

A theory of fine structure image models with an application to detection and classification of dementia

William O'Neill¹, Richard Penn^{1,2}, Michael Werner³, Justin Thomas¹

¹Bioengineering Department, University of Illinois at Chicago, Chicago, Illinois 60607, USA; ²Department of Neurosurgery, Rush University and Hospital, Chicago, Illinois 60607, USA; ³Department of Electrical and Computer Engineering, University of Illinois at Chicago, Chicago, Illinois 60607, USA

Correspondence to: William O'Neill, Bioengineering Department, University of Illinois at Chicago, 851 S. Morgan St., Chicago, Illinois 60607, USA. Email: wonell@uic.edu.

Background: Estimation of stochastic process models from data is a common application of time series analysis methods. Such system identification processes are often cast as hypothesis testing exercises whose intent is to estimate model parameters and test them for statistical significance. Ordinary least squares (OLS) regression and the Levenberg-Marquardt algorithm (LMA) have proven invaluable computational tools for models being described by non-homogeneous, linear, stationary, ordinary differential equations.

Methods: In this paper we extend stochastic model identification to linear, stationary, partial differential equations in two independent variables (2D) and show that OLS and LMA apply equally well to these systems. The method employs an original nonparametric statistic as a test for the significance of estimated parameters.

Results: We show gray scale and color images are special cases of 2D systems satisfying a particular autoregressive partial difference equation which estimates an analogous partial differential equation. Several applications to medical image modeling and classification illustrate the method by correctly classifying demented and normal OLS models of axial magnetic resonance brain scans according to subject Mini Mental State Exam (MMSE) scores. Comparison with 13 image classifiers from the literature indicates our classifier is at least 14 times faster than any of them and has a classification accuracy better than all but one.

Conclusions: Our modeling method applies to any linear, stationary, partial differential equation and the method is readily extended to 3D whole-organ systems. Further, in addition to being a robust image classifier, estimated image models offer insights into which parameters carry the most diagnostic image information and thereby suggest finer divisions could be made within a class. Image models can be estimated in milliseconds which translate to whole-organ models in seconds; such runtimes could make real-time medicine and surgery modeling possible.

Keywords: Image regression models; dementia classification; MRI classification information; parameter classification information

Submitted Nov 09, 2014. Accepted for publication Feb 13, 2015.

doi: 10.3978/j.issn.2223-4292.2015.03.11

View this article at: <http://dx.doi.org/10.3978/j.issn.2223-4292.2015.03.11>

Introduction

When viewing a complete, as opposed to a region of interest, medical image such as an axial MRI brain scan, the diagnostician typically makes an overall image assessment and then studies from region to region seeking elements of normal structure and pathology. The eye cannot readily assess the thousands of small image regions for comparison. Thus, image fine structure information at the individual pixel level, of necessity, has been visually ignored. Whether such fine structure of closely spaced pixels, of any size image, will be useful for medical purposes is an open question since it has been difficult to obtain such data. Further, in the problem of diagnosing neurological diseases, our bioengineering application addresses a long recognized clinical need of dementia diagnosis: MR image quantification has been, and still is, an arduous task, currently requiring considerable processing time and operator attention (1). In Alzheimer's disease (AD), for example, it is difficult to distinguish normal aging changes of increased ventricular and sulcus size from AD degeneration.

The definitive diagnosis of dementia is by autopsy but of the ten recognized cognitive tests, the most popular in use is the Mini Mental State Examination (MMSE) which has the best correlation with autopsy outcomes and is the most useful in following diagnosed dementia over time (2). Therefore we use subject MMSE scores to categorize normal or demented subjects. To support this definition of subject cognitive state, we seek a brain scan image model whose parameters not only capture discriminating image information but also indicate which parameters of the image carry the most discriminating information as measured by their inferred confidence intervals (3).

Materials and methods

Our original computational method is based on ordinary least squares (OLS) regression because we discovered a simple matrix transformation that enables images to be represented as linear combinations of row and column spatial lags of the original image. These row and column lagged images give rise to the "fine structure" description of the images we model because the row and column spatially lagged images are visually indistinct from the original. The key step in our analysis is to transform any two dimensional (2D) image pixel array to a one dimensional (1D) pixel array. Next we show that transformed images can be modeled by traditional OLS methods. Therefore, under verified distributional conditions,

the OLS model parameters can be tested for statistical significance. This is an important model property as it means recognized neurological conditions can be tested for causality by parameter significance. In addition, the most significant parameters carry the most discrimination information in implementing image classifications.

Image model development

To reduce a 2D pixel array (an image) to a 1D pixel array (a column) a transformation called vectorization is applied by stacking the left-most column of image pixels onto its neighbor pixel column and these two columns onto the third and so on, until all pixel columns are stacked (4). For example, an image A of 208 rows and 176 columns of pixels becomes a single column of $[208] \times [176]=36,608$ pixels. We call this column the vectorized form of A and write it as $\text{vec}(A)$. The next step is to recognize that any image $y(i,j)$, whose intensity y at pixel location (i,j) , can be hypothesized to be a solution of the linear autoregressive (AR) model (5).

$$y(i, j) = \sum_{k=0}^p \sum_{l=0}^q \beta_{kl} y(i-k, j-l) + u(i, j) \quad [1]$$

$$k+l > 0$$

In Eq. [1], $u(i,j)$ is a random process error term and we note Eq. [1] is a discrete approximation to a general, linear, causal, partial differential equation in two space dimensions.

The final step is to show, see Appendix A_1 , that if c_i and A_i are real constants and matrices then:

$$\text{vec}\left(\sum_{i=1}^k c_i A_i\right) = \sum_{i=1}^k c_i \text{vec}(A_i) \quad [2]$$

Define $y_o = \text{vec}[y(i,j)]$, $x_1 = \text{vec}[y(i,j-1)]$, $x_2 = \text{vec}[y(i-1,j)]$, ..., $x_{p+q+qp} = \text{vec}[y(i-p,j-q)]$, and $u_o = \text{vec}[u(i,j)]$. By Eq. [2], the model in Eq. [1], with β as a column vector with $p+q+qp$ elements β_{kl} has the familiar OLS regression form.

$$y_o = (x_1 x_2 \dots x_{p+q+qp}) \beta + u_o = X \beta + u_o \quad [3]$$

In Eq. [3], $X \beta$ is called a minimum variance representation (MVR) of the vectorized image y_o if a β can be found such that $u_o^T u_o$ is minimized. A necessary and sufficient condition that such a β exists is:

MVR existence theorem: [Th1].

An image of pixel intensity $y(i,j)$ has a MVR in the form of

$$X \beta = \text{vec}\left[\sum_{k=0}^p \sum_{l=0}^q \beta_{kl} y(i-k, j-l)\right],$$

if and only if $k+l > 0$, the image column vectors $x_1 x_2 \dots x_{p+q+qp}$ are linearly independent.

