Using Non-Linear Mixed Models for Agricultural Data

Fernando E. Miguez

Energy Biosciences Institute
Crop Sciences
University of Illinois, Urbana-Champaign
miguez@illinois.edu

Oct 8th, 2008

Outline

Objectives of Statistical Modeling

Objectives

1. Develop the simplest model which still captures the structure of the data
2. Interpret the model (give meaning to the parameters)
3. Generate predictions (validation)

Objectives of Statistical Modeling

Objectives

1. Develop the simplest model which still captures the structure of the data
2. Interpret the model (give meaning to the parameters)
3. Generate predictions (validation)

Objectives of Statistical Modeling

Objectives

1. Develop the simplest model which still captures the structure of the data
2. Interpret the model (give meaning to the parameters)
3. Generate predictions (validation)

Non-Linear and Mixed Models

Non-Linear Models

1. Parsimony
2. Interpretability
3. Model the mean structure

Mixed Models

1. Flexibility
2. Hierarchy
3. Model the error structure

Non-Linear and Mixed Models

Non-Linear Models

1. Parsimony
2. Interpretability
3. Model the mean structure

Mixed Models

1. Flexibility
2. Hierarchy
3. Model the error structure

Non-Linear and Mixed Models

Non-Linear Models

1. Parsimony
2. Interpretability
3. Model the mean structure

Mixed Models

1. Flexibility
2. Hierarchy
3. Model the error structure

Non-Linear and Mixed Models

Non-Linear Models

1. Parsimony
2. Interpretability
3. Model the mean structure

Mixed Models

1. Flexibility
2. Hierarchy
3. Model the error structure

Non-Linear and Mixed Models

Non-Linear Models

1. Parsimony
2. Interpretability
3. Model the mean structure

Mixed Models

1. Flexibility
2. Hierarchy
3. Model the error structure

Non-Linear and Mixed Models

Non-Linear Models

1. Parsimony
2. Interpretability
3. Model the mean structure

Mixed Models

1. Flexibility
2. Hierarchy
3. Model the error structure

Outline

Barley N response trials

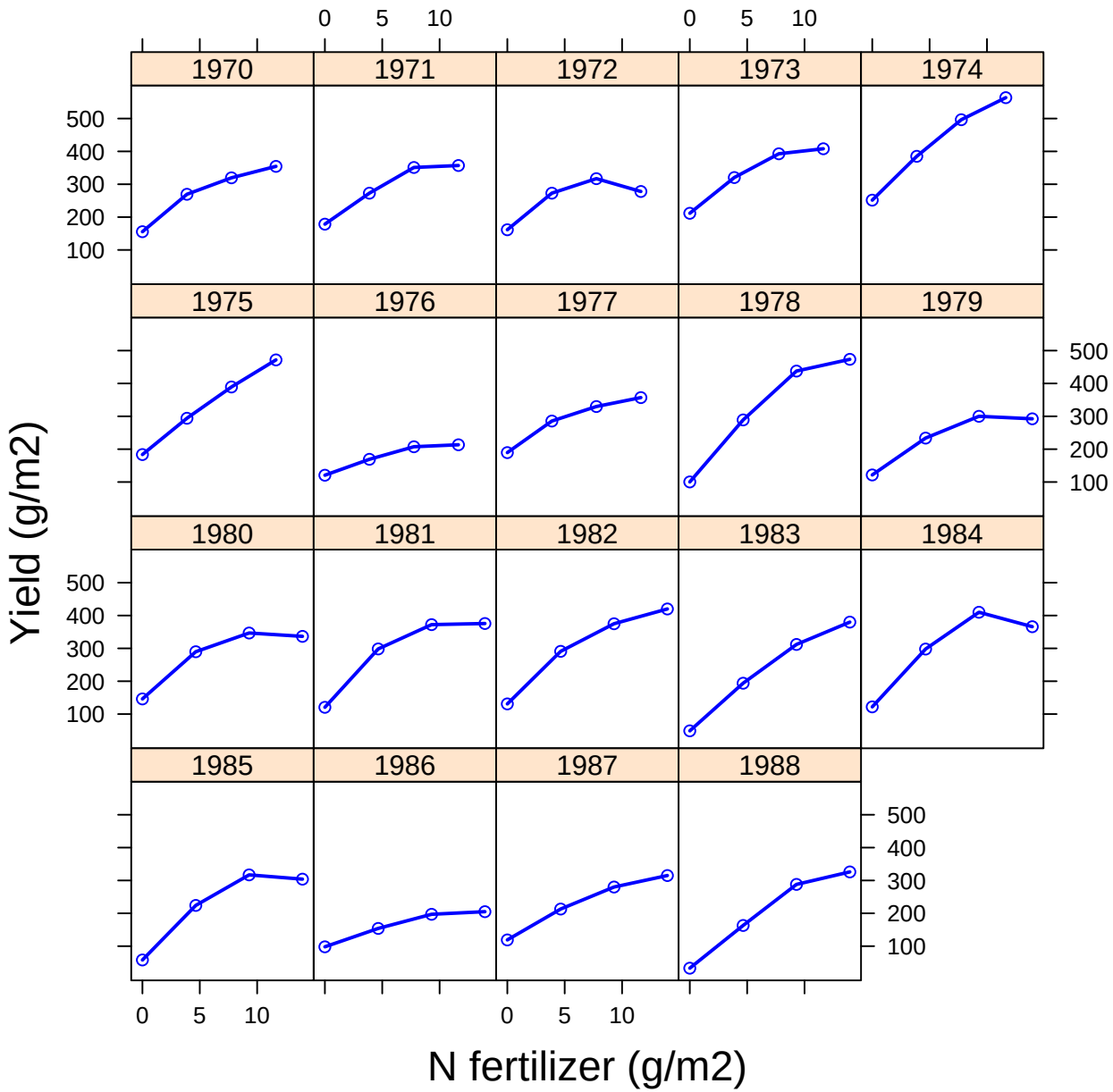
Aril Vold (1998). A generalization of ordinary yield response functions. Ecological Applications. 108:227-236.

