Published in final edited form as:

Stat Med. 2015 January 15; 34(1): 145–161. doi:10.1002/sim.6324.

Selecting a Separable Parametric Spatiotemporal Covariance Structure for Longitudinal Imaging Data

Brandon George, MS^{a,*} and Inmaculada Aban, PhD^a

^aDepartment of Biostatistics, University of Alabama at Birmingham

Abstract

Longitudinal imaging studies allow great insight into how the structure and function of a subject's internal anatomy changes over time. Unfortunately, the analysis of longitudinal imaging data is complicated by inherent spatial and temporal correlation: the temporal from the repeated measures, and the spatial from the outcomes of interest being observed at multiple points in a patient's body. We propose the use of a linear model with a separable parametric spatiotemporal error structure for the analysis of repeated imaging data. The model makes use of spatial (exponential, spherical, and Matérn) and temporal (compound symmetric, autoregressive-1, Toeplitz, and unstructured) parametric correlation functions.

A simulation study, inspired by a longitudinal cardiac imaging study on mitral regurgitation patients, compared different information criteria for selecting a particular separable parametric spatiotemporal correlation structure as well as the effects on Type I and II error rates for inference on fixed effects when the specified model is incorrect. Information criteria were found to be highly accurate at choosing between separable parametric spatiotemporal correlation structures. Misspecification of the covariance structure was found to have the ability to inflate the Type I error or have an overly conservative test size, which corresponded to decreased power.

An example with clinical data is given illustrating how the covariance structure procedure can be done in practice, as well as how covariance structure choice can change inferences about fixed effects.

Keywords

spatiotemporal; separable; longitudinal; spatial; imaging

1. Introduction

One of the most common approaches to analyzing correlated data is the use of linear regression models with correlated errors. Through this method the investigator can specify the covariance structure between observations that can be assumed to be related. For example, in longitudinal studies the successive observations on a subject are considered to have temporal correlation, with observations closer in time being more highly correlated

than those further apart. Similarly, in imaging studies the observations from different parts of the same individual are considered to have spatial correlation that follows the same pattern of decreasing correlation with distance. Therefore longitudinal imaging studies need to be analyzed using methods that account for the spatiotemporal correlation in the data. Note that here we refer to longitudinal imaging studies as those that take a handful of successive images days, months, or even years apart; we do not refer to functional imaging studies that take hundreds of images seconds or fractions of a second apart.

In spatial statistics, this relationship between distance and correlation is typically defined by a parametric covariance structure. Commonly used spatial structures include the exponential, spherical, and Matérn structures. For temporal correlation, typically the autoregressive-1 (AR-1) and Toeplitz structures are used to define decreasing correlation with time. The compound symmetric structure can also be used for distance-independent correlation, while the unstructured correlation model allows for the most precise fit to the data. Note that when separability between space and time is assumed, these temporal and spatial structures can be combined in a spatiotemporal model [1]. Even if the correlation between observations is a nuisance and only fixed effect parameters are of interest, use of the proper covariance matrix is essential to prevent the estimates of the standard errors of the parameters from being erroneous [2].

1.1. Selecting a Covariance Structure

In practice, however, the investigator will not know *a priori* which covariance function will best represent the data. A three-step approach, proposed by Diggle [3] and refined by Wolfinger [4], starts by (1) choosing fixed effects for the model, then (2) fitting different covariance structures, and (3) finally choose between covariance structures using either formal testing or examination of the variogram. The variogram is a plot of covariance between observations as a function of distance and is the spatial statistician's preferred option for spatial covariance structure selection [5]. Unfortunately, in spatiotemporal models it can be difficult to 'eyeball' which of several semivariograms would be appropriate since space and time would be on different axes or different graphs. Furthermore, this procedure does not take the complexity of the model into consideration. Most spatial covariance functions will have five or fewer parameters so overfitting is not as much of a concern in that field, but when the unstructured covariance matrix is brought into consideration one must be acutely aware of model parsimony.

Early on, formal testing for choosing covariance structures had been done using likelihood ratio tests (LRT). For example, Schaalje *et al.* [6] looked at an example of how one could choose between common longitudinal correlation structures using the LRT, and Grady and Helms [7] used the LRT to decide between multiple covariance structures and random effects. However, as Grady and Helms noted the LRT is only a valid test for nested models which works decently well in longitudinal structures where everything is nested within the unstructured model and compound symmetry and AR-1 are nested within Toeplitz. Unfortunately, compound symmetry and AR-1 are not nested within one another, making the LRT useless for choosing directly between the two. The problem of non-nested covariance structures is greater in spatial statistics, where beyond exponential within Matérn

there is very little nesting of models. Thus, the likelihood ratio test is not a valid choice for choosing between spatiotemporal covariance structures.

A more helpful tool for choosing a spatiotemporal structure is the use of information criteria (IC). As noted by Wolfinger [4], information criteria can be used instead of the LRT in the third step of Diggle's algorithm [3]. The most commonly used information criteria are AIC, BIC, HQIC, CAIC, and AICC, whose forms and preference for parsimony will be discussed later. In general, these information criteria provide a metric that quantifies how well the model fits the data with a penalty that increases with model complexity. They can be used to choose between covariance structures [2] or between different parameterizations involving random effects [8] with no requirement that the covariance matrices be nested.

Several other methods have also been previously studied for their ability to choose between covariance structures in longitudinal data analysis. Keselman *et al.* [9] investigated a whole battery of statistical tests for model selection, while Wang and Schaalje [10] examined the use of predictive criteria such as the R^2 statistic and predictive error sum of squares (PRESS). Ibrahim *et al.* [11] looked into smooth clipped absolute deviation (SCAD) and the adaptive least absolute shrinkage and selection operator (ALASSO) as model selection tools.

A large amount of work has been done to assess the accuracy of information criteria in the selection of the true longitudinal covariance structure [8, 12, 13, 14, 15, 16, 17]. In addition to comparing the different information criteria against one another, many of these simulation studies have examined how study parameters and peculiarities in the data affect accuracy. Some of these parameters and conditions include sample size, number of follow-up visits, balance between treatment groups, skewness of the outcome, degree of correlation, and what sorts of covariance structures generated the simulated data and which were candidates for selection. Although much has been done with choosing temporal correlation structures for longitudinal studies, there has been little work in choosing spatiotemporal covariance structures like those seen in longitudinal imaging studies. Simpson *et al.* [18] considered information criteria for the selection of correlation functions with an end goal of application to longitudinal imaging data, but their simulation studies on IC accuracy were limited to one dimension of correlation. Therefore we sought to investigate which information criterion is best suited for choosing between spatiotemporal covariance structures and how accurate we can expect it to be.

1.2. Effects of Covariance Structure on Fixed Effect Inference

Since selection methods are not perfect, it is necessary to assess how statistical inference changes when the covariance structure is misspecified. There has been a large amount of research into how different factors and covariance structures affect the Type I error rate and power of tests of fixed effects as well as the bias and standard error of the corresponding parameter estimates [9, 12, 13, 16, 18, 19, 20]. Much like the studies of accuracy, these studies examined the effects of sample size, treatment group balance, skewness, and temporal covariance structure misspecification on inference. Many of these studies overlapped with investigations of information criteria, and reported the Type I error rate and bias observed when either AIC or BIC was used to choose a covariance structure. Again,

these studies have simulated purely longitudinal data and so there remains a need to examine Type I and II error rates in a spatiotemporal model under similar conditions.

