A Semiparametric Stochastic Mixed Model for Increment-Averaged Data with Application to Carbon Sequestration in Agricultural Soils

F. Jay Breidt
Colorado State University

Joint work with Nan-Jung Hsu, National Tsing Hua University and
Stephen Ogle, Colorado State University

The work reported here was developed under STAR Research Assistance Agreements CR-829095 awarded by the U.S. Environmental Protection Agency (EPA) to Colorado State University. This presentation has not been formally reviewed by EPA. EPA does not endorse any products or commercial services mentioned in this report.

Space-**T**ime **A**quatic **R**esources **M**odeling & **A**nalysis **P**rogram

- **STARMAP:** EPA-funded STAR grant
  - study probability surveys of aquatic resources
  - CSU focuses on modeling side
  - partners with sister program at OSU (design side)
- Um, agricultural soils $\neq$ aquatic resources?
  - spin-off from other work on nonparametric survey regression estimation

Greenhouse Effect

- Solar energy transmitted to earth as visible and ultraviolet radiation

	Atmosphere	Surface
Reflected	25%	5%
Absorbed	25%	45%

- Radiation absorbed by surface gets re-radiated as infrared
- **Greenhouse Gases (GHGs)**
  - pass visible and UV, but trap infrared
  - include water vapor, CO_2, methane, nitrous oxide, others

Kyoto Protocol

- December 1997 meeting in Kyoto, Japan
- Resulted in Kyoto Protocol
  - signed by over 170 nations, including US
- Binding commitment for US to reduce emissions of six GHGs
  - reduce 7% from 1990 levels during 2008–2012
- US never ratified Kyoto
  - no commitment by developing nations
  - potentially large economic impact

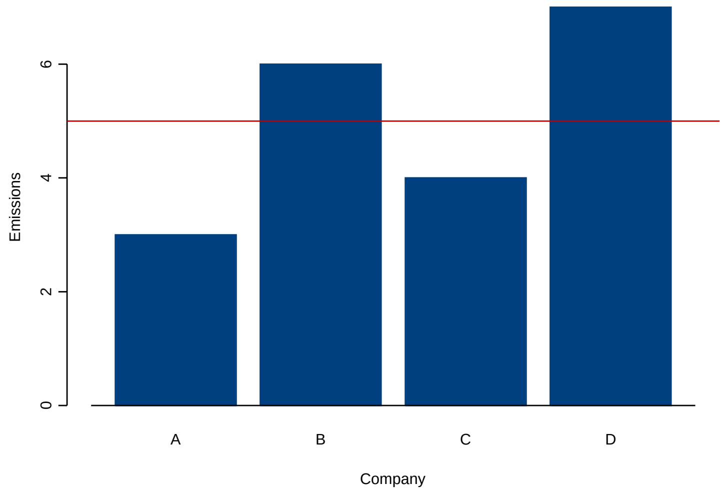
Chicago Climate Exchange

- www.chicagoclimatex.com
- Participants set voluntary limits on GHG emissions
  - legally binding commitments for reductions

	Commitments			
Baseline	2003	2004	2005	2006
avg. 1998–2001	-1%	-2%	-3%	-4%

Why Submit to Binding Commitments?

- Creates a "cap-and-trade" market
  - can make own reductions
  - can buy credits from others that have extra reductions to sell



Why Participate?

- Participants include Ford, Motorola, DuPont, City of Chicago, Waste Management, . . .
- Reduce long-term costs of carbon reduction
- Get financial benefits (e.g., reduced energy costs)
- Enhance environmental leadership reputation
- Build trading skills / help create market rules

Examples of GHG Mitigation and Offset Projects

- Renewal energy systems
  - wind power, solar power
- Energy efficient process innovations
- Recovery / use of landfill and agricultural methane
- Carbon sequestration
  - tree biomass
  - agricultural soils: **no-till agriculture**

Agriculture 001: Tillage

- Traditional Tillage:
  - after harvest, field contains crop residues
  - tillage turns over the soil to bury residues
  - often repeated several times prior to planting
- Conservation Tillage:
  - Reduced-Till: limited tillage; substantial crop residues on surface
  - No-Till (**NT**): doesn't use tillage; all crop residues left on surface
  - Economically feasible due to new technology

Advantages of No-Till Management

- NT results in lower production costs
  - fewer management steps
  - cheaper, lower horsepower tractors
- NT leaves crop residue on soil surface
  - reduces soil loss due to wind and water erosion
  - reduces flow of sediments, nutrients, and pesticides into surface waters
  - improves soil fertility; enhances soil organic matter
  - **sequesters carbon**

Carbon Sequestration via No-Till

- Original soil carbon content has been reduced 50% since invention of the steel plow
  - aeration due to tillage leads to faster decomposition
  - decomposition releases CO_2 to atmosphere
- No-till significantly slows decomposition
  - widespread adoption could restore original carbon within 40 years
  - could cut projected growth in US CO_2 emissions by 20%

Show Me the Money

- Carbon as a cash crop?
- Government could subsidize switch to no-till
- Corporations faced with caps on GHG emissions could buy **carbon credits**
  - some Canadian companies have already paid Iowa farmers for carbon credits
- Estimates of value range from \$4–\$40 per acre

