This research is funded by
U.S.EPA – Science To Achieve Results (STAR) Program
Cooperative Agreement # CR - 829095

A Semiparametric 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 Steve 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.

Outline

- Carbon sequestration in agricultural soils
  - greenhouse gases and carbon credits
  - soil sampling and increment-averaged data
- Semiparametric mixed models
  - integrated cubic splines
  - penalized splines
- Analysis using mixed model formulation
  - fitting with standard software
  - automatic smoothing parameter selection
  - results and extensions

Greenhouse Effect

- Solar energy transmitted to earth as visible and ultraviolet radiation

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

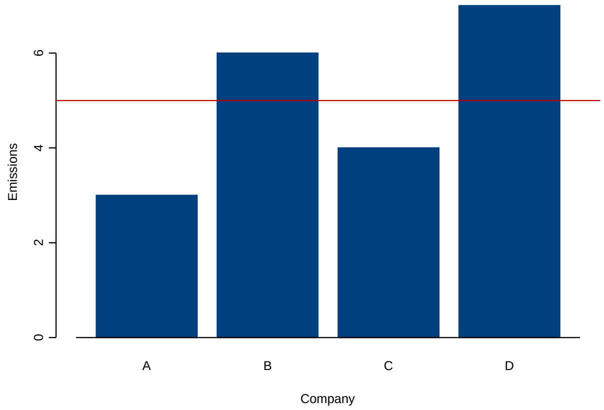
- Radiation absorbed by surface gets re-radiated as infrared
- Greenhouse Gases (GHGs)
  - pass visible and UV, but trap infrared
  - include water vapor, CO_2, methane, nitrous oxide, etc.

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

www.chicagoclimatex.com

- Participants voluntarily commit to GHG reductions
- Creates a "cap-and-trade" market
  - can make own reductions
  - can buy credits from others with extra reductions to sell



Extra Reductions from No-Till Agriculture

- 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

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** by slowing decomposition

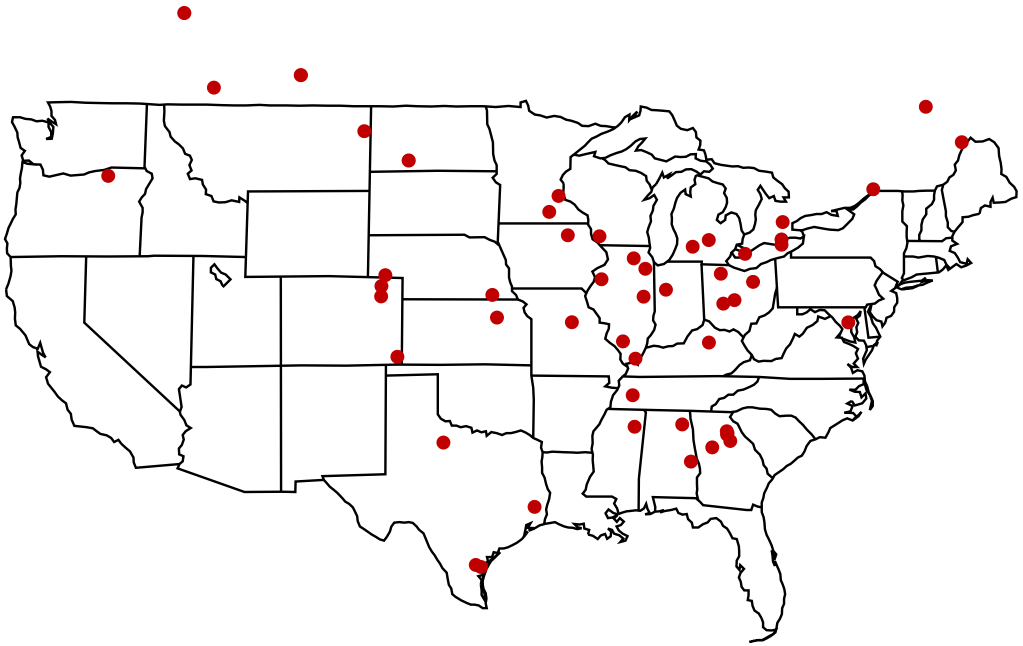
Key Question for Carbon Credits

- Corporations faced with caps on GHG emissions could buy **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