1.3. Motivation

Our work here has been motivated by a multi-aim longitudinal cardiac imaging study that looked at the effects of medical therapy on patients with mitral regurgitation (MR) [21]. These patients randomized to treatment or placebo at baseline and were observed every six months for up to two years. At baseline and every follow-up visit, the patients underwent three-dimensional cardiac magnetic resonance (CMR) imaging to observe the geometry and function of the left ventricle. The data from these 3-D images was condensed into 16 data values corresponding to the segments defined by the American Heart Association's model for the left ventricle [22]. Due to the spatial relationship between these segments and the repeated measures over time, a spatiotemporal model is needed to properly model the entirety of the data. Also, since this arm of the study has a small sample size characteristic of imaging studies a separable parametric spatiotemporal model is preferred in lieu of an unstructured covariance matrix. The goal of this arm of the study was to determine if two different treatment strategies affected the remodeling of the left ventricle post-surgery; in other words, we wish to know if the time course post-surgery is different for the different treatment strategies. In terms of a statistical model, we are interested in making inferences about the treatment-by-time fixed effect term. Previously, longitudinal studies of CMR imaging data have focused on global outcomes such as ejection fraction, which removes the concern with spatial correlation, or have taken summary values from the segments and analyzed them separately [23]. In our example we wish to investigate an outcome (end-systolic wall thickness) at all of the segments of the left ventricle simultaneously, since we feel that the time course of the outcome may be different at different locations in the heart. The outcome has also been observed to differ between the base and apex of the left ventricle, so we wish to utilize a linear regression framework to control for and possibly make inferences on the regions of the heart as fixed effects.

Previous methodological work into cardiac imaging has fallen short of a true spatiotemporal model. Bowman and Waller [24] constructed a mixed model that incorporated spatial correlation as well as covariance between related conditions, but the spatial aspect did not use parametric spatial covariance structures and the conditions did not follow that typical longitudinal form of several follow-up visits. Simpson *et al.* [25] applied a separable parametric spatiotemporal model to neuroimaging data, but their work looked at only a limited set of parametric correlation functions and did not report simulation results as to how that choice affects statistical inference. Seals [27] built a spatial model for the left ventricle based on the 16-segment model and cardiac imaging data. That study also examined how accurate AIC and BIC are at choosing the correct covariance structure as well as how the choice of covariance structure influences the inference of fixed effects. Ultimately, the spatial model by Seals provides the foundation for this spatiotemporal model.

The goal of this paper is to propose and examine a spatiotemporal model appropriate for analyzing longitudinal imaging data. We will present a linear model and describe how a separable, parametric covariance structure can be used to capture the inherent spatial and

temporal correlation within such studies. The results of simulations will also be discussed, where the accuracy of various information criteria used to select a spatiotemporal covariance structure were evaluated. The simulations also looked at the effects on Type I and Type II error rates when the fitted covariance structure was different than the one used to generate the data. Finally, we will present an example with real longitudinal imaging data to illustrate how the covariance structure selection can be done and how fitting different covariance structures can affect inferences on fixed effects.

2. Proposed Statistical Model, Covariance Structures, and Information Criteria

2.1. Linear Model with Correlated Errors

We propose the use of a linear model with a spatiotemporal error structure. First, consider a dataset with no missing values where N subjects were imaged at J points in time and each image contains K outcome values. Let us define Y_{ijk} as the observed outcome of subject i at time j and location k such that $i = 1, \dots, N$, $j = 1, \dots, J$, and $k = 1, \dots, K$. Let us also define $\mathbf{Y}_i$ as the $JK \times 1$ vector of observations from subject i and $\mathbf{Y}_{ij\bullet}$ as subject i 's K observed outcomes at time j , so that

$$\mathbf{Y}_i = \mathbf{Y}_{i\bullet\bullet} = \{\mathbf{Y}_{i1\bullet}, \mathbf{Y}_{i2\bullet}, \dots, \mathbf{Y}_{iJ\bullet}\}, \quad \mathbf{Y}_{ij\bullet} = \{Y_{ij1}, Y_{ij2}, \dots, Y_{ijK}\}.$$

Therefore, the dataset contains NJK observations divided up into correlated blocks of size JK . Using the framework described by Jennrich and Schluchter [28], the linear model has the form

$$\mathbf{Y}_i = \mathbf{X}_i \boldsymbol{\beta} + \boldsymbol{\varepsilon}_i \quad (1)$$

where $\mathbf{X}_i = \mathbf{X}_{i\bullet\bullet}$ is a $JK \times p$ matrix containing the values of the fixed effects variables for subject i , $\boldsymbol{\beta}$ are the corresponding p regression parameters to be estimated, and $\boldsymbol{\varepsilon}_i \sim MVN_{JK}(0, \sigma^2 \boldsymbol{\Sigma})$ and is independent and identically distributed for all N subjects, given no missing data. The σ^2 is the variance of the errors and $\boldsymbol{\Sigma}$ is their correlation matrix, which assumes homogeneity of variance across repeated observations. If there are missing observations, one could define the errors to have a correlation matrix $\boldsymbol{\Sigma}_i$ which is a subset of $\boldsymbol{\Sigma}$. In this application, we wish for the correlation matrix $\boldsymbol{\Sigma}$ to follow a separable parametric spatiotemporal structure; the parameters of this matrix are contained in the vector $\boldsymbol{\theta}$ so that we can refer to the correlation matrix as $\boldsymbol{\Sigma}(\boldsymbol{\theta})$.

When $\boldsymbol{\theta}$ and thus $\boldsymbol{\Sigma}$ is known, the fixed effect parameters $\boldsymbol{\beta}$ are easily estimated by solving the normal equations where $\mathbf{V}$ is a block diagonal matrix comprised of N blocks of $\sigma^2 \boldsymbol{\Sigma}$:

$$(\mathbf{X}' \mathbf{V}^{-1} \mathbf{X}) \hat{\boldsymbol{\beta}} = \mathbf{X}' \mathbf{V}^{-1} \mathbf{Y} \quad (2)$$

with the property that the estimator $\hat{\boldsymbol{\beta}}$ is the best linear unbiased estimator of $\boldsymbol{\beta}$ [29]. This estimator comes from maximizing the log-likelihood ℓ of the form [14]

$$\ell(\beta|\mathbf{Y}) = -\frac{NJK}{2}\log(2\pi) - \frac{1}{2}\sum_{i=1}^N \log|\sigma^2\mathbf{\Sigma}| - \frac{1}{2}\sum_{i=1}^N (\mathbf{Y}_{i\bullet\bullet} - \mathbf{X}_{i\bullet\bullet}\beta)' (\sigma^2\mathbf{\Sigma})^{-1} (\mathbf{Y}_{i\bullet\bullet} - \mathbf{X}_{i\bullet\bullet}\beta). \quad (3)$$

In the case where $\mathbf{\Sigma}$ is known, inferences can be made on β using either a Wald's test or, preferably, a likelihood ratio test.

Unfortunately, in practice we do not know the values of θ and must estimate them from the data. Although maximum likelihood estimation is possible, the estimator for θ is biased as it fails to account for the degrees of freedom lost in estimating β [30]. There are also concerns with the likelihood function being multi-modal in maximum likelihood estimation [31]. Therefore, it is more common to see $\{\sigma^2, \theta\}$ estimated with restricted maximum likelihood (REML) which greatly reduces the bias. It does so by projecting the data into a space where all of the fixed effects of $\mathbf{X}\beta$ are removed from the likelihood function. The $\hat{\theta}$ are then found by maximizing the restricted log-likelihood

$$\begin{aligned} REML_1(\sigma^2, \theta|\mathbf{Y}) = & -\frac{NJK-p}{2}\log(2\pi) + \frac{1}{2}\log \left| \sum_{i=1}^N \mathbf{X}'_{i\bullet\bullet} \mathbf{X}_{i\bullet\bullet} \right| - \frac{1}{2}\log \left| \sum_{i=1}^N \mathbf{X}'_{i\bullet\bullet} [\sigma^2\mathbf{\Sigma}(\theta)]^{-1} \mathbf{X}_{i\bullet\bullet} \right| \\ & - \frac{1}{2}\sum_{i=1}^N \log|\sigma^2\mathbf{\Sigma}(\theta)| - \frac{1}{2}\sum_{i=1}^N (\mathbf{Y}_{i\bullet\bullet} - \mathbf{X}_{i\bullet\bullet}\hat{\beta})' [\sigma^2\mathbf{\Sigma}(\theta)]^{-1} (\mathbf{Y}_{i\bullet\bullet} - \mathbf{X}_{i\bullet\bullet}\hat{\beta}) \end{aligned} \quad (4)$$

where the constant term has been modified and the second and third terms have been added when compared to the full likelihood function [14]. Note that in many software implementations the second term is dropped when reporting the maximum restricted likelihood as it is not a function of θ ; this quantity is referred to as $REML_2$ and is defined as

$$REML_2(\sigma^2, \theta|\mathbf{Y}) = REML_1(\sigma^2, \theta|\mathbf{Y}) - \frac{1}{2}\log \left| \sum_{i=1}^N \mathbf{X}'_{i\bullet\bullet} \mathbf{X}_{i\bullet\bullet} \right|. \quad (5)$$

If inferences about θ are of interest, it was noted by Harville [29] that $\hat{\theta}$ follows an asymptotically normal distribution.