Details

- 19 years of data, Norway
- N rates (0, 3.38, 7.76 and 11.69 g N m^{-2}) raised by 20% in 1978

Agronomic Questions

1. How does it respond to N?
2. How does it vary among years?


1970
1971
1972
1973
1974
1975
1976
1977
1978
1979
1980
1981
1982
1983
1984
1985
1986
1987
1988
Yield (g/m2)
N fertilizer (g/m2)
0
5
10
100
200
300
400
500

Outline

Basics of Statistical Models

$$y = f(x, \theta) + \epsilon$$

where,

$y =$ observed

$f =$ mean structure

$x =$ input

$\theta =$ parameters

$\epsilon =$ error

Basics of Statistical Models

$$y = f(x, \theta) + \epsilon$$

$$\mathcal{D} = \mathcal{M} + \mathcal{E}$$

where,

$y =$ observed

$f =$ mean structure

$x =$ input

$\theta =$ parameters

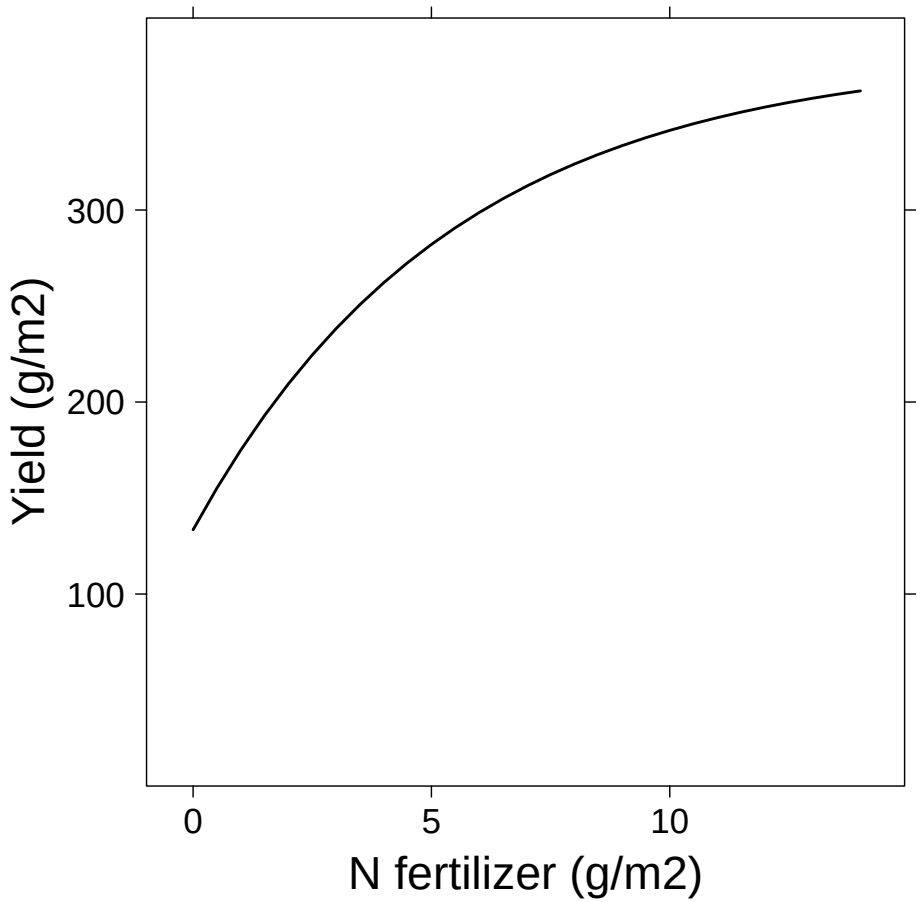
$\epsilon =$ error

Choosing the Mean Structure

Asymptotic Regression Model

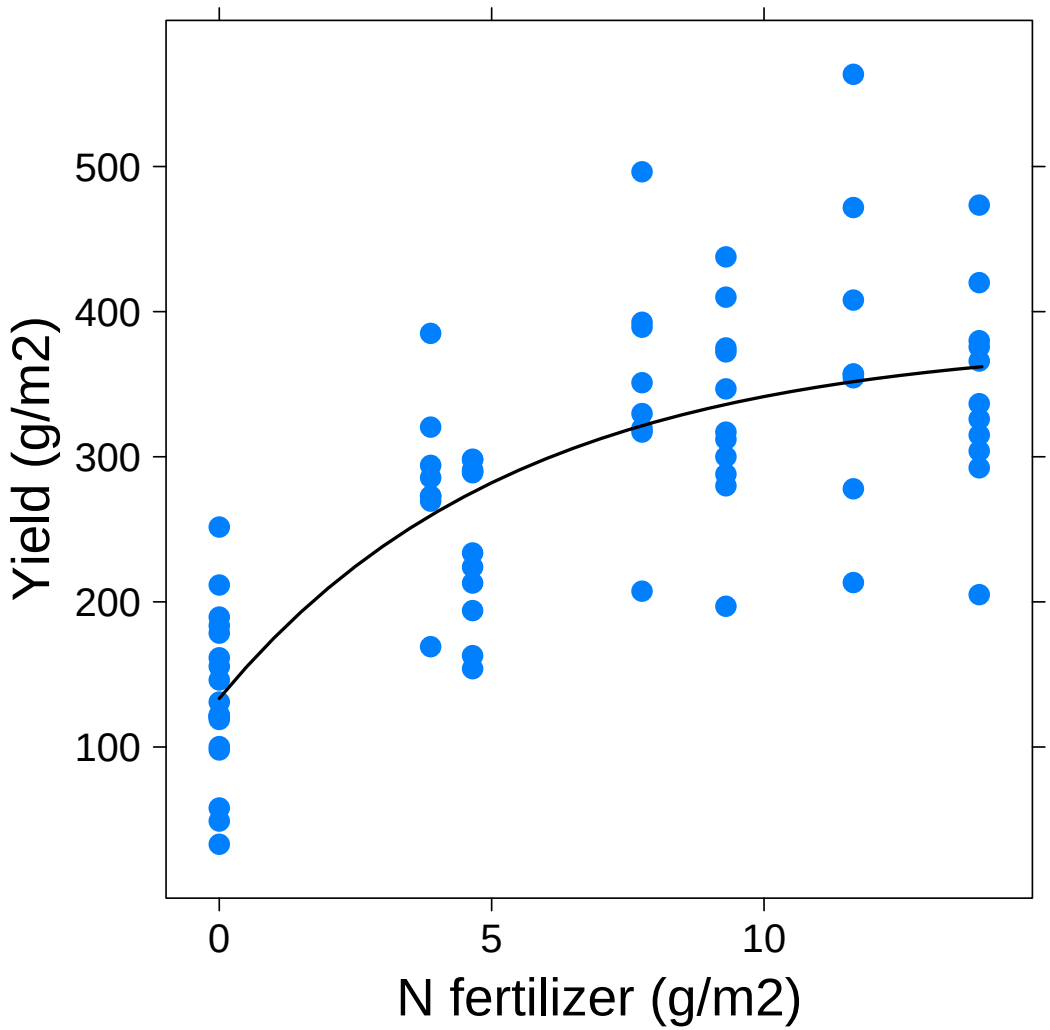
$$y = \theta_1 + (\theta_2 - \theta_1) \times \exp(-\exp(\theta_3) \times x)$$

where,
θ_1 is the maximum value of y
θ_2 is the value of y for $x = 0$.
θ_3 is the growth rate of y