When this condition is met the minimizing β satisfies:

$$\beta = (X^T X)^{-1} X^T y_o \tag{4}$$

and the image MVR is:

$$X\beta = X(X^T X)^{-1} X^T y_o \tag{5}$$

The finite sample form of Eq. [3] employed to estimate β is

$$y_o = Xb + e_o \tag{6}$$

where b is the OLS estimate of the population (true) value β and the vector e_o estimates the random residual vector u_o . Using the format of Eq. [6], each image of the Alzheimer's Disease Neuroimaging Initiative (ADNI) and Open Access Series of Imaging Studies (OASIS) data sets had an estimated MVR, denoted as Xb , minimizing $e_o^T e_o$ and computed as:

$$b = (X^T X)^{-1} X^T y_o \tag{7}$$

$$Xb = X(X^T X)^{-1} X^T y_o \tag{8}$$

The integer p, q orders are determined by increasing p and q so that $(X^T X)^{-1}$ exists for all image data. For both the ADNI and OASIS data this was for $p=q=2$. Larger p or q made $(X^T X)$ ill-conditioned. If the image to be modeled is m by n pixels then the minimum variance realization Xb of the image is a column vector of $(m-p) \times (n-q)$ by 1 pixels. To realize the 2D picture of Xb it must be unstacked into 2D image model starting at the top of Xb .

Only in the rarest of cases will the residual vector e_o in Eq. [6] test to be white Gaussian noise which would permit Student's t -tests of parameter statistical significance. So we have developed, see Appendix A₂, a nonparametric significance statistic, w_{kl} , based on the Chebychev inequality (6). $|w_{kl}|$ measures the width, normalized by b_{kl} , of a $1-\alpha$ confidence interval of the true β_{kl} . The smaller $|w_{kl}|$ is for a given b_{kl} , the more significant is the estimated parameter b_{kl} multiplying lagged image $\text{vec}[y(i-k, j-l)]$. For the diagnostician the confidence interval of β_{kl} is the most important property derived from its estimate b_{kl} since it tells him the probability with which the true parameter he seeks, β_{kl} , is estimated by what he estimates, b_{kl} .

The X matrix in Eq. [6] has eight column vectors, eight spatially lagged, vectorized images of $y(i-k, j-l)$ for $k=0,1,2, l=0,1,2, k+l>0$. However, we found that adding a column of ones to X to account for image sample means slightly increases the likelihood of correct classification as developed in the following section. Relative to the theoretical model in Eq. [1], this amounts to adding a constant to the right-hand side. The column of ones is crucial because it assures the

residual vector e_o in Eq. [6] and is estimated without bias, a bias which would also induce a bias in the estimated vector b .

Optimal image classification

In this section we derive a discriminator by which a given subject's MMSE score identifies the subject as normal or demented and also gives a ranking statistic to compare with other subjects in the same class. A range of MMSE scores is chosen and subjects with scores in that range are categorized as members of the demented class. Another unique MMSE range is chosen for those subjects categorized as normal subjects. Let N_d and N_n be the respective number of subjects in each class. For each subject Eq. [7] estimates a b vector of size $p+q+pq+l=9$. These are arranged in parameter matrices B_d , 9 by N_d and B_n , 9 by N_n and the pooled parameter matrix B_p which is 9 by $N_d + N_n$. Let $\hat{\Sigma}_d, \hat{\Sigma}_n, \hat{\Sigma}_p$ be estimates of the covariance matrices of B_d, B_n , and B_p . The projections onto the scalar z axis of $z_n = B_n v, z_d = B_d v$, and $z_p = B_p v$ are maximally separated in sample means if the projection vector v satisfies the Fisher linear discriminant eigenvector identity (7),

$$(n_1 \hat{\Sigma}_p - n_2 \hat{\Sigma}_n - n_3 \hat{\Sigma}_d) v = \gamma_1 (n_2 \hat{\Sigma}_n + n_3 \hat{\Sigma}_d) v \tag{9}$$

for the only nonzero eigenvalue γ_1 . In Eq. [9], $n_1 = N_d + N_n - 1, n_2 = N_n - 1, n_3 = N_d - 1$. For even modest sample sizes, N_d and $N_n \geq 10$, a Kolmogorov-Smirnov test (K-S test) confirms the hypothesis the z projections are normally distributed (8).

Since z_n and z_d are normally distributed with estimated densities $f_n(z)$ and $f_d(z)$, the Kullback-Leibler discrimination (KLD) statistic for a zero threshold, $L_o(z)$, is a convenient means of comparing projections for "strength of class membership" since $|L_o(z)|$ is monotonic with membership probability (9).

$$\begin{aligned} L_o(z) &= \ln [f_d(z) / f_n(z)] \\ &= \ln (\sigma_n / \sigma_d) + (z - \mu_n)^2 / 2\sigma_n^2 - (z - \mu_d)^2 / 2\sigma_d^2 \end{aligned} \tag{10}$$

In Eq. [10], means and variances are estimated from the sample z projections. $L_o(z) > 0$ classifies a z -projection as demented and $L_o(z) < 0$ as a normal.

The KLD statistics provide a graphical display of TP, TN, FP, and FN measures of the classification algorithm.

Large-sample ADNI dataset

Our large sample study employed ADNI archive axial MR image slices. The images were 1.5 T, T1 weighted using the MP Rage sequence, calibration scans for B₁ correction,

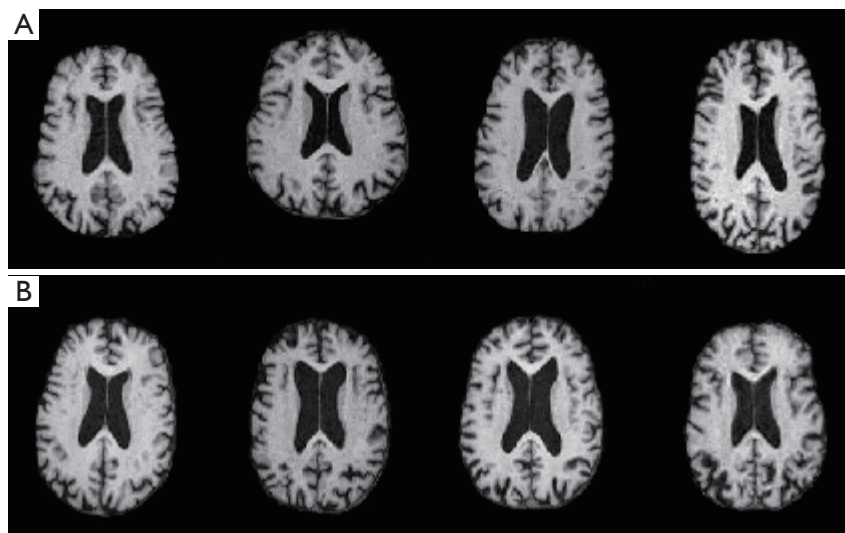


Figure 1 (A) Typical MMSE-designated normal MR axial images from a sample of size 50 from the ADNI national archive; (B) typical MMSE-designated demented MR axial images from a sample of 47 from the ADNI national archive. Subjects matched for age and gender. MMSE, Mini Mental State Examination; ADNI, Alzheimer's Disease Neuroimaging Initiative.

