, S.C. A regional covariance approach to the analysis of functional patterns in positron emission tomographic data. *J. Cereb. Blood Flow Metab.* **11** (2), A121-135 (1991).
2. Alexander, G.E. & Moeller, J.R. Application of the scaled subprofile model to functional imaging in neuropsychiatric disorders: A principal component approach to modeling brain function in disease. *Hum. Brain Mapp.* **2**, 1-16 (1994).
3. Habeck, C.G. Basics of multivariate analysis in neuroimaging data. *J. Vis. Exp.* (41), e1988, doi:10.3791/1988 (2010).
4. Moeller, J.R., Strother, S.C., Sidtis, J.J., & Rottenberg, D.A. Scaled subprofile model: a statistical approach to the analysis of functional patterns in positron emission tomographic data. *J. Cereb. Blood Flow Metab.* **7** (5), 649-658 (1987).

5. Spetsieris, P.G. & Eidelberg, D. Scaled subprofile modeling of resting state imaging data in Parkinson's disease: methodological issues. *Neuroimage*. **54** (4), 2899-2914 (2011).
6. Dhawan, V., Tang, C.C., Ma, Y., Spetsieris, P., & Eidelberg, D. Abnormal network topographies and changes in global activity: Absence of a causal relationship. *Neuroimage*. **63** (4), 1827-1832 (2012).
7. Eidelberg, D. Metabolic brain networks in neurodegenerative disorders: a functional imaging approach. *Trends Neurosci*. **32** (10), 548-557 (2009).
8. Habeck, C., Foster, N.L., *et al.* Multivariate and univariate neuroimaging biomarkers of Alzheimer's disease. *Neuroimage*. **40** (4), 1503-1515 (2008).
9. Efron, B. & Tibshirani, R. *An introduction to the bootstrap*. CRC Press, LLC, New York, (1994).
10. Ma, Y., Tang, C., Spetsieris, P., Dhawan, V., & Eidelberg, D. Abnormal metabolic network activity in Parkinson's disease: test-retest reproducibility. *J. Cereb. Blood Flow & Metab*. **27** (3), 597-605 (2007).
11. Ma, Y. & Eidelberg, D. Functional imaging of cerebral blood flow and glucose metabolism in Parkinson's disease and Huntington's disease. *Mol. Imaging Biol*. **9** (4), 223-233 (2007).
12. Tang, C.C., Poston, K.L., *et al.* Differential diagnosis of parkinsonism: a metabolic imaging study using pattern analysis. *Lancet Neurol*. **9** (2), 149-158 (2010).
13. Spetsieris, P.G., Ma, Y., Dhawan, V., & Eidelberg, D. Differential diagnosis of parkinsonian syndromes using PCA-based functional imaging features. *Neuroimage*. **45** (4), 1241-1252 (2009).
14. Tang, C.C., Poston, K.L., Dhawan, V., & Eidelberg, D. Abnormalities in metabolic network activity precede the onset of motor symptoms in Parkinson's disease. *J. Neurosci*. **30** (3), 1049-1056 (2010).
15. Mure, H., Hirano, S., *et al.* Parkinson's disease tremor-related metabolic network: characterization, progression, and treatment effects. *Neuroimage*. **54** (2), 1244-1253 (2011).
16. Niethammer, M. & Eidelberg, D. Metabolic brain networks in translational neurology: concepts and applications. *Ann. Neurol.*, (2012).
17. Petersson, K.M., Nichols, T.E., Poline, J.B., & Holmes, A.P. Statistical limitations in functional neuroimaging. I. Non-inferential methods and statistical models. *Philos. Trans. R. Soc. Lond. B. Biol. Sci*. **354** (1387), 1239-1260 (1999).
18. Habeck, C. & Stern, Y. Multivariate data analysis for neuroimaging data: overview and application to Alzheimer's disease. *Cell Biochem. Biophys*. **58** (2), 53-67 (2010).