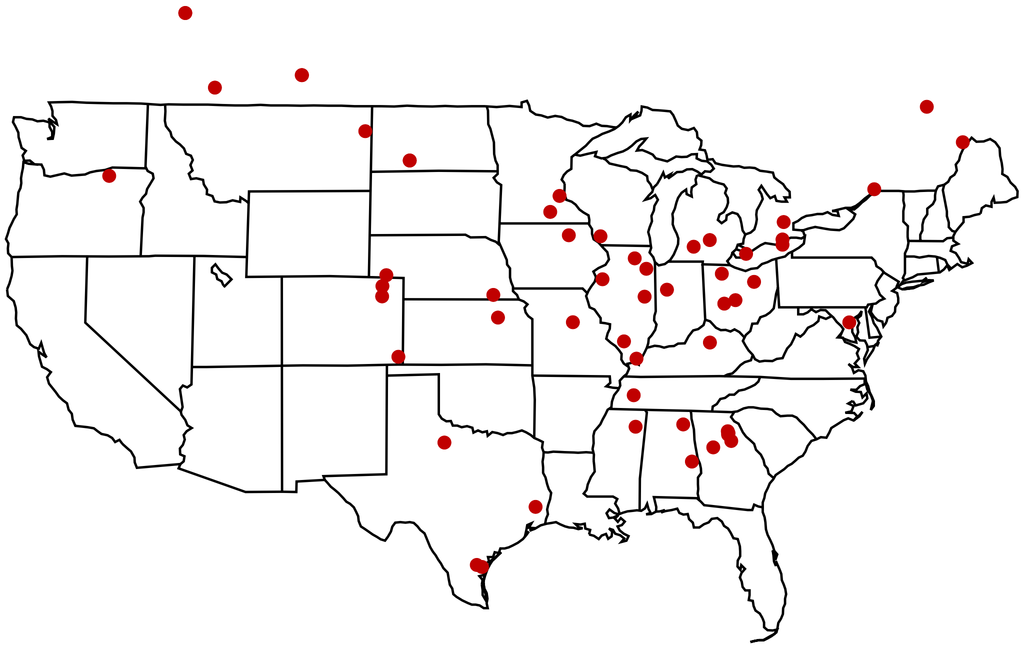
Key Question for Carbon Credits

- How much carbon is sequestered in switching to no-till?
- Several studies have compared no-till with traditional tillage on paired fields
- Measure carbon difference after one or more years since management change:

$$Y = (\text{no-till carbon}) - (\text{traditional tillage carbon})$$

Available North American Studies

- 63 studies compare tillage types on paired fields



Soil Core Samples

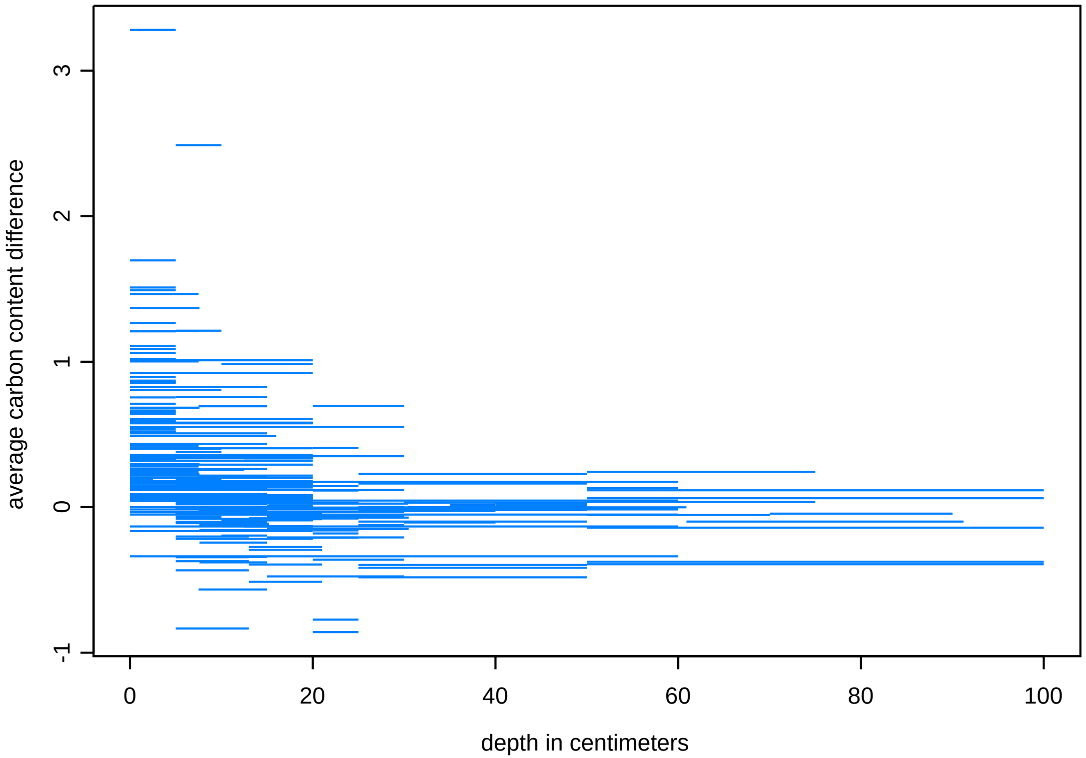
- Use probe to select one or more cores
- Separate cores into increments
  - may be fixed increments (e.g., 0–15 cm, 15–30 cm)
  - may be determined by soil profile (e.g., plow layer, A horizon, B horizon, C horizon)
- Mix matching increments in a bucket
- Bag a subsample and send to the lab

Aside on Core Data

- Like a time series, with depth replacing time
  - specifically, a **flow**: defined on interval (income, expenditure, precipitation)
  - not a **stock**: instantaneous (interest rate, temp)
- Other "cores":
  - ice cores: increment represents one year of snowfall
  - vertical ozone profiles
  - like time series, tend to be many, regularly spaced, relatively narrow increments
- Not true for soil cores!