dual fast spin echo and were 256 by 170 gray scale pixels. All ADNI images were selected from one slice axially above the anterior commissure brain structure. ADNI subject inclusion criteria varied over four classes: (I) normal subjects had no subjective memory complaints; (II) normal memory function stratified by educational attainment; (III) MMSE scores between 24 and 30; and (IV) a Clinical Dementia Rating (CDR) of 0. EMCI subjects had subjective memory complaints, abnormal memory function stratified by age, MMSE 24 to 30, a CDR of 0.5, and AD could not be diagnosed at time of screening visit. LMCI subjects had the same criteria as the EMCI subjects but educational attainment was lower. AD subjects had memory complaints, abnormal memory function, MMSE scores 20 to 27, CDR of 0.5 to 1.0, and satisfied a NINCDS criteria for probable AD. The 50 subjects we took as normal were screened as males, ages 65 to 84 years with MMSE scores of 29 to 30. The 47 subjects we took as demented were of ages 65 to 84 years with MMSE scores of 17 to 27. Therefore the 97 subjects were identified as normal or demented according to the range of their MMSE scores. The 97 subjects so identified exhausted the four classes cited above for the constraints we specified. It is likely that about 70% of the categorized demented subjects suffered from AD (10). The chosen MMSE score ranges are typical for normal/demented studies with 20 year age spans (2). All ADNI axial images had cranium and dura artifacts manually cropped. *Figure 1* displays typical ADNI images.

Small-sample OASIS data set

Our small sample study employed OASIS axial MR image slices screened as right-handed males of ages 65 to 75 years with MMSE scores of 29 to 30. The database produced 14 such images we refer to as the normal class. Screening for right-handed males of ages 67 to 84 years with MMSE scores of 17 to 27 produced 12 images we refer to as the demented class. These images came from a set of 416 subjects aged 18 to 96. For each subject, 3 or 4 individual T1-weighted MRI scans obtained in single scan sessions were averaged to get the images, 208 by 176 pixels, for our study. A total of 100 of the included subjects over the age of 60 had been clinically diagnosed with very mild to moderate AD, in addition, a reliability data set of 20 non-demented subjects imaged on a subsequent visit within 90 days of their initial session were available as normal subjects. OASIS images had dura and cranium artifacts manually cropped. All axial images from the OASIS archive were at a fixed, unspecified number of 1.5 mm slices from the anterior commissure brain structure. *Figure 2* are typical OASIS images.

Results

Typical image OLS models, parameters, and statistics

All 123 images were modeled using the OLS Eqs. [7] and [8] and the $|w_{kl}|$ significance statistics (for a 95% confidence

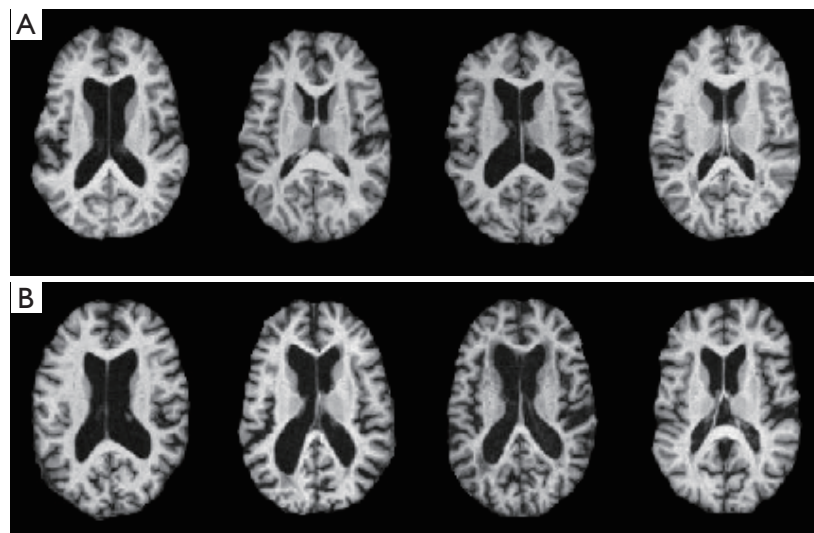


Figure 2 (A) Typical MMSE-designated normal MR axial images from a sample of size 14 from the OASIS archive; (B) typical MMSE-designated demented MR axial images from a sample of size 12 from the OASIS archive. Subjects are matched for age and gender. MMSE, Mini Mental State Examination; OASIS, Open Access Series of Imaging Studies.

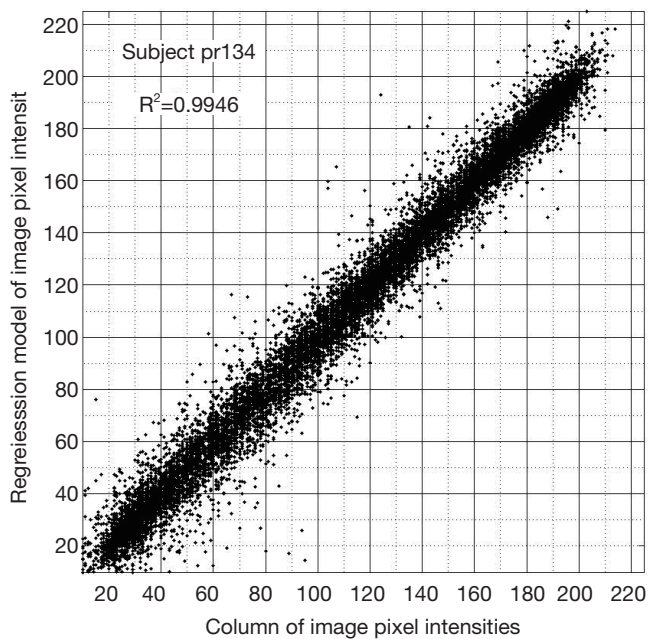


Figure 3 A scatter diagram of a typical regression determining a MVR of a demented subject from the OASIS archive. Horizontal axis is the scale for the pixel intensities of the vectorized image y_o . The vertical axis is the scale for the vectorized image model pixel intensities Xb estimated by the regression formula Eq. [8]. Regression models of all 123 subjects from both archives had R^2 exceeding 99%. MVR, minimum variance representation; OASIS, Open Access Series of Imaging Studies.

interval) computed for all b_{kl} parameters. *Figure 3* is a typical OLS scatter diagram, this for a demented subject from the OASIS dataset. The regression R^2 for this model is typical of image regressions for both datasets as they all exceeded 99%. *Figure 4* illustrates typical estimation results for two OASIS subjects and *Table 1* lists the OLS b_{kl} parameters and $|w_{kl}|$ statistics for the images in *Figure 4*. Typical normal and demented images and models of ADNI subjects are shown in *Figure 5*; b_{kl} parameters and $|w_{kl}|$ statistics for these two ADNI subjects are in *Table 4*.

