Modeling Random Effects

Maximum Likelihood

- Find most likely solutions for unknowns given the data at hand
- The likelihood function is log-linear
  - Maximizing the log of the function is equivalent to solving the actual function
- Take the first derivative of log-likelihood
  - Set equal to zero, do a little matrix algebra, and you get the equations to solve for all effects

Log-likelihood

$$2\sigma_e^2\pi^{-\frac{n}{2}}e^{-\frac{1}{2\sigma_e^2}(\mathrm{y}-\mathrm{X}\hat{\mathrm{b}})'(\mathrm{y}-\mathrm{X}\hat{\mathrm{b}})}$$

$$-\frac{n}{2}\ln(\sigma_e^2)-\frac{n}{2}\ln(2\pi)-\frac{1}{2\sigma_e^2}(\mathrm{y}-\mathrm{X}\hat{\mathrm{b}})'(\mathrm{y}-\mathrm{X}\hat{\mathrm{b}})$$

$$-\frac{1}{2}(\mathrm{y}-\mathrm{X}\hat{\mathrm{b}})'\mathrm{R}^{-1}(\mathrm{y}-\mathrm{X}\hat{\mathrm{b}})$$

MLE and OLS

OLS $$\hat{\mathrm{b}} = (\mathrm{X}'\mathrm{X})^{-1}\mathrm{X}'\mathrm{y}$$

MLE $$\hat{\mathrm{b}} = (\mathrm{X}'\mathrm{R}^{-1}\mathrm{X})^{-1}\mathrm{X}'\mathrm{R}^{-1}\mathrm{y}$$

Mixed model

- Has both fixed and random effects

$$y = Xb + Zu + e$$

Where $u \sim N(0, G)$ and $e \sim N(0, R)$. Given u and e are independent we can use expectations to determine the mean and variance of y

Mixed model

$$E[y] = E[Xb + Zu + e]$$

$$E[y] = E[Xb] + E[Zu] + E[e]$$

$$E[y] = XE[b] + ZE[u] + E[e]$$

$$E[y] = Xb + Z0 + E0 \longrightarrow E[y] = Xb$$

Mixed model

$$V(y) = V(Xb + Zu + e)$$

$$V(y) = V(Xb) + V(Zu) + V(e)$$

$$V(y) = XV(b)X' + ZV(u)Z' + V(e)$$

$$V(y) = X0X' + ZGZ' + R \longrightarrow V(y) = ZGZ' + R$$

Mixed Model Equations

$$y \sim N(Xb, ZGZ'+R)$$

Using maximum likelihood the equations for mixed models are derived as:

$$\begin{bmatrix} X'R^{-1}X & X'R^{-1}Z \\ Z'R^{-1}X & Z'R^{-1}Z + G^{-1} \end{bmatrix} \begin{bmatrix} b \\ u \end{bmatrix} = \begin{bmatrix} X'R^{-1}y \\ Z'R^{-1}y \end{bmatrix}$$

Mixed Model Equations

$$\underbrace{\begin{bmatrix} X'R^{-1}X & X'R^{-1}Z \\ Z'R^{-1}X & Z'R^{-1}Z + G^{-1} \end{bmatrix}}_{\text{Left Hand Side of the equations (LHS)}} \begin{bmatrix} b \\ u \end{bmatrix} = \underbrace{\begin{bmatrix} X'R^{-1}y \\ Z'R^{-1}y \end{bmatrix}}_{\text{Right Hand Side of the equations (RHS)}}$$

$$\begin{bmatrix} \hat{\mathbf{b}} \\ \hat{\mathbf{u}} \end{bmatrix} = \mathbf{LHS}^{-1}\mathbf{RHS}$$

Variance Component Estimation

- Means Squared
  - Unbiased Estimates of Variance
  - Requires Balanced Data
- MLE
  - Works well with unbalanced and correlated data
  - Produces biased estimates
- REML
  - Uses contrast of the observed data to remove fixed effects
  - Works well with unbalanced and correlated data
  - Produces unbiased estimates

Basic Operations –Kronecker Product

Complex covariance structures are typically specified as the **Kronecker Product** of multiple simple covariance structures

$$\mathrm{A} = \begin{bmatrix} a_{11} & a_{12} & a_{13} \\ a_{21} & a_{22} & a_{23} \\ a_{31} & a_{32} & a_{33} \end{bmatrix} \qquad \mathrm{B} = \begin{bmatrix} b_{11} & b_{12} & b_{13} \\ b_{21} & b_{22} & b_{23} \\ b_{31} & b_{32} & b_{33} \end{bmatrix}$$