Increment-Averaged Data

- Difference in metric tons C ha^{-1} vs. depth



Ad Hoc Methods: Midpoint Assignment

- True model is

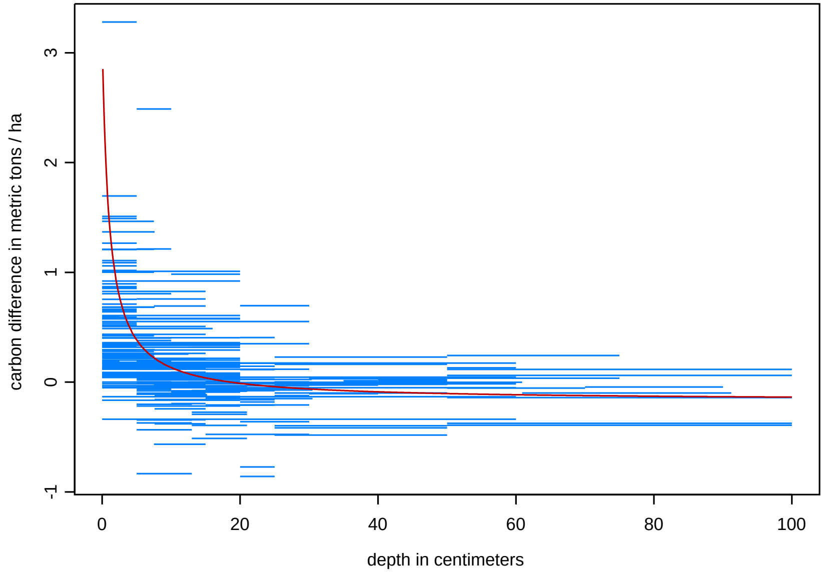
$$\begin{aligned} Y_{ij} &= \frac{1}{d_{ij} - d_{i,j-1}} \int_{d_{i,j-1}}^{d_{ij}} \sum_k \alpha_k g_k(t)\, dt + \epsilon_{ij} \\ &= \sum_k \alpha_k \frac{1}{d_{ij} - d_{i,j-1}} \int_{d_{i,j-1}}^{d_{ij}} g_k(t)\, dt + \epsilon_{ij} \\ &= \sum_k \alpha_k g_k(t^*_{ijk}) + \epsilon_{ij} \text{ where } t^*_{ijk} \in (d_{i,j-1}, d_{ij}) \end{aligned}$$

- But we regress on $g_k(m_{ij}) \neq g_k(t^*_{ijk})$: measurement error problem
- Least squares estimators are biased and inconsistent

Do We Need to Worry?

- Simulate from the fitted model

$$Y_{ij} = \frac{1}{d_{ij} - d_{i,j-1}} \int_{d_{i,j-1}}^{d_{ij}} (\alpha_0 + \alpha_1/(1+t))\ dt + \epsilon_{ij}$$



Simulation Results

- Simulate from true model, then estimate with midpoint assignment
  - use actual increments from data set
  - resample if necessary

sample size	**reps**	$\boldsymbol{\alpha_0 = -0.17}$	$\boldsymbol{\alpha_1 = 3.32}$
218	10,000	$\mathrm{E}[\hat{\alpha}_0] = -0.17$	$\mathrm{E}[\hat{\alpha}_1] = 4.21$
10,000	1	$\hat{\alpha}_0 = -0.17$	$\hat{\alpha}_1 = 4.18$

- Over 25% relative bias in slope estimate

Ad Hoc Methods: Adjustment of Increments

- Study 1 measures

$$Y_{11} = \text{carbon stock change over increment 0–15 cm}$$
$$Y_{12} = \text{carbon stock change over increment 15–30 cm}$$

- Study 2 measures

$$Y_{21} = \text{carbon stock change over increment 0–50 cm}$$

- "Adjustment" of Y-values

$$Y_1^* := Y_{11} + Y_{12} \text{ represents 0–30 cm}$$
$$Y_2^* := \frac{30}{50} Y_{21} \text{ represents 0–30 cm}$$

Ad Hoc Methods for Increments

- Midpoint assignment leads to bias, inconsistency
- "Adjustment" leads to loss of information, likely bias
- One final method: simply drop studies with non-matching increments
  - obvious loss of information
- Need to recognize the increment nature of the data

Key Data Features

- Increment averaging
  - irregular, wide
  - difficult to specify parametric model
- Within-core dependence
  - increments within same core may be correlated
- Other effects
  - time since change to no-till
  - climate regime
  - soil type

Semiparametric Stochastic Mixed Model: Longitudinal

- Zhang, Lin, Raz, Sowers (1998) *JASA*:

$$Y_{ij} = \boldsymbol{X}_{ij}^T \boldsymbol{\beta} + g(t_{ij}) + \boldsymbol{Z}_{ij}^T \boldsymbol{b}_i + U_i(t_{ij}) + \epsilon_{ij}$$