Sample means of model parameters and their significance statistics

Table 3 tabulates sample means of estimated parameters and sample means of their $|w_{kl}|$ statistics for the OASIS dataset and *Table 2* tabulates these sample means for the ADNI dataset. With sample sizes of 50 and 47, the entries in *Table 2* are highly significant. For example, an interpretation of the mean ($|w_{kl}|$) value of b_{01} is: with a probability of 0.95 the average true demented parameter β_{01} is in the interval 1.0812 ± 0.0232 ; for normal subject b_{01} the 95% interval of β_{01} is 1.1174 ± 0.0228 .

Table 3 entries for the OASIS (26 samples) dataset do not differ dramatically from the ADNI *Table 2* entries but there is a subtle shift of the most significant b_{kl} parameter from b_{10} for the OASIS data to b_{01} for the ADNI data. This shift is

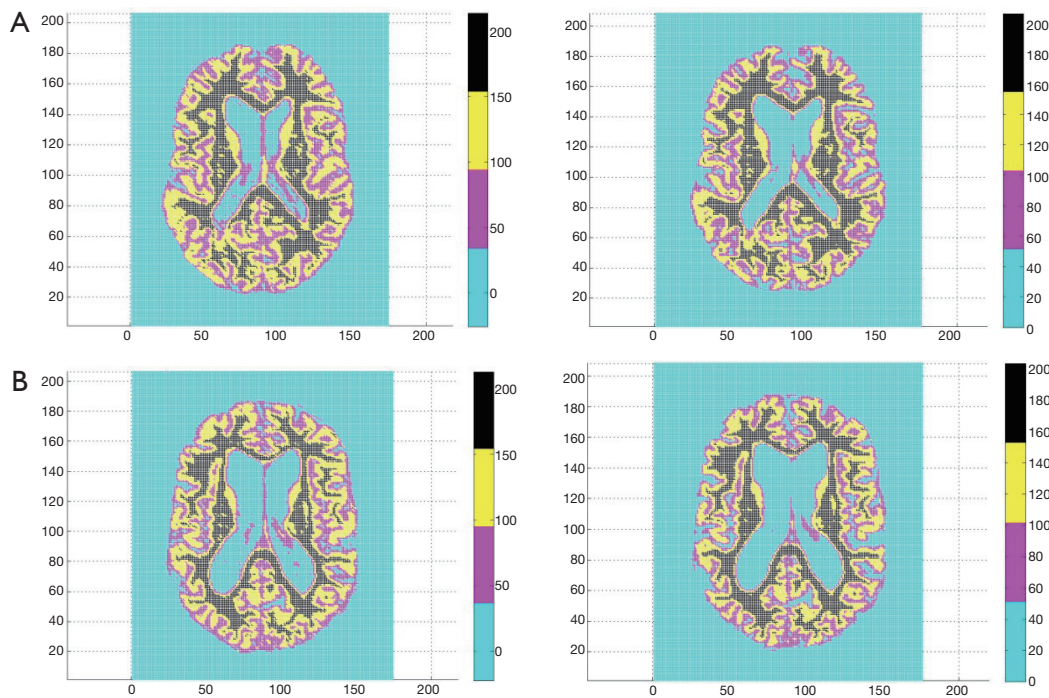


Figure 4 (A) A MMSE-designated normal image from the OASIS archive and a 9 MVR of that image; (B) a MMSE-designated demented image from the OASIS archive and a 9 MVR of that image. The b_{kl} parameters and their $|w_{kl}|$ parameter significance statistics for the models of *Figure 4* are listed in *Table 1*. With $|w_{01}| = 0.0412$ for the normal subject, b_{01} is the most significant parameter for the normal image model and $|w_{01}| = 0.0432$ for b_{01} of the demented subject makes this parameter the most significant for the demented subject. MMSE, Mini Mental State Examination; OASIS, Open Access Series of Imaging Studies; MVR, minimum variance representation.

Table 1 A listing of ordinary least squares estimates of the b_{kl} parameters for the normal and demented OASIS brain scans of *Figure 4*. The $|w_{kl}|$ statistics measure the normalized width of a 95% confidence interval of the true parameter β_{kl} . The smaller $|w_{kl}|$ the more significant is the parameter in determining the image model. For both of these images b_{01} , relating $y(i,j)$ and $y(i,j-1)$, is the most significant parameter

Xb	b_0	$+b_{01}x_1$	$+b_{02}x_2$	$+b_{10}x_3$	$+b_{11}x_4$	$+b_{12}x_5$	$+b_{20}x_6$	$+b_{21}x_7$	$+b_{22}x_8$
Normal b_{kl}	0.4375	1.1139	-0.3471	0.9891	-0.9051	0.2361	-0.2781	0.2425	-0.0590
Normal $ w_{kl} $	0.8304	0.0412	0.1284	0.0479	0.0845	0.2778	0.1640	0.2841	0.7996
Demented b_{kl}	0.4937	1.0616	-0.3233	1.0130	-0.8535	0.2121	-0.2752	0.1910	-0.0380
Demented $ w_{kl} $	0.7274	0.0432	0.1373	0.0461	0.0896	0.3125	0.1641	0.3531	1.2440

OASIS, Open Access Series of Imaging Studies.

interpreted later.

Optimal image classification

Figure 6 shows the inferred probability density functions of the ADNI z-projections resulting from projecting the b_{kl} parameters with the eigenvector solution of Eq. [9]. These probability densities passed K-S tests for normality at level 10^{-5} which decisively supports the classification probability

inference of the KLD statistics, computed from Eq. [10] and also shown in *Figure 6*. *Figure 7* shows the inferred z-projection probability density functions for the OASIS subjects. These densities passed K-S tests for normality at level 10^{-3} . An interpretation of the KLD statistics for the demented (normal) projections is that the larger (smaller) the KLD statistic is, then the more likely the sample is correctly classified. For example, in *Figure 7* the leftmost z-projection (near 114.4) corresponds to demented