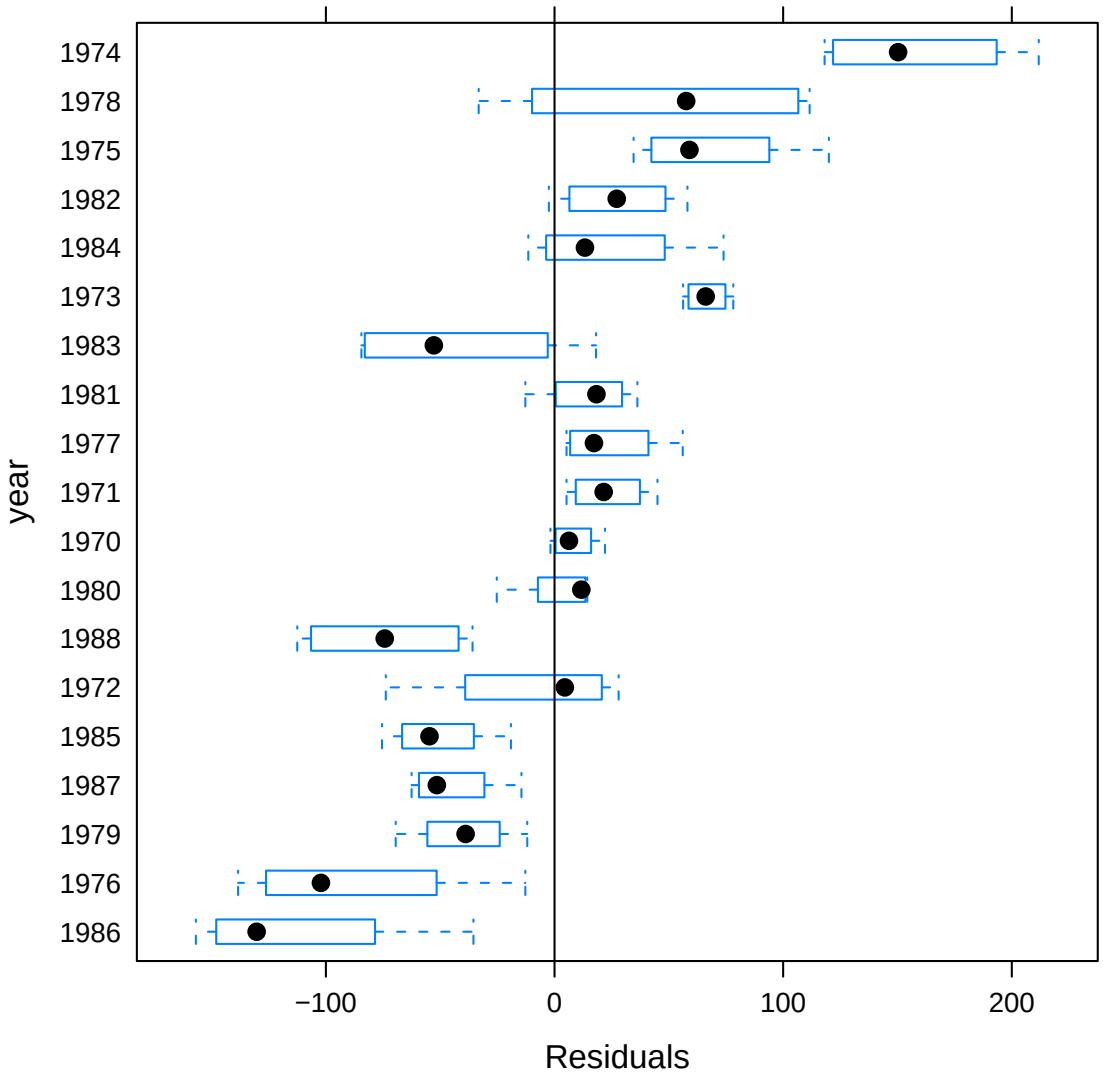
Barley N response trials

Non-linear regression with years combined

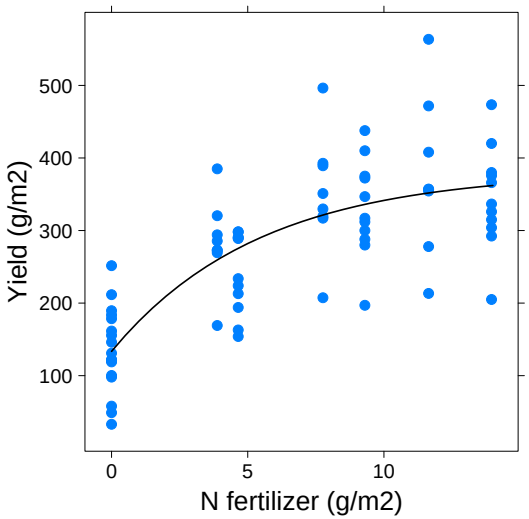


Barley N response trials

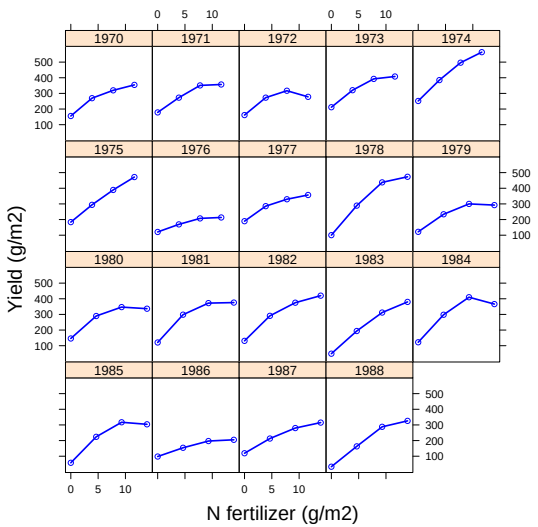