  - $\boldsymbol{\beta}$: p unknown regression coefficients
  - $\boldsymbol{X}_{ij}, \boldsymbol{Z}_{ij}$: known covariates
  - $g(t)$: twice-differentiable function of time
  - $\boldsymbol{b}_i$: independent q-vectors of random effects
  - $U_i(t)$: independent stochastic processes
  - ϵ_{ij}: independent errors
- Does not handle increment averages

Semiparametric Stochastic Mixed Model: Increments

- Increment average in ith core, jth increment

$$Y_{ij} = \boldsymbol{X}_{ij}^T \boldsymbol{\beta} + \frac{1}{d_{ij} - d_{i,j-1}} \int_{d_{i,j-1}}^{d_{ij}} g(t)\, dt + \boldsymbol{Z}_{ij}^T \boldsymbol{b}_i + U_{ij} + \epsilon_{ij}$$

  – $\boldsymbol{\beta}$: p unknown regression coefficients
  – $\boldsymbol{X}_{ij}, \boldsymbol{Z}_{ij}$: known covariates
  – $g(t)$: twice-differentiable function of depth
  – $\boldsymbol{b}_i$: independent q-vectors of random effects
  – U_{ij}: increment-averaged stochastic process
  – ϵ_{ij}: independent errors

Assumptions for Semiparametric Stochastic Mixed Model

- Define

$$\boldsymbol{U}_i = \left(U_{i1}, U_{i2}, \ldots, U_{in_i}\right)^T$$

- Assume
  - $\boldsymbol{U}_i$ are independent normal$(\mathbf{0}, \boldsymbol{\Gamma}_i)$ where $\boldsymbol{\Gamma}_i = \text{cov}\left(\boldsymbol{U}_i, \boldsymbol{U}_i; \boldsymbol{\xi}\right)$
  - $\boldsymbol{b}_i$ are independent normal$(\mathbf{0}, \boldsymbol{D}(\boldsymbol{\phi}))$ where $\boldsymbol{D}$ is a positive definite matrix
  - ϵ_{ij} are independent normal$(0, \sigma^2)$
  - $\boldsymbol{b}_i$, $\boldsymbol{U}_i$, and ϵ_{ij} are mutually independent

Integrated Stochastic Process Specification

- Increment-averaged form:

$$U_{ij} = \frac{1}{d_{ij} - d_{i,j-1}} \int_{d_{i,j-1}}^{d_{ij}} U_i(t)\, dt,$$

  where $U_i(t)$ are independent, mean zero, Gaussian stochastic processes

- Choices for instantaneous process:
  - Wiener process
  - integrated Wiener process
  - non-homogeneous Ornstein-Uhlenbeck process

Nonhomogeneous Ornstein-Uhlenbeck Process

- Covariance structure for instantaneous process:

$$\begin{aligned} \operatorname{var}(U_i(t)) &= \xi(t) = \exp\{\xi_0 + \xi_1 t\} \\ \operatorname{corr}(U_i(s), U_i(t)) &= \exp\{-\alpha|s-t|\} \end{aligned}$$

- Covariances for increment-averaged process:

$$\begin{aligned} &\operatorname{cov}(U_{ij}, U_{ik}) \\ &\quad = \frac{e^{\xi_0}}{(d_{ij} - d_{i,j-1})(d_{ik} - d_{i,k-1})} \\ &\quad \times \frac{e^{d_{ij}(\frac{\xi_1}{2}+\alpha)} - e^{d_{i,j-1}(\frac{\xi_1}{2}+\alpha)}}{\frac{\xi_1}{2}+\alpha} \frac{e^{d_{ik}(\frac{\xi_1}{2}-\alpha)} - e^{d_{i,k-1}(\frac{\xi_1}{2}-\alpha)}}{\frac{\xi_1}{2}-\alpha} \end{aligned}$$

Integrated Nonparametric Function Specification

- "Parameter" $g(t)$ is infinite-dimensional
- Assume $g(t)$ is **natural cubic spline**
  - knots at distinct right-hand endpoints
  - cubic polynomial between knots
  - g, g', g'' continuous at knots (hence everywhere)
  - linear before first knot, after last knot

Splitting Up the Integral

- Distinct right-hand endpoints $= \boldsymbol{t} = (t_1, t_2, \ldots, t_r)^T$
- Increment total:

$$\begin{aligned}
&\int_{d_1}^{d_2} g(t)\, dt \\
&= \int_{t_k=d_1}^{t_{k+1}} g(t)\, dt + \int_{t_{k+1}}^{t_{k+2}} g(t)\, dt + \cdots + \int_{t_{k+p-1}}^{t_{k+p}=d_2} g(t)\, dt \\
&= (0, \ldots, 0, 1, \ldots, 1, 0, \ldots, 0)\boldsymbol{G}
\end{aligned}$$

for some non-negative integer k and some positive integer p, where

$$\boldsymbol{G} = \left(\int_{t_0}^{t_1} g(t)\, dt, \int_{t_1}^{t_2} g(t)\, dt, \ldots, \int_{t_{r-1}}^{t_r} g(t)\, dt \right)^T$$