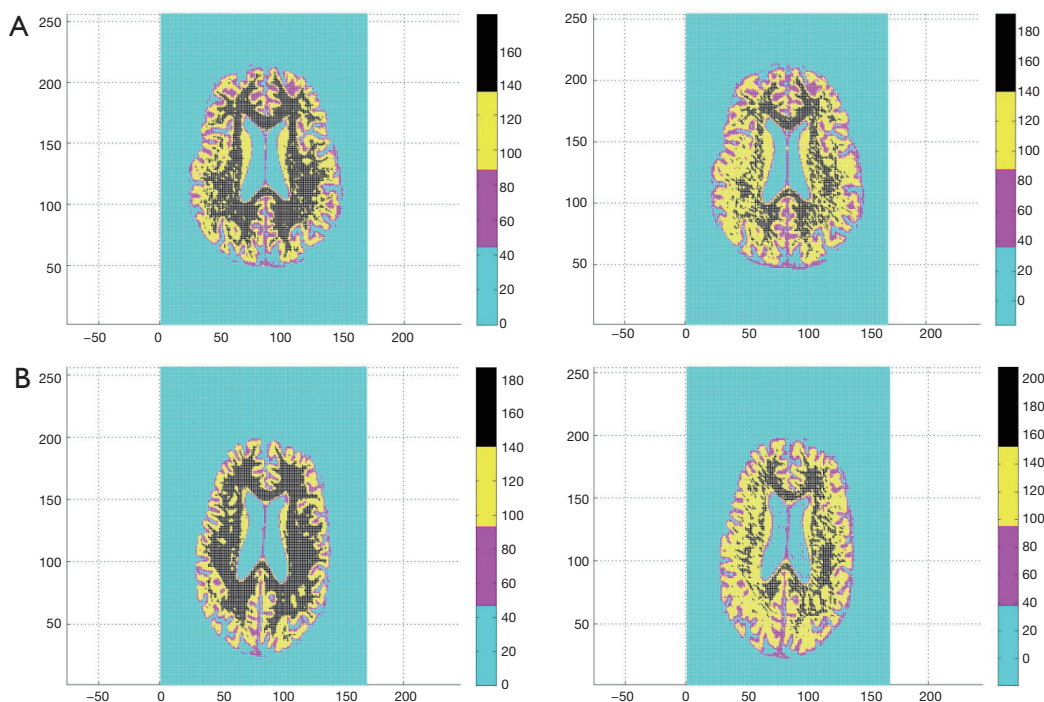


Figure 5 (A) Typical MMSE-designated normal MR axial images from a sample of size 50 from the ADNI archive; (B) typical MMSE-designated demented MR axial images from a sample of size 47 from the ADNI archive. Subjects are matched for age and gender. For these images b_{01} is the most significant parameter (see Table 2), while for the OASIS images (see Figure 4), b_{10} is the most significant (see Table 3). MMSE, Mini Mental State Examination; ADNI, Alzheimer’s Disease Neuroimaging Initiative; OASIS, Open Access Series of Imaging Studies.

Table 2 A listing of b_{kl} and $|w_{kl}|$ sample means for 47 demented and 50 normal ADNI subjects. Unlike the OASIS models, for each class b_{01} is on average the most significant parameter (minimum $|w_{kl}|$ sample means) and numerically largest parameter

Xb	b_0	$+b_{01}x_1$	$+b_{02}x_2$	$+b_{10}x_3$	$+b_{11}x_4$	$+b_{12}x_5$	$+b_{20}x_6$	$+b_{21}x_7$	$+b_{22}x_8$
Normal									
Mean(b_{kl})	0.3803	1.1174	-0.3568	1.0528	-0.9609	0.2596	-0.2990	0.2192	-0.0392
Mean($ w_{kl} $)	0.8215	0.0409*	0.1253	0.0442	0.0812	0.2586	0.1502	0.3144	1.2033
Demented									
Mean(b_{kl})	0.4572	1.0812	-0.3212	1.0337	-0.8859	0.2019	-0.2815	0.1755	-0.0115
Mean($ w_{kl} $)	0.7724	0.0431*	0.1395	0.0456	0.0872	0.3328	0.1615	0.3937	4.0859

*, indicates minimum. ADNI, Alzheimer’s Disease Neuroimaging Initiative; OASIS, Open Access Series of Imaging Studies.

sample number 7 with KLD of 20. On the other hand, the demented projection near 115.4 corresponds to sample number 2 which has a negative KLD and this sample is misclassified.

We also performed a cross validation test for the ADNI images by randomly selecting 12 normal and 12 demented subjects to be test subjects for a training classifier based on 38 normal and 35 demented subjects. The results of that

test are shown in Figure 8. In that figure the accuracy of the training classifier (inferred z-projection densities shown) was 69/73=94.5%, while the test subjects were classified with an accuracy of 22/24=91.7%.

Comparison with other medical image classifiers

The first comparison study is selected because it is

Table 3 A listing of b_{ki} and $|w_{ki}|$ sample means for 12 demented and 14 normal OASIS subjects. For each class b_{10} is on average the most significant parameter (minimum $|w_{ki}|$ sample means) and numerically largest parameter

Xb	b_0	$+b_{01}x_1$	$+b_{02}x_2$	$+b_{10}x_3$	$+b_{11}x_4$	$+b_{12}x_5$	$+b_{20}x_6$	$+b_{21}x_7$	$+b_{22}x_8$
Normal									
Mean (b_{ki})	0.1624	0.9222	-0.2480	1.1688	-0.896	0.2131	-0.3702	0.2421	-0.0378
Mean ($ w_{ki} $)	1.0643	0.0467	0.1684	0.0353*	0.0788	0.3015	0.1087	0.2392	1.1431
Demented									
Mean (b_{ki})	0.2378	0.8494	-0.2262	1.1589	-0.7900	0.2040	-0.3701	0.2205	-0.0532
Mean ($ w_{ki} $)	1.0401	0.0514	0.1855	0.0357*	0.0876	0.3149	0.1082	0.2559	0.8121

*, indicates minimum. OASIS, Open Access Series of Imaging Studies.

Table 4 A listing of ordinary least squares estimates of the b_{ki} parameters for the normal and demented ADNI brain scans of Figure 5. The $|w_{ki}|$ statistics measure the normalized width of a 95% confidence interval of the true parameter β_{ki} . The smaller $|w_{ki}|$ the more significant is the parameter in determining the image model. For both of these images b_{10} , relating $y(i,j)$ and $y(i-1,j)$, is the most significant parameter

Xb	b_0	$+b_{01}x_1$	$+b_{02}x_2$	$+b_{10}x_3$	$+b_{11}x_4$	$+b_{12}x_5$	$+b_{20}x_6$	$+b_{21}x_7$	$+b_{22}x_8$
Normal b_{ki}	0.2453	0.7594	-0.1110	1.1928	-0.7366	0.0769	-0.4121	0.2426	-0.0183
Normal $ w_{ki} $	1.1042	0.0588	0.3889	0.0341	0.0936	0.8538	0.0961	0.2245	2.3603
Demented b_{ki}	0.2335	0.8030	-0.2136	1.2374	-0.7847	0.1854	-0.4232	0.2195	-0.0294
Demented $ w_{ki} $	1.0471	0.0547	0.1978	0.0327	0.0893	0.3597	0.0927	0.2515	1.4738

ADNI, Alzheimer's Disease Neuroimaging Initiative.