Box-plots of residuals for each year



Barley N response trials



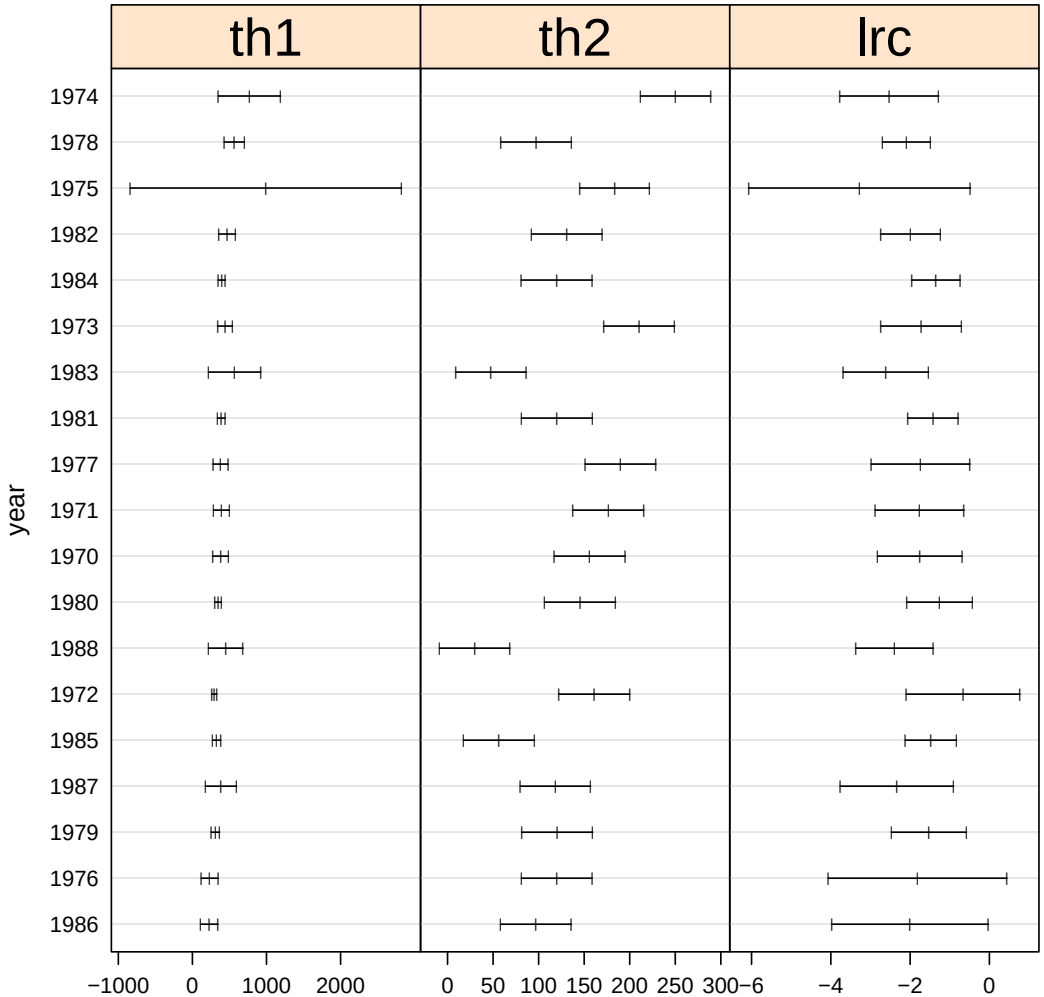
- One single regression to all the data
  - Wide confidence intervals
  - Ignores the structure of the data