Value-Second Derivative Representation

- Define $g_i = g(t_i)$ and $\gamma_i = g''(t_i)$ for $i = 1, 2, \ldots, r$ $(\gamma_1 = \gamma_r = 0)$
- For $h_i = t_{i+1} - t_i$, we have

$$G_1 = \int_0^{t_1} g(t)\,dt = \left(t_1 + \frac{t_1^2}{2h_1}\right) g_1 - \frac{t_1^2}{2h_1} g_2 + \frac{h_1}{6}\frac{t_1^2}{2}\gamma_2$$

and

$$G_{i+1} = \int_{t_i}^{t_{i+1}} g(t)\,dt = \frac{h_i}{2}(g_i + g_{i+1}) - \frac{h_i^3}{24}(\gamma_i + \gamma_{i+1})$$

for $i = 1, 2, \ldots, r-1$

Ah, Linear

- So $\boldsymbol{G}$ is a linear function of $\boldsymbol{g} = (g_1, \ldots, g_r)^T$ and $\boldsymbol{\gamma} = (\gamma_2, \ldots, \gamma_{r-1})^T$:

$$\boldsymbol{G} = [\boldsymbol{A}_1, \boldsymbol{A}_2] \begin{bmatrix} \boldsymbol{g} \\ \boldsymbol{\gamma} \end{bmatrix} = [\boldsymbol{A}_1, \boldsymbol{A}_2] \begin{bmatrix} \boldsymbol{I} \\ \boldsymbol{R}^{-1}\boldsymbol{Q}^T \end{bmatrix} \boldsymbol{g},$$

  where $\boldsymbol{A}_1$, $\boldsymbol{A}_2$, $\boldsymbol{R}$ and $\boldsymbol{Q}$ are known matrices (depending only on $\{h_i\}$)

- In fact, $\boldsymbol{G}$ is linear function of $\boldsymbol{g}$ alone

Finite-Dimensional Model

- Integrals of infinite-dimensional parameter now reduced to r unknown constants:

$$
\begin{aligned}
\boldsymbol{Y}_i &= \boldsymbol{X}_i\boldsymbol{\beta} + \left[\frac{\int_{d_{i,j-1}}^{d_{ij}} g(t)\,dt}{d_{ij} - d_{i,j-1}}\right]_{j=1}^{n_i} + \boldsymbol{Z}_i\boldsymbol{b}_i + \boldsymbol{U}_i + \boldsymbol{\epsilon}_i \\
&= \boldsymbol{X}_i\boldsymbol{\beta} + \begin{bmatrix} \boldsymbol{c}_{i1}^T \\ \vdots \\ \boldsymbol{c}_{in_i}^T \end{bmatrix} \boldsymbol{G} + \boldsymbol{Z}_i\boldsymbol{b}_i + \boldsymbol{U}_i + \boldsymbol{\epsilon}_i \\
&= \boldsymbol{X}_i\boldsymbol{\beta} + \boldsymbol{N}_i\boldsymbol{g} + \boldsymbol{Z}_i\boldsymbol{b}_i + \boldsymbol{U}_i + \boldsymbol{\epsilon}_i
\end{aligned}
$$

Likelihood

- Log-likelihood given covariance parameters $\boldsymbol{\phi}$, $\boldsymbol{\xi}$, and σ^2

$$\ell(\boldsymbol{\beta}, \boldsymbol{g}; \boldsymbol{Y}) = -\frac{1}{2}\log|\boldsymbol{V}|$$
$$-\frac{1}{2}(\boldsymbol{Y} - \boldsymbol{X}\boldsymbol{\beta} - \boldsymbol{N}\boldsymbol{g})^T\boldsymbol{V}^{-1}(\boldsymbol{Y} - \boldsymbol{X}\boldsymbol{\beta} - \boldsymbol{N}\boldsymbol{g})$$

where

– $\boldsymbol{V} = \mathrm{diag}\{\boldsymbol{V}_1, \boldsymbol{V}_2, \ldots, \boldsymbol{V}_m\}$

– $\boldsymbol{V}_i = \boldsymbol{Z}_i\boldsymbol{D}(\boldsymbol{\phi})\boldsymbol{Z}_i^T + \boldsymbol{\Gamma}_i + \sigma^2\boldsymbol{I}$

Penalized Likelihood

- Log-likelihood given covariance parameters $\boldsymbol{\phi}$, $\boldsymbol{\xi}$, and σ^2

$$\ell(\boldsymbol{\beta}, \boldsymbol{g}; \boldsymbol{Y}) = -\frac{1}{2}\log|\boldsymbol{V}|$$
$$-\frac{1}{2}(\boldsymbol{Y} - \boldsymbol{X}\boldsymbol{\beta} - \boldsymbol{N}\boldsymbol{g})^T \boldsymbol{V}^{-1}(\boldsymbol{Y} - \boldsymbol{X}\boldsymbol{\beta} - \boldsymbol{N}\boldsymbol{g})$$

