Linear Models with Generalized AR(1) Covariance Structure for Longitudinal and Spatial Data

Sean L. Simpson, Lloyd J. Edwards, Keith E. Muller, Pranab K. Sen,
Martin A. Styner

Outline

I. Motivation

II. GAR Covariance Model

III. Data Analysis Example

IV. Kronecker Product GAR Covariance Model

V. Conclusions

VI. Future Research

I. Motivation

- Accurately model covariance in cross-sectional and longitudinal imaging studies.
  - Accurate fixed effect inference heavily dependent on proper covariance model specification (Muller et al. (2007)).

- Many repeated measures settings have within-subject correlation decreasing exponentially in time or space.
  - AR(1) model most commonly used. Observed correlations often decay at a slower or faster rate than that imposed by model.

- Generalized Autoregressive (GAR) covariance model accommodates these decay patterns with just 3 parameters (2 for correlation structure)
  - Very attractive for HDLSS situations.

II. GAR Covariance Model

Consider the following general linear model for repeated measures data with the GAR covariance structure:

$$\boldsymbol{y}_i = \boldsymbol{X}_i \boldsymbol{\beta} + \boldsymbol{e}_i \tag{1}$$

where

- $\boldsymbol{y}_i$ is a $p_i \times 1$ vector of p_i observations on the i^{th} subject $i = 1, ..., N$,
- $\boldsymbol{\beta}$ is a $q \times 1$ vector of fixed and unknown population parameters,
- $\boldsymbol{X}_i$ is a $p_i \times q$ fixed and known design matrix corresponding to the fixed effects, $\boldsymbol{\beta}$,
- $\boldsymbol{e}_i$ is a $p_i \times 1$ vector of random error terms.

II. GAR Covariance Model

The following assumptions are made:

$\boldsymbol{e}_i \sim N_{p_i}(\mathbf{0}, \boldsymbol{\Sigma}_{ei})$ and independent for $i \neq i'$

$\Rightarrow \boldsymbol{y}_i \sim N_{p_i}(\boldsymbol{X}_i\boldsymbol{\beta}, \boldsymbol{\Sigma}_{ei})$ and independent for $i \neq i'$,

$$\boldsymbol{\Sigma}_{ei} = \{\sigma_{ei;jk}\}$$

with

$$\sigma_{ei;jk} = \mathcal{V}(y_{ij}, y_{ik}) = \sigma_e^2 \begin{cases} \rho_e^{1+[(d(t_{ij},t_{ik})-1)\delta_e/(D-1)]} & j \neq k \\ 1 & j = k \end{cases},$$

$0 \leq \rho_e < 1, 0 \leq \delta_e, D > 1,$ and $\det(\boldsymbol{\Sigma}_{ei}) > 0$.