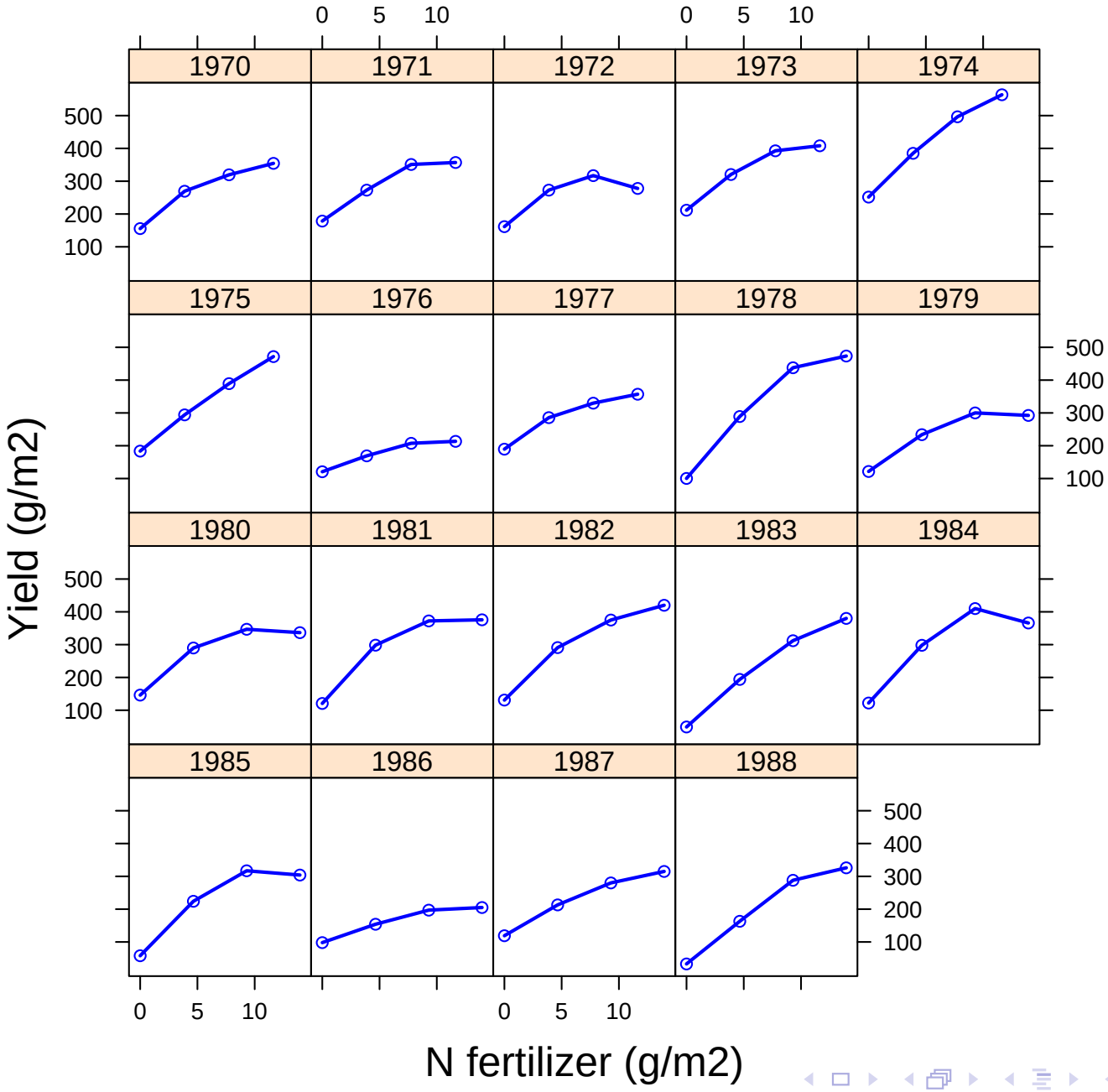


- Fitting one function for each separate year
  - Over-parameterized model
  - 3 parms $\times$ 19 y $=$ 57 parms

Barley N response trials

Confidence Intervals for Non-linear regressions for each year




1970
1971
1972
1973
1974
1975
1976
1977
1978
1979
1980
1981
1982
1983
1984
1985
1986
1987
1988
Yield (g/m2)