Kronecker Product

$$\begin{bmatrix} a_{11} & a_{12} & a_{13} \\ a_{21} & a_{22} & a_{23} \\ a_{31} & a_{32} & a_{33} \end{bmatrix} \otimes \begin{bmatrix} b_{11} & b_{12} & b_{13} \\ b_{21} & b_{22} & b_{23} \\ b_{31} & b_{32} & b_{33} \end{bmatrix} = \begin{bmatrix} a_{11}\mathrm{B} & a_{12}\mathrm{B} & a_{13}\mathrm{B} \\ a_{21}\mathrm{B} & a_{21}\mathrm{B} & a_{23}\mathrm{B} \\ a_{31}\mathrm{B} & a_{32}\mathrm{B} & a_{33}\mathrm{B} \end{bmatrix}$$

9x9 matrix

Constructing Covariance Matrices

$\mathrm{V}(\mathbf{y}) = \mathbf{V}_{nxn}$ Where *n* is the number of observations

When there is no covariance between values of **y** then:

$$\mathbf{V}_{nxn} = \sigma^2 \mathbf{I}_{nxn}$$

For complex structures it is helpful to think of $\mathbf{V}_{nxn}$ as the kronecker product of smaller matrices:

$$\mathbf{V}_{nxn} = \mathbf{A} \otimes \mathbf{B} \otimes \mathbf{C}$$

In this example $\mathbf{A}$, $\mathbf{B}$, and $\mathbf{C}$ may represent different processes impacting the covariance of **y**. When constructing $\mathbf{V}_{nxn}$ careful attention must be given to the order of values in **y** and the order of $\mathbf{A}$, $\mathbf{B}$, and $\mathbf{C}$ in the Kronecker product

Correlated Residuals - Repeated Measures

When the same measurements are taken on the same experimental unit, residuals associated with those measurement can no longer be considered independent.

In such cases it is important to account for the correlation structures between observations.

Repeated measures – Technical Replication

The use of technical replications in experiments is common when the technology being used to take measurements is error prone and the cost of including more and there are limitations including additional experimental units.

Examples:

- Conducting an experiment in a growth chamber that has limits on the number of plants.
- Conducting a field experiment but there is limited seed for planting plots

If I were collecting gene expression data (which can be noisy) it may increase the power of my experiment to take repeated measurements

Repeated measures – Time Series

Some studies may be focused on modeling an effect and how it changes through time. In these studies, repeated measurements of the experiment units are taken through time.

Examples:

- Experiments looking a disease resistance/susceptibility. In such studies it is common to take measurements at multiple time points following inoculation.

- Developmental biology. Experiments may be interested in understanding how expression of certain genes changes through time. In such cases each experimental unit may be measure at different development stages.

Repeated measures

In the case of technical replication and time series experiments it is reasonable to assume that residual error from measurements taken on the same experimental unit may be correlated.

Residual Correlation Structures

Residual correlation structures represent assumptions of the underlying factors that lead to the correlations.

The full covariance structure may be the product of multiple correlation and variance structures.

Technical Replicates – Uniform Correlation

Simple Example – 3 repeated measurements on 5 experimental units

$$\begin{bmatrix} \sigma^2 & & & & \\ & \sigma^2 & & & \\ & & \sigma^2 & & \\ & & & \sigma^2 & \\ & & & & \sigma^2 \end{bmatrix} \otimes \begin{bmatrix} 1 & \rho & \rho \\ \rho & 1 & \rho \\ \rho & \rho & 1 \end{bmatrix}$$

Technical Replicates – Uniform Correlation

Simple Example – 3 repeated measurements on 5 experimental units

In this case of a fixed effect model $\quad y \sim N(Xb, R)$

$$R = \begin{bmatrix} \sigma^2 & & & & \\ & \sigma^2 & & & \\ & & \sigma^2 & & \\ & & & \sigma^2 & \\ & & & & \sigma^2 \end{bmatrix} \otimes \begin{bmatrix} 1 & \rho & \rho \\ \rho & 1 & \rho \\ \rho & \rho & 1 \end{bmatrix}$$

Technical Replicates – Mixed Models

Simple Example – 3 repeated measurements on 5 experimental units

What if we ran a mixed model with a random effect for experimental unit and treated the residuals as i.i.d?

$$\mathrm{u} \sim \mathrm{N}(0, \sigma_u^2 * \mathbf{I}) \qquad \mathrm{e} \sim \mathrm{N}(0, \sigma_e^2 * \mathbf{I})$$

What is the covariance structure for **y**?

Problem Set: Genotype by Environment Interactions

- In plant breeding it is common to observe genotype by environment interactions (GxE)
  - GxE happens when the relative performance of plant varieties change under different environmental conditions.
  - When modeling data collected in multiple environments how do we account for the fact that the relative performance of lines may change in different locations?
  - What are the implications of model specification on the covariance between observations collected under different environmental conditions?

