OPEN ACCESS Freely available online

PLOS ONE

Kronecker Product Linear Exponent AR(1) Correlation Structures for Multivariate Repeated Measures

Sean L. Simpson[1,2]*, Lloyd J. Edwards[2], Martin A. Styner[3], Keith E. Muller[4]

1 Department of Biostatistical Sciences, Wake Forest School of Medicine, Winston-Salem, North Carolina, United States of America, **2** Department of Biostatistics, University of North Carolina at Chapel Hill, Chapel Hill, North Carolina, United States of America, **3** Departments of Psychiatry and Computer Science, University of North Carolina at Chapel Hill, Chapel Hill, North Carolina, United States of America, **4** Department of Health Outcomes and Policy, University of Florida, Gainesville, Florida, United States of America

Abstract

Longitudinal imaging studies have moved to the forefront of medical research due to their ability to characterize spatio-temporal features of biological structures across the lifespan. Credible models of the correlations in longitudinal imaging require two or more pattern components. Valid inference requires enough flexibility of the correlation model to allow reasonable fidelity to the true pattern. On the other hand, the existence of computable estimates demands a parsimonious parameterization of the correlation structure. For many one-dimensional spatial or temporal arrays, *the linear exponent autoregressive* (LEAR) correlation structure meets these two opposing goals in one model. The LEAR structure is a flexible two-parameter correlation model that applies to situations in which the within-subject correlation decreases exponentially in time or space. It allows for an attenuation or acceleration of the exponential decay rate imposed by the commonly used continuous-time AR(1) structure. We propose the Kronecker product LEAR correlation structure for multivariate repeated measures data in which the correlation between measurements for a given subject is induced by two factors (e.g., spatial and temporal dependence). Excellent analytic and numerical properties make the Kronecker product LEAR model a valuable addition to the suite of parsimonious correlation structures for multivariate repeated measures data. Longitudinal medical imaging data of caudate morphology in schizophrenia illustrates the appeal of the Kronecker product LEAR correlation structure.

Citation: Simpson SL, Edwards LJ, Styner MA, Muller KE (2014) Kronecker Product Linear Exponent AR(1) Correlation Structures for Multivariate Repeated Measures. PLoS ONE 9(2): e88864. doi:10.1371/journal.pone.0088864

Editor: Xi-Nian Zuo, Institute of Psychology, Chinese Academy of Sciences, China

Received June 12, 2013; **Accepted** January 14, 2014; **Published** February 21, 2014

Funding: Sean Simpson's support includes the Translational Scholar Award from the Translational Science Institute of Wake Forest School of Medicine and NIBIB K25 EB012236-01A1. Lloyd Edwards was supported by the National Center for Research Resources and the National Center for Advancing Translational Sciences, National Institutes of Health, through Grant Award Number UL1TR000083. Keith Muller's support includes NIDCR R01 DE020832, R01 DE020832-01A1S1, U54 DE019261, NICHD P01 HD065647, NCRR 3UL1RR029890-03S1, NHLBI R01 HL091005, NIAAA R01 AA016549, NIDA R01 DA031017, and NIDDK R01 DK072398. The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

Competing Interests: The authors have declared that no competing interests exist.

* E-mail: slsimpso@wakehealth.edu

Introduction

Multivariate repeated measures studies are characterized by data that have more than one set of correlated outcomes or repeated factors. Spatio-temporal data fall into this more general category, since the outcome variables repeat in both space and time. Valid analysis requires accurately modeling the correlation pattern. Muller et al. and Gurka et al. showed that under-specifying the correlation structure can severely inflate test size of tests of fixed effects in the general linear mixed model [1,2]. Modeling the correlation pattern separately for each repeated factor with multivariate repeated measures data has substantial advantages. Most important, the approach allows for the choosing and tuning of each model separately, which improves accuracy and makes model fitting easier. Furthermore, the approach uses fewer parameters than an unstructured model. The Kronecker product combines the factor-specific correlation structures into an overall correlation model.

Separable Correlation Models

Galecki [3] gave a detailed treatment of Kronecker product covariance structures, also known as separable covariance models. A covariance matrix is *separable* if and only if it can be written as $\boldsymbol{\Sigma}=\boldsymbol{\Gamma}\otimes\boldsymbol{\Omega}$, where $\boldsymbol{\Gamma}$ and $\boldsymbol{\Omega}$ are factor-specific covariance matrices (e.g. the covariance matrices for the temporal and spatial dimensions of spatio-temporal data respectively). A key advantage of the model is the ease of interpretation of the independent contribution of every repeated factor to the overall within-subject error covariance matrix. Galecki, Naik and Rao, and Mitchell et al. [3–5] detailed the computational advantages of the Kronecker product covariance structure. The partial derivatives, inverse, and Cholesky decomposition of the overall covariance matrix can be performed more easily on the factor-specific models because they have much smaller dimensions.

While separable covariance models are commonly used in the spatial statistics literature [6], they have rarely been used in multivariate longitudinal (and more generally, multivariate repeated measures) data analysis. To our knowledge, no commonly used statistical packages provide a flexible framework for implementing the structures, limiting their use to those with the appropriate programming skills. For example, SAS version 9.3 [7] has only three Kronecker product covariance structures (unstructured matrix paired with either an unstructured, compound

symmetric, or discrete-time AR(1) matrix). Given the advantages of separable models, extending software to allow general implementation is important for researchers in a variety of areas. For example, longitudinal group-randomized controlled trials often have within-group correlations (e.g., by school for youth based studies) and within-subject, longitudinally induced correlations [8]. Such data can be modeled effectively with the Kronecker product of a compound symmetric and LEAR correlation structure.

Brain morphology research is another area in which separable models would be effective because both the shape of brain structures (spatial correlation) and how they change over time (temporal correlation) must be analyzed. Our work concerns temporal changes in caudate morphology in schizophrenics. Schizophrenia is characterized by disabling impairments in the perception or expression of reality. Pathological changes in brain morphology in schizophrenics may be progressive and associated with clinical outcome. Much recent work has focused on the effect of antipsychotic drugs on brain morphology [9,10]. The drugs were tested on the caudate (shown in **Figure 1**), an important part of the brain's learning and memory system. Assessing drug efficacy requires proper analysis of temporal shape changes in the caudate, which can be modeled with separable correlation structures.

Limitations of separable models have been noted by various authors. Most important, as mentioned in [11], patterns of interaction among the various factors cannot be modeled when utilizing a Kronecker product structure. Within a given subject, all factors must have *consistently-spaced* measurements. In the context of spatio-temporal data, this means that at each time point a given subject must have the same number of measurements taken at the same spatial locations.

An Appealing and Flexible Separable Correlation Model

Fitting a Kronecker product structure requires choosing models for each of the factors. In medical imaging, repeated measures dimensions typically have within-subject correlation decreasing exponentially in time or space. The continuous-time, first-order autoregressive correlation structure, denoted AR(1), is often used in longitudinal settings. This model was briefly examined in [13] and is a special case of the model described in [14]. Despite its wide use, the AR(1) structure often poorly gauges within-subject correlations that decay at a slower or faster rate than required by the AR(1) model. The *linear exponent autoregressive* (LEAR) correlation model, defined in **Table 1** (reproduced from [15]) and equations 1 and 2 below, overcomes this limitation by allowing an attenuation or acceleration of the exponential decay rate imposed by the AR(1) structure [15]. **Table 1** also defines the AR(1) model