The actual REML estimation is more complicated than it may seem since the form of $REML$ contains $\hat{\beta}$ but $\hat{\beta}$ needs $\sigma^2\mathbf{\Sigma}(\theta)$ to be estimated before it can be calculated. Thus, iterative algorithms are needed to estimate all of the parameters in our model. One possibility is the EM algorithm, most famously described by Laird and Ware [30], which has a low computational burden per iteration but may require many iterations. Another common choice is the Newton-Raphson algorithm, described by Jennrich and Schluchter [28] and refined by Lindstrom and Bates [32], which has a greater computational burden per step in calculating matrix derivatives but ultimately requires far fewer steps than the EM algorithm. A third option is the Fisher scoring algorithm, also discussed by Jennrich and Schluchter [28], which provides the base of the model estimation software we utilized [33]. These algorithms address the cyclic nature of estimating $\hat{\beta}$ and $\hat{\theta}$ by switching between the two estimations, using the most recent estimate of one to estimate the other and repeating until

convergence. In this approach, the typical initial step is to estimate $\hat{\beta}$ using least-squares estimation then using that to begin the REML estimation of $\hat{\theta}$.

One limitation of REML estimation is that its estimates are specific to a certain structure of $\mathbf{X}$. If one wished to compare a full model to one with a reduced number of fixed effects, the REML procedure will project the data into two different spaces and the likelihoods cannot be compared. Thus, traditional likelihood ratio tests cannot be used to compare nested fixed effects models when REML is used. In practice inferences about fixed effects under REML estimation have been restricted to a hybrid of ML and REML estimation [4] or, more commonly, Wald's tests. Previous research has suggested that when a Wald's test is used it is highly advisable to use some form of degrees of freedom correction such as Kenward-Roger's adjustment [4, 12, 13, 20]. There have been efforts made by Welham and Thompson [34] to use LRTs in REML estimation by altering the full model's projection to go to the same space as the reduced model, but to our knowledge their method has not been implemented in common statistical software.

2.2. Separable Parametric Covariance Structures

Once an initial set of fixed effects has been chosen, the parametric structure of the correlation matrix must be determined. Ultimately, the only restriction on $\Sigma(\theta)$ is that it must be positive definite, though it is practically always symmetric as well. The main idea behind parametric covariance functions is that 'close' observations are more highly related than 'far' observations. There is also the consideration that two pairs of observations with identical 'distances' apart should have similar if not identical correlations. We will discuss below which parametric covariance models we considered.

When choosing a spatiotemporal model, one must think about how space and time relate in the problem at hand. A simplifying assumption is that the spatial and temporal correlations are independent; in this case one can construct a separable model where the correlation at between two observations is the product of the spatial and temporal correlation functions. Using matrix algebra, one can create a full correlation matrix from two independent sources of correlation by taking the direct/Kronecker product (denoted $\otimes$) of the two correlation matrices [35, 36]. Separable covariance structures for repeated measures imaging data have been used previously by Simpson *et al.* [25], who have also investigated a test for separability [26].

Nonseparable spatiotemporal models exist, and frequently rely on Taylor's hypothesis that spatial and temporal correlation use the same function and are connected through a velocity [1]. This kind of spatiotemporal structure is most common in meteorology, where the difference in weather between points A and B is the same as the difference at point A now and point A at a later time once the front over point B has moved to A. While this form may be useful in medical imaging that examines signal transduction, for imaging studies such as ours that look at anatomical structure the Taylor hypothesis is entirely inappropriate. Thus, we will focus on investigating a separable spatiotemporal model which is reasonable based on the nature of the problem as well as successful previous use of the assumption [25].

We investigated four temporal and three spatial correlation functions, chosen for how they are frequently fitted in practice and appear in most common statistical software. The three spatial structures are exponential (EXP), spherical (SPH), and Matérn (MAT). The temporal structures are compound symmetry (CS), autoregressive-1 (AR-1), Toeplitz (TOEP), and the unstructured model (UN). The details of these correlation functions are given in Table 1. Note that we have assumed the spatial structures are isotropic: the correlation between two points depends only on the magnitude of the distance between them and not the direction/orientation.

2.3. Information Criteria

In analyzing longitudinal imaging data with our model, one would need to choose a particular combination of spatial and temporal correlation structures when fitting the data. Studies on longitudinal methods had seen promising results in the use of information criteria to select temporal correlation structures, and so we wish to extend their use to spatiotemporal models. The most common information criteria (AIC [37], AICC [38], CAIC [39], BIC [40], HQIC [41]) all have the form of the log-likelihood function, ℓ , penalized by the number of parameters. The information criteria that came after the AIC also weight the penalty according to the sample size. However, the choice of likelihood and sample size is not immediately clear when REML is part of the model estimation. Should the maximum likelihood be used for ℓ , or should $REML_1$ or $REML_2$? Should the sample size penalty n^* be based on the total number of observations minus the number of fixed effects be used ($NJK - p$, the dimension of ℓ in REML), or should one instead weight the penalties with the number of independent observational units (N)? Numerous studies have looked for answers to these questions in longitudinal studies [10, 14, 16, 17, 18] but not in a true spatiotemporal framework.

The details of the five information criteria are given in Table 2, presented such that the smaller information criterion is the ‘better’ model [14, 17]. In a broad sense, each information criterion could have six forms: ML with $n^* = NJK$, ML with $n^* = N$, $REML_1$ with $n^* = NJK - p$, $REML_1$ with $n^* = N$, $REML_2$ with $n^* = NJK - p$, and $REML_2$ with $n^* = N$. However, since we ultimately used these information criteria to choose between parametric covariance structures and not between models with different fixed effects, making the X matrices identical between prospective models, $REML_1$ and $REML_2$ would give identical results. The estimation procedure used was purely based on REML and did not support maximum likelihood estimation, so we just concerned ourselves with the full restricted likelihood $REML_1$ for ℓ . Thus, we investigated nine information criteria: AIC; and forms of $n^* = N$ or $NJK - p$ for AICC, BIC, CAIC, and HQIC.

3. Simulation Study

In order to assess the accuracy of these information criteria as well as the effects of covariance structure misspecification on statistical inference, we performed a simulation study which observed those properties under a variety of conditions. Specifically, we estimated the Type I and Type II error rates for testing a treatment-by-time fixed effect when the covariance structure used to fit the model either matched or did not match the covariance structure used to generate the data. Under the same conditions, we estimated how often

different information criteria would correctly select the covariance structure used to generate the data when presented with a variety of possible structures. Furthermore, the effects of sample size and the degree of spatial or temporal correlation were investigated for how they changed IC accuracy or Type I and Type II error rates. In the design and reporting of these simulations, we endeavored to adhere to the guidelines proposed by Burton *et al.* [42] but for the sake of space many details have been reserved for the supplementary materials.

3.1. Simulation Details

The generated data is based on the structure of the data collected in the cardiac imaging study reported on by Schiros *et al.* [21]. The outcome variable is multivariate normal and was collected from two treatment groups at five time points, evenly spaced six months apart ($J=5$). At each observation, cardiac imaging was done where the outcome was observed at 16 points within the left ventricle ($K=16$). Thus, each subject had 80 observed outcomes. The layout of the spatially observed locations was modeled using the AHA's 17 segment model presented by Cerqueria *et al.* [22], where the left ventricle is laid out in a circular pattern with concentric rings corresponding to the different levels of the left ventricle (base, mid, and apex), much as if one is looking down from the left atrium. The coordinates were defined as the center of each segment laid out on a unit circle shown in Figure 1 and are specifically defined in the supplementary materials [27].