Some Common Structures

Independent and Identically Distributed

$$\mathbf{I} * \sigma^2 = \begin{bmatrix} \sigma^2 & 0 & \vdots & 0 \\ 0 & \sigma^2 & \vdots & 0 \\ \cdots & \cdots & \ddots & \cdots \\ 0 & 0 & \vdots & \sigma^2 \end{bmatrix}$$

Some Common Structures

Uniform Correlation/Covariance

$$\begin{bmatrix} 1 & \rho & \vdots & \rho \\ \rho & 1 & \vdots & \rho \\ \cdots & \cdots & \ddots & \cdots \\ \rho & \rho & \vdots & 1 \end{bmatrix} * \sigma^2 = \begin{bmatrix} \sigma^2 & \sigma & \vdots & \sigma \\ \sigma & \sigma^2 & \vdots & \sigma \\ \cdots & \cdots & \ddots & \cdots \\ \sigma & \sigma & \vdots & \sigma^2 \end{bmatrix}$$

$$\sigma = \rho * \sigma^2$$

Some Common Structures

Unstructured

$$\begin{bmatrix} \sigma_1^2 & \sigma_{12} & \vdots & \sigma_{1n} \\ \sigma_{21} & \sigma_2^2 & \vdots & \sigma_{2n} \\ \cdots & \cdots & \ddots & \cdots \\ \sigma_{n1} & \sigma_{n2} & \vdots & \sigma_n^2 \end{bmatrix}$$

Places no restrictions on the covariance structure

Not very scalable

$$\# \, parameters = \frac{n * (n-1)}{2} + n$$

Longitudinal Data – Uniform Correlation

- When working with longitudinal data the term permanent environmental effect is sometimes used. It represents residual effects that are consistent through time, as opposed to residual effects associate with a specific point in time.

$$\begin{bmatrix} \sigma_{pe}^2 + \sigma_{te}^2 & \sigma_{pe}^2 & \sigma_{pe}^2 \\ \sigma_{pe}^2 & \sigma_{pe}^2 + \sigma_{te}^2 & \sigma_{pe}^2 \\ \sigma_{pe}^2 & \sigma_{pe}^2 & \sigma_{pe}^2 + \sigma_{te}^2 \end{bmatrix} \otimes \mathbf{I}_{\boldsymbol{n}}$$

Longitudinal Data

When taking multiple measurements through time it is reasonable to assume that residuals from measurements taken closer together in time may have higher correlations than measurements taken further apart.

In this scenario a uniform correlation structure may not be appropriate.

Longitudinal Data - Toeplitz

It is reasonable to assume that when dealing

$$\begin{bmatrix} \sigma^2 & \sigma_1 & \sigma_2 \\ \sigma_1 & \sigma^2 & \sigma_1 \\ \sigma_2 & \sigma_1 & \sigma^2 \end{bmatrix} \otimes \mathbf{I}_{\boldsymbol{n}}$$

Longitudinal Data - Autoregressive

Unlike the banded structure of the Toeplitz matrix, and autoregressive function has a single parameter that decays with time/distance between points:

$$\sigma^2 * \mathbf{I}_{\boldsymbol{n}} \otimes \begin{bmatrix} 1 & \rho & \rho^2 & \rho^3 & \rho^4 \\ \rho & 1 & \rho & \rho^2 & \rho^3 \\ \rho^2 & \rho & 1 & \rho & \rho^2 \\ \rho^3 & \rho^2 & \rho & 1 & \rho \\ \rho^4 & \rho^3 & \rho^2 & \rho & 1 \end{bmatrix}$$

Longitudinal Data

As with all covariances matrices modeling an underlying process that results in correlated effects or residuals, when in doubt there is always the unstructured covariance matrix:

$$\begin{bmatrix} \sigma_{t1}^2 & \sigma_{t12} & \sigma_{t13} \\ \sigma_{t21} & \sigma_{t2}^2 & \sigma_{t23} \\ \sigma_{31} & \sigma_{t32} & \sigma_{t3}^2 \end{bmatrix} \otimes \mathbf{I}_{\boldsymbol{n}}$$