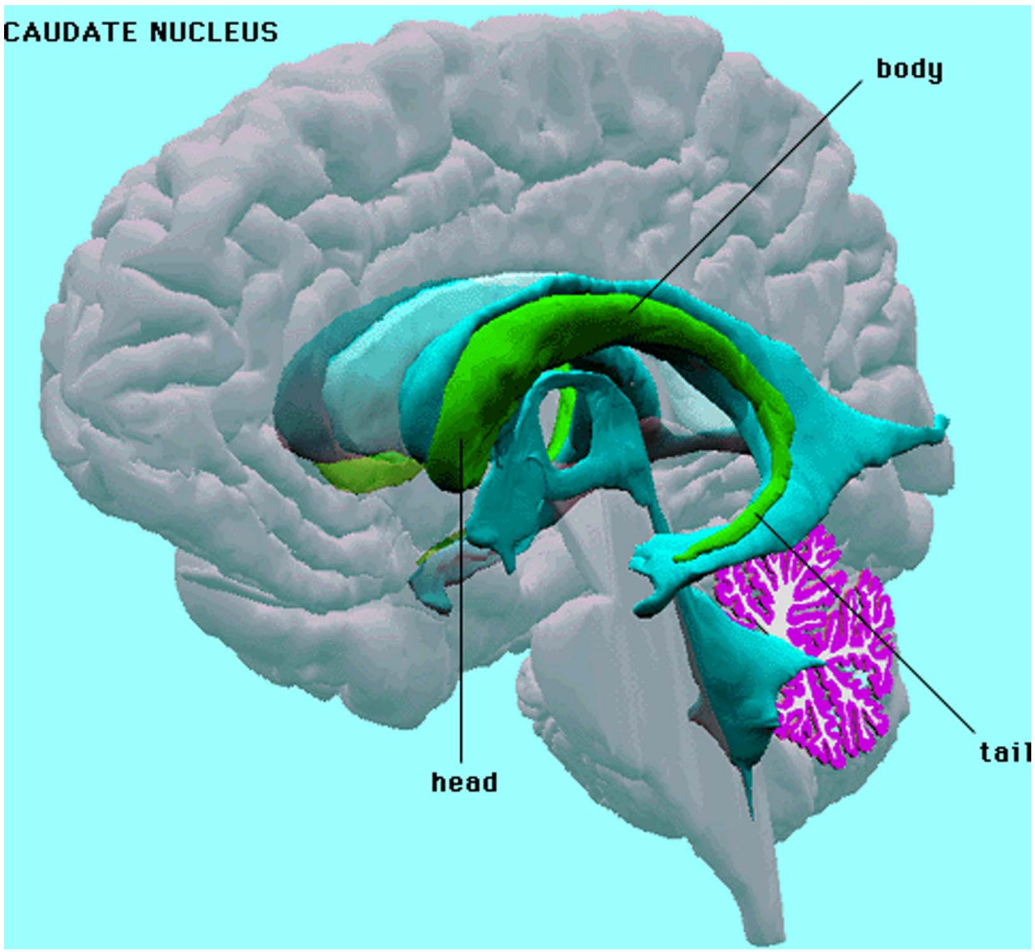


Figure 1. The Caudate Nuclei in the Human Brain.
doi:10.1371/journal.pone.0088864.g001

along with other stationary correlation structures that are continuous functions of distance. The focus on stationary models reflects the desire to maintain parsimony across a variety of data types. Moreover, the greater complexity of non-stationary models does not seem necessary for the limited applications of interest. The exponential model defined in Table 1, discussed almost exclusively in the spatial statistics literature, is in fact equivalent to the continuous-time AR(1) model with $\phi = -(1/\ln \rho)$. As proposed in [15], we believe that the AR(1) and damped exponential (DE) models serve as the most relevant competitors to the LEAR structure. Special cases of both the LEAR and DE families include the AR(1), compound symmetry, and first-order moving average (MA(1)) correlation structures.

The advantages of employing a LEAR model for each component led us to consider a Kronecker product LEAR correlation structure for multivariate repeated measures data in which the correlation between measurements for a given subject is induced by two factors. We allow for an imbalance in both dimensions across subjects, i.e., an unequal numbers of observations. The LEAR model also accommodates any arbitrary spacing within a dimension. We use maximum likelihood estimation of the general linear model with Gaussian errors to illustrate the benefits of the structure. All other common estimation methods for linear and nonlinear models could also be used with a Kronecker product LEAR structure.

Materials and Methods

Ethics Statement

This analysis involved the application of a new method to extant data. The original study was conducted from March 1, 1997 to July 31, 2001 at 14 academic medical centers. All subjects gave written informed consent in accordance with the Declaration of Helsinki. Each site's institutional review board (University of North Carolina School of Medicine, Chapel Hill, NC; Emory University School of Medicine, Atlanta, GA; McLean Hospital, Harvard Medical School, Belmont, MA; John Umstead Hospital, Duke University Health System, Butner, NC; University of Florida, Gainesville, FL; Massachusetts Mental Health Center, Harvard Medical School, Boston, MA; University of Massachusetts Medical Center, Worchester, MA; University of Pennsylvania Medical Center, Philadelphia, PA; University of Toronto School of Medicine, Toronto; University of Cincinnati, Cincinnati, OH; Stanford University School of Medicine, Stanford, CA; University of Michigan Medical Center, Ann Arbor, MI; University Hospital Utrecht, Utrecht, the Netherlands; Maudsley Hospital, Institute of Psychiatry, London) approved the study.

Table 1. Stationary Correlations Structures That are Continuous Functions of Distance.

Structure	(j,k)th element[a], $j \neq k$	Params	Data Types[b]
LEAR	$\rho^{d_{\min}+\delta[(d_{jk}-d_{\min})/(d_{\max}-d_{\min})]}$	2	L/T,S,O
AR(1)	$\rho^{d_{jk}}$	1	L/T,S,O
DE	$\rho^{d_{jk}^{\nu}}$	2	L/T,S,O
GAR(1)	$\rho^{d_{jk}}\frac{\Gamma(d_{jk}+\delta)F(\delta,d_{jk}+\delta;d_{jk}+1;\rho^2)}{\Gamma(\delta)\Gamma(d_{jk}+1)F(\delta,\delta;1;\rho^2)}$	2	L/T
Exponential	$\exp(-d_{jk}/\phi)$	1	S
Gaussian	$\exp\left(-d_{jk}^2/\phi^2\right)$	1	S
Linear	$(1-\phi d_{jk})I(\phi d_{jk}\leq 1)$	1	S
Matern	$[1/\Gamma(\nu)](d_{jk}/2\phi)^{\nu}2K_{\nu}(d_{jk}/\phi)$	2	S
Spherical	$\left[1-(3d_{jk}/2\phi)+\left(d_{jk}^3/2\phi^3\right)\right]I(d_{jk}\leq\phi)$	1	S

NOTE: [41] and [42] detail the DE and GAR(1) structures respectively. See [43] for further details regarding the spatial structures.
[a]elements of $\mathbf{\Gamma}_i$ and/or $\mathbf{\Omega}_i$ from equation 3.
d_{jk} – distance between j^{th} and k^{th} measurement of i^{th} subject.
$\Gamma(\cdot)$ – gamma function.
$F(\theta_1,\theta_2;\theta_3;\theta)$ – hypergeometric function.
$K_{\nu}(\cdot)$ – modified Bessel function of the second kind of (real) order $\nu>0$.
[b]L/T: Longitudinal/Time Series.
S: Spatial.
O: Other.
doi:10.1371/journal.pone.0088864.t001

Example Data: Schizophrenia and Caudate Morphology

Our data came from longitudinal MRI scans of the left caudate for 240 schizophrenia patients and 56 controls. The surface of each object was parameterized via the medial representation (m-rep) method as described in [16]. The caudate shape was determined as a 3×7 grid of medial mesh points (spherical nodes) (see **Figure 2**). Data were reduced to one outcome measure: *radius* in cm as a measure of local object width at an m-rep node (21 locations per caudate). This measure is represented in **Figure 2** by the length of the spokes emanating from the spherical nodes to the surface of the object. The distance between two radii for a given subject was calculated as the mean Euclidean distance over all images. The schizophrenia patients were randomized to either haloperidol (a conventional antipsychotic) or olanzapine (an atypical antipsychotic). Scans were taken up to 47 months post-baseline with the median and maximum number of scans per subject being three and seven respectively. The two treatment groups were combined in order to avoid undermining ongoing research, as the final trial results comparing the treatments have not yet been published. The other covariates of interest were age, gender, and race. Preliminary analyses (based on the same test discussed in the Results section, namely the residual approximation of the F-test for a Wald statistic) showed that the shape of the caudate, and thus the radii, differs significantly at baseline between schizophrenics and controls. The study hypothesized that the neuroprotective effect of the drugs would lead to no overall differences in shape between the patients and controls. An example subset of one subject's data is given in **Table 2**.