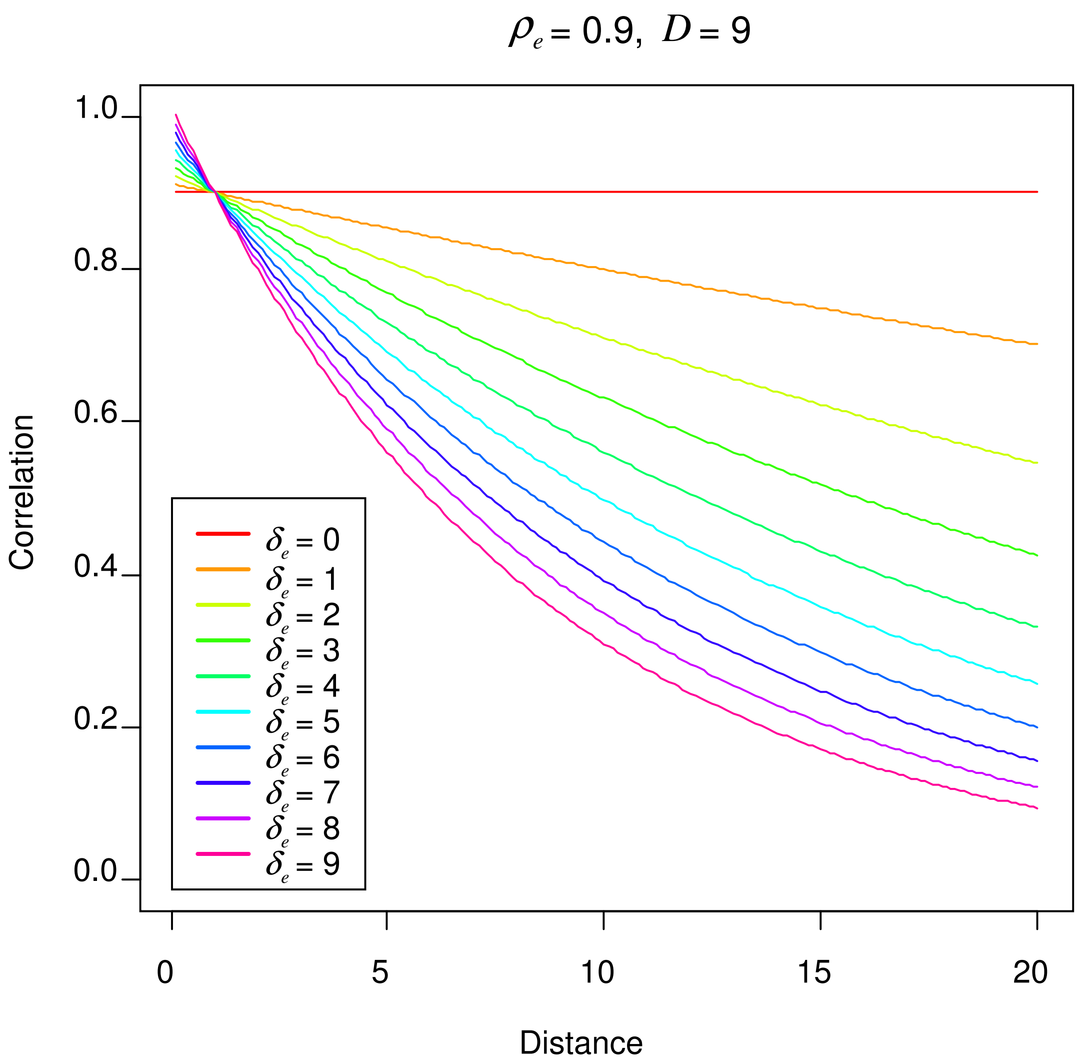
II. GAR Covariance Model

where

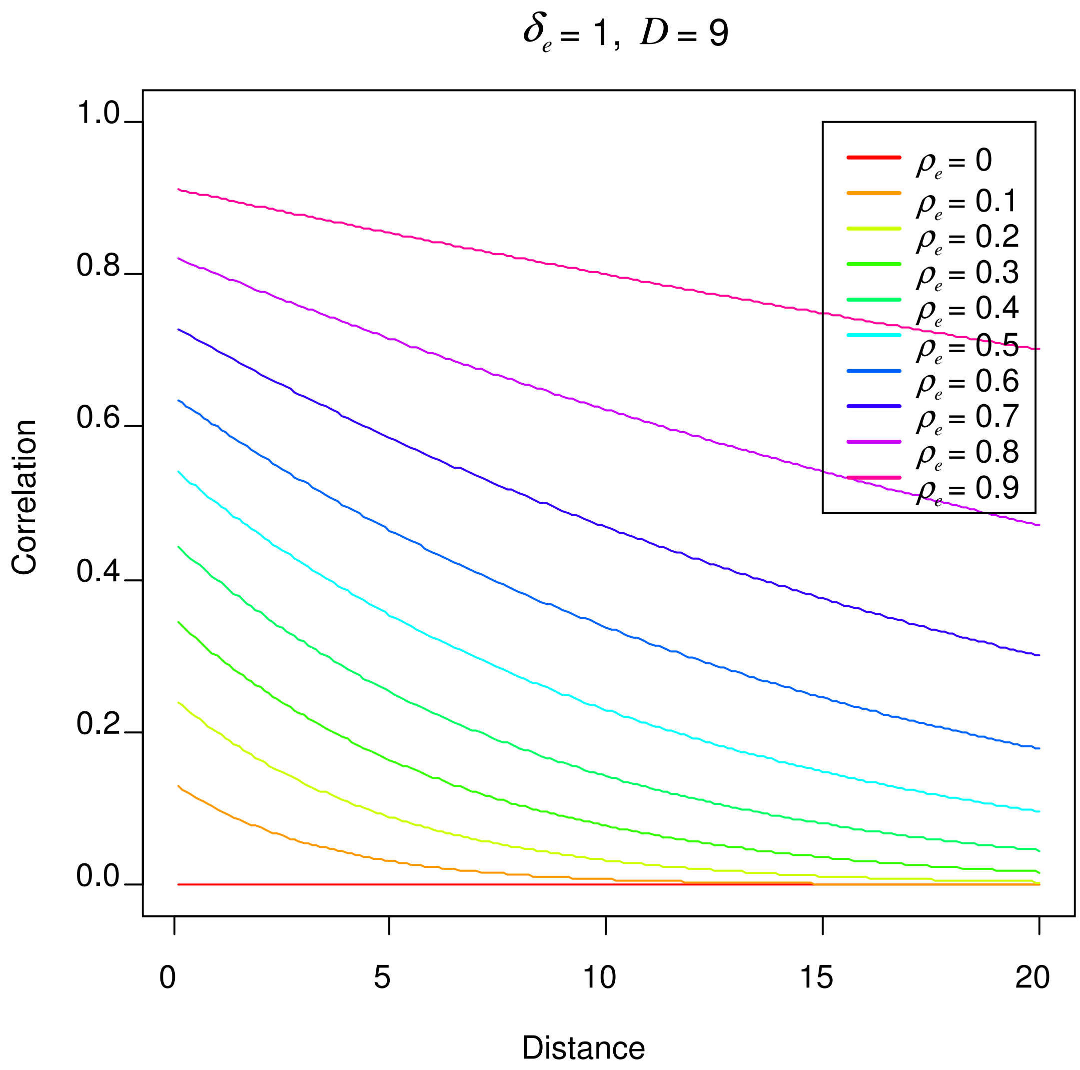
$d(t_{ij}, t_{ik})$	is the distance between measurement times or locations,
D	is a constant that can be specified, by default it is set to the maximum number of distance units,
σ_e^2	is the variability of the measurements at each time or location,
ρ_e	is the correlation between observations separated by one unit of time or distance,
δ_e	is the decay speed.