- Penalized likelihood:

$$\ell(\boldsymbol{\beta}, \boldsymbol{g}; \boldsymbol{Y}) - \frac{\lambda}{2}\int_0^{t_r} \left[g''(t)\right]^2 \, dt$$

where $\lambda > 0$ controls smoothness

– $\lambda \to 0$ implies interpolating

– $\lambda \to \infty$ implies global regression

Penalized Likelihood, Continued

- Penalized likelihood can be rewritten:

  $$\ell(\boldsymbol{\beta}, \boldsymbol{g}; \boldsymbol{Y}) - \frac{\lambda}{2} \int_0^{t_r} \left[g''(t) \right]^2 \, dt = \ell(\boldsymbol{\beta}, \boldsymbol{g}; \boldsymbol{Y}) - \frac{\lambda}{2} \boldsymbol{g}^T \boldsymbol{K} \boldsymbol{g},$$

  where $\boldsymbol{K} = \boldsymbol{Q} \boldsymbol{R}^{-1} \boldsymbol{Q}^T$ is known

- Now differentiate with respect to $\boldsymbol{\beta}$ and $\boldsymbol{g}$

Maximum Penalized Likelihood Estimators

- Assume λ as well as $\boldsymbol{\phi}$, $\boldsymbol{\xi}$, and σ^2 are given
- MPLEs solve the following linear equations:

$$\begin{bmatrix} \boldsymbol{X}^T\boldsymbol{V}^{-1}\boldsymbol{X} & \boldsymbol{X}^T\boldsymbol{V}^{-1}\boldsymbol{N} \\ \boldsymbol{N}^T\boldsymbol{V}^{-1}\boldsymbol{X} & \boldsymbol{N}^T\boldsymbol{V}^{-1}\boldsymbol{N} + \lambda\boldsymbol{K} \end{bmatrix} \begin{bmatrix} \boldsymbol{\beta} \\ \boldsymbol{g} \end{bmatrix} = \begin{bmatrix} \boldsymbol{X}^T\boldsymbol{V}^{-1}\boldsymbol{Y} \\ \boldsymbol{N}^T\boldsymbol{V}^{-1}\boldsymbol{Y} \end{bmatrix}$$

- Unique solution?

Uniqueness of MPLEs

- By Theorem 4.1 of Green and Silverman (1994), unique solution if $[\boldsymbol{X}, \boldsymbol{N}\mathbf{1}, \boldsymbol{N}\boldsymbol{t}]$ has full column rank
- **Result:** If $\theta_0 \mathbf{1} + \theta_1 \boldsymbol{m} \notin \mathrm{Col}(\boldsymbol{X})$, where $\boldsymbol{m}$ is vector of increment midpoints, then

$$\begin{bmatrix} \boldsymbol{X}^T \boldsymbol{V}^{-1} \boldsymbol{X} & \boldsymbol{X}^T \boldsymbol{V}^{-1} \boldsymbol{N} \\ \boldsymbol{N}^T \boldsymbol{V}^{-1} \boldsymbol{X} & \boldsymbol{N}^T \boldsymbol{V}^{-1} \boldsymbol{N} + \lambda \boldsymbol{K} \end{bmatrix} \begin{bmatrix} \boldsymbol{\beta} \\ \boldsymbol{g} \end{bmatrix} = \begin{bmatrix} \boldsymbol{X}^T \boldsymbol{V}^{-1} \boldsymbol{Y} \\ \boldsymbol{N}^T \boldsymbol{V}^{-1} \boldsymbol{Y} \end{bmatrix}$$

  has a unique solution $(\hat{\boldsymbol{\beta}}, \hat{\boldsymbol{g}}) = \mathrm{MPLE}$

Yet Another Representation: Linear Mixed Model

- From Green (1987) *ISR*, it turns out that

$$\boldsymbol{g} = (\mathbf{1}, \boldsymbol{t})\boldsymbol{\delta}_{2\times 1} + \boldsymbol{B}\boldsymbol{a}_{(r-2)\times 1},$$

  where $\boldsymbol{B}$ is a known matrix derived from $\boldsymbol{K}$

- Linear mixed model representation:

$$\begin{aligned}\boldsymbol{Y} &= [\boldsymbol{X}, \boldsymbol{N}\boldsymbol{T}]\begin{bmatrix}\boldsymbol{\beta} \\ \boldsymbol{\delta}\end{bmatrix} + \boldsymbol{N}\boldsymbol{B}\boldsymbol{a} + \boldsymbol{Z}\boldsymbol{b} + \boldsymbol{U} + \boldsymbol{\epsilon} \\ &= \boldsymbol{X}_*\boldsymbol{\beta}_* + \boldsymbol{N}\boldsymbol{B}\boldsymbol{a} + \boldsymbol{Z}\boldsymbol{b} + \boldsymbol{U} + \boldsymbol{\epsilon},\end{aligned}$$

  where $\boldsymbol{a}$ is normal$(\mathbf{0}, \tau\boldsymbol{I})$ and $\tau = 1/\lambda$

  – Tricky: *not* the data-generating mechanism

- "BLUE" of $\boldsymbol{\beta}$ and "BLUP" of $\boldsymbol{g}$ from mixed model = MPLEs