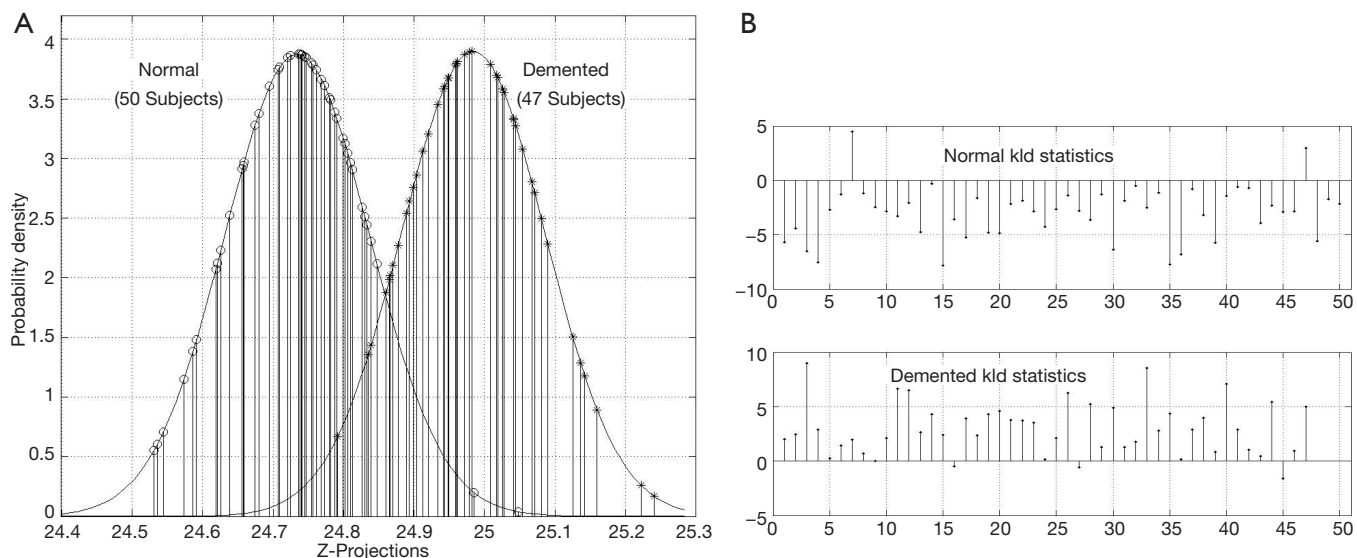


Figure 6 (A) Estimated z-projection probability densities resulting from the Fisher linear discriminant projections of the normal and demented b_{ki} parameter matrices B_n and B_d for all of the ADNI image models; (B) the KLD statistics calculated from Eq. [10]. KLD statistics measure the likelihood a given subject is correctly classified; more negative normal and more positive demented KLD statistics imply higher likelihood of correct classification. For example, the largest demented KLD statistic is 9 for subject number 3, the right-most sample in Figure 6A. This demented subject is clearly the most likely to be correctly classified. ADNI, Alzheimer's Disease Neuroimaging Initiative; KLD, Kullback-Leibler discrimination.

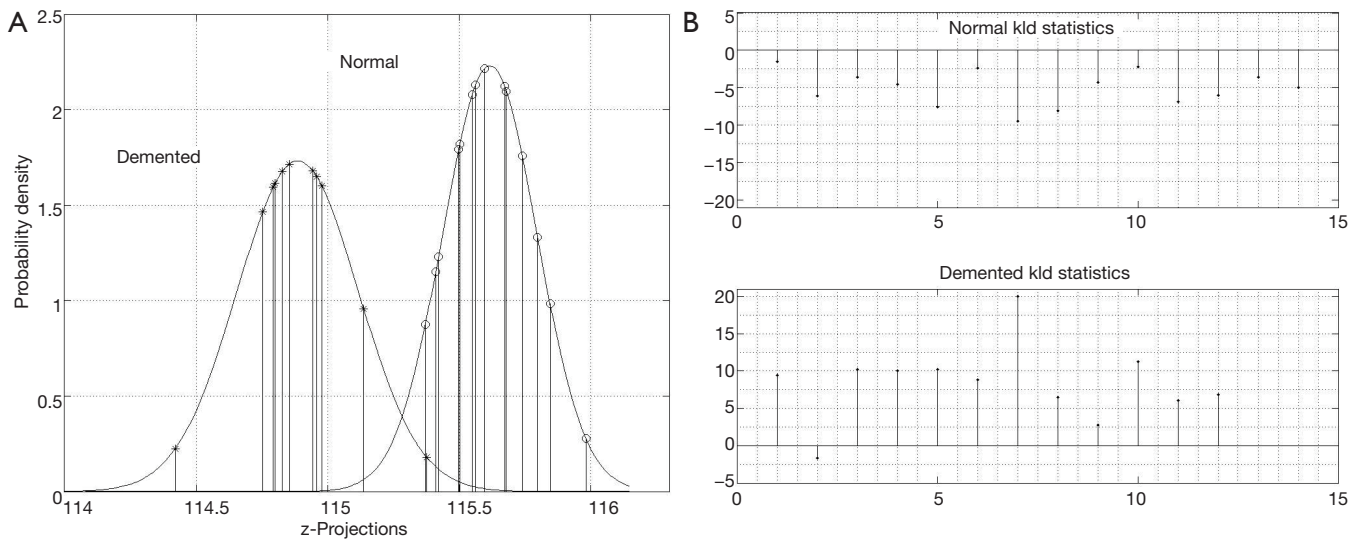


Figure 7 (A) Estimated z-projection probability densities resulting from the Fisher linear discriminant projections of the normal and demented b_{kl} parameter matrices B_n and B_d of the OASIS image models; (B) the KLD statistics are computed as described in the *Figure 6* caption. Demented subject number 7 has a KLD of 20 and a z-projection in *Figure 7A* of 114.4. OASIS, Open Access Series of Imaging Studies; KLD, Kullback-Leibler discrimination.

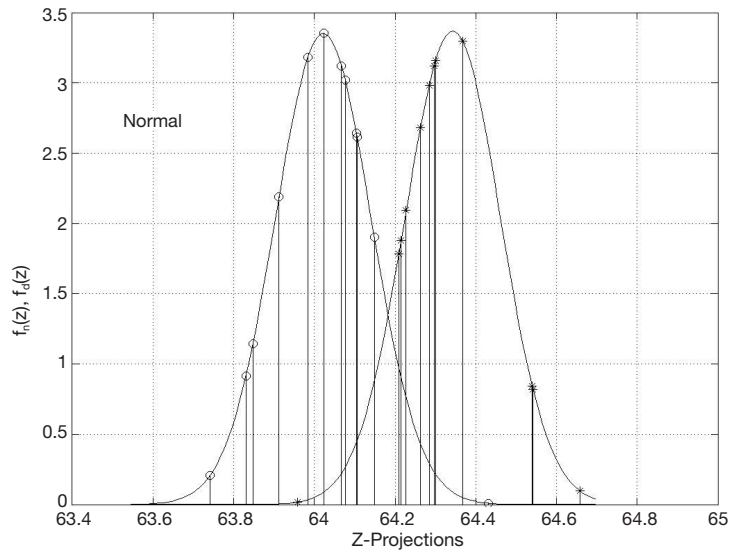


Figure 8 Z-projection probability density functions estimated from 38 normal and 35 demented ADNI subject images. Density functions represent a classification accuracy of 94.5%. Accuracy for the 24 test images is 91.7%. ADNI, Alzheimer's Disease Neuroimaging Initiative.

comprehensive in feature extractions (gray level histogram, gray level co-occurrence matrix, shape invariant moment, and FFT frequency), uses an average of 408 CT and MR images (102 testing for a 25% testing-to-training ratio) and tests 4 popular classification methods (C4.5, SVM, naive Bayes, KNN) (11). The averages over feature extractions shown in *Table 5* are to be compared to our OLS method's

91.7% accuracy using a 24.2% testing-to-training ratio. The authors give no computational time or computer specification results.