$\delta_e = 0$:	compound symmetric covariance model
$0 < \delta_e < D - 1$:	within-subject decay slower than AR(1)
$\delta_e = D - 1$:	AR(1) covariance model
$\delta_e > D - 1$:	within-subject decay faster than AR(1)
$\delta_e \to \infty$:	approaches MA(1) covariance model

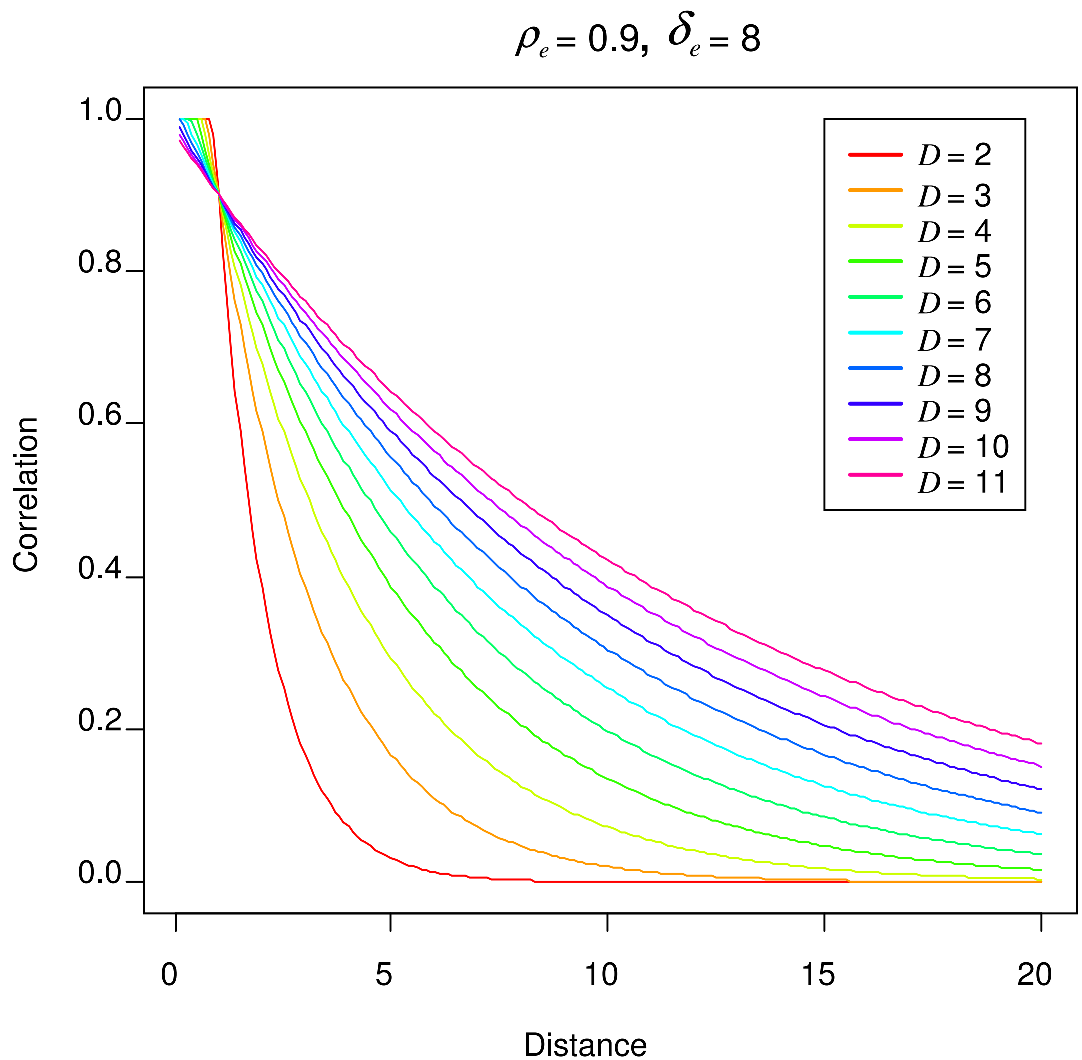
II. GAR Covariance Model



II. GAR Covariance Model



II. GAR Covariance Model



II. GAR Covariance Model

Estimation Steps

1) σ_e^2 is first profiled out of the log-likelihood.
   - Better convergence properties.

2) 1^{st} and 2^{nd} partials derived to compute estimates via Newton-Raphson.

3) Resulting estimates then used to compute the value and variance of $\widehat{\sigma}_e^2$.

III. Data Analysis Example

- DTI scans of fibers of the cortico-spinal tracts associated with motor functioning for 46 control neonates.

- FA values at 20 locations, 3 millimeters apart, along fiber tracts.