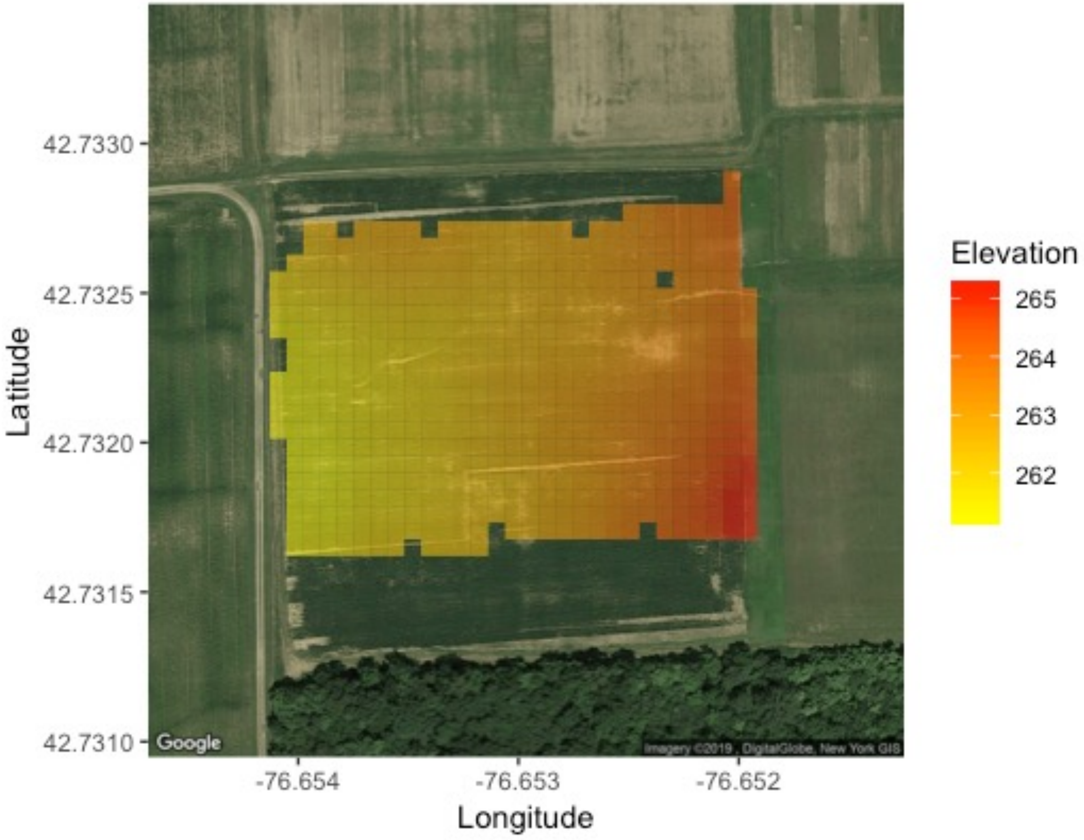
In this small example there is not a big difference between US than Toeplitz structure previously covered; however, the number of parameters for Toeplitz increases linearly with the number of timepoints unlike US.

Spatial Effects

When conducting field experiments spatial variation is common and can have a significant impact on phenotypic measures.

Factors that lead to spatial variation:

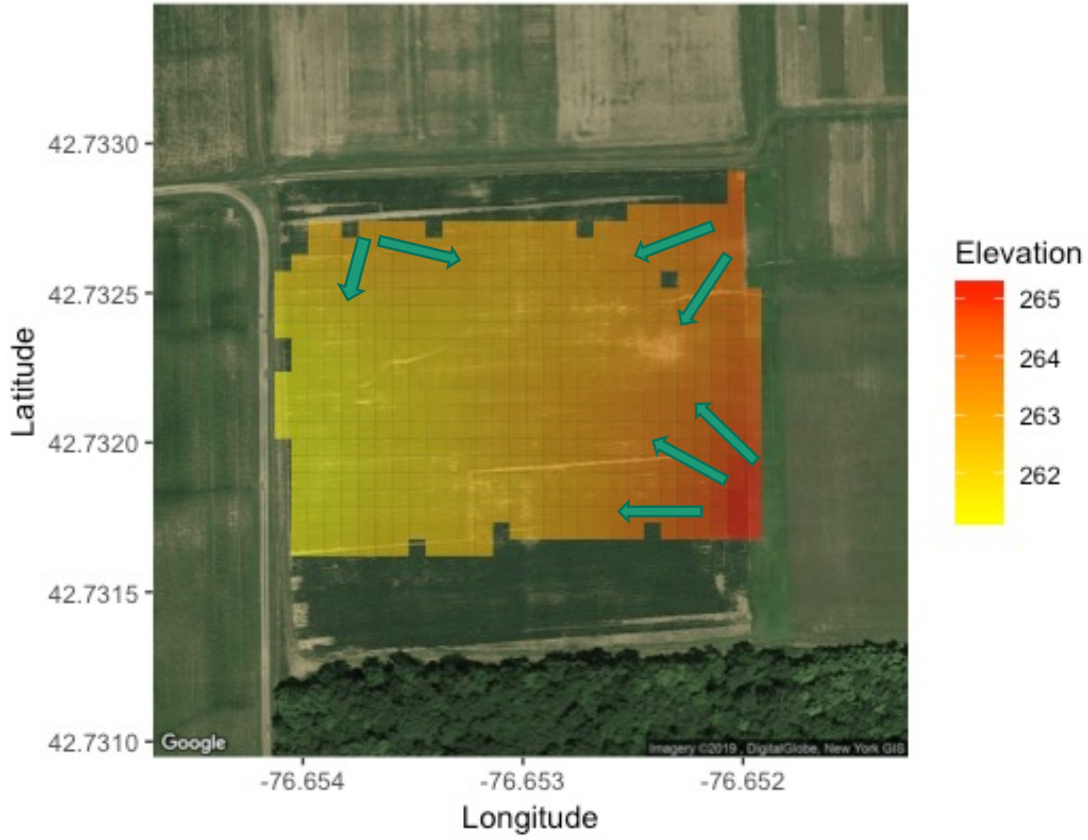
- Soil composition
- Change in elevation across the field
- Soil fertility
- Disease pressure