The second comparison study is selected because it tests an advanced form of the random forest (RF) classifier, uses a test-to-training ratio of 33% on 300 region of interest MR images of 256x216 pixels each (12). They report a

Table 5 Average accuracy of 12 categories of medical X-ray images (average sample size 306 training, 102 testing). Feature types are gray-level histogram, gray level co-occurrence matrix, shape invariant moment, FFT frequency (11)

Classifier types	C4.5	SVM	Naive Bayes	KNN
Average accuracy over feature types	84.71%	90.45%	76.03%	86.61%

computation time of 304 seconds average with 86.4% accuracy using a 2.66 GHz machine of unspecified RAM.

The final comparison study is selected because it tests three forms of the RF classifier, uses 10 fold cross-validation, and gives comprehensive computation times using a 3.4 GHz, 12 GB RAM computer (13). Images are 3 color of 174 colon biopsies. The grid size modeled is a rather crude 49 blocks. The original feature set produced an accuracy of 91.3% averaged over the three classifiers. Increasing features from 20 to 50 boosted the average accuracy to 93.0% with run times of 104 seconds averaged over the three RF classifiers. The authors quote five other comparable colon studies whose average accuracy is 85.43%.

The dramatic difference between our OLS method and those cited above is computation time. Using a 3.00 GHz computer with 3.72 RAM our all-training result, *Figure 6*, required 0.9014 seconds, *Figure 7* took 0.2516 seconds while *Figure 8* required 0.6466 seconds. The RF classifier of ROI images (of a comparable number of pixels) cited above required 224 seconds on a machine 13% slower than ours. Derating 224 seconds by 13% and increasing our run time by 300/73 to compensate for sample sizes implies our runtime is $195/2.66=73.3$ times faster than the cited classification of ROI images.

The colon biopsy classifiers must classify red, green and blue images so we must triple our runtimes and compensate for sample sizes by $174/73$. This gives our estimated comparable runtime of $0.6466 \times 3 \times 2.38 = 4.62$ seconds which is $104/4.62 = 22.5$ times faster than the RF image classifiers. Individually, the red, green and blue images required 34, 48, and 22 seconds respectively. OLS equivalent time is $2.38 \times 0.6466 = 1.54$ seconds for each color.

Summary

Including the 5 colon studies cited in (13), we have compared our OLS classifier to 13 classifiers from the literature and our training-test accuracy of 91.7% is better than all but one average.

In the runtime domain, none of the cited classifiers is

competitive with the OLS classifier. The lowest runtime cited for the reviewed classifiers is a minimum of 14 times slower than the OLS classifier.

The implied superiority of the OLS method over the comparison image classification methods in accuracy and computation times results from two properties of the OLS method: (I) the OLS image models are rigorously defined in terms of statistical estimation of partial difference equation solutions; (II) the `vec(.)` transformation and its theoretical preservation of the OLS format for the partial difference equation model and [Th1], is complemented by a runtime-optimized `vec(.)` function on Matlab.

Discussion

The innovation of OLS image modeling

Modeling of discrete time series, representing native discrete time systems or discretized ordinary differential equations, has advanced to the stage that transfer function theory and ubiquitous software to model data are commonplace (14-16). OLS is the estimation method of choice for such linear system models (also called 1D system because they employ one independent variable). In our past work we have found the estimated parameters of such models are capable of classifying the subject data they model and can provide insight into physical system interpretations (17-21).

The AR model hypothesis in Eq. [1], the vector transformation in Eq. [2], and the MVR theorem [Th1] make available to 2D modeling all of the methods cited above for 1D systems. Further, images are a special case of 2D systems which include image modalities of X-ray, US, MRI, PET, and SPECT.

Interpretation of image model parameters for large brain scans

While OLS models of medical image parameters are capable of robust image classification, an equally important utility lies in their diagnostic potential.

Diagnosis is taken to be at two image sizes: (I) entire

brain scans, for example the axial brain scans of this study; (II) and ROI size images which are not investigated in this study. An interpretation of average image parameters we found in image height to width ratios. This ratio calculated for all ADNI and OASIS images averages 24.9% greater for the ADNI images. Thus, the subtle shift, noted above, in the most significant parameters b_{10} in Table 3 to b_{01} in Table 2 is due to the ADNI images having more informative pixel columns than rows, thereby favoring correlations between $y(i,j)$ and $y(i,j-1)$ as measured by b_{01} . Analogously, b_{10} , relating $y(i,j)$ and $y(i-1,j)$, is more significant for the OASIS images whose pixel rows are more informative than their columns.

Image parameter medical interpretation addresses the difficulty in distinguishing which of seven possible dementia types a given subject is suffering. Having nine parameter image models and statistics to measure individual parameter significance presents the opportunity to design a study in which cognitive test scores and post mortem dementia identification outcomes can be used to identify parameters or groups of parameters specific to dementia types.

Figures 1,2 and 4,5 indicate a large within-sample variation in ventricle size as a percent of total brain scan. To determine the possible classifying effect this variation might have on z-projections, a 20% image random sample from both datasets was used to cross-correlate ventricle area with z-projection size. The correlations were not significant at the 10% level. Further, even within-class correlation was not evident, two OASIS demented subjects differed 67% in ventricle area and only 2.2×10^{-4} percent in z-projection size. The reason for this outcome lies in the observation that ventricles are interpreted by OLS as zero pixel intensity background similar to that which fills the rectangle surrounding the brain scan.

Interpretation of image model parameters for region of interest scans

ROI scan models can have more diagnostic potential than those of entire image scans. This derives from the fact that ROI regions can be modeled by generalized least squares (GLS). A smaller number of pixels make GLS computationally tractable and estimated parameters normally distributed. Consequently, one-way classification of regression models becomes a powerful diagnostic tool (22). In a preliminary study we showed that a 40×40 pixel ROI of an OASIS slice and a similar sized ROI one 1.5 mm slice above the first were significantly different at level 0.029 by

a Snedecor F test. Clearly, such analysis also is applicable to therapy progress.

Three color RGB images, such as prostate biopsies, require three OLS models which combine for the color model. Such models present no additional estimation problems but we still must work out how to compare images represented by three mutually exclusive groups of nine parameters each.