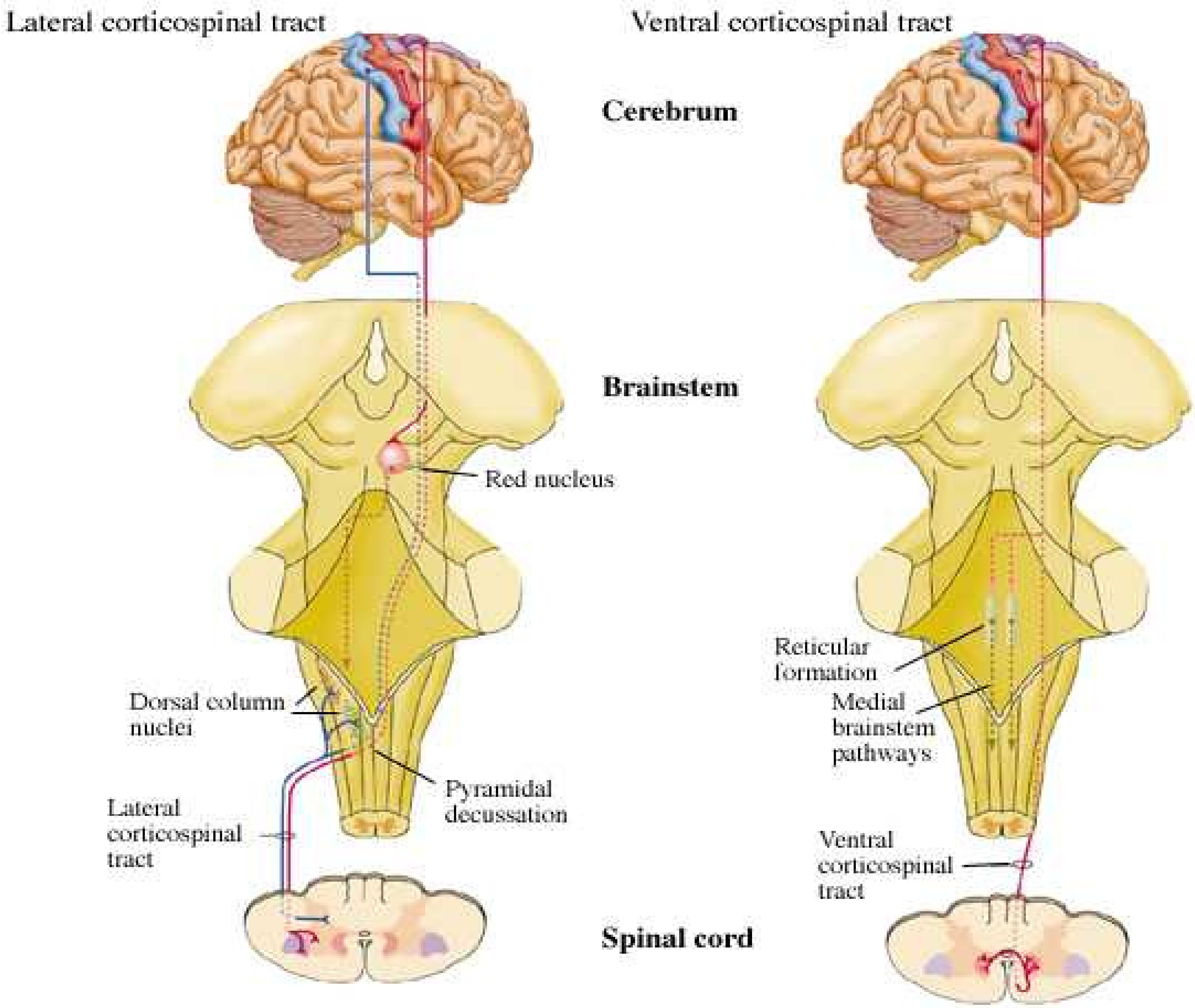
  - FA values can theoretically range between 0 and 1, with higher values representing a more mature nerve cell, but are typically between 0 and 0.6 for neonates.

- Covariates: location, gender, race, birth weight, gestational age at birth, gestational age at time of the scan.

- Hypothesis: Older neonates would have higher FA values.

III. Data Analysis Example

Lateral and Ventral Cortico-spinal Tracts In the Human Nervous System



III. Data Analysis Example

The initial full model is as follows:

$$\boldsymbol{y}_i = \beta_0 + \beta_1 \boldsymbol{X}_{i,\mathrm{loc}} + \beta_2 \boldsymbol{X}^2_{i,\mathrm{loc}} + \beta_3 \boldsymbol{X}_{i,\mathrm{gab}} + \beta_4 \boldsymbol{X}_{i,\mathrm{gas}} + \beta_5 \boldsymbol{X}_{i,\mathrm{gen}} + \beta_6 \boldsymbol{X}_{i,\mathrm{rac}} + \beta_7 \boldsymbol{X}_{i,\mathrm{bwt}} + \boldsymbol{e}_i.$$

where

$\boldsymbol{y}_i$ FA values for 20 locations for subject i,

$\boldsymbol{X}_i$'s covariate vectors

$\boldsymbol{e}_i \sim N_{p_i}(\mathbf{0}, \boldsymbol{\Sigma}_{ei})$.

- GAR covariance model fit better than AR(1) according to AIC & BIC.
  - Other comparable structure, DE, did not converge when given same starting values as others.