19. McKeown, M.J., Hansen, L.K., & Sejnowski, T.J. Independent component analysis of functional MRI: what is signal and what is noise? *Current Opinion in Neurobiology*. **13** (5), 620-629 (2003).
20. Stone, J.V. Independent component analysis: an introduction. *Trends Cogn. Sci*. **6** (2), 59-64 (2002).
21. McIntosh, A.R., Bookstein, F.L., Haxby, J.V., & Grady, C.L. Spatial pattern analysis of functional brain images using partial least squares. *Neuroimage*. **3** (3 Pt. 1), 143-157 (1996).
22. Habeck, C., Krakauer, J.W., *et al.* A new approach to spatial covariance modeling of functional brain imaging data: ordinal trend analysis. *Neural Comput*. **17** (7), 1602-1645 (2005).
23. Habeck, C. & Moeller, J.R. Intrinsic functional-connectivity networks for diagnosis: just beautiful pictures? *Brain Connect*. **1** (2), 99-103 (2011).
24. Huang, C., Tang, C., *et al.* Changes in network activity with the progression of Parkinson's disease. *Brain*. **130** (Pt. 7), 1834-1846 (2007).
25. Feigin, A., Tang, C., *et al.* Thalamic metabolism and symptom onset in preclinical Huntington's disease. *Brain*. **130** (Pt. 11), 2858-2867 (2007).
26. Poston, K.L., Tang, C.C., *et al.* Network correlates of disease severity in multiple system atrophy. *Neurology*. **78** (16), 1237-1244 (2012).
27. Biswal, B., Yetkin, F.Z., Haughton, V.M., & Hyde, J.S. Functional connectivity in the motor cortex of resting human brain using echo-planar MRI. *Magnetic Resonance in Medicine: Official Journal of the Society of Magnetic Resonance in Medicine / Society of Magnetic Resonance in Medicine*. **34** (4), 537-541 (1995).
28. Greicius, M.D., Krasnow, B., Reiss, A.L., & Menon, V. Functional connectivity in the resting brain: a network analysis of the default mode hypothesis. *Proceedings of the National Academy of Sciences of the United States of America*. **100** (1), 253-258 (2003).
29. Friston, K.J., Frith, C.D., Liddle, P.F., Frackowiak, R.S. Functional connectivity: the principal component analysis of large (PET) data sets. *J. Cereb. Blood Flow Metab*. **13**, 5-14 (1993).
30. Strother, S.C., Kanno, I. & Rottenberg, D.A. Commentary and opinion: I. Principal component analysis, variance partitioning, and "functional connectivity". *Journal of Cerebral Blood Flow and Metabolism: Official Journal of the International Society of Cerebral Blood Flow and Metabolism*. **15** (3), 353-360 (1995).
31. Ekstrom, A. How and when the fMRI BOLD signal relates to underlying neural activity: the danger in dissociation. *Brain Research Reviews*. **62** (2), 233-244 (2010).
32. Logothetis, N.K. What we can do and what we cannot do with fMRI. *Nature*. **453** (7197), 869-878 (2008).
33. Lin, T.P., Carbon, M., *et al.* Metabolic correlates of subthalamic nucleus activity in Parkinson's disease. *Brain*. **131** (Pt. 5), 1373-1380 (2008).
34. Hirano, S., Asanuma, K., *et al.* Dissociation of metabolic and neurovascular responses to levodopa in the treatment of Parkinson's disease. *J. Neurosci*. **28** (16), 4201-4209 (2008).
35. Feigin, A., Antonini, A., *et al.* Tc-99m ethylene cysteinate dimer SPECT in the differential diagnosis of parkinsonism. *Mov. Disord*. **17** (6), 1265-1270 (2002).
36. Eckert, T., Van Laere, K., *et al.* Quantification of Parkinson's disease-related network expression with ECD SPECT. *Eur. J. Nucl. Med. Mol. Imaging*. **34** (4), 496-501 (2007).