Model Definition

Suppose $\boldsymbol{Y}_i$ is a $T_i \times S_i$ matrix of observations (e.g., T_i temporal measurements and S_i spatial measurements) on the i^{th} subject $i \in \{1,\ldots,N\}$. Let $\boldsymbol{y}_i = vec(\boldsymbol{Y}_i')$ be the $T_iS_i \times 1$ vector of the T_iS_i observations. Here, $\mathcal{C}(y_{ijl},y_{ikl}) = \rho_{i\gamma;jk}$ and $\mathcal{C}(y_{ijl},y_{ijm}) = \rho_{i\omega;lm}$ represent the temporal (or factor 1) and spatial (or factor 2) correlations, respectively, for $\mathcal{C}(\cdot)$ the correlation operator. Then for $\mathbf{\Gamma}_i = \{\rho_{i\gamma;jk}\}$ (the temporal/factor 1 correlation matrix) and $\mathbf{\Omega}_i = \{\rho_{i\omega;lm}\}$ (the spatial/factor 2 correlation matrix), the factor-specific *linear exponent autoregressive* (LEAR) correlation structures are

$$\rho_{i\gamma;jk} = \mathcal{C}(y_{ijl},y_{ikl}) = \begin{cases} \rho_{\gamma}^{d_{t;\min}+\delta_{\gamma}\left[(d(t_{ijl},t_{ikl})-d_{t;\min})/(d_{t;\max}-d_{t;\min})\right]} & j \neq k \\ 1 & j = k \end{cases}, \quad (1)$$

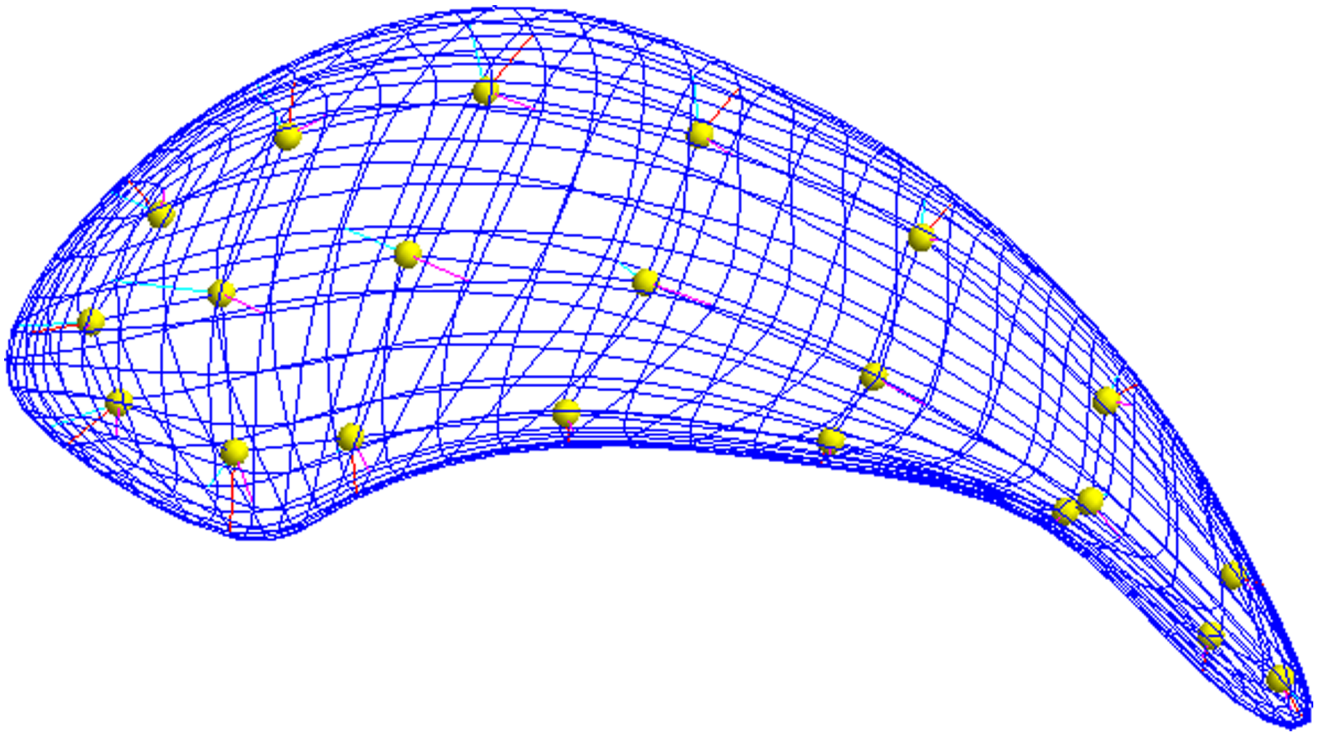


Figure 2. M-rep shape representation model of the caudate.
doi:10.1371/journal.pone.0088864.g002

$$\rho_{i\omega;lm} = \mathcal{C}(y_{ijl}, y_{ijm}) = \begin{cases} \rho_\omega^{\,d_{s;\min} + \delta_\omega \left[(d(s_{ijl}, s_{ijm}) - d_{s;\min}) / (d_{s;\max} - d_{s;\min}) \right]} & l \neq m \\ 1 & l = m \end{cases}. \quad (2)$$

The Kronecker product LEAR correlation structure is

$$\mathbf{\Gamma}_i \otimes \mathbf{\Omega}_i, \quad (3)$$

where $d(t_{ijl}, t_{ikl})$ and $d(s_{ijl}, s_{ikl})$ are the distances between measurement times and locations respectively. In turn, $(d_{t;\min}, d_{s;\min})$ and $(d_{t;\max}, d_{s;\max})$ are computational constants equal to the minimum and maximum number of temporal and spatial distance units across all subjects. Parameters ρ_γ and ρ_ω are the correlations between observations separated by one unit of time and distance respectively, and δ_γ and δ_ω are the decay speeds. We assume $0 \leq \rho_\gamma, \rho_\omega < 1$ and $0 \leq \delta_\gamma, \delta_\omega$. The $(d_{t;\min}, d_{s;\min})$ and $(d_{t;\max}, d_{s;\max})$ constants allow the model to adapt to the data and scale distance such that the multiplier of the decay speeds δ_γ and δ_ω, $(d(t_{ijl}, t_{ikl}) - d_{t;\min}) / (d_{t;\max} - d_{t;\min})$ and $(d(s_{ijl}, s_{ijm}) - d_{s;\min}) / (d_{s;\max} - d_{s;\min})$, is between 0 and 1 for computational purposes. One could also consider tuning the constants if necessary to address convergence issues. Simpson et al. [15] gave details on setting the distance constants. Ensuring that the factor-specific matrices $\mathbf{\Gamma}_i$ and $\mathbf{\Omega}_i$ are positive definite (as discussed in [15]) is sufficient for ensuring the positive definiteness of $(\mathbf{\Gamma}_i \otimes \mathbf{\Omega}_i)$ (Theorem 7.10, [17]). Note that each factor-specific LEAR model can also be reparameterized as the Hadamard product of an equal correlation and a continuous-time AR(1) model, as detailed in [15].

Graphical depictions of the Kronecker product LEAR structure help to provide insight into the types of correlation patterns that can be modeled. A correlation pattern in which both of the factor-specific matrices (e.g. spatial and temporal matrices) have decay rates slower than that of the AR(1) model is illustrated in **Figure 3A**. **Figures 3B and 3C** exhibit patterns with dual AR(1) and faster than AR(1) decay rates respectively.