The outcome was calculated as the result of a linear model with correlated errors. Specifically,

$$Y_{ijk} = \beta_0 + \beta_1 Time_{ij} + \beta_2 Group_i + \beta_3 Mid_k + \beta_4 Apex_k + \beta_5 Time_{ij} * Group_i + \varepsilon_{ijk} \quad (6)$$

where $Time_{ij}$ was the continuous time of subject i 's j^{th} observation; $Group_i$ was the treatment group for subject i ; Mid_k and $Apex_k$ were indicator variables for whether the ijk^{th} observation was from the mid or apex of the left ventricle, respectively. The collected error terms for subject i were independent and identically distributed for all subjects and were distributed $\varepsilon_{i\bullet\bullet} = [\varepsilon_{i,1,1}, \varepsilon_{i,1,2}, \dots, \varepsilon_{i,1,16}, \varepsilon_{i,2,1}, \dots, \varepsilon_{i,5,16}] \sim MVN_{80}(0, \sigma^2 \Sigma)$, where σ^2 was the variance of the outcome (assumed to be equal across all observations) and Σ is an 80-by-80 separable parametric spatiotemporal correlation matrix. For the sake of simplicity we assumed $\beta_0 = \beta_1 = \beta_2 = \beta_3 = \beta_4 = 1$ and $\sigma^2 = 1$.

The parts of the linear model itself that were changed between conditions include β_5 and Σ . Note that β_5 represents the time-by-treatment interaction and if $\beta_5 = 0$ then there exists a difference in the time course between the two groups, which was the original research hypothesis. $\beta_5 = 0$ allowed for the examination of the Type I error rate, and the other values were chosen to create sigmoidal power curves. The generating Σ was varied between different parametric structures and degrees of freedom. Total sample size was set to either 50 or 100 subjects. The complete listing of simulation conditions is given in Table 3.

For each condition described in Table 3, 5000 independent simulated datasets were run and 12 separable covariance structures corresponding to the combinations of three spatial (exponential, spherical, and Matérn) and four temporal (compound symmetric,

autoregressive-1, Toeplitz, and unstructured) working correlation structures. From each model fit, we looked at the p -value for the Wald's test against the null hypothesis of $H_0 : \beta_5 = 0$ as well as the maximum restricted likelihood. A conditional F-test with a Kenward-Roger adjustment for denominator degrees of freedom was considered but did not alter the conclusions. We calculated the values of the 9 information criteria for each of the 12 covariance models using the restricted log-likelihood, and defined the 'chosen' model for that dataset as the one with the smallest criterion. The selection probabilities for each information criterion for a certain condition was calculated as the proportion of the 5000 datasets where the information criterion chose the covariance structure that had been used to generate the dataset.

The Type I error rates were calculated as the proportion of the 5000 datasets where a model with a certain covariance structure rejected the null hypothesis of $\beta_5 = 0$ at an $\alpha = 0.05$ level when the data was generated with β_5 truly equal to zero. Using the convention suggested by Bradley [43], we considered the Type I error rate to be conserved when it fell between 0.75α and 1.25α , or 0.0375 and 0.625 for our chosen α level. The value of 1.25α was chosen to be moderate, neither highly conservative (1.1α) nor liberal (1.5α) as Bradley described them. Thus, we classified the fit of a certain covariance structure on data generated under a given covariance structure to be overconservative if the empirical Type I error rate $\hat{\alpha}$ was below 0.0375 and inflated if the error rate was above 0.0625.

The Type II error rates (the power curves) were found by checking the proportion of times the Wald's test for $\beta_5 = 0$ was rejected on datasets generated with $\beta_5 > 0$. Note that the combinations of fitted and generating covariance structures that were deemed to have inflated Type I error rates ($\hat{\alpha} > 0.0625$) did not have their power investigated as they were deemed to be unreliable.

3.2. Results

3.2.1. Selection Accuracy of Information Criteria—The five different information criteria (AIC, AICC, BIC, CAIC, and HQIC), along with their different forms for the sample size penalties, were evaluated for their accuracy in selecting the true covariance structure that was used to generate simulated data when presented with a variety of possible structures to choose from; the summarized results are given in Figure 3 and the full results are in the supplementary materials. The information criteria were highly accurate as a whole, with the lowest value for any condition being 68.4% accuracy and the highest being over 99.9%.

In general, the consistent information criteria (BIC, CAIC, HQIC) seemed to have a higher accuracy than the efficient criteria (AIC, AICC). CAIC seemed to be the most accurate information criterion, followed closely by BIC then HQIC, with AICC further behind that. AIC was the least accurate of the information criteria evaluated in terms of covariance structure selection. As for the sample size adjustment, AICC seemed to be more accurate using the number of subjects (N) while the consistent criteria were slightly more accurate using the total number of observations minus the number of fixed effects ($NJK - p$).

When choosing between temporal correlation structures, all information criteria were able to perfectly distinguish between CS and AR-1 structures. Incorrect choices were the result of

picking a more complex temporal function (Toeplitz, unstructured) which was more common among the efficient criteria. The degree of temporal correlation had no noticeable effect on selection accuracy. Conversely, the degree of spatial correlation had a large effect on accuracy of choosing the true spatial structure. In particular, accuracy dropped for all criteria under a high degree of spatial correlation compared to a low degree. The reason for this behavior can be guessed at from Figure 2; the spatial correlation structures were much more distinct and different under the low-correlation condition than the high-correlation condition.

The larger sample size had a subtle effect on selection accuracy. Increasing the total number of subjects from 50 to 100 did not simply result in an across-the-board increase in accuracy. Under low spatial correlation, the larger sample size had a slight improvement for consistent criteria, no noticeable effect on AIC, and an apparent decrease in accuracy for AICC. When the spatial correlation was high, increasing the sample size to 100 seemed to mitigate some of the accuracy loss seen in the $N = 50$ groups, especially for exponential or spherical simulated data.

3.2.2. Effects of Correlation Structure Specification on Type I Error Rates—In addition to examining the accuracy of information criteria, we looked at the empirical Type I error rates for unadjusted Wald's tests under different combinations of generating and fitted covariance structures. The chosen α level for these tests was 0.05. The summarized results of these simulations are given in Figures 4 and 5, with the full results given in supplementary tables. As stated before, we considered the Type I error rate to be inflated if it was greater than 0.0625 (1.25α) and overly conservative if it was less than 0.0375 (0.75α).

We found that when the spatiotemporal covariance structure used to fit the data was the same as the one used to generate the data the Type I error rate is conserved. This result held regardless of the degree of spatial or temporal correlation, whether the sample size was 50 or 100, or what the actual structures were. Specifically, all of the Type I error rates for the conditions with matching generating and fitted covariance structures fell between 0.0448 and 0.0598 with a median of 0.0516. This conservation of the error rate held when more complex correlation functions were fit to special cases such as Matérn fit to exponential data and Toeplitz or unstructured fit to compound symmetric or autoregressive-1 data.

The simulation showed that misspecification of the correlation structure can cause the Type I error rate to be inflated or overly conservative, sometimes extremely so. In cases where the fitted temporal correlation structure was correctly specified, the effects of spatial structure misspecification are shown in Figure 4. When the degree of spatial correlation was low, the mismatch of spherical structures and exponential/Matérn structures resulted in the Type I error not being conserved. However, when the degree of spatial correlation was high there was no problem with the Type I error rate with the mismatch of those structures. We also found that fitting an exponential structure to Matérn data can result in the Type I rate being overly conservative. On the temporal side, when the spatial structure was correctly specified, the results were much more straightforward as seen in Figure 5. Fitting compound symmetric to autoregressive-1 data greatly inflated the Type I error, while fitting

autoregressive-1 to compound symmetric data resulted in a very conservative Type I error rate.