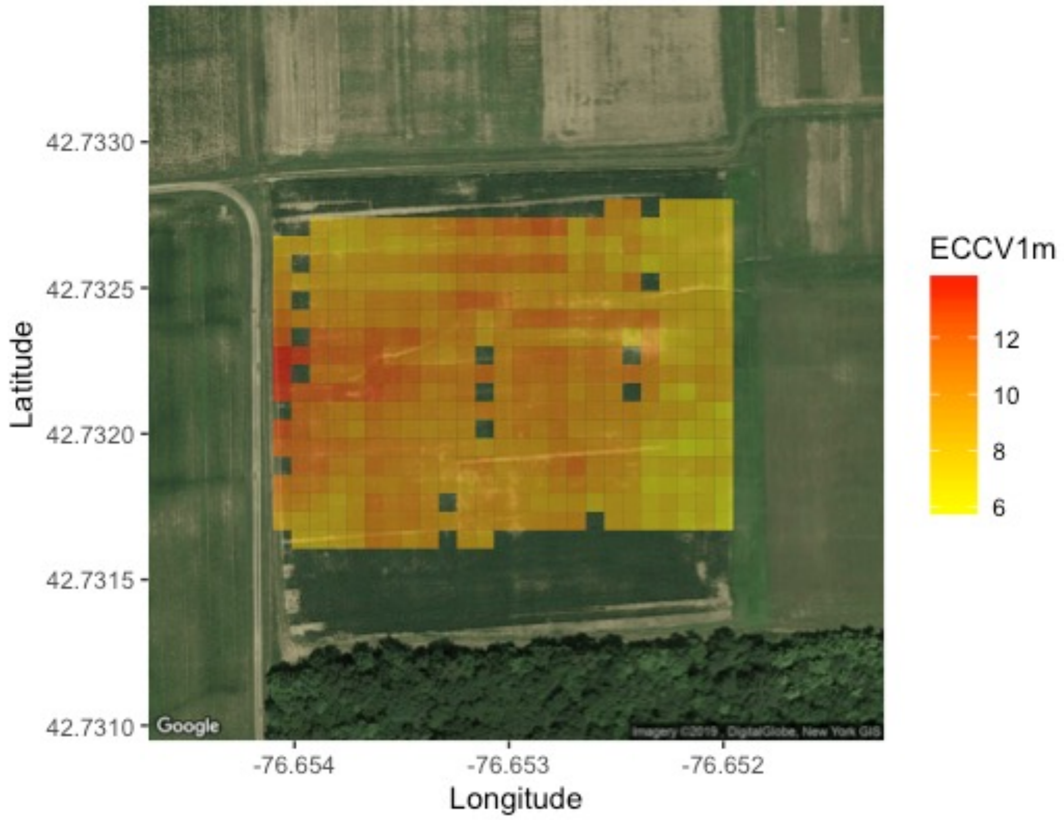
Spatial Effects

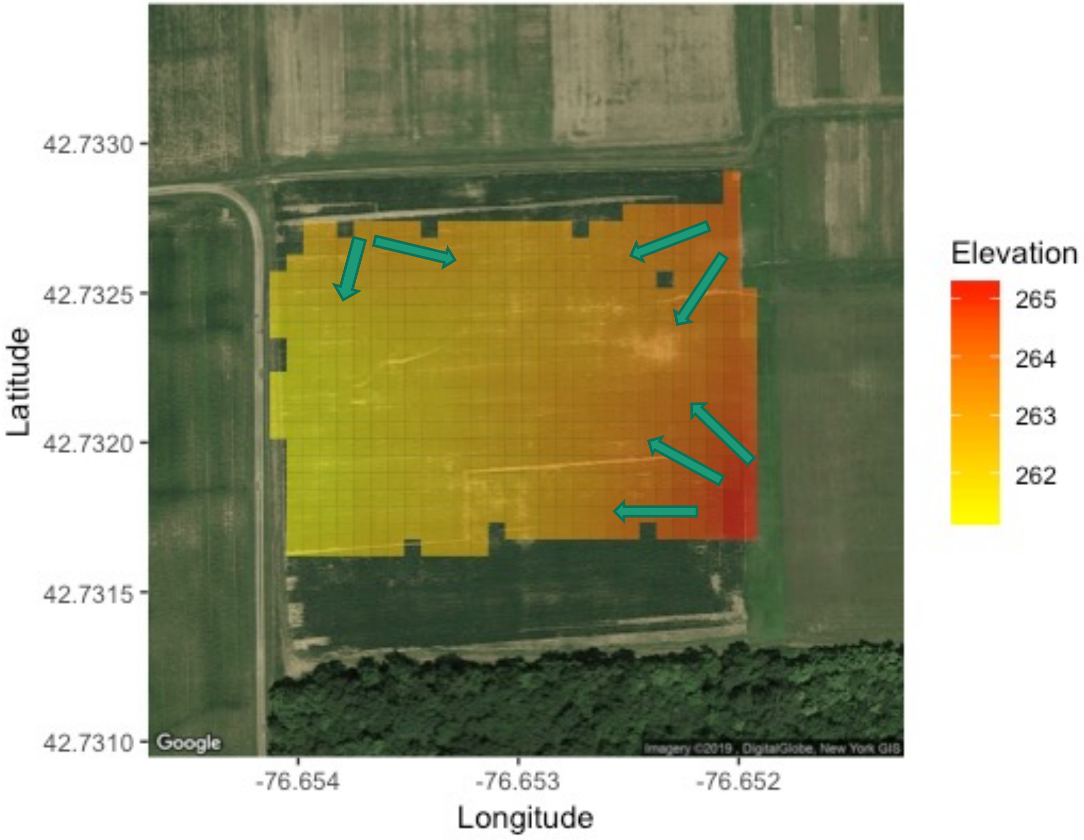
When conducting field experiments spatial variation is common and can have a significant impact on phenotypic measures.