N fertilizer (g/m2)
0
5
10
100
200
300
400
500

Non-Linear Mixed Model

Asymptotic regression with random effects

$$y_{ij} = (\theta_1 + b_{1i}) + ((\theta_2 + b_{2i}) - (\theta_1 + b_{1i})) \times \exp(-\exp(\theta_3 + b_{3i}) \times x_{ij}) + \epsilon_{ij}$$

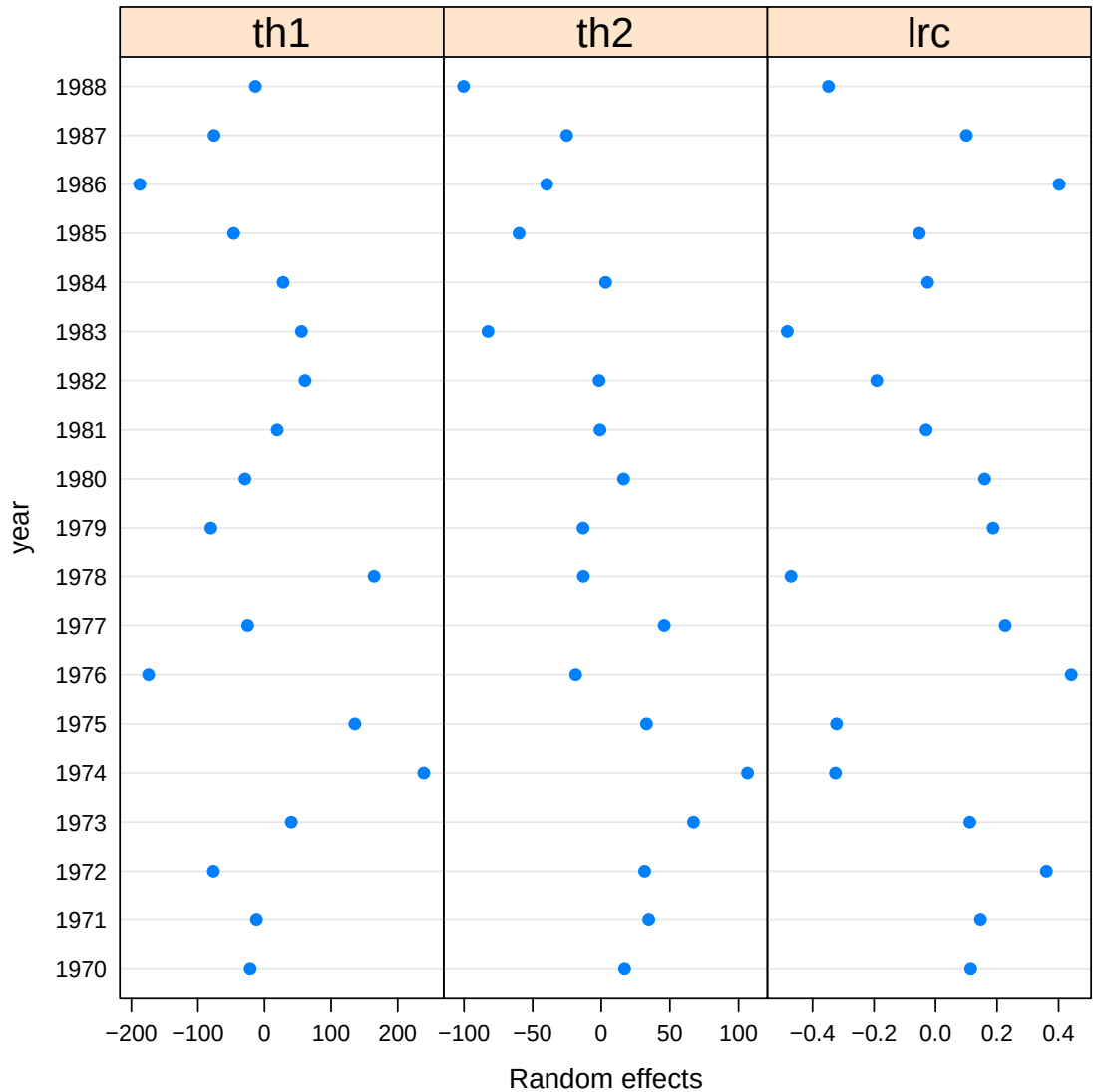
$i =$ the year (or experimental unit)
$j =$ the N rate