Given the advantages of the Kronecker product covariance model, we believe that it has been underutilized in practice. As alluded to in the Introduction, the model has great computational properties and simplifies interpretation. It also reduces the dimension of the calculations, sometimes drastically (e.g., having two 30×30 matrices vs. a 900×900 matrix), while allowing complex factor-specific correlation structures. These inherent qualities make the Kronecker product covariance model an appealing and useful tool (among the suite of tools) in many High Dimensional, Low Sample Size contexts common in medical imaging and various kinds of "-omics" data. Modeling the factor-specific matrices with the LEAR structure is especially attractive due to the increased flexibility, parsimony, and numerical stability resulting from this combination.

Here, we adopt the technique of modeling the correlation and variance structures separately as seen in [18] and others. With the assumptions of covariance model separability and homoscedasticity, an equal variance Kronecker product structure has great appeal. The overall within-subject error covariance matrix is then defined as

$$\mathbf{\Sigma}_i = \sigma^2 (\mathbf{\Gamma}_i \otimes \mathbf{\Omega}_i) \quad (4)$$

for the i^{th} subject or independent sampling unit. The formulation

Table 2. Example subset of one subject's data (Treatment - Olanzapine, Gender - M, Age - 20).

	$\log_2$(radius)				
Node #	Baseline	Month 3	Month 12	Month 24	Month 47
1	−5.92	−5.96	−5.95	−5.98	−5.98
2	−5.58	−5.61	−5.65	−5.76	−5.68
3	−5.21	−5.17	−5.10	−5.15	−5.11
4	−4.72	−4.67	−4.67	−4.77	−4.75
5	−4.53	−4.43	−4.53	−4.49	−4.53
6	−4.81	−4.89	−4.80	−4.84	−4.82
7	−5.00	−5.14	−4.86	−5.16	−4.91
8	−5.67	−5.82	−5.73	−5.75	−5.78
9	−5.01	−5.09	−5.09	−5.12	−5.16
10	−4.43	−4.47	−4.51	−4.46	−4.52
11	−3.94	−4.04	−4.01	−4.05	−4.05
12	−3.85	−3.82	−3.84	−3.88	−3.82
13	−3.66	−3.58	−3.64	−3.67	−3.71
14	−4.59	−4.63	−4.52	−4.60	−4.54
15	−5.68	−5.76	−5.56	−5.61	−5.83
16	−5.88	−5.91	−5.86	−5.89	−5.90
17	−5.76	−5.79	−5.76	−5.78	−5.80
18	−5.76	−5.76	−5.70	−5.72	−5.78
19	−5.08	−5.06	−5.07	−5.14	−5.08
20	−4.42	−4.46	−4.37	−4.34	−4.47
21	−4.57	−4.52	−4.53	−4.59	−4.62

doi:10.1371/journal.pone.0088864.t002

has several advantages. The reduction in the number of parameters leads to computational benefits. The model is also identifiable since $\boldsymbol{\Gamma}_i$ and $\boldsymbol{\Omega}_i$ will necessarily be correlation matrices. When heteroscedasticity is present, σ^2 can be thought of as an aggregate variance parameter for the two factors. The robustness of the equal variance model to deviations from homoscedasticity likely depends on the accuracy and flexibility of the specified correlation matrices ($\boldsymbol{\Gamma}_i$ and $\boldsymbol{\Omega}_i$) and its examination will be left to future work. We focus on modeling the correlation and assume an equal variance structure for the application of interest.

Model Estimation

The Kronecker product LEAR structure can be imbedded within various modeling and estimation methods. The best approach may vary by context. With linear structured, factor-specific matrices, the noniterative approach in [19] has appeal. However, the approach is not appropriate for the LEAR structure given its nonlinear nature. Many have used maximum likelihood methods for parameter estimation in a Kronecker product model [4,5,12,20–24], but none of their approaches allow for data that are unbalanced in both dimensions. As noted in [25] and others, the Kenward-Roger approach with REML estimation is preferable for small sample estimation and inference. Non- and semiparametric approaches, as used in [18,26], may prove beneficial for non-Gaussian data or when a nonparametric variance function is appropriate (which could still be coupled with the parametric LEAR correlation functions). However, fully parametric covariance functions (parametric variances and correlations) may provide more informative results and yield more powerful inference. The Kronecker product LEAR model may also serve as a plausible working correlation structure in a generalized estimating equation (GEE) framework. We focus on Gaussian data with moderately large sample sizes, and leave the examination of the Kronecker product model in other contexts to future work.

We consider maximum likelihood (ML) estimation of the general linear model where the Gaussian errors have a Kronecker product LEAR correlation structure. We allow for an imbalance in both dimensions of the data. Our moderately large sample context precludes the need for REML estimation.

Consider the following general linear model for multivariate repeated measures data with the Kronecker product LEAR correlation structure:

$$\boldsymbol{y}_i = \boldsymbol{X}_i\boldsymbol{\beta} + \boldsymbol{e}_i. \tag{5}$$

Again, $\boldsymbol{y}_i = vec(\boldsymbol{Y}_i')$, with $\boldsymbol{Y}_i$ being a $T_i \times S_i$ matrix of observations (e.g., T_i temporal measurements and S_i spatial measurements) on the i^{th} subject $i \in \{1,\ldots,N\}$. Thus, $\boldsymbol{y}_i$ is a $T_iS_i \times 1$ vector of the T_iS_i observations, $\boldsymbol{\beta}$ is a $q \times 1$ vector of fixed and unknown population parameters, $\boldsymbol{X}_i$ is a $T_iS_i \times q$ fixed and known design matrix corresponding to the fixed effects, and $\boldsymbol{e}_i$ is a $T_iS_i \times 1$ vector of random error terms. We assume $\boldsymbol{e}_i \sim N_{T_iS_i}(\boldsymbol{0}, \boldsymbol{\Sigma}_i = \sigma^2[\boldsymbol{\Gamma}_i \otimes \boldsymbol{\Omega}_i])$ is independent of $\boldsymbol{e}_{i'}$ for $i \neq i'$, where $\boldsymbol{\Gamma}_i$ and $\boldsymbol{\Omega}_i$ are defined in equations 1 and 2.

Setting $\boldsymbol{\Sigma}_i = \sigma^2[\boldsymbol{\Gamma}_i(\tau_\gamma) \otimes \boldsymbol{\Omega}_i(\tau_\omega)]$, where $\boldsymbol{\tau} = \{\boldsymbol{\tau}_\gamma; \boldsymbol{\tau}_\omega\} = \{\delta_\gamma, \rho_\gamma; \delta_\omega, \rho_\omega\}$, the log-likelihood function of the parameters given the data under the model is:

$$\begin{aligned} & l(\boldsymbol{y}; \boldsymbol{\beta}, \sigma^2, \boldsymbol{\tau}) \\ &= -\frac{n}{2}\ln(2\pi) - \frac{1}{2}\sum_{i=1}^{N}\ln|\sigma^2\boldsymbol{\Gamma}_i \otimes \boldsymbol{\Omega}_i| \\ &\quad - \frac{1}{2\sigma^2}\sum_{i=1}^{N}\boldsymbol{r}_i(\boldsymbol{\beta})'(\boldsymbol{\Gamma}_i \otimes \boldsymbol{\Omega}_i)^{-1}\boldsymbol{r}_i(\boldsymbol{\beta}) \\ &= -\frac{n}{2}\ln(2\pi) - \frac{1}{2}\sum_{i=1}^{N}\left(T_iS_i\ln(\sigma^2) + \ln|\boldsymbol{\Gamma}_i \otimes \boldsymbol{\Omega}_i|\right) \\ &\quad - \frac{1}{2\sigma^2}\sum_{i=1}^{N}\boldsymbol{r}_i(\boldsymbol{\beta})'(\boldsymbol{\Gamma}_i \otimes \boldsymbol{\Omega}_i)^{-1}\boldsymbol{r}_i(\boldsymbol{\beta}), \end{aligned} \tag{6}$$

where $n = \sum_{i=1}^{N} T_iS_i$ and $\boldsymbol{r}_i(\boldsymbol{\beta}) = \boldsymbol{y}_i - \boldsymbol{X}_i\boldsymbol{\beta}$. The ML estimates are derived following the approach employed in [15]. After profiling σ^2 out of the likelihood, the profile log-likelihood is given by

$$\begin{aligned} l_p(\boldsymbol{y}; \boldsymbol{\beta}, \boldsymbol{\tau}) = & -\frac{n}{2}\ln(2\pi) - \frac{1}{2}\sum_{i=1}^{N}\ln|\boldsymbol{\Gamma}_i \otimes \boldsymbol{\Omega}_i| \\ & - \frac{1}{2}n\ln\left[\sum_{i=1}^{N}\boldsymbol{r}_i(\boldsymbol{\beta})'(\boldsymbol{\Gamma}_i^{-1} \otimes \boldsymbol{\Omega}_i^{-1})\boldsymbol{r}_i(\boldsymbol{\beta})\right] + \frac{1}{2}n\ln n - \frac{n}{2} \end{aligned} \tag{7}$$