Factors that lead to spatial variation:
- Soil composition
- Change in elevation across the field
- Soil fertility
- Disease pressure



Spatial Effects





Spatial Effects – 2D Separable Autoregressive Structure

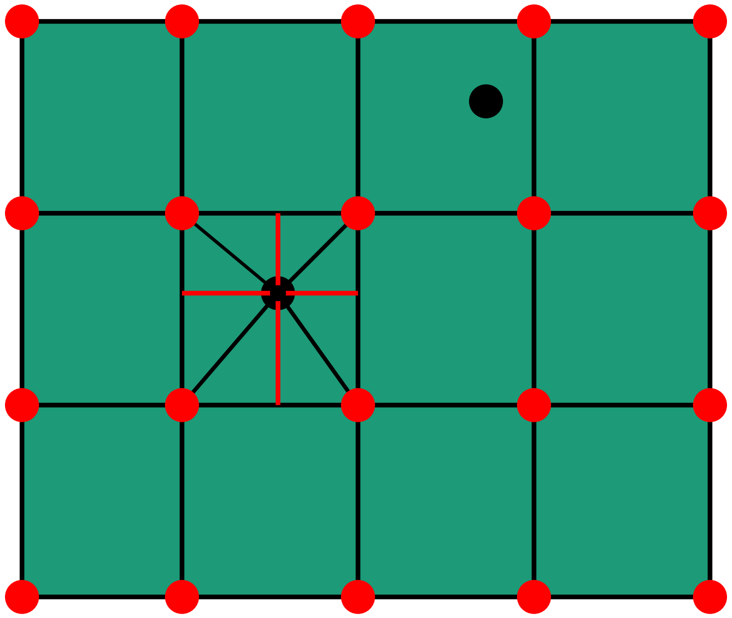
In this structure correlations are modeled in two-dimensions, both row and column

$$\sigma^2 * \begin{bmatrix} 1 & \rho_r & \rho_r^2 & \rho_r^3 & \rho_r^4 \\ \rho_r & 1 & \rho_r & \rho_r^2 & \rho_r^3 \\ \rho_r^2 & \rho_r & 1 & \rho_r & \rho_r^2 \\ \rho_r^3 & \rho_r^2 & \rho_r & 1 & \rho_r \\ \rho_r^4 & \rho_r^3 & \rho_r^2 & \rho_r & 1 \end{bmatrix} \otimes \begin{bmatrix} 1 & \rho_c & \rho_c^2 & \rho_c^3 & \rho_c^4 \\ \rho_c & 1 & \rho_c & \rho_c^2 & \rho_c^3 \\ \rho_c^2 & \rho_c & 1 & \rho_c & \rho_c^2 \\ \rho_c^3 & \rho_c^2 & \rho_c & 1 & \rho_c \\ \rho_c^4 & \rho_c^3 & \rho_c^2 & \rho_c & 1 \end{bmatrix}$$

Spatial Effects – Moving Average

$$\sigma^2 * \begin{bmatrix} 1 & \phi_{r1} & \phi_{r2} & \phi_{r3} & 0 \\ \phi_{r1} & 1 & \phi_{r1} & \phi_{r2} & \phi_{r3} \\ \phi_{r2} & \phi_{r1} & 1 & \phi_{r1} & \phi_{r2} \\ \phi_{r3} & \phi_{r2} & \phi_{r1} & 1 & \phi_{r1} \\ 0 & \phi_{r3} & \phi_{r2} & \phi_{r1} & 1 \end{bmatrix} \otimes \begin{bmatrix} 1 & \phi_{c1} & \phi_{c2} & 0 & 0 \\ \phi_{c1} & 1 & \phi_{c1} & \phi_{c2} & 0 \\ \phi_{c2} & \phi_{c1} & 1 & \phi_{c1} & \phi_{c2} \\ 0 & \phi_{c2} & \phi_{c1} & 1 & \phi_{c1} \\ 0 & 0 & \phi_{c2} & \phi_{c1} & 1 \end{bmatrix}$$