One very important observation from the simulations is that the sample size increase from 50 to 100 did not noticeably change situations where the Type I error rate was not conserved, suggesting that increasing sample size is an ineffective way to conserve the Type I error rate when the spatiotemporal structure is misspecified. A more effective strategy of conserving the Type I error rate seems to be proper selection of the covariance structure. As we have seen, commonly used information criteria are very accurate when it comes to picking the true parametric spatiotemporal covariance structure. To test this approach, we looked at the empirical Type I error rate when the covariance structure chosen by the information criterion was used; the detailed results are given in a supplementary table. We found that when the working covariance structure was chosen by one of the nine information criteria the Type I error rate was very well conserved ([0.0446,0.0600] for all IC and conditions; median=0.0516), even when using the least accurate AIC ([0.0448,0.0600], median=0.0514). This suggests that the strategy of using an information criterion to choose the working correlation structure is one that, on average, will conserve the α -level of a Wald's test of fixed effects.

3.2.3. Effects of Correlation Structure Misspecification on Power—In order to explore the relationship between fitted and actual spatiotemporal correlation structures with regard to their effect on Type II error rates, more simulation studies were run. As with the Type I error rate, the test of interest is an unadjusted Wald's test for a group-by-time interaction (β_5 in Equation 6). An example of the results is given in Figure 6 for data generated under spherical-by-compound symmetric, and the results for other generating structures are shown in the supplementary materials.

The first thing to note is that statistical power seems to be maximized when the true correlation structure is used to fit the data. That is, when the working correlation structure is the same as the one used to generate the data the resulting power is as large or larger than any other working correlation structure we considered. Other working covariance structures were able to match that level of power, though. In particular, the more complex working covariance structures that contained the true covariance structure (Matérn containing exponential, Toeplitz and unstructured containing both autoregressive-1 and compound symmetry) matched the true correlation structure in terms of power despite requiring additional degrees of freedom.

When the correlation structure was misspecified in a pairing that had an overly conservative Type I error rate, the resulting power was lower than that of the true correlation structure. For example, when an autoregressive-1 structure was fit to compound symmetric data the power was greatly reduced, reflecting the extremely low $\hat{\alpha}$ that was observed. Mismatched spatial structures that were overly conservative when spatial correlation was low had similarly lower power. Those pairings whose Type I error rate returned to normal for a high degree of spatial correlation had their power increase to roughly the maximum as well.

We did observe that power was higher for a high degree of temporal correlation compared to a low degree of temporal correlation, and that power seemed to be lower when spatial correlation was high versus low. Increased spatial correlation means that there is less unique information amongst the multiple observations from a subject, which may explain the power loss. As for the gain in power under high temporal correlation, it is probable that this result only holds for inference on parameters related to time such as the treatment-by-time interaction we tested.

4. Illustrating the Use of Covariance Structures in a Longitudinal Cardiac Imaging Study

The data used in this example analysis was first reported in the paper by Ahmed et al. [23], which contains the details of data collection and study design. In short, this study was a randomized controlled phase IIb trial for the use of medical therapy in the treatment of patients with chronic degenerative mitral regurgitation. The intent was that treatment would prevent the adverse left ventricular remodeling typically seen in MR patients. Over time the left ventricle balloons outward in these patients resulting in a more spherical chamber compared to the ‘bullet’ shape seen in healthy patients; a sample MRI scans of the hearts of a healthy subject and a MR patient, both enrolled in the study, are given in Figure 7. In the previous work global or summary measures were taken from the MRI scans, while this analysis makes use of outcomes from all 16 segments. The particular outcome analyzed here is the end-systolic (ES) wall thickness of the left ventricle.

A linear model similar to the one used in the simulation study was fitted to the data, where the treatment-by-time interaction was of interest. In this model, shown below, higher order interaction terms were included up through the three-way time/level/treatment interaction. These terms represent how the treatment would affect the time course of LV remodeling, but it would affect the different levels in different ways. The parentheses represent the terms relating to ventricular level.

$$Y_{ijk} = \beta_0 + \beta_1 Time_{ij} + \beta_2 Group_i + (\beta_3 Mid_k + \beta_4 Apex_k) + \beta_5 Time_{ij} * Group_i + (\beta_6 Time_{ij} * Mid_k + \beta_7 Time_{ij} * Apex_k) + (\beta_8 Group_i * Mid_k + \beta_9 Group_i * Apex_k) + (\beta_{10} Group_i * Time_{ij} * Mid_k + \beta_{11} Group_i * Time_{ij} * Apex_k) + \varepsilon_{ijk} \quad (7)$$

Like before, Y_{ijk} represents the wall thickness of the k^{th} segment of the i^{th} subject at the j^{th} time point and the error terms for subject i were independent and identically distributed for all subjects and were distributed $\varepsilon_{i\bullet\bullet} = [\varepsilon_{i,1,1}, \varepsilon_{i,1,2}, \dots, \varepsilon_{i,1,16}, \varepsilon_{i,2,1}, \dots, \varepsilon_{i,5,16}] \sim MVN_{80}(0, \sigma^2 \Sigma)$, where σ^2 was the variance of the outcome (assumed to be equal across all observations) and Σ is an 80-by-80 separable parametric spatiotemporal correlation matrix. Twelve models were fit where Σ was varied to the structures discussed earlier. Note that only complete cases were used, giving a sample size of 26.

The log-likelihood of the fitted models was then used to calculate the nine information criteria. The BIC and CAIC with $n^* = NJK - p$ chose MAT $\otimes$ CS to be the best covariance structure, while the others all picked EXP $\otimes$ UN. When there is a discrepancy between criteria, it is best to look at how the estimated correlation functions compare to the observed

correlation; the easiest way is to plot the correlation versus distance, as seen in Figure 8. The unstructured model works best for comparing the estimated functions to the ‘true’ correlation, but this approach is limited by the number of correlated outcomes. Here, five temporal observations made fitting an unstructured model simple while the 16 spatial locations meant that a spatial unstructured model could not be estimated. The values in Figure 8 were calculated by taking pairs of segments from every time point and calculating the correlation with an unstructured model in space and time; this strategy was inspired by the method used in Simpson et al. [26].

From the plot we can see that the compound symmetric model fits the temporal correlation well, but there is a bit of heterogeneity that could give the unstructured model the nod when additional parameters are not too harshly penalized. On the other hand, the correlation between the 16 segments is extremely heterogeneous when considered in terms of distance between segments. None of the three functions fit the correlations particularly well, but with such a high amount of heterogeneity it is doubtful any one parametric function could fit the data closely. The exponential model may fit the middle distances better, but Matérn seems to fit the far distances better; the discrepancy between the information criteria in choosing a spatial structure can likely be explained by this.

Since we have chosen a covariance structure, we are now able to draw inferences about the fixed effects in the model. The results of the Wald’s and conditional F-tests for the three-way interaction terms is given in Table 4. The simple Wald’s χ^2 test and the conditional F-tests with the Kenward-Roger adjustment did not appear give noticeably different results. Another thing to note is that the choice of covariance structure did affect the test statistics and p -values, although the effect was minor in this case. However, if one considers the adjusted residual degrees of freedom in the F-test it is clear that choosing a spatiotemporal covariance structure with more parameters can substantially reduce the degrees of freedom available for testing.

We can also see the importance of selecting a proper correlation structure. From the simulation results, it would appear that if the data was truly exponential- or Matérn-by-compound symmetric then a spherical-by-autoregressive-1 model would be entirely inappropriate and be extremely conservative. This result is borne out when such a SPH $\otimes$ AR-1 model is fitted, as the test statistic was much smaller and the p -value much larger than when the chosen structures were fitted.

Since the three-way interaction was not statistically significant at an $\alpha = 0.05$ level, the next logical step is to remove it and test the two-way treatment-by-time interaction in a reduced model. Here we can see how the choice of correlation structures can be dependent on the fixed effects. When the twelve models were fitted and the information criteria compared, all nine criteria chose the exponential-by-unstructured model. This is a different result than the conflicting choices from the full model, suggesting that the chosen structure should be reevaluated when the fixed effects change.