Acknowledgements

Authors' contributions: W O'Neill wrote the manuscript with R Penn who also provided neurological interpretations of modeling results. M Werner and J Thomas gathered, categorized, and prepared data for analysis. They also ran the regression and subject classification programs.

Disclosure: The authors declare no conflicts of interest.

References

1. Neuroimaging Work Group of the Alzheimer's Association, The Use of MRI and PET for Clinical Diagnosis of Dementia and Cognitive Impairment: A Consensus Report, 2005.
2. Lin JS, O'Connor E, Rossom RC, Perdue LA, Eckstrom E. Screening for cognitive impairment in older adults: A systematic review for the U.S. Preventive Services Task Force. *Ann Intern Med* 2013;159:601-12.
3. Azimi-Sadjadi MR. Estimation and identification for 2-D block Kalman filtering. *Signal Processing, IEEE Trans* 1991;39:1885-89.
4. Magnus JR. eds. *Linear Structures*. New York: Oxford U. Press, 1988.
5. Jain AK. Partial differential equations and finite-difference methods in image processing, part 1: Image representation. *J Optim Theory Appl* 1977;23:65-91.
6. Wilks SS. eds. *Mathematical Statistics*. New York: Wiley, 1963:75.
7. Wilks SS. eds. *Mathematical Statistics*. New York: Wiley, 1963:573.
8. Wilks SS. eds. *Mathematical Statistics*. New York: Wiley, 1963:454.
9. Kullback S. eds. *Information Theory and Statistics*. Mineola, NY: Dover Publications, 1997:6.
10. Plassman BL, Langa KM, Fisher GG, Heeringa SG, Weir DR, Ofstedal MB, Burke JR, Hurd MD, Potter GG, Rodgers WL, Steffens DC, Willis RJ, Wallace RB.

- Prevalence of dementia in the United States: the aging, demographics, and memory study. *Neuroepidemiology* 2007;29:125-32.
11. Li B, Li W, Zhao D. Multi-scale feature based medical image classification. *IEEE 3rd International Conference on Computer Science and Network Technology*. Chengdu, China, 2013:1182-6.
 12. Mahapatra D. Analyzing training information from random forests for improved image segmentation. *IEEE Trans Image Process* 2014;23:1504-12.
 13. Rathore S, Iftikhar MA, Hussain M, Jalil A. Classification of colon biopsy images based on novel structural features. *IEEE Int Conf on Emerging Technologies*. Islamabad, Pakistan, 2013:1-6.
 14. Box G, Jenkins G, Reinsel G. eds. *Time Series Analysis*, 3rd ed. Upper Saddle River NJ: Prentice Hall, 1994.
 15. Fuller W. eds. *Introduction to Statistical Time Series*. New York: Wiley & Son, 1996.
 16. Ljung L. eds. *System Identification: Theory for the User*. Upper Saddle River, NJ: Prentice Hall, 1999.
 17. O'Neill WD, Oroujeh AM, Merritt SL. Pupil noise is a discriminator between narcoleptics and controls. *IEEE Trans Biomed Eng* 1998;45:314-22.
 18. O'Neill WD, Zimmerman S. Neurological interpretations and the information in the cognitive pupillary response. *Methods Inf Med* 2000;39:122-4.
 19. O'Neill WD, Trick KP. The narcoleptic cognitive pupillary response. *IEEE Trans Biomed Eng* 2001;48:963-8.
 20. O'Neill W, Mercer P, Sheldon S, Kotsos T. Detection of juvenile sleep deprivation by stochastic optimization of pupillographic records. *Methods Inf Med* 2003;42:282-6.
 21. Shuhaiber J, O'Neill W. The Diagnostic Efficacy of Left Ventricle Hemodynamic Parameters: Classifying Normal Mitral Valves Against Prosthetic Disease Proxies. *Cardiovasc Eng and Tech* 2012;3:112-22.
 22. Kullback S. eds. *Information Theory and Statistics*. Mineola, NY: Dover, 1997:219-22.

Cite this article as: O'Neill W, Penn R, Werner M, Thomas J. A theory of fine structure image models with an application to detection and classification of dementia. *Quant Imaging Med Surg* 2015;5(3):356-367. doi: 10.3978/j.issn.2223-4292.2015.03.11

Appendix

A₁—Derivation of the vec(A) transformation

We need to show $\text{vec}\left(\sum_{i=1}^k c_i A_i\right) = \sum_{i=1}^k c_i \text{vec}(A_i)$

or equivalently (I) $\text{vec}(cA) = c\text{vec}(A)$ and (II) $\text{vec}(A+B) = \text{vec}(A) + \text{vec}(B)$. (I) is trivial since a matrix can be scaled by rows or columns. If A and B are m by n then $A = [a_1 \ a_2 \ \dots \ a_n]$ and $B = [b_1 \ b_2 \ \dots \ b_n]$ so $\text{vec}(A+B) = [a_1^T + b_1^T \ a_2^T + b_2^T \ \dots \ a_n^T + b_n^T]^T = [a_1^T \ a_2^T \ \dots \ a_n^T]^T + [b_1^T \ b_2^T \ \dots \ b_n^T]^T = \text{vec}(A) + \text{vec}(B)$.

A₂—Derivation of the w_{kl} statistic

The well-known result using Eq. [3] in Eq. [7] is $E(b) = \beta$. Then for every k, l , and defining $\sigma_{kl}^2 = \text{var}(b_{kl})$, the Chebyshev inequality is:

$$\Pr\left(|b_{kl} - \beta_{kl}| \geq \varepsilon\right) \leq \sigma_{kl}^2 / \varepsilon^2 = \alpha \quad [\text{A1}]$$

Therefore a $1-\alpha$ confidence interval for the true β_{kl} satisfies:

$$\Pr\left(|b_{kl} - \beta_{kl}| \leq \varepsilon\right) \geq 1 - \alpha \quad [\text{A2}]$$

b and β have the same scale so it is clearly better to have a large estimated b_{kl} and a small confidence interval for β_{kl} than conversely.

Thus, we normalize the confidence interval width 2ε by b_{kl} , and define $w_{kl} = 2\varepsilon/b_{kl}$. The coefficient of variation of the estimate b_{kl} is $c_{kl} = \sigma_{kl} / b_{kl}$ and $\sigma_{kl} = \varepsilon\sqrt{\alpha}$, so:

$$w_{kl} = 2c_{kl} / \sqrt{\alpha} \quad [\text{A3}]$$

Calculating w_{kl} for a given model requires an estimate of σ_{kl}^2 , the k, l diagonal element of the covariance matrix $\text{cov}(b)$. From Eq. [3] in Eq. [7]:

$$\text{cov}(b) = E[(b - \beta)(b - \beta)^T] = (X^T X)^{-1} X^T \text{cov}(u_o) X (X^T X)^{-1} \quad [\text{A4}]$$

and $\text{cov}(u_o)$ is readily estimated from the residual vector e_o in Eq. [6].