Spatial Effects – 2D Spline



Correlated Dependent/Response Variables

- There are many cases in which multiple phenotypes are measured on the same experimental unit.
- In some cases, the phenotypes are correlated due to common effects influencing the phenotypes.
  - This can result in covariance between effects of interest or residual error.
- When collected on the same experimental unit are correlated Multi-trait Models can provide increased accuracy and power

Multi-Trait Models

- Multiple Trait Models (MTM)
  - Model (co)variance between traits
  - Both Residual and genetic

$$\mathbf{g} \sim N\left(0, \mathbf{I}_g \otimes \begin{bmatrix} \sigma_{g1}^2 & \sigma_{g12} \\ \sigma_{g12} & \sigma_{g2}^2 \end{bmatrix}\right) \text{ or } g \sim N\left(0, \mathbf{G} \otimes \begin{bmatrix} \sigma_{g1}^2 & \sigma_{g12} \\ \sigma_{g12} & \sigma_{g2}^2 \end{bmatrix}\right)$$

$$\mathbf{e} \sim N\left(0, \mathbf{I}_n \otimes \begin{bmatrix} \sigma_{e1}^2 & \sigma_{e12} \\ \sigma_{e12} & \sigma_{e2}^2 \end{bmatrix}\right)$$

Extra Slides

Restricted Maximum Likelihood (REML)

- REML maximizes the likelihood of contrasts of y. Contrasts are set up so that fixed effects are removed.

$$\mathrm{K'X} = 0 \text{ and } \mathrm{rank(K')} = \mathrm{N} - \mathrm{rank(X)} \qquad \mathrm{K(K'K)^{-1}K} = \mathrm{I} - \mathrm{X(X'X)^{-1}X}$$

V=ZGZ'+R

$$L(\mathbf{K'y}) = (2\pi)^{-.5(N-rank(\mathbf{X}))}|\mathbf{K'VK}|^{-.5}\exp(-.5(\mathbf{K'y}-\mathbf{K'Xb})'(\mathbf{K'VK})^{-1}(\mathbf{K'y}-\mathbf{K'Xb}))$$

$$L(\mathbf{K'y}) = (2\pi)^{-.5(N-rank(\mathbf{X}))}|\mathbf{K'VK}|^{-.5}\exp(-.5(\mathbf{K'y})'(\mathbf{K'VK})^{-1}(\mathbf{K'y}))$$

$$L(\mathbf{K'y}) = -.5\ln(|\mathbf{K'VK}| - .5(\mathbf{K'y})'(\mathbf{K'VK})^{-1}(\mathbf{K'y}) \Longleftrightarrow L_1 = -.5\ln(|\mathbf{V}|) - .5\ln(|\mathbf{X'V^{-1}X}|) - .5(\mathbf{y}-\mathbf{X\hat{b}})'\mathbf{V}^{-1}(\mathbf{y}-\mathbf{X\hat{b}})$$

ML vs REML

$$L_1 = -.5\ln(|\mathbf{V}|) - .5(\mathbf{y} - \mathbf{Xb})'\mathbf{V}^{-1}(\mathbf{y} - \mathbf{Xb})$$

$$L_2 = -.5\ln(|\mathbf{V}|) - .5\ln(|\mathbf{X}'\mathbf{V}^{-1}\mathbf{X}|) - .5(\mathbf{y} - \mathbf{X}\hat{\mathbf{b}})'\mathbf{V}^{-1}(\mathbf{y} - \mathbf{X}\hat{\mathbf{b}})$$

EM Algorithm

E step: solve using current estimate of R and G

$$\begin{bmatrix} X'X & X'Z \\ Z'X & Z'Z + \alpha A^{-1} \end{bmatrix} \begin{bmatrix} b \\ u \end{bmatrix} = \begin{bmatrix} X'y \\ Z'y \end{bmatrix}$$

M step: estimate new R and G using current solutions from MME

$$\hat{\sigma}_e^2 = {y'y - \widehat{b}'X'y - \widehat{u}'Z'Y} \Big/ {(N - rank(X))}$$

$$\hat{\sigma}_a^2 = {(\hat{u}'A^{-1}\hat{u} + trace(A^{-1}C_{22})\hat{\sigma}_e^2)} \Big/ {k} \qquad \begin{bmatrix} X'X & X'Z \\ Z'X & Z'Z + \alpha A^{-1} \end{bmatrix}^{-1} = \begin{bmatrix} C_{11} & C_{12} \\ C_{21} & C_{22} \end{bmatrix}$$

EM Implementation

- Step 1: Solve for $\widehat{\mathrm{b}}$ and $\hat{\mathrm{u}}$ given the current solutions of R and G
- Step 2:
  - Solve $$\hat{\sigma}_e^2 = {\mathrm{y}'\mathrm{y} - \widehat{\mathrm{b}'}\mathrm{X}'\mathrm{y} - \widehat{\mathrm{u}'}\mathrm{Z}'\mathrm{Y}} \Big/ {(N - rank(\mathrm{X}))}$$
  - Solve $$\hat{\sigma}_a^2 = {(\hat{\mathrm{u}}'\mathrm{A}^{-1}\hat{\mathrm{u}} + trace(\mathrm{A}^{-1}\mathrm{C}_{22})\hat{\sigma}_e^2)} \Big/ {k}$$
- Repeat steps 1 and 2 until solutions converge.