- 211 tillage comparisons on paired fields from 63 studies

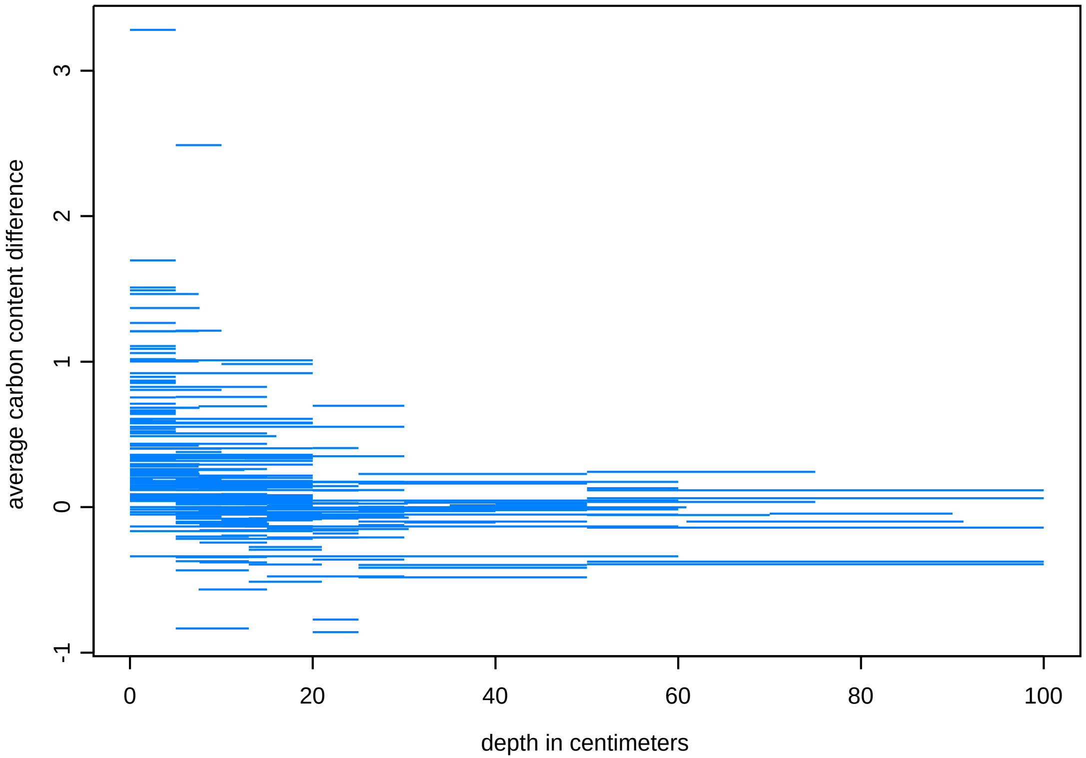


Soil Core Samples

- 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

Increment-Averaged Data

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



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-site dependence
  - increments within same site may be correlated
- Other effects
  - time since change to no-till
  - climate regime
  - soil type

Semiparametric Mixed Model: Longitudinal

- Observation on subject i at jth time point

$$Y_{ij} = \boldsymbol{x}_{ij}^T \boldsymbol{\beta} + g(t_{ij}; \boldsymbol{w}_i) + \boldsymbol{z}_{ij}^T \boldsymbol{b}_i + \epsilon_{ij}$$

  - e.g., Zhang, Lin, Raz, Sowers (1998) *JASA*
  - $\boldsymbol{x}_{ij}, \boldsymbol{w}_i, \boldsymbol{z}_{ij}$: known covariates
  - $g(t; \boldsymbol{w}_i)$: smooth function of time
  - $\boldsymbol{b}_i$: independent q-vectors of random effects
  - ϵ_{ij}: independent errors
- Does not handle increment averages