III. Data Analysis Example

The final model after reduction via backward selection (α=0.10) is

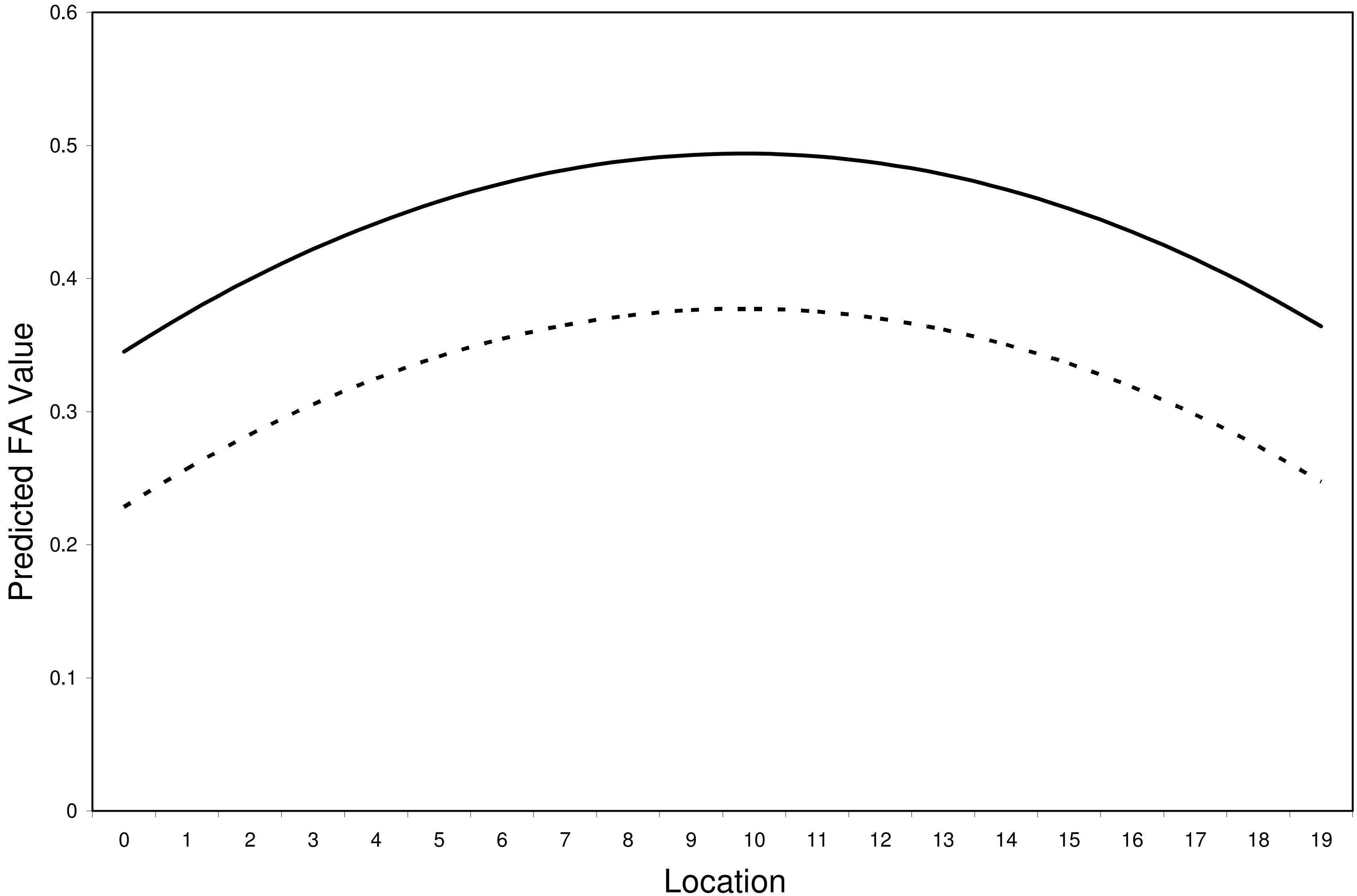
$$\boldsymbol{y}_i = \beta_0 + \beta_1 \boldsymbol{X}_{i,\text{loc}} + \beta_2 \boldsymbol{X}^2_{i,\text{loc}} + \beta_4 \boldsymbol{X}_{i,\text{gas}} + \boldsymbol{e}_i.$$

Covariate	Parameter Estimate	P-value
Location	0.0303	< 0.0001
Location^2	– 0.0015	< 0.0001
Gestational Age at Scan	0.0137	< 0.0001

- As expected, neonates who are older at the time of the scan have significantly higher FA values, and thus are likely to have more developed cortico-spinal fiber tracts.