37. Ma, Y., Huang, C., *et al.* Parkinson's disease spatial covariance pattern: noninvasive quantification with perfusion MRI. *J. Cereb. Blood Flow Metab*. **30** (3), 505-509 (2010).
38. Melzer, T.R., Watts, R., *et al.* Arterial spin labelling reveals an abnormal cerebral perfusion pattern in Parkinson's disease. *Brain*. **134** (Pt. 3), 845-855 (2011).
39. Skidmore, F., Spetsieris, P., *et al.* Diagnosis of Parkinson's disease using resting state fMRI In: *15th International congress of Parkinson's disease and movement disorders.*, LB22, (2011).
40. Peng, S., Wu, T., *et al.* A comparison study of Parkinson's disease-related patterns between FDG PET and fMRI at rest state. *Neuroimage*. **61** (Suppl. 1), 5610 (2012).
41. Brickman, A.M., Habeck, C., Zarahn, E., Flynn, J., & Stern, Y. Structural MRI covariance patterns associated with normal aging and neuropsychological functioning. *Neurobiol. Aging*. **28** (2), 284-295 (2007).

42. Bergfield, K.L., Hanson, K.D., *et al.* Age-related networks of regional covariance in MRI gray matter: reproducible multivariate patterns in healthy aging. *Neuroimage*. **49** (2), 1750-1759 (2010).
43. Steffener, J., Brickman, A.M., Habeck, C.G., Salthouse, T.A., & Stern, Y. Cerebral blood flow and gray matter volume covariance patterns of cognition in aging. *Human Brain Mapping*, (2012).
44. Cangelosi, R. & Goriely, A. Component retention in principal component analysis with application to cDNA microarray data. *Biol. Direct*. **2**, 2 (2007).
45. Akaike, H. A new look at the statistical model identification. *IEEE Transactions on Automatic Control*. **19** (6), 716-723 (1974).
46. Eidelberg, D., Moeller, J.R., *et al.* Assessment of disease severity in parkinsonism with fluorine-18-fluorodeoxyglucose and PET. *J. Nucl. Med.* **36** (3), 378-383 (1995).
47. Ma, Y., Tang, C., Moeller, J.R., & Eidelberg, D. Abnormal regional brain function in Parkinson's disease: truth or fiction? *Neuroimage*. **45** (2), 260-266 (2009).
48. Habeck, C., Steffener, J., Rakitin, B., & Stern, Y. Can the default-mode network be described with one spatial-covariance network? *Brain Res.* **1468** 38-51 (2012).
49. Jolliffe, I.T. *Principal Components Analysis, 2nd edition, Springer Series in Statistics.*, Springer-Verlag, New York, (2002).
50. Limpert, E., Stahel, W.A., & Abbt, M. Log-normal distributions across the sciences: keys and clues. *BioScience*. **51** (5), 341-352 (2001).
51. Huang, C., Mattis, P., *et al.* Metabolic abnormalities associated with mild cognitive impairment in Parkinson disease. *Neurology*. **70** (16 Pt. 2), 1470-1477 (2008).
52. Mattis, P.J., Tang, C.C., Ma, Y., Dhawan, V., & Eidelberg, D. Network correlates of the cognitive response to levodopa in Parkinson disease. *Neurology*. **77** (9), 858-865 (2011).
53. Feigin, A., Kaplitt, M.G., *et al.* Modulation of metabolic brain networks after subthalamic gene therapy for Parkinson's disease. *Proc. Natl. Acad. Sci. U.S.A.* **104** (49), 19559-19564 (2007).
54. Mure, H., Tang, C.C., *et al.* Improved sequence learning with subthalamic nucleus deep brain stimulation: evidence for treatment-specific network modulation. *The Journal of Neuroscience: The Official Journal of the Society for Neuroscience*. **32** (8), 2804-2813 (2012).