To avoid computational issues it is best to use the equality

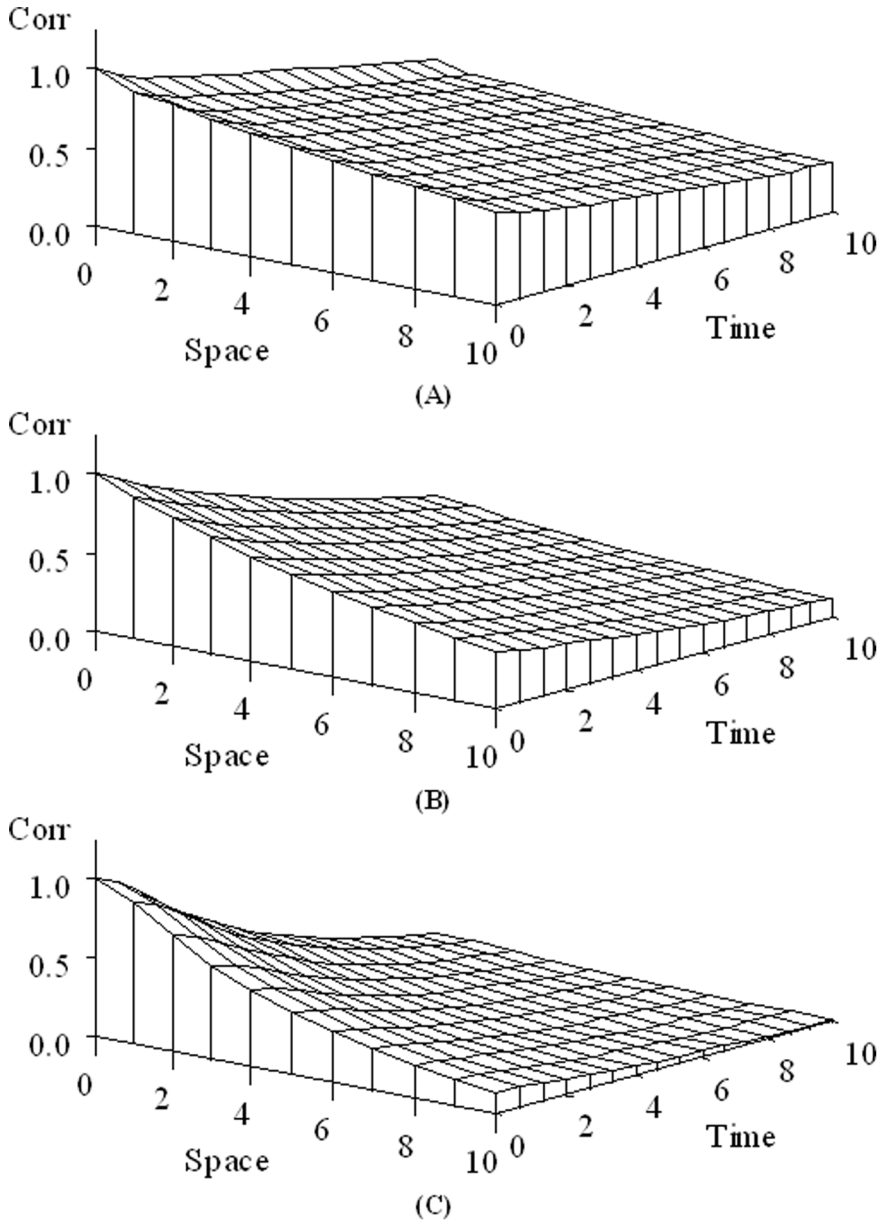


Figure 3. Plot of correlation as a function of spatial and temporal distance. (**A**) Both factor specific matrices have a decay rate that is slower than that of the AR(1) model with correlation parameters $\{\rho_\gamma = 0.9, \rho_\omega = 0.9\}$ and $\{\delta_\gamma/(d_{t;\max} - d_{t;\min}) = 0.5, \delta_\omega/(d_{s;\max} - d_{s;\min}) = 0.5\}$. (**B**) Both factor specific matrices have an AR(1) decay rate with correlation parameters $\{\rho_\gamma = 0.9, \rho_\omega = 0.9\}$and $\{\delta_\gamma/(d_{t;\max} - d_{t;\min}) = 1, \delta_\omega/(d_{s;\max} - d_{s;\min}) = 1\}$. (**C**) Both factor specific matrices have a decay rate that is faster than that of the AR(1) model with correlation parameters $\{\rho_\gamma = 0.9, \rho_\omega = 0.9\}$ and $\{\delta_\gamma/(d_{t;\max} - d_{t;\min}) = 2, \delta_\omega/(d_{s;\max} - d_{s;\min}) = 2\}$.
doi:10.1371/journal.pone.0088864.g003

$$\ln|\boldsymbol{\Gamma}_i \otimes \boldsymbol{\Omega}_i| = S_i \ln|\boldsymbol{\Gamma}_i| + T_i \ln|\boldsymbol{\Omega}_i|$$

in case $|\boldsymbol{\Gamma}_i \otimes \boldsymbol{\Omega}_i|$ is close to zero.

The ML estimates of the model parameters may be computed with the Newton-Raphson algorithm, which requires the first and second partial derivatives of the profile log-likelihood. The derivations of the first partial derivatives are available from the authors. The second partial derivatives of the parameters, which are employed to determine the asymptotic variance-correlation matrix of the estimators, may be approximated by finite difference formulas. The derivative approximations are detailed in [27,28]. The 15 analytic second derivatives can be derived explicitly as in [15]. However, the approximations have proved very accurate.

After getting the estimates of $\boldsymbol{\beta}$ and $\boldsymbol{\tau}$ utilizing the Newton-Raphson algorithm, an estimate of σ^2 is calculated by substituting the estimates into $\hat{\boldsymbol{\sigma}}^2{}_{ML}(\boldsymbol{\beta},\boldsymbol{\tau}) = n^{-1}\sum_{i=1}^{N} \boldsymbol{r}_i(\boldsymbol{\beta})'(\boldsymbol{\Gamma}_i^{-1} \otimes \boldsymbol{\Omega}_i^{-1})\boldsymbol{r}_i(\boldsymbol{\beta})$, which is the expression resulting from the initial profiling of σ^2 out of the likelihood. An estimator of the variance for $\hat{\boldsymbol{\sigma}}^2{}_{ML}(\boldsymbol{\beta},\boldsymbol{\tau})$, assuming that $\boldsymbol{\beta}$ and $\boldsymbol{\tau}$ are known, is then

Table 3. AIC values for all combinations of factor specific correlation models.