$$\boldsymbol{b_i} \sim \mathcal{N}(\mathbf{0}, \boldsymbol{\Psi}), \quad \epsilon_{ij} \sim \mathcal{N}(0, \sigma^2)$$

$$\boldsymbol{\Psi} = \begin{bmatrix} \sigma_{11} & \sigma_{12} & \sigma_{13} \\ \sigma_{21} & \sigma_{22} & \sigma_{23} \\ \sigma_{31} & \sigma_{32} & \sigma_{33} \end{bmatrix}$$

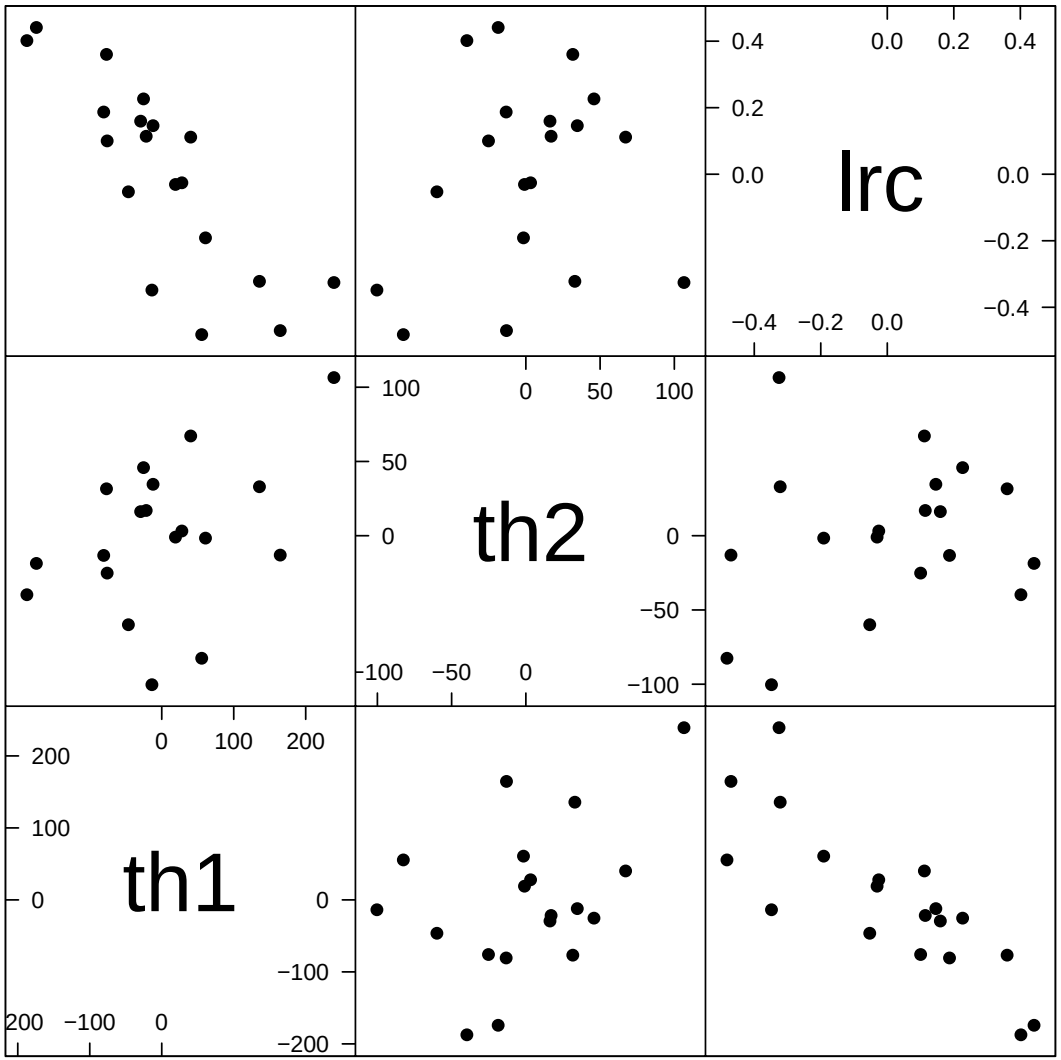
Random Effects

Dot plot for the random effects



Random Effects

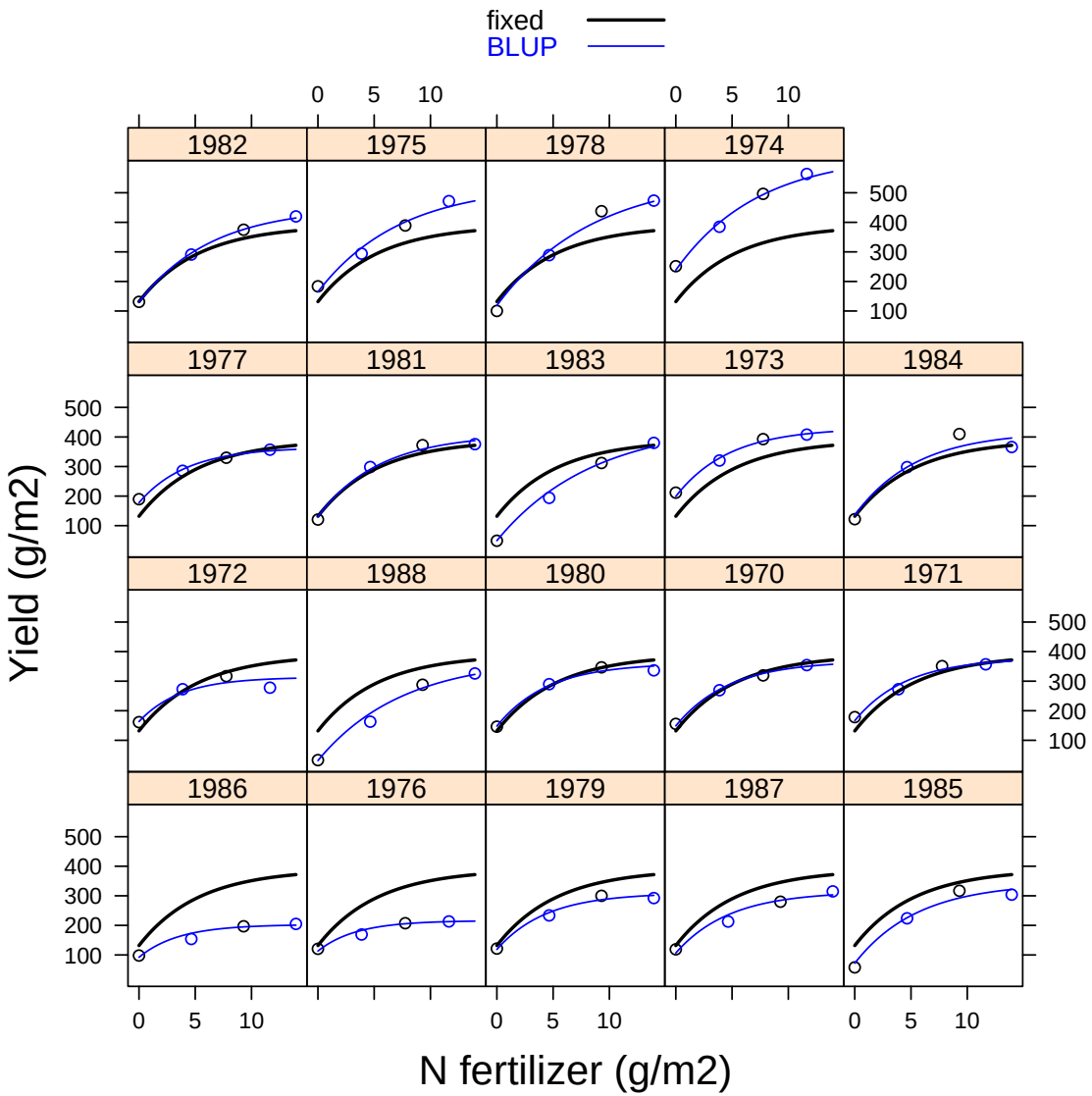
Scatter plot matrix for the random effects



Scatter Plot Matrix

Non-Linear Mixed Model

Fixed and BLUP



Comparison of NLS and NLME

Estimate, and 95% confidence intervals for the three parameters of the asymptotic regression model (NLS) and the mixed-effects model (NLME).

Fixed term	**Estimate**	**Lower**	**Upper**
θ_1 NLS	381	335	507
θ_1 NLME	390	337	443
θ_2 NLS	133	101	166
θ_2 NLME	132	107	157
lrc NLS	-1.7	-2.7	-1.1
lrc NLME	-1.7	-1.9	-1.4
$\hat{\sigma}$ NLS	71.2		
$\hat{\sigma}$ NLME	18.8	13.8	25.6

Summary: Using NLME

- NLME are able to accomodate the mean and error structure
- NLME produce a parsimonious and easy to interpret model
- The NLME estimates are more accurate and the confidence intervals are narrower

Summary: Using NLME

- NLME are able to accomodate the mean and error structure
- NLME produce a parsimonious and easy to interpret model
- The NLME estimates are more accurate and the confidence intervals are narrower

Summary: Using NLME

- NLME are able to accomodate the mean and error structure
- NLME produce a parsimonious and easy to interpret model
- The NLME estimates are more accurate and the confidence intervals are narrower

Outline

Application to Meta-analysis

Meta-analysis of the effects of management factors on Miscanthus x giganteus growth and biomass production. Miguez et al (2008) Agricultural and Forest Meteorology. 148:1280-1292.

R Code and Data

E-mail: miguez@illinois.edu

Website: https://netfiles.uiuc.edu/miguez/www

Questions?