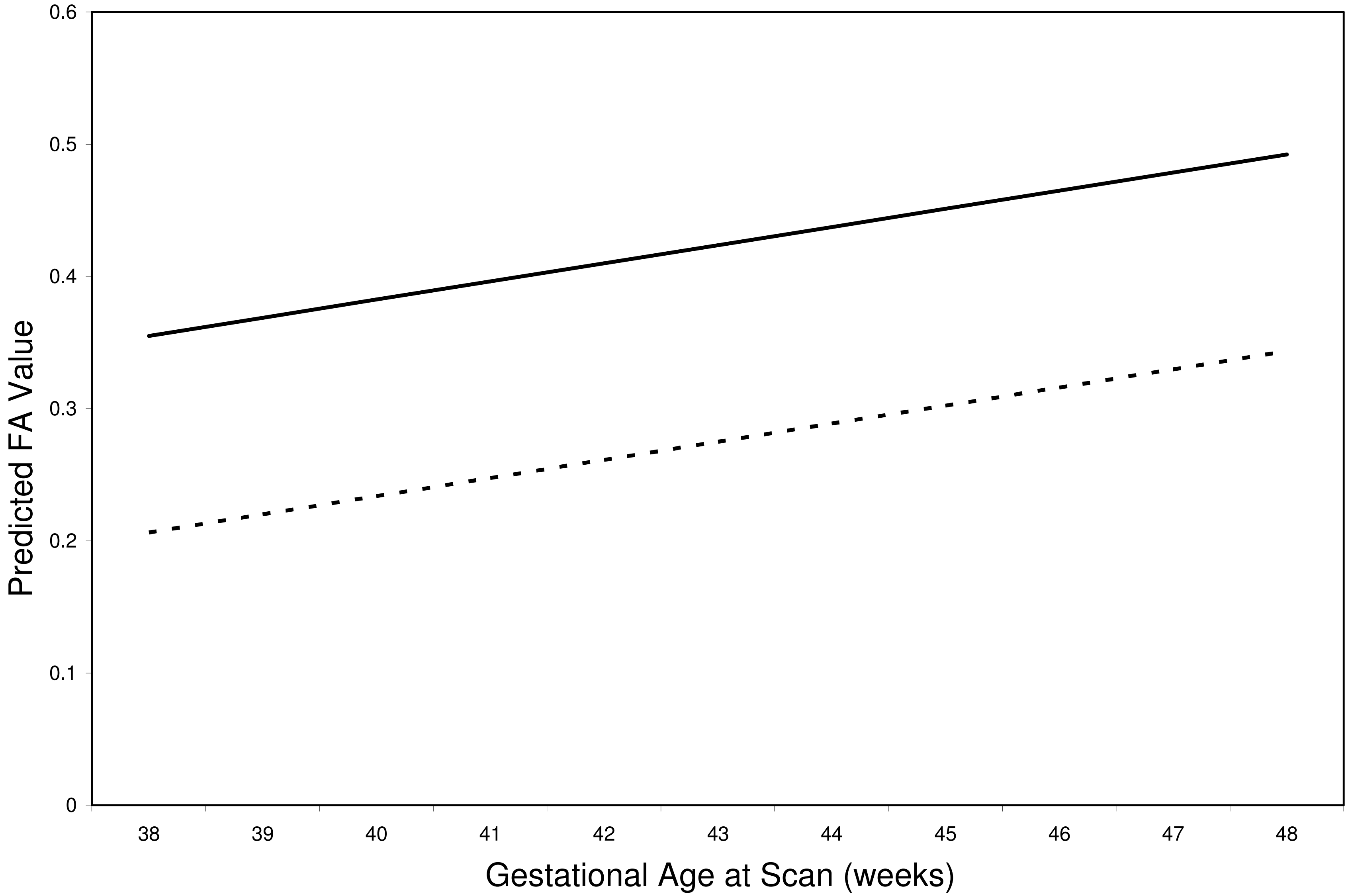
III. Data Analysis Example

Predicted FA values for the neonates by location at the minimum (dashed line) and maximum (solid line) gestational ages at the time of the scan