Estimation of Covariance and Smoothing Parameters

- Use REML = Restricted Maximum Likelihood
  - standard method for linear mixed models
- Define $\boldsymbol{V}_* = \tau(\boldsymbol{N}\boldsymbol{B})(\boldsymbol{N}\boldsymbol{B})^T + \boldsymbol{V}$
- REML criterion for $(\boldsymbol{\phi}^T, \boldsymbol{\xi}^T, \sigma^2, \tau)$ is

$$\begin{aligned}
&\ell_R((\boldsymbol{\phi}^T, \boldsymbol{\xi}^T, \sigma^2, \tau); \boldsymbol{Y}) \\
&\quad = -\frac{1}{2}\log|\boldsymbol{V}_*| - \frac{1}{2}\log|\boldsymbol{X}_*^T\boldsymbol{V}_*^{-1}\boldsymbol{X}_*| \\
&\qquad -\frac{1}{2}(\boldsymbol{Y} - \boldsymbol{X}_*\hat{\boldsymbol{\beta}}_*)^T\boldsymbol{V}_*^{-1}(\boldsymbol{Y} - \boldsymbol{X}_*\hat{\boldsymbol{\beta}}_*)
\end{aligned}$$

- Given REML estimates, MPLEs are immediate

Back to the Carbon Sequestration Data

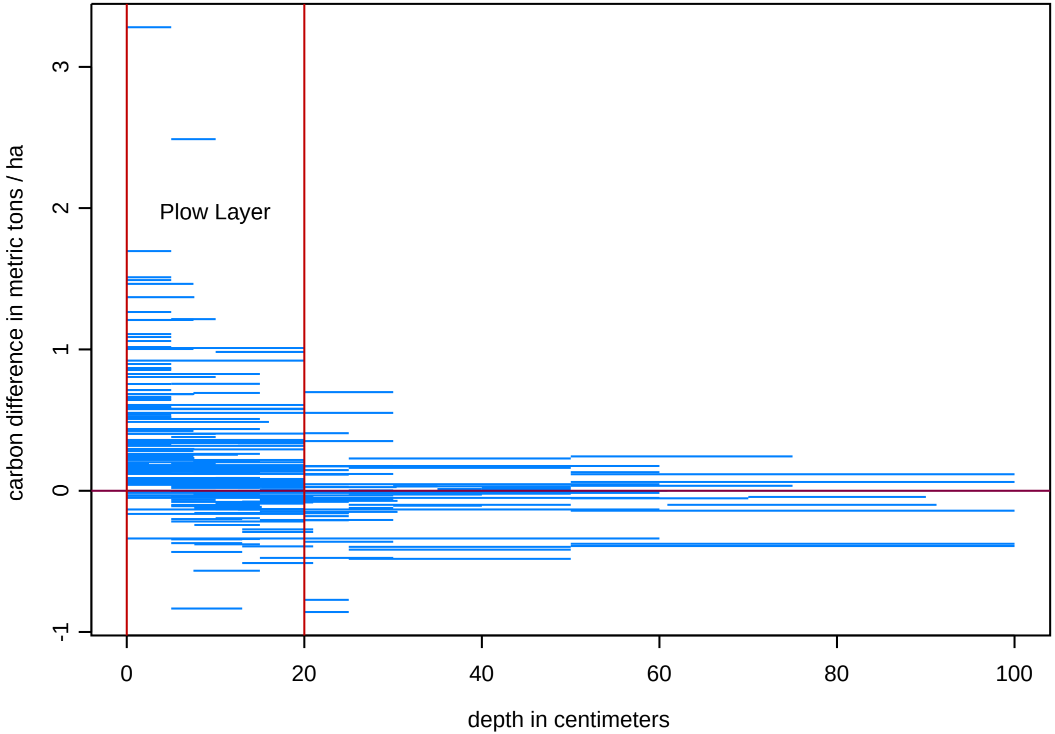
- Now have all the tools needed for inference in the carbon data set
- **Goals:**
  - identify important fixed effects
  - identify important sources of variation and correlation
  - estimate depth function
  - estimate expected carbon sequestered due to no-till:

$$\mathrm{IPCC} = \int_0^{30} \left(\boldsymbol{X}_{00}^T \boldsymbol{\beta} + g(t) \right) dt,$$

  where $\boldsymbol{X}_{00}$ = covariates at 20 years after management change

The Plow Layer

- Top 15–20 cm of agricultural soil



Model Specification

- Fixed effects:
  - wet/dry climate
  - aquic/non-aquic soils
  - years since management change
  - parametric or nonparametric year*depth interactions
- Random effect: soil core
- Stochastic process: non-homogeneous OU