Newton-Raphson

Based on Taylor's series expansion

$$f(\theta_t) = f(\theta_{t-1}) + \frac{f'(\theta_{t-1})}{1!}(\theta_t - \theta_{t-1}) + \frac{f''(\theta_{t-1})}{2!}(\theta_t - \theta_{t-1})^2 + \cdots + \frac{f^{n\prime}(\theta_{t-1})}{n!}(\theta_t - \theta_{t-1})^n$$

Newton-Raphson

Based on Taylor's series expansion

$$f(\theta_t) = f(\theta_{t-1}) + \frac{f'(\theta_{t-1})}{1!}(\theta_t - \theta_{t-1}) + \frac{f''(\theta_{t-1})}{2!}(\theta_t - \theta_{t-1})^2 + \cdots + \frac{f^{n\prime}(\theta_{t-1})}{n!}(\theta_t - \theta_{t-1})^n$$

This approach is useful for approximating values that are difficult or impossible to solve for directly.

A common example is approximating the value of the mathematical constant *e* using a Taylor series example.

Newton-Raphson

Based on Taylor's series expansion

$$f(\theta_t) = f(\theta_{t-1}) + \frac{f'(\theta_{t-1})}{1!}(\theta_t - \theta_{t-1}) + \frac{f''(\theta_{t-1})}{2!}(\theta_t - \theta_{t-1})^2 + \cdots + \frac{f^{n\prime}(\theta_{t-1})}{n!}(\theta_t - \theta_{t-1})^n$$

This approach is useful when $f(\theta)$ is undefined at θ_t . We can choose some value θ_{t-1} for which $f(\theta)$ is defined and use it to approximate $f(\theta_t)$.

In the case of mixed models our solutions $\hat{b}$ and $\hat{u}$ are undefined because we don't know the variance components

Newton-Raphson

We want solutions that maximize the likelihood function, or more specifically the log-likelihood function.

$$\underbrace{0 = f(\theta_{t-1})}_{\text{First derivative of the log-likelihood function}} + \underbrace{\frac{f'(\theta_{t-1})}{1!}}_{\text{Second derivative of the log-likelihood function}} \overbrace{(\theta_t - \theta_{t-1})}^{\Delta\theta}$$

$$\Delta\theta = -f(\theta_{t-1})f'(\theta_{t-1})^{-1}$$

$$\theta_t - \theta_{t-1} = -f(\theta_{t-1})f'(\theta_{t-1})^{-1}$$

$$\theta_t = \theta_{t-1} - f(\theta_{t-1})f'(\theta_{t-1})^{-1}$$

$$\theta_t = \theta_{t-1} + \Delta\theta$$

Newton-Raphson

In our case θ is a vector of unknowns so:

$$f(\theta) = \begin{bmatrix} \frac{\partial L}{\partial \theta_1} & \frac{\partial L}{\partial \theta_1 \partial \theta_2} & \frac{\partial L}{\partial \theta_1 \partial \theta_n} \\ & \frac{\partial L}{\partial \theta_2} & \frac{\partial L}{\partial \theta_2 \partial \theta_n} \\ & & \frac{\partial L}{\partial \theta_n} \end{bmatrix}$$

$$f'(\theta) = \begin{bmatrix} \frac{\partial^2 L}{\partial^2 \theta_1} & \frac{\partial^2 L}{\partial^2 \theta_1 \partial^2 \theta_2} & \frac{\partial^2 L}{\partial^2 \theta_1 \partial^2 \theta_n} \\ & \frac{\partial^2 L}{\partial^2 \theta_2} & \frac{\partial^2 L}{\partial^2 \theta_2 \partial^2 \theta_n} \\ & & \frac{\partial^2 L}{\partial^2 \theta_n} \end{bmatrix}$$

Hessian Matrix

Quasi-Newton Methods: Use an approximation of the Hessian

Fischer's Scoring Method: Replace the Hessian with Fischer Information Matrix/Expected Information Matrix

Newton-Raphson

Step 1:

Set initial value as θ_{t-1}

Step2:

Calculate $\Delta\theta = -f(\theta_{t-1})f'(\theta_{t-1})^{-1}$

Step 3:

Calculate $\theta_t = \theta_{t-1} + \Delta\theta$

Step 4:

Set new θ_{t-1} equal to θ_t

Repeat until convergence