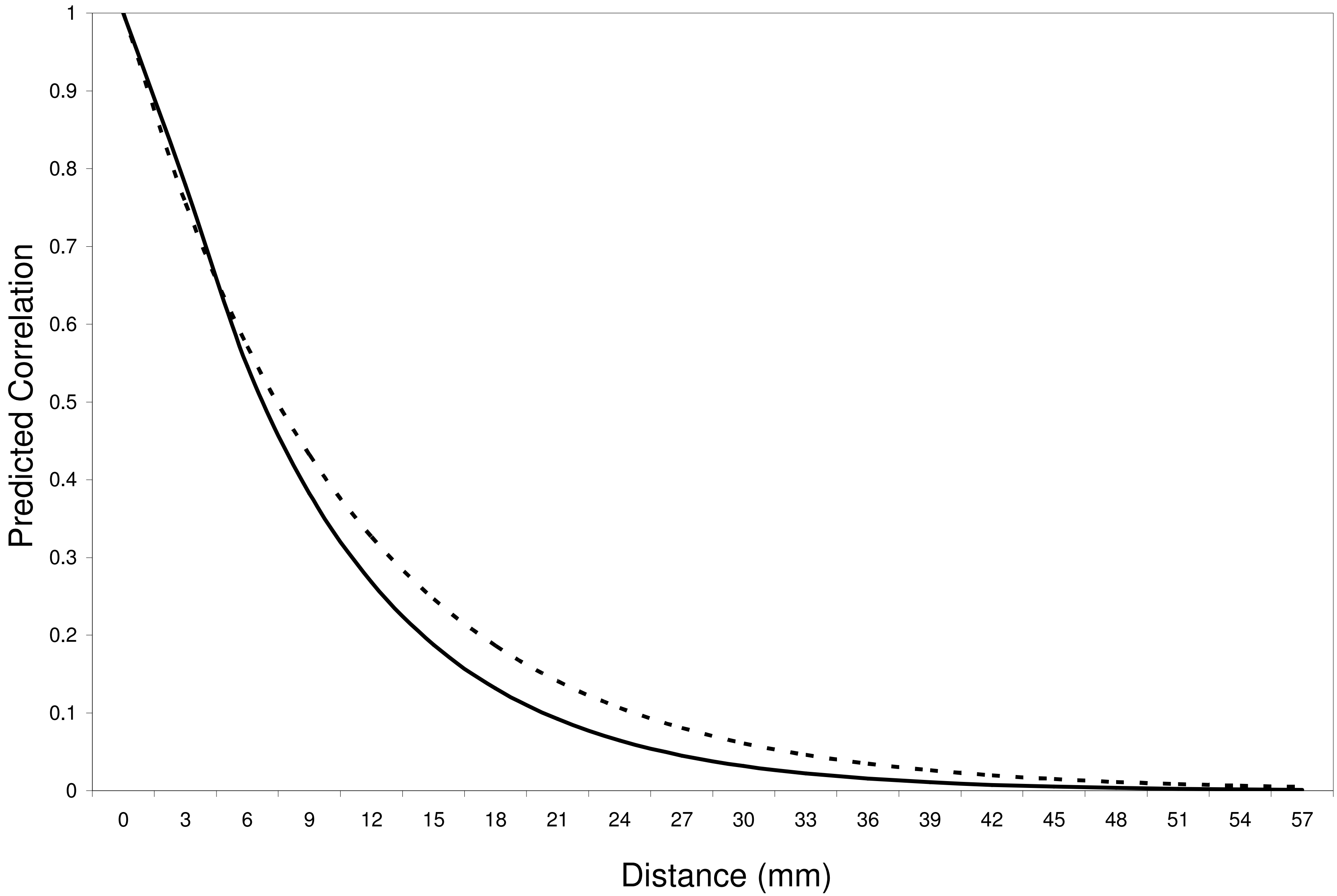
III. Data Analysis Example

Predicted FA values for the neonates by the gestational age at the time of the scan at the first (dashed line) and middle (solid line) locations



III. Data Analysis Example

Predicted correlation curve for the GAR model (solid line) and worse fitting AR(1) model (dashed line) as a function of the distance between measurements



IV. Kronecker Product GAR Covariance Model

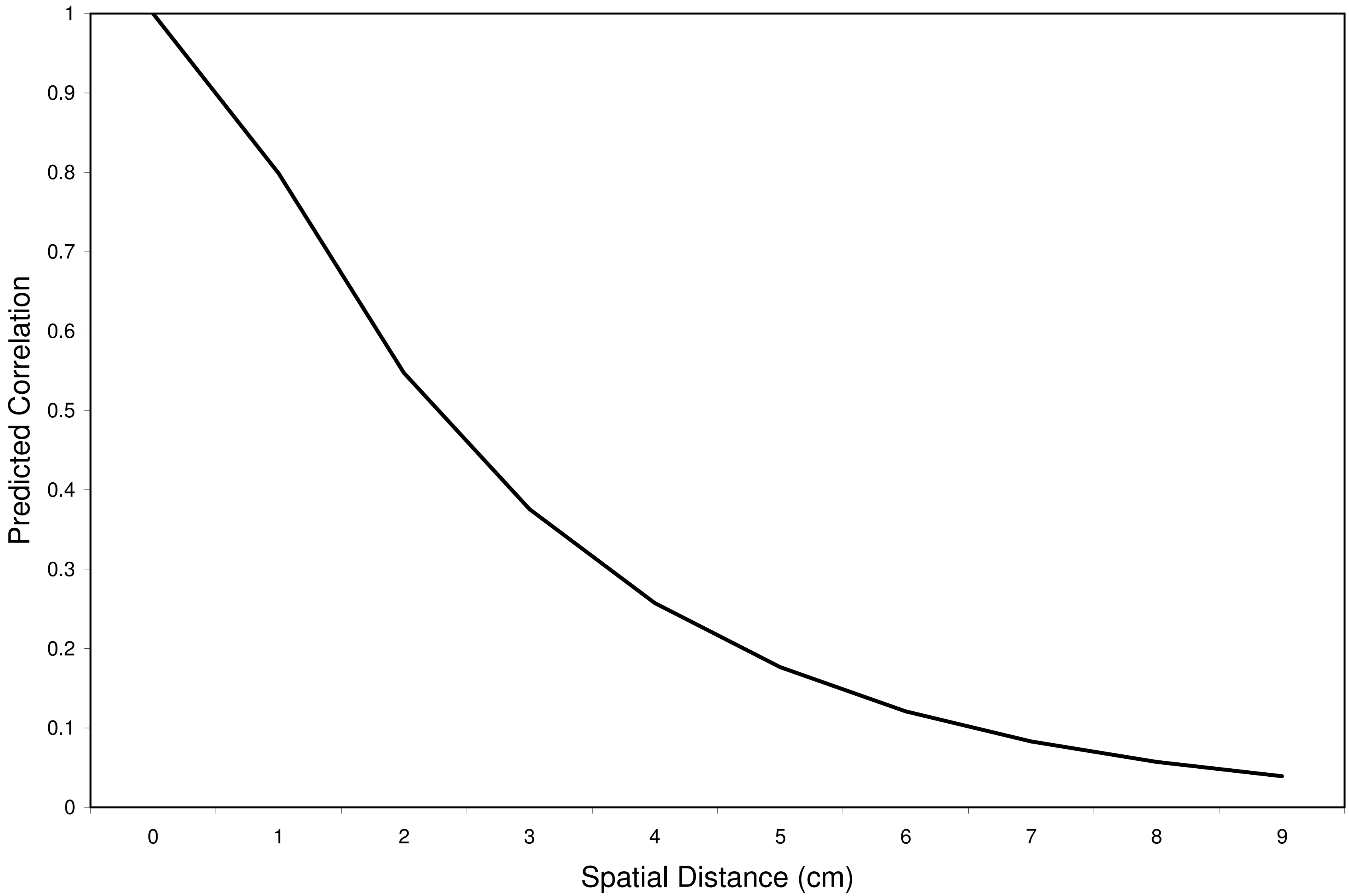
- Allows modeling of data where within-subject correlation induced by 2 factors.
  - Spatio-temporal, 2-dimensions of m-rep, etc.

- Now have $\boldsymbol{e}_i \sim N_{t_i s_i}(\mathbf{0}, \boldsymbol{\Sigma}_{ei} = \sigma_e^2[\boldsymbol{\Gamma}_{ei} \otimes \boldsymbol{\Omega}_{ei}])$,

  where $t_i \times t_i$ correlation matrix $\boldsymbol{\Gamma}_{ei}$ and the $s_i \times s_i$ correlation matrix $\boldsymbol{\Omega}_{ei}$ have GAR structure.