	Initial Caudate Data Model			Final Caudate Data Model		
	Spatial Model			Spatial Model		
Temporal Model	LEAR	DE	AR(1)	LEAR	DE	AR(1)
LEAR	−14,298	−13,903	−12,717	−14,307	−13,912	−12,722
DE	−14,295	−13,900	−12,714	−14,304	−13,909	−12,719
AR(1)	−10,377	− 9,983	− 8,768	−10,386	−9,992	−8,774

doi:10.1371/journal.pone.0088864.t003

$$\widehat{\mathcal{V}}\left[\widehat{\boldsymbol{\sigma}}^{2}{}_{ML}(\boldsymbol{\beta},\boldsymbol{\tau})\right]=2\widehat{\sigma}^{4}/n. \tag{8}$$

The derivation of this estimator is available from the authors. A SAS IML [7] program implementing this estimation procedure for the general linear model with a Kronecker product LEAR correlation structure is also available upon request, and a more general macro is in development.

A possible complication when implementing the Kronecker product LEAR correlation model is that the proposed estimation method can produce negative variance estimates for the correlation parameters. This may occur for the parameters of either one or both of the factor-specific matrices when there is a faster decay rate than that imposed by the AR(1) model coupled with a "small" ρ_{γ} and/or ρ_{ω}. The instability of the second order derivatives of the objective function, resulting from the small, quickly decaying correlation(s), leads to this problem. More specifically, when the log likelihood is shallow, approximately quadratic in the parameter being considered, the second derivatives will be near zero analytically, and therefore indistinguishable from zero numerically, so finite precision arithmetic can lead to negative values with nonzero probability.

An alternate approach is to implement an estimation method which only uses first-order derivatives such as a quasi-Newton procedure. An efficient modification of Powell's [29–32] Variable Metric Constrained WatchDog (VMCWD) algorithm is often used. A quadratic programming subroutine updates and downdates the Cholesky factor, as detailed in [33]. However, quasi-Newton approaches generally have less stability and worse convergence properties than the Newton-Raphson method. Another approach is to recognize this complication as a diagnostic tool. Since a correlation matrix of this nature (one with most off-diagonal elements being close to zero given the small correlation and fast decay rate) is approximately equal to the identity matrix, an independence model may be the best fit for the factor specific structure in this situation.

Results

We model the caudate data with the general linear model for multivariate repeated measures data defined in the previous section (all modeling assumptions were assessed and met as discussed in [34]). The initial full mean model is as follows:

$$\begin{aligned}\boldsymbol{y}_i=&\beta_0+\beta_1\boldsymbol{X}_{i,\mathrm{trt}}+\beta_2\boldsymbol{X}_{i,\mathrm{age}}+\beta_3\boldsymbol{X}_{i,\mathrm{gen}}+\beta_4\boldsymbol{X}_{i,\mathrm{af\_race}}\\&+\beta_5\boldsymbol{X}_{i,\mathrm{oth\_race}}+\boldsymbol{e}_i.\end{aligned} \tag{9}$$

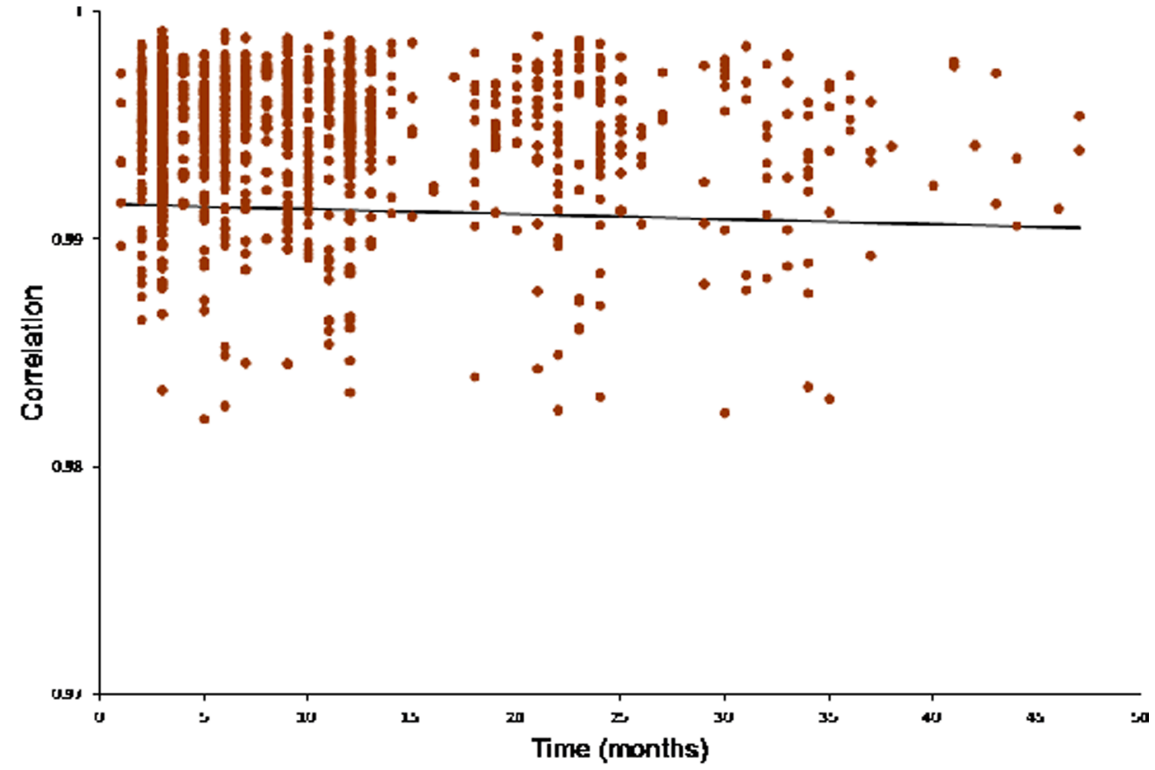
The $\log_2$(radius) values for each of the $S_i=S=21$ locations (spatial factor) and T_i images (temporal factor) for each subject are contained in $\boldsymbol{y}_i$ $(T_i\cdot 21\times 1)$. The vectors $\boldsymbol{X}_{i,\mathrm{trt}}$, $\boldsymbol{X}_{i,\mathrm{gen}}$, $\boldsymbol{X}_{i,\mathrm{af\_race}}$ and $\boldsymbol{X}_{i,\mathrm{oth\_race}}$ indicate the treatment group (patients and controls), gender, and race (African-American, Other, and White–reference group) of the i^{th} subject respectively. The ages at baseline are contained in $\boldsymbol{X}_{i,\mathrm{age}}$.

We first assume a separable covariance and model the temporal and spatial factor-specific correlations of the model errors with continuous-time AR(1), DE, and LEAR structures in order to assess the best model via the AIC $[\mathrm{AIC}=-2l(\boldsymbol{y}_i;\boldsymbol{\beta},\boldsymbol{\Sigma}_i)+2(q+w)$, where q is the number of fixed effect parameters and w is the number of unique covariance parameters]. **Table 3** contains the AIC values for all nine possible correlation model fits with the initial full mean model. Modeling both the temporal and spatial correlations with the LEAR structure provides the best model fit of the nine combinations. The BIC $[\mathrm{BIC}=-2l(\boldsymbol{y}_i;\boldsymbol{\beta},\boldsymbol{\Sigma}_i)+(q+w)\ln(n)$, where n is the total number of observations] corroborates the differences in fits. The