5. Discussion

In this paper, we have presented a spatiotemporal linear model for data correlated through space and time and have considered the relationship separable parametric covariance structure choice has on statistical inference.

We have found that when the chosen covariance structure matches the true structure of the data (or has the true structure as a special case), statistical inference works ‘best’ as expected. Specifically, the Type I error rate is conserved and tests about fixed effects have their power maximized. When the chosen structure is misspecified, however, there is the potential for the Type I error rate to be inflated or too conservative which makes statistical inference unreliable or inefficient, respectively. Furthermore, misspecifications that are overly conservative in Type I error rate were found to also have noticeably reduced power compared to the true structure.

The best determinant of whether a given working covariance structure will have a negative effect on inference is whether or not it can closely approximate the data’s structure. When one tries to use an exponentially decreasing function like AR-1 with a flat function like compound symmetry, the resulting statistical properties are irredeemably bad. Similarly, fitting a mostly linear function like the spherical structure to exponential data generally causes problems with inference; however, if all of the correlations are high (such as in Figure 2) then there is little difference in the functions and statistical inferences are reasonable. In practice it is advisable to plot the estimated correlation functions versus the unstructured values [25], or to consult a semivariogram [5]. This finding is similar to the one by Gurka *et al.* [19] that found that underfitting the correlation often led to the Type I error rate not being conserved.

Overall, it seems that the use of more flexible and complex spatial and temporal correlation structures result in reliably high power regardless of the degree of correlation, assuming that the complex model can fully or at least reasonably approximate the true model. Note that this conclusion may be entirely reliant on the data having a high dimension per subject: despite the low sample size of 50, when one considers the effective degrees of freedom to be at most 3994 ($NJK - p = 50 * 5 * 16 - 6$) then it makes sense that spending a few more degrees of freedom to estimate additional covariance structure parameters would have a negligible effect in terms of power loss.

In addition, we have found that information criteria are highly accurate at choosing between separable spatiotemporal covariance models. Even the simplest AIC was at least 68% accurate. The consistent information criteria (BIC, CAIC, and HQIC) seemed to be more accurate than efficient (AIC, AICC) criteria, with CAIC being the most accurate. The CAIC’s accuracy was generally over 90%, and often went over 99%. Although with REML estimation the sample size adjustment of N worked much better for AICC, HQIC performed better with $NJK - p$ and CAIC/BIC were largely invariant to which of the two were used. As a whole, we would recommend the use of CAIC for choosing a spatiotemporal covariance structure; if it is unavailable, though, BIC should work admirably as well. One possible caveat to this high accuracy was that only one variance was considered in our simulation

study; simulations reported by Gurka [14] suggest that accuracy of information criteria can sharply decrease at higher variances, particularly for efficient criteria. Gurka also found that $REML_2$ was better for efficient criteria and $REML_1$ for consistent criteria for choosing fixed effects, but in our simulations the fixed effects were constant so there was inherently no difference between the two formulations; Additional work would be needed to address the appropriateness of using information criteria to choose fixed effects and covariance structures simultaneously.

In this study we only considered three spatial and four temporal correlation structures for model fitting, and only two temporal for data generation. The actual number of structures one may wish to choose from will likely be much larger in practice, and previous studies suggest that increasing the number of choices reduces the accuracy of information criteria monotonically [12, 13, 17]. We suggest that an analyst should consider the goodness-of-fit of spatial and temporal correlation structures when making a decision about which model to use for drawing inferences, and that graphical checks can be very useful as a supplement to information criteria. It should also be remembered that choosing a covariance structure should be redone if the predictors of the model change as the ‘best’ correlation function may change as well. This was observed in the above applied example, and is a direct consequence of the statistical theory; changing the fixed effects results in the restricted likelihood being projected into a different space which directly alters the estimation of the covariance parameters. The amount which correlation structure selection and overall goodness-of-fit was not quantified in our simulation study, but would be useful to the field and a logical next step. A practical approach would be to employ backwards selection from a model with all potential predictors, choosing a new covariance model at each step back.

One of the most important assumptions in our model was that the temporal and spatial correlation were separable. It is an extremely useful assumption that we feel is reasonable, but still one that must be checked. Simpson et al. [26] recently proposed a likelihood ratio test for whether a separable parametric structure should be rejected when compared to a fully unstructured covariance matrix. Such an unstructured matrix was estimated by looking at subsets of temporal and spatial observations to fill out the area around the main diagonal, with the assertion that correlation quickly decays to zero with distant observations. Neither spatial nor temporal correlation quickly decayed with distance in our example, which agrees with Simpson’s suggestion that a more general yet still computationally efficient method is needed. It should be noted that they did report that for their application in caudate morphology the separability assumption was found to be invalid, so one cannot simply state that separable structures are always valid for longitudinal imaging data as there may be an interaction between space and time. Unfortunately, as mentioned before the vast majority of structured nonseparable correlation models rely on Taylor’s hypothesis, which does not seem valid for such an application. Cressie and Huang [45] developed a framework for a wide variety of nonseparable parametric spatiotemporal functions, but work has not been done as to the appropriateness of those for longitudinal imaging studies. Work also remains to be done to examine the impact of violating the separability assumption in our model as it pertains to inference.

There are other considerations that fell outside the scope of this study that merit further investigation. For drawing statistical inferences, we looked at an unadjusted Wald's test as well as an F-test with a Kenward-Roger adjustment to the denominator degrees of freedom; one could argue for the use of an additional Kenward-Roger adjustment to the covariance matrix [4, 12, 13, 20] or of the REML-friendly likelihood ratio test [34] which would necessitate similar inquiries as to how covariance structure specification effects statistical inference. Another possible direction would be to consider the sensitivity of our model to the assumption of multivariate normality, and to look at how it behaves when faced with skewness or kurtosis in the outcome.

Lastly, it is worth considering how our model and an information criteria-based covariance structure selection approach handle some challenges seen in the analysis of real studies, such as unbalanced designs or missing data. These are issues that may encourage the use of summary methods which have seen much discussion and research [46, 47, 48, 49]. Therefore a direct comparison of such summary methods with our model is currently planned.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgments

Contract/grant sponsor: National Heart, Lung, and Blood Institute

We wish to thank Drs. Louis Dell'Italia, Tom Denney, Jr., Himanshu Gupta, and Chun Schiros of the SCCOR study for their support and for the cardiac MRI data and images they provided. Predoctoral funding was provided by NHLBI T32HL079888. The SCCOR study was supported by National Institutes of Health Specialized Center of Clinically Oriented Research in Cardiac Dysfunction P50-HL077100.

References

1. Gelfand, AE.; Diggle, PJ.; Guttorp, P.; Furntes, M. Handbook of Spatial Statistics. CRC Press; Boca Raton: 2010.
2. Littell RC, Pendergast J, Natarajan R. Modelling covariance structure in the analysis of repeated measures data. *Statistics in Medicine*. 2000; 19:1793–1819. [PubMed: 10861779]
3. Diggle PJ. An approach to the analysis of repeated measurements. *Biometrics*. 1988; 44(4):959–971. [PubMed: 3233259]
4. Wolfinger RD. Covariance Structure Selection in General Mixed Models. *Communications in Statistics - Simulation and Computation*. 1993; 22(4):1079–1106.
5. Diggle PJ, Tawn JA, Moyeed RA. Model-based geostatistics. *Journal of the Royal Statistical Society*. 1998; 47(3):299–350.
6. Schaalje B, Zhang J, Pantula SG, Pollock KH. Analysis of Repeated-Measurements Data from Randomized Block Experiments. *Biometrics*. 1991; 47(3):813–824.
7. Grady JA, Helms RW. Model Selection Techniques for the Covariance Matrix for Incomplete Longitudinal Data. *Statistics in Medicine*. 1995; 14:1397–1416. [PubMed: 7481180]
8. Liu S, Rovine MJ, Molenaar P. Selecting a linear mixed model for longitudinal data: repeated measures analysis of variance, covariance pattern model, and growth curve approaches. *Psychological Methods*. 2012; 17(1):15–30.10.1037/a0026971 [PubMed: 22251268]