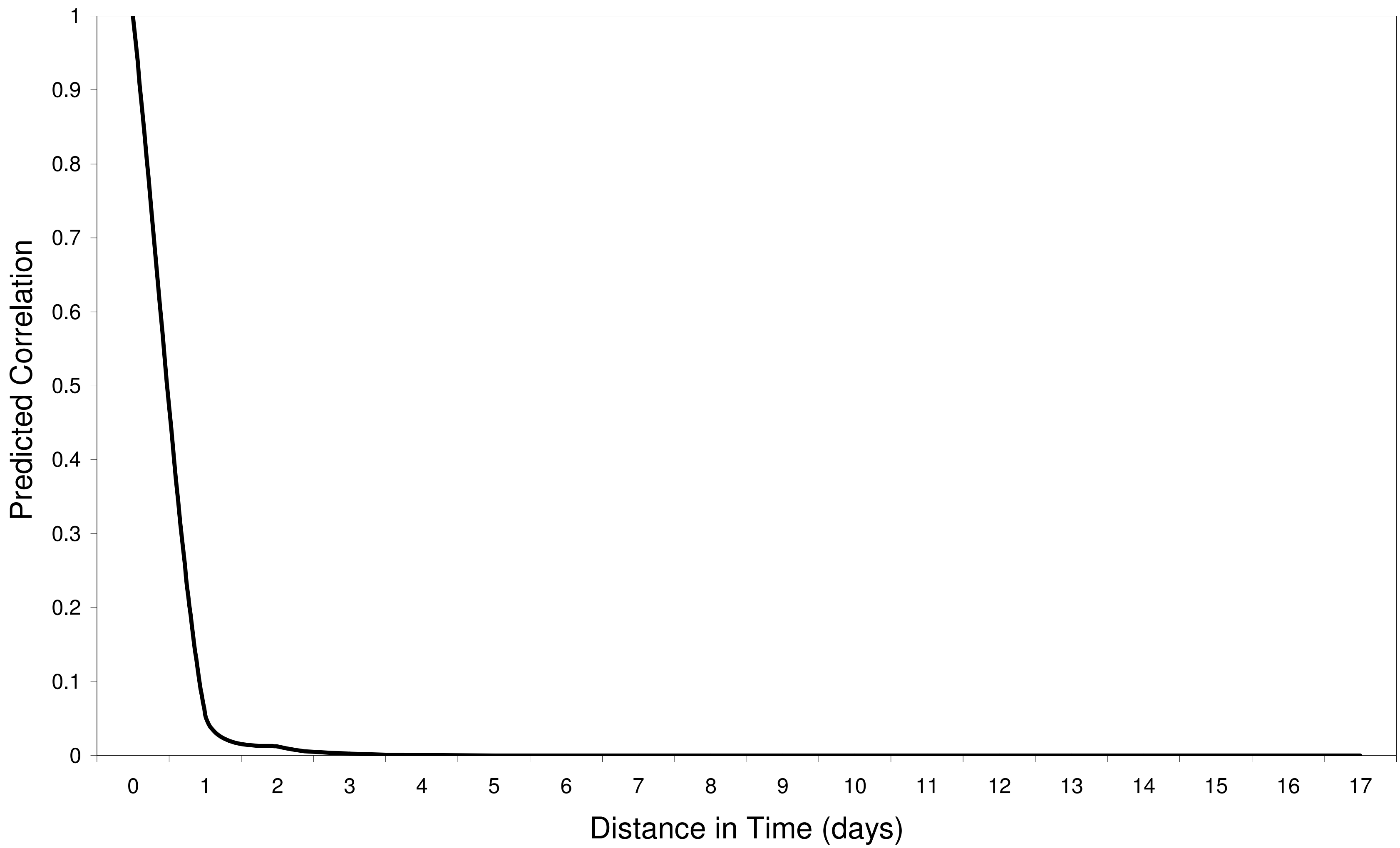
IV. Kronecker Product GAR Covariance Model

Predicted correlation curve for Log_2(radius) in bladder as a function of atom distance (OLD m-rep data)



IV. Kronecker Product GAR Covariance Model

Predicted correlation curve for Log_2(radius) in bladder as a function of distance in time (OLD m-rep data)



V. Conclusions

- GAR covariance model is a flexible, parsimonious structure that allows for a wide range of exponentially decaying correlation patterns
  - The classic temporal covariance structures for longitudinal data are special cases of this model.
  - Applicable to spatial, longitudinal, and other correlated data.
  - Good for HDLSS data.

- Kronecker product GAR covariance model allows for the modeling of doubly multivariate data (i.e., correlation induced by two factors).
  - Spatio-temporal, 2-dimensions of m-rep data.

VI. Future Research

- Nonstationary GAR covariance model (variance and/or correlations change as a function of absolute time or distance)
  - Extremely useful in neuroimaging studies of the developing brain variability of brain characteristics tends to change over time.

- Triply multivariate Kronecker Product GAR structure
  - e.g., for m-reps would have $\mathbf{\Sigma} = \sigma^2[\mathbf{\Gamma}_{\text{time}} \otimes \mathbf{\Omega}_{\text{atom}} \otimes \mathbf{\Phi}_{\text{feature}}]$