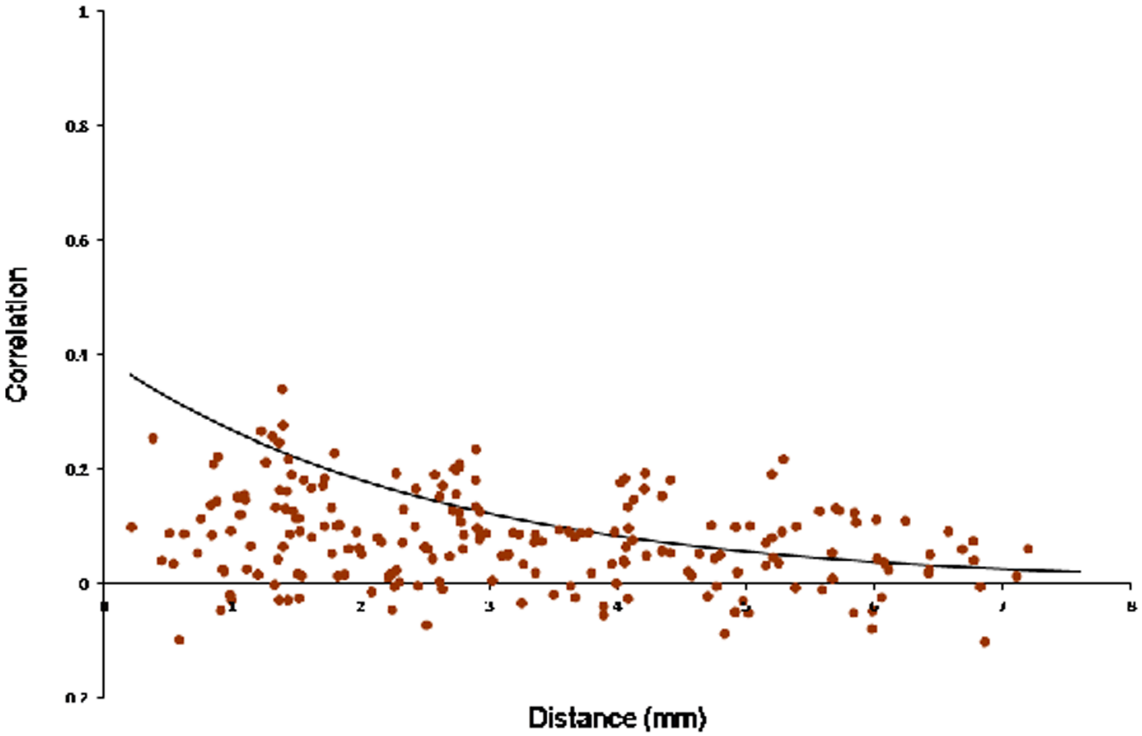
Table 4. Initial full mean model estimates, standard errors, and p-values.

	LEAR⊗LEAR			AR(1)⊗AR(1)			DE⊗DE		
Parameter	Estimate	SE	P-value	Estimate	SE	P-value	Estimate	SE	P-value
β_0	−5.017	0.046	0.000	−4.922	0.030	0.000	−4.985	0.044	0.000
β_1	0.004	0.041	0.931	0.001	0.027	0.969	0.003	0.039	0.946
β_2	−0.002	0.003	0.549	−0.003	0.002	0.240	−0.002	0.003	0.493
β_3	−0.036	0.038	0.346	−0.040	0.025	0.111	−0.038	0.037	0.299
β_4	−0.009	0.034	0.783	−0.013	0.022	0.560	−0.010	0.033	0.748
β_5	0.016	0.054	0.763	0.014	0.035	0.678	0.016	0.051	0.756

doi:10.1371/journal.pone.0088864.t004



(A)



(B)

Figure 4. Observed (dots) vs. predicted (curve) correlation. (**A**) as a function of the time between images; (**B**) as a function of the distance between radius locations.
doi:10.1371/journal.pone.0088864.g004

resulting parameter estimates and p-values (based on the residual approximation of the F-test for a Wald statistic [34]) associated with each of the covariates are presented in **Table 4** for three of the correlation model fits: LEAR⊗LEAR, AR(1)⊗AR(1), and DE⊗DE. Although, for our particular example, there is no difference in fixed effect (mean model) inference among the models; the covariate p-values for the better fitting LEAR⊗LEAR model are uniformly larger for β_2 (age), β_3 (gender), and β_4 and β_5 (race). Thus, the results illuminate the difference in fixed effect inference that could occur if there were a covariate of borderline significance. The better fit of the Kronecker product LEAR structure instills more confidence in the results. Therefore, we continue the analysis using the Kronecker product LEAR correlation model.

In order to obtain a parsimonious model, the full model defined in equation 9 is reduced via backward selection with $\alpha = 0.20$. At each reduction step the covariance parameters are re-estimated and the fixed effect covariate with the largest p-value is removed if it is non-significant at the $\alpha = 0.20$ level based on the residual approximation of the F-test for a Wald statistic. The final model after reduction is

$$\boldsymbol{y}_i = \beta_0 + \boldsymbol{e}_i. \tag{10}$$

There is no evidence of a difference in caudate shape between the treated schizophrenics and the controls when taking into account all images taken over time. To ensure that the Kronecker product LEAR correlation model still provides the better fit for the final model, we again model the temporal and spatial factor-specific correlations of the model errors with the continuous-time AR(1), DE, and LEAR structures. **Table 3** contains the AIC values for the final model fits. The Kronecker product LEAR correlation model remains the better correlation structure for the final data model. The BIC corroborates the differences in fits.

The residual variance estimate and correlation parameter estimates of the Kronecker product LEAR structure (defined in equation 4) for the final data model are given in **Table 5**. Graphical depictions of these estimates are exhibited in **Figures 4A and 4B**, which show the observed vs. predicted correlation patterns as a function of the months between images and millimeters between radii respectively, starting with the minimum temporal and spatial distances for the data. As evidenced by **Figure 4A**, the temporal factor-specific LEAR correlation structure is able to model a correlation function in which the correlation remains high regardless of how far apart in time the images are taken. The fact that the correlation estimates in the time dimension are close to 1.0might be considered a problem if the presence of a unit root is expected. This would also present a problem for the competing DE and AR(1) factor-specific models. While much work has been done on the development of unit root tests for time series data [35–39], to our knowledge, none are applicable to unbalanced, inconsistently-spaced multivariate repeated measures data modeled with a Kronecker product covariance structure. A test for a unit root might be useful, but developing one is beyond the scope of the present work. The spatial correlations, shown in **Figure 4B**, are modest for radii that are close, and decay slowly toward zero as they become farther apart. The predicted correlation curve appears to slightly overestimate the spatial correlations for small distances. This may be due to the restriction $0 \le \rho_\omega < 1$, since the model cannot accurately incorporate the negative correlations. One solution may be to add an offset parameter to the model in order to account for negative correlations, i.e.,

$$\rho_{0_\omega} + \rho_\omega^{d_{s;\min} + \delta_\omega \left[(d(s_{ijl}, s_{ijm}) - d_{s;\min}) / (d_{s;\max} - d_{s;\min}) \right]},$$

under the condition that

$$-1 - \rho_\omega^{d_{s;\min} + \delta_\omega \left[(d(s_{ijl}, s_{ijm}) - d_{s;\min}) / (d_{s;\max} - d_{s;\min}) \right]} < \rho_{0_\omega} < 1 - \rho_\omega^{d_{s;\min} + \delta_\omega \left[(d(s_{ijl}, s_{ijm}) - d_{s;\min}) / (d_{s;\max} - d_{s;\min}) \right]}.$$

An examination of this approach, and others, will be left for future research.

Table 5. Final Kronecker product LEAR structure correlation model estimates for caudate data.

Factor	Parameter	Estimate	SE
–	σ^2	0.405	0.005
Time	ρ_γ	0.992	0.000
	$\delta_\gamma/(D_\gamma - 1)$	0.003	0.001
Space	ρ_ω	0.381	0.011
	$\delta_\omega/(D_\omega - 1)$	0.040	0.004

doi:10.1371/journal.pone.0088864.t005

Discussion

The Kronecker product LEAR correlation model allows modeling and understanding two factor-specific correlation patterns. Excellent analytic and numerical properties make the structure especially attractive for High Dimensional, Low Sample Size settings that are common in longitudinal medical imaging and various kinds of longitudinal "-omics" data. The structure is able to model a wide variety of correlation patterns with just four parameters. Analysis of the caudate data illustrates the interpretability of the model in a complex context.

An assessment of model fit and inference accuracy in higher dimensional settings is a priority for future research on Kronecker product LEAR correlation models. Also, introducing a nonstationary Kronecker product LEAR correlation or variance structure may prove extremely useful in neuroimaging, since the variability of brain characteristics tends to vary spatially and temporally. Comparing the implementation of the Kronecker product LEAR structure with various modeling and estimation methods will prove valuable. For data that have within-subject correlations induced by three or more factors, as in longitudinal imaging data represented via the m-rep method ([40] has details), the generalization of the Kronecker product LEAR correlation model to F repeated factors would be beneficial.