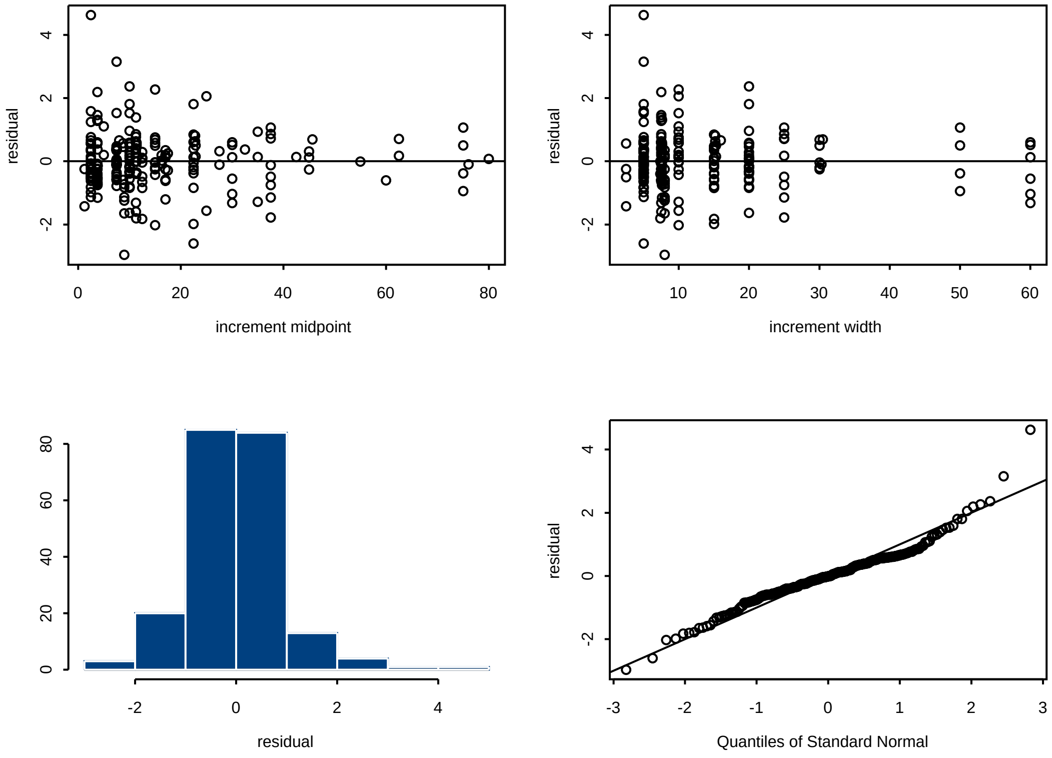
Some Results

- Estimation results from REML and MPLE:

	Model 1		Model 2	
Parameter	Estimate	Standard Error	Estimate	Standard Error
aquic	-0.1383	0.0587	-0.1371	0.0592
wet	0.1725	0.0593	0.1720	0.0596
years	0.0157	0.0047	0.0031	0.0032
year*depth	-0.0003	0.0001		
ϕ	0.0078	0.0077	0.0094	0.0093
σ^2	0.0073	0.0091	0.0064	0.0102
ξ_0	-0.8260	0.2368	-0.7911	0.2463
ξ_1	-0.0601	0.0205	-0.0643	0.0219
α	0.2607	0.0447	0.2693	0.0490
$\tau = \lambda^{-1}$	0.0326	0.0046	0.0189	0.0026
$\tau_x = \lambda_x^{-1}$			0.00005	0.000008

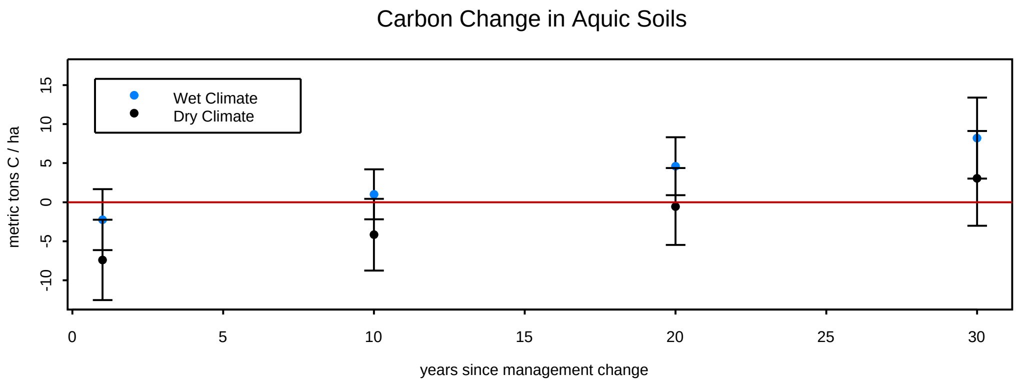
Residual Diagnostics

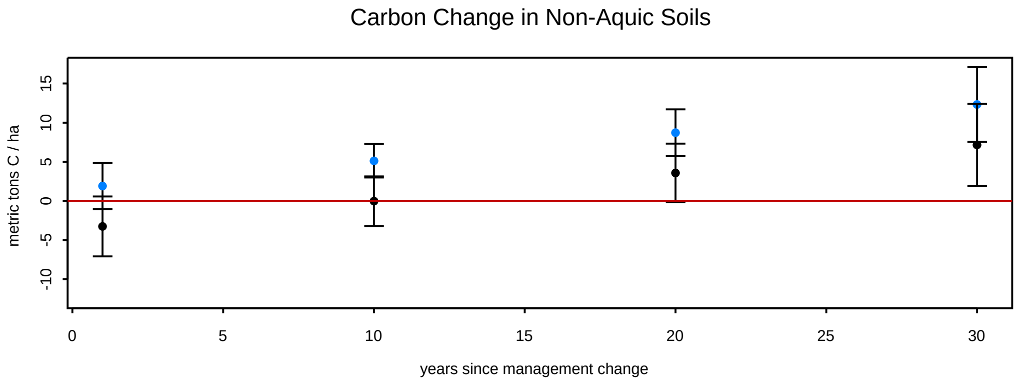
- Accounting for OU dependence structure



Intergovernmental Panel on Climate Change Integrals

- Estimated IPCC integrals





Summary and Future Research

- Increment averaging cannot be ignored in soil core data
- Semiparametric stochastic mixed model
  - flexibly models increment averages
  - handles fixed and random effects
  - fits in standard linear mixed model framework
- Further work:
  - more modeling and diagnostics for carbon data
  - extension to generalized linear mixed models