9. Keselman HJ, Algina J, Kowalchuk RK, Wolfinger RD. A comparison of recent approaches to the analysis of repeated measurements. *British Journal of Mathematical and Statistical Psychology*. 1999; 52(1):63–78.10.1348/000711099158964
10. Wang J, Schaallje B. Model Selection for Linear Mixed Models Using Predictive Criteria. *Communications in Statistics - Simulation and Computation*. 2009; 38(4):788–801.10.1080/03610910802645362
11. Ibrahim JG, Zhu H, Garcia RI, Guo R. Fixed and random effects selection in mixed effects models. *Biometrics*. 2011; 67(2):495–503.10.1111/j.1541-0420.2010.01463.x [PubMed: 20662831]
12. Gomez EV, Schaallje GB, Fellingham GW. Performance of the Kenward-Roger Method when the Covariance Structure is Selected Using AIC and BIC. *Communications in Statistics - Simulation and Computation*. 2005; 34(2):377–392.10.1081/SAC-200055719
13. Guerin, L.; Stroup, WW. A Simulation Study to Evaluate PROC MIXED Analysis of Repeated Measures Data. *Proceedings of the 2000 Conference on Applied Statistics in Agriculture*; Kansas State University; 2000. p. 170-203.
14. Gurka MJ. Selecting the Best Linear Mixed Model Under REML. *The American Statistician*. 2006; 60(1):19–26.10.1198/000313006X90396
15. Keselman HJ, Algina J, Kowalchuk RK, Wolfinger RD. A Comparison of Two Approaches For Selecting Covariance Structures in The Analysis of Repeated Measurements. *Communications in Statistics - Simulation and Computation*. 1998; 27(3):591–604.
16. Vallejo G, Ato M, Valdés T. Consequences of Misspecifying the Error Covariance Structure in Linear Mixed Models for Longitudinal Data. *Methodology: European Journal of Research Methods for the Behavioral and Social Sciences*. 2008; 4(1):10–21.10.1027/1016-9040.12.1.10
17. Vallejo G, Fernández MP, Livacic-Rojas PE, Tuero-Herrero E. Selecting the best unbalanced repeated measures model. *Behavior Research Methods*. 2011; 43(1):18–36.10.3758/s13428-010-0040-1 [PubMed: 21287107]
18. Simpson SL, Edwards LJ, Muller KE, Sen PK, Styner MA. A linear exponent AR(1) family of correlation structures. *Statistics in Medicine*. 2010; 29:1825–1838. [PubMed: 20658550]
19. Gurka MJ, Edwards LJ, Muller KE. Avoiding bias in mixed model inference for fixed effects. *Statistics in Medicine*. 2011; 30(22):2696–2707.10.1002/sim.4293 [PubMed: 21751227]
20. Schaallje B, McBride JB, Fellingham GW. Adequacy of approximations to distributions of test statistics in complex mixed linear models. *Journal of Agricultural, Biological, and Environmental Statistics*. 2002; 7(4):512–524.
21. Schiros CG, Dell'Italia LJ, Gladden JD, Clark D, Aban I, Ahmed MI. Magnetic resonance imaging with 3-dimensional analysis of left ventricular remodeling in isolated mitral regurgitation: implications beyond dimensions. *Circulation*. 2012; 125(19):2334–2342.10.1161/CIRCULATIONAHA.111.073239 [PubMed: 22496130]
22. Cerqueria MD, Weissman NJ, Dilsizian V, Jacobs AK, Kaul S, Laskey WK, Verani MS. Standardized Myocardial Segmentation and Nomenclature for Tomographic Imaging of the Heart: A Statement for Healthcare Professionals From the Cardiac Imaging Committee of the Council on Clinical Cardiology of the American Heart Association. *Circulation*. 2002; 105:539–542.10.1161/hc0402.102975 [PubMed: 11815441]
23. Ahmed MI, Aban I, Lloyd SG, Gupta H, Howard G, Inusah S, Dell'Italia LJ. A Randomized Controlled Phase IIb Trial of Beta-1-Receptor Blockade for Chronic Degenerative Mitral Regurgitation. *Journal of the American College of Cardiology*. 2012; 60(9):833–838.10.1016/j.jacc.2012.04.029 [PubMed: 22818065]
24. Bowman FD, Waller LA. Modeling of cardiac imaging data with spatial correlation. *Statistics in Medicine*. 2004; 23(6):965–985.10.1002/sim.1741 [PubMed: 15027083]
25. Simpson SL, Edwards LJ, Muller KE, Styner MA. Kronecker Product Exponent AR(1) Correlation Structures for Multivariate Repeated Measures. *PLoS ONE*. 2013; 9:e88864. [PubMed: 24586419]
26. Simpson SL, Edwards LJ, Styner MA, Muller KE. Separability tests for high-dimensional, low-sample size multivariate repeated measures data. *Journal of Applied Statistics*. 2014.10.1080/02664763.2014.919251
27. Seals, S. Spatial analysis of cardiovascular MRI data (dissertation). Birmingham, AL: University of Alabama at Birmingham; 2013.

28. Jennrich RI, Schluchter MD. Models with Unbalanced Structured Covariance Matrices. *Biometrics*. 1986; 42(4):805–520. [PubMed: 3814725]
29. Harville DA. Maximum Likelihood Approaches to Variance Component Estimation and to Related Problems. *Journal of the American Statistical Association*. 1977; 72(358):320–338.
30. Laird NM, Ware JH. Random-effects models for longitudinal data. *Biometrics*. 1982; 38(4):963–974. [PubMed: 7168798]
31. Warnes JJ, Ripley BD. Problems with likelihood estimation of covariance functions of spatial Gaussian processes. *Biometrika*. 1987; 74(3):640–642.
32. Lindstrom MJ, Bates DM. Newton-Raphson and EM Algorithms for Linear Mixed-Effects Models for Repeated-Measures Data. *Journal of the American Statistical Association*. 1988; 83(404):1014–1022.
33. Butler, D.; Cullis, BR.; Gilmour, AR.; Gogel, BJ. [Accessed 4/11.2014] ASReml-R reference manual (Release 2.00, 2007) [Software]. Available from <http://www.vsni.co.uk/software/asrem>
34. Welham SJ, Thompson R. Likelihood Ratio Tests for Fixed Model Terms using Residual Maximum Likelihood. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*. 1997; 59(3):701–714.10.1111/1467-9868.00092
35. Galecki AT. General class of covariance structures for two or more repeated factors in longitudinal data analysis. *Communications in Statistics - Theory and Methods*. 1994; 23(11):3105–3119.
36. Parks RW. Efficient Estimation of a System of Regression Equations when Disturbances are Both Serially and Contemporaneously Correlated. *Journal of the American Statistical Association*. 1967; 62(318):500–509.
37. Akaike H. A New Look at the Statistical Model Identification. *IEEE Transactions on Automatic Control*. 1974; 19(6):716–723.
38. Hurvich CM, Tsai C. Regression and time series model selection in small samples. *Biometrika*. 1989; 76(2):297–307.
39. Bozdogan H. Model Selection and Akaike's Information Criterion (AIC): The General Theory and Its Analytical Extensions. *Psychometrika*. 1987; 52(3):345–370.
40. Schwarz G. Estimating the Dimension of a Model. *Annals of Statistics*. 1978; 6(2):461–464.
41. Hannan EJ, Quinn BG. The Determination of the Order of an Autoregression. *Journal of the Royal Statistical Society, Series B*. 1979; 41(2):190–195.
42. Burton A, Altman DG, Royston P, Holder RL. The design of simulation studies in medical statistics. *Statistics in Medicine*. 2006; 25(24):4279–4292.10.1002/sim.2673 [PubMed: 16947139]
43. Bradley JV. Robustness? *British Journal of Mathematical and Statistical Psychology*. 1978; 31:144–152.
44. Waller, LA.; Gotway, CA. *Applied Spatial Statistics for Public Health Data*. Wiley; Hoboken: 2004.
45. Cressie N, Huang H. Classes of nonseparable, spatio-temporal stationary covariance functions. *Journal of the American Statistical Association*. 1999; 94:1330–1340.
46. Albert PS. Longitudinal Data Analysis (Repeated Measures) in Clinical Trials. *Statistics in Medicine*. 1999; 18:1707–1732. [PubMed: 10407239]
47. Fitzmaurice, G.; Davidian, M.; Verbeke, G.; Molenberghs, G. *Longitudinal Data Analysis*. CRC Press; Boca Raton: 2008.
48. Gibbons R, Hedeker D, Waternaux C, Davis JM. Random Regression Models: A Comprehensive Approach to the Analysis of Longitudinal Psychiatric Data. *Psychopharmacology Bulletin*. 1988; 24(3):438–443. [PubMed: 3153505]
49. Zucker DM, Manor O, Gubman Y. Power comparison of summary measure, mixed model, and survival analysis methods for analysis of repeated-measures trials. *Journal of Biopharmaceutical Statistics*. 2012; 22(3):519–534.10.1080/10543406.2010.550702 [PubMed: 22416838]

Figure 1.