Acknowledgments

An earlier version of this manuscript can be found at arxiv.org (Simpson et al., arXiv: 1010.4471v1 stat.AP). The authors thank the editor and referees for their comments that considerably improved the paper. The authors also thank Amy Scharmann from the University of Florida at Gainesville for her suggested edits that greatly improved the presentation of material.

Author Contributions

Performed the experiments: SS LE MS KM. Analyzed the data: SS LE MS KM. Contributed reagents/materials/analysis tools: SS LE MS KM. Wrote the paper: SS LE KM.

References

1. Muller KE, Edwards LJ, Simpson SL, Taylor DJ (2007) Statistical tests with accurate size and power for balanced linear mixed models. Statistics in Medicine 26: 3639–3660.
2. Gurka MJ, Muller KE, Edwards LJ (2011) Avoiding bias in mixed model inference for fixed effects. Statistics in Medicine 30: 2696–2707.
3. Galecki AT (1994) General class of correlation structures for two or more repeated factors in longitudinal data analysis. Communications In Statistics-Theory and Methods 23: 3105–3119.
4. Naik DN, Rao SS (2001) Analysis of multivariate repeated measures data with a kronecker product structured covariance matrix. Journal of Applied Statistics 28: 91–105.
5. Mitchell MW, Genton MG, Gumpertz ML (2006) A likelihood ratio test for separability of covariances. Journal of Multivariate Analysis 97: 1025–1043.
6. Genton MG (2007) Separable approximations of space-time covariance matrices. 681–695 Environmetrics 18.
7. SAS Institute (2002) SAS/IML, Version 9. SAS Institute, Inc.: Cary, NC.
8. Komro KA, Maldonado-Molina MM, Tobler AL, Bonds JR, Muller KE (2007) Effects of home access and availability of alcohol on young adolescents' alcohol use. Addiction 102: 1597–1608.
9. Lieberman JA, Tollefson GD, Charles C, Zipursky R, Sharma T, et al. (2005) Antipsychotic drug effects on brain morphology in first-episode psychosis. Archives of General Psychiatry 62: 361–370.
10. McClure RK, Styner M, Maltbie E, Lieberman JA, Gouttard S, et al. (2013) Localized differences in caudate and hippocampal shape are associated with schizophrenia but not antipsychotic type. Psychiatry Research: Neuroimaging 211: 1–10.
11. Cressie N, Huang H (1999) Classes of nonseparable, spatio-temporal stationary covariance functions. Journal of the American Statistical Association 94: 1330–1340.
12. Huizenga HM, de Munck JC, Waldorp LJ, Grasman RP (2002) Spatiotemporal EEG/MEG source analysis based on a parametric noise covariance model. IEEE Transactions on Biomedical Engineering 49: 533–539.
13. Louis TA (1988) General methods for analyzing repeated measures. Statistics in Medicine 7: 29–45.
14. Diggle PJ (1988) An approach to the analysis of repeated measures. Biometrics 44: 959–971.
15. Simpson SL, Edwards LJ, Muller KE, Sen PK, Styner MA (2010) A linear exponent AR(1) family of correlation structures. Statistics In Medicine 29: 1825–1838.
16. Styner MA, Gerig G (2001) Three-dimensional medial shape representation incorporating object variability. Computer Vision and Pattern Recognition CVPR: 651–656.
17. Schott JR (1997) Matrix Analysis for Statistics. John Wiley & Sons: New York.
18. Fan J, Huang T, Li R (2007) Analysis of longitudinal data with semiparametric estimation of covariance function. Journal of the American Statistical Association 102: 632–641.
19. Werner K, Jansson M, Stoica P (2008) On estimation of covariance matrices with kronecker product structure. IEEE Transactions on Signal Processing 56: 478–491.
20. Lu N, Zimmerman DL (2005) The likelihood ratio test for a separable covariance matrix. Statistics and Probability Letters 73: 449–457.
21. Roy A, Khattree R (2003) Tests for mean and covariance structures relevant in repeated measures based discriminant analysis. Journal of Applied Statistical Science 12: 91–104.
22. Roy A, Khattree R (2005a) On implementation of a test for kronecker product covariance structure for multivariate repeated measures data. Statistical Methodology 2: 297–306.
23. Roy A, Khattree R (2005b) Testing the hypothesis of a kronecker product covariance matrix in multivariate repeated measures data. Proceedings of the 30th Annual SAS Users Group International Conference (SUGI).
24. Roy A, Leiva R (2008) Likelihood ratio tests for triply multivariate data with structured correlation on spatial repeated measurements. Statistics and Probability Letters 78: 1971–1980.
25. Edwards LJ, Muller KE, Wolfinger RD, Qaqish BF, Schabenberger O (2008) An R statisti2 c for fixed effects in the linear mixed model. Statistics in Medicine 27: 6137–6157.
26. Wang N (2003) Marginal nonparametric kernel regression accounting for within-subject correlation. Biometrika 90: 43–52.
27. Abramowitz M, Stegun IA (1972) Handbook of Mathematical Functions. Dover Publications, Inc.: New York.
28. Dennis JE, Schnabel RB (1983) Numerical Methods for Unconstrained Optimization and Nonlinear Equations. Prentice-Hall: New Jersey.
29. Powell JMD (1978a) A fast algorithm for nonlinearly constrained optimization calculations. Numerical Analysis, Dundee 1977, Lecture Notes in Mathematics 630. G. A. Watson (ed.), Springer-Verlag: Berlin, 144–175.
30. Powell JMD (1978b) Algorithms for nonlinear constraints that use lagrangian functions. Mathematical Programming 14: 224–248.
31. Powell JMD (1982a) Extensions to subroutine VF02AD. Systems Modeling and Optimization, Lecture Notes In Control and Information Sciences 38. R. F. Drenick and F. Kozin (eds.), Springer-Verlag: Berlin, 529–538.
32. Powell JMD (1982b) VMCWD: A fortran subroutine for constrained optimization. DAMTP 1982/NA4. Cambridge, England.
33. Gill EP, Murray W, Saunders MA, Wright MH (1984) Procedures for optimization problems with a mixture of bounds and general linear constraints. ACM Transactions on Mathematical Software 10: 282–298.
34. Muller KE, Stewart PW (2006) Linear Model Theory: Univariate, Multivariate, and Mixed Models. John Wiley & Sons, Inc: New Jersey.
35. Im KS, Pesaran MH, Shin Y (2003) Testing for unit roots in heterogeneous panels. Journal of Econometrics 115: 53–74.
36. Baltagi BH, Bresson G, Pirotte A (2007) Panel unit root tests and spatial dependence. Journal of Applied Econometrics 22: 339–360.
37. Moon HR, Perron B (2012) Beyond panel unit root tests: Using multiple testing to determine the nonstationarity properties of individual series in a panel. Journal of Econometrics 169: 29–33.
38. Lin JH (2013) A Monte Carlo comparison of panel unit root tests under factor structure. Applied economics letters 20: 288.
39. Westerlund J, Larsson R (2012) Testing for a unit root in a random coefficient panel data model. Journal of Econometrics 167: 254–273.
40. Pizer SM, Fletcher T, Thall A, Styner M, Gerig G, et al. (2002) Object models in multiscale intrinsic coordinates via m-reps. Proceedings of Generative Model Based Vision GMBV.
41. Munoz A, Carey V, Schouten JP, Segal M, Rosner B (1992) A parametric family of correlation structures for the analysis of longitudinal data. Biometrics 48: 733–742.
42. Peiris MS (2003) Improving the quality of forecasting using generalized AR models: an application to statistical quality control. Statistical Methods 5: 156–171.
43. Schabenberger O, Gotway CA (2005). Statistical Methods for Spatial Data Analysis. Chapman & Hall: Boca Raton, FL.