Plot of the 16 segments of the left ventricle. The outer ring corresponds to the base, the middle ring to the mid, and the inner circle to the apex [22, 27]. The numbers correspond to the segment's index as defined in Table A.1.

Figure 2.

Plots of the covariance functions used to generate the spatiotemporal data. The spatial structures (exponential, spherical, and Matérn) are on the left and the temporal structures (compound symmetric and autoregressive-1) are on the right. The functions used to generate data with a low degree of correlation are on the top, and those generating a high degree are on the bottom.

Figure 3. Box plots of the accuracies of the nine information criteria for selecting the generating correlation structure from 12 working structures. The values plotted are aggregated from 48 simulation conditions.

Figure 4.

Plot of the empirical Type I error rates for all combinations of generating and working spatial and the correct temporal correlation structures. The endpoints denote the minimum and maximum observed rates when the temporal and spatial structures, respectively, were correctly specified. Magenta circles denote a low degree of spatial correlation and green squares a high degree. Smaller markers denote a sample size of N=50, while the larger ones denote N=100.

Figure 5.

Plot of the empirical Type I error rates for all combinations of generating and working temporal and correct spatial correlation structures. The endpoints denote the minimum and maximum observed rates when the temporal and spatial structures, respectively, were correctly specified. Magenta circles denote a low degree of temporal correlation and green squares a high degree. Smaller markers denote a sample size of $N=50$, while the larger ones denote $N=100$.

Figure 6.

Plot of power curves for a generating covariance structure of $\text{SPH} \otimes \text{CS}$ under high and low degrees of correlation in space and time. The working covariance structure corresponding to the true structure is denoted with a dot and solid black line. Note that working covariance structures with approximately equivalent power are grouped in the legend and share a common line style and color.

Figure 7. Representative end-systolic (ES) cardiac MRI scan with the four chamber view from a control subject and a MR patient. Note how remodeling has caused the MR patient to have a more spherical left ventricle (circled) compared with the healthy subject.

Figure 8.

Plot of the estimated correlation functions for the full model of the end-systolic wall thickness in the SCCOR data. On the left are the spatial functions when the temporal correlation is unstructured, while on the right are the temporal structures for a Matérn spatial structure. The unstructured spatial correlations were estimated from subset pairs of the 16 segments.

Table 1

Table of the spatial and temporal correlation functions examined in this paper. Temporally, consider t_a to be the discrete time of the a^{th} observation in time, with $a = 1, \dots, J$, and $Y(t_a)$ be the observed outcome at that time point. Let us assume that all observations are evenly spaced, so that $t_a - t_{a-1} = t_b - t_{b-1}$ for any $a, b \geq 2$. Spatially, consider d_{ab} as the Euclidian spatial distance between observations Y_a and Y_b . [2, 44]

Type	Correlation Fcn.	$Corr(Y_a, Y_b)$	Parameter Space	$Length(\theta)$
Spatial	Exponential	$\rho^{d_{ab}}$	$\rho \in (0, 1)$	1
	Spherical	$1 - \frac{3}{2} \left(\frac{d_{ab}}{\phi} \right) + \frac{1}{2} \left(\frac{d_{ab}}{\phi} \right)^3$	$\phi \in (0, \max(d_{ab}))$	1
	Matérn	$\frac{1}{2^{\nu-1}\Gamma(\nu)} \left(\frac{d_{ab}}{\phi} \right)^{\nu} K_{\nu} \left(\frac{d_{ab}}{\phi} \right)$	$\phi > 0, \nu > 0$	2
Temporal	Compound Symmetric	ρ	$\rho \in (0, 1)$	1
	Autoregressive-1	$\rho^{ t_b - t_a }$	$\rho \in (0, 1)$	1
	Toeplitz	$\rho_c, c = a - b $	$\rho_c \in (0, 1)$	$J - 1$
	Unstructured	ρ_{ab}	$\rho_{ab} \in (0, 1)$	$\frac{J(J-1)}{2}$

Table 2

Table of the function forms of the different information criteria (IC) under REML, presented so that ‘smaller is better.’ The ℓ is the log-likelihood, s is the number of parameters in the covariance structure, and n^* is the sample size penalty of either the number of subjects (N) or the dimension of the log-likelihood in REML ($NJK - p$, number of observations minus the number of fixed effects).

IC	Source	Form with $n^* = N$	Form with $n^* = NJK - p$
AIC	Akaike (1974)	$-2\ell + 2s$	$-2\ell + 2s$
BIC	Schwarz (1978)	$-2\ell + s \log(N)$	$-2\ell + s \log(NJK - p)$
HQIC	Hannan & Quinn (1979)	$-2\ell + 2s \log(\log(N))$	$-2\ell + 2s \log(\log(NJK - p))$
CAIC	Bozdogan (1987)	$-2\ell + s(\log(N) + 1)$	$-2\ell + s(\log(NJK - p) + 1)$
AICC	Hurvich & Tsai (1989)	$-2\ell + 2s \left(\frac{N}{N-s-1} \right)$	$-2\ell + 2s \left(\frac{NJK-p}{NJK-s-1} \right)$

Table 3

Total number of conditions for the simulation study which included varying the sample size, temporal and spatial structures used to generate the data, the degree of spatial and temporal correlation used in data generation, and the values of β_S used.

Sample Size	Σ_T Structures	Degree of Σ_T	Σ_S Structures	Degree of Σ_S	Values of β_S	Total Number of Conditions
N=50	2 (CS,AR-I)	2 (High, Low)	3 (Exp.Sph,Mat)	2 (High, Low)	$\beta_S = \{0, 0.05, 0.10, 0.15, 0.20\}$	120
N=100	2 (CS,AR-I)	2 (High, Low)	3 (Exp.Sph,Mat)	2 (High, Low)	$\beta_S = \{0\}$	24
						144

Table of the statistical inferences for the treatment-by-time-by-level and treatment-by-time interactions for end-systolic wall thickness as the outcome.

Table 4

Model and Variable Tested	Chosen $\Sigma_s \otimes \Sigma_T$	IC Choosing Σ	Wald χ^2 Test		Conditional F-Test	
			Test stat.	p-value	Test statistic	p-value
Full Model, Time*Treatment*Level	MAT $\otimes$ CS	BIC, CAIC	$\chi^2_2=1.66$	0.4370	$F_{2,1113.0}=0.828$	0.4373
	EXP $\otimes$ UN	All others	$\chi^2_2=1.60$	0.4593	$F_{2,413.2}=0.778$	0.4599
	MAT $\otimes$ UN	None	$\chi^2_2=1.64$	0.4397	$F_{2,366.4}=0.822$	0.4406
	SPH $\otimes$ AR-1	None	$\chi^2_2=0.70$	0.7057	$F_{2,1231.4}=0.348$	0.7059
Reduced Model, Time*Treatment	EXP $\otimes$ UN	All IC	$\chi^2_1=0.40$	0.5364	$F_{1,133.1}=0.382$	0.5375
	MAT $\otimes$ UN	None	$\chi^2_1=0.34$	0.5593	$F_{1,111.1}=0.341$	0.5605