Semiparametric Stochastic Mixed Model: Increments

- Increment average in ith core, jth increment

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

  - $\boldsymbol{x}_{ij}, \boldsymbol{w}_i, \boldsymbol{z}_{ij}$: known covariates
  - $g(t; \boldsymbol{w}_i)$: smooth function of depth
  - $\boldsymbol{b}_i$: independent q-vectors of random effects
  - ϵ_{ij}: independent errors

Additive Varying Coefficient Models

- Smooth component functions of depth:

$$\alpha_\ell(t) = \alpha_{0\ell} + \alpha_{1\ell} t + \sum_{k=1}^{K} a_{k\ell}(t - \kappa_k)_+$$

  - truncated line basis with knots $\{\kappa_k\}_{k=1}^{K}$
  - penalize unnecessary slope changes via $\lambda_\ell^2 \sum_{k=1}^{K} a_{k\ell}^2$

- Additive depth function with varying coefficients:

$$g(t; \boldsymbol{w}) = \sum_\ell \alpha_\ell(t) w_\ell$$

  - (Hastie and Tibshiranie, 1993; Ruppert, Wand, and Carroll, 2003)

Mixed Model Formulation of Penalized Splines

- Integrated version of varying-coefficient model:

$$Y_{ij} = \boldsymbol{x}_{ij}^T\boldsymbol{\beta} + \sum_{\ell}\left(\alpha_{0\ell} + \alpha_{1\ell}\frac{d_{i,j-1} + d_{ij}}{2}\right) w_{\ell i}$$
$$+ \sum_{\ell}\left(\sum_{k=1}^{K}\frac{a_{k\ell}}{2}\left\{(d_{ij} - \kappa_k)_+^2 - (d_{i,j-1} - \kappa_k)_+^2\right\}\right) w_{\ell i}$$
$$+ \boldsymbol{z}_{ij}^T\boldsymbol{b}_i + \epsilon_{ij}$$

- Regard spline as linear mixed model with $a_{k\ell}$ iid $\mathrm{N}(0, \sigma_{a\ell}^2)$, and $\sigma_{a\ell}^2 = \sigma^2/\lambda_\ell^2$
  - automatic choice of penalty constants λ_ℓ via REML

Back to the Carbon Sequestration Data

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

$$\text{IPCC} = \int_0^{30} \left(\boldsymbol{x}^T \boldsymbol{\beta} + g(t; \boldsymbol{w}) \right) \, dt,$$

$\boldsymbol{x}, \boldsymbol{w} =$ covariates at 20 years after management change

Model Specification

- Available data: 211 paired increments from 63 studies
- Covariates:
  - `study`: study/site identifier
  - `wet`: wet/dry climate
  - `aquic`: aquic/non-aquic soils
  - `years` since management change
- Nonparametric functions:
  - parametric term $\boldsymbol{x}_{ij}^T\boldsymbol{\beta}$ may not be so useful
  - focus on varying coefficient models (parametric-by-nonparametric interactions)

Model Selection: Focus on Interaction Models

depth	aquic*depth	wet*depth	years*depth	AIC	r
	X			323	0.609
	X	X		226	0.797
			X	226	0.789
X	X			218	0.818
X	X	X	X	217	0.854
X				216	0.809
X	X		X	216	0.835
	X	X		216	0.812
		X		214	0.808
X	X	X		212	0.834
X			X	211	0.825
X		X		209	0.828
X		X	X	200	0.846
	X	X	X	193	0.851
		X	X	191	0.846

Final Model

- AIC suggests the following model:

$$
\begin{aligned}
g(t;\boldsymbol{w}_i) &= \Big(\alpha_{01} + \alpha_{11} t + \textstyle\sum_{k=1}^{K} a_{k1}(t-\kappa_k)_+\Big)\,\texttt{wet}_i \\
&\quad + \Big(\alpha_{02} + \alpha_{12} t + \textstyle\sum_{k=1}^{K} a_{k2}(t-\kappa_k)_+\Big)\,\texttt{years}_i \\
Y_{ij} &= \frac{1}{d_{ij} - d_{i,j-1}} \int_{d_{i,j-1}}^{d_{ij}} g(t;\boldsymbol{w}_i)\,dt + b_i + \epsilon_{ij}
\end{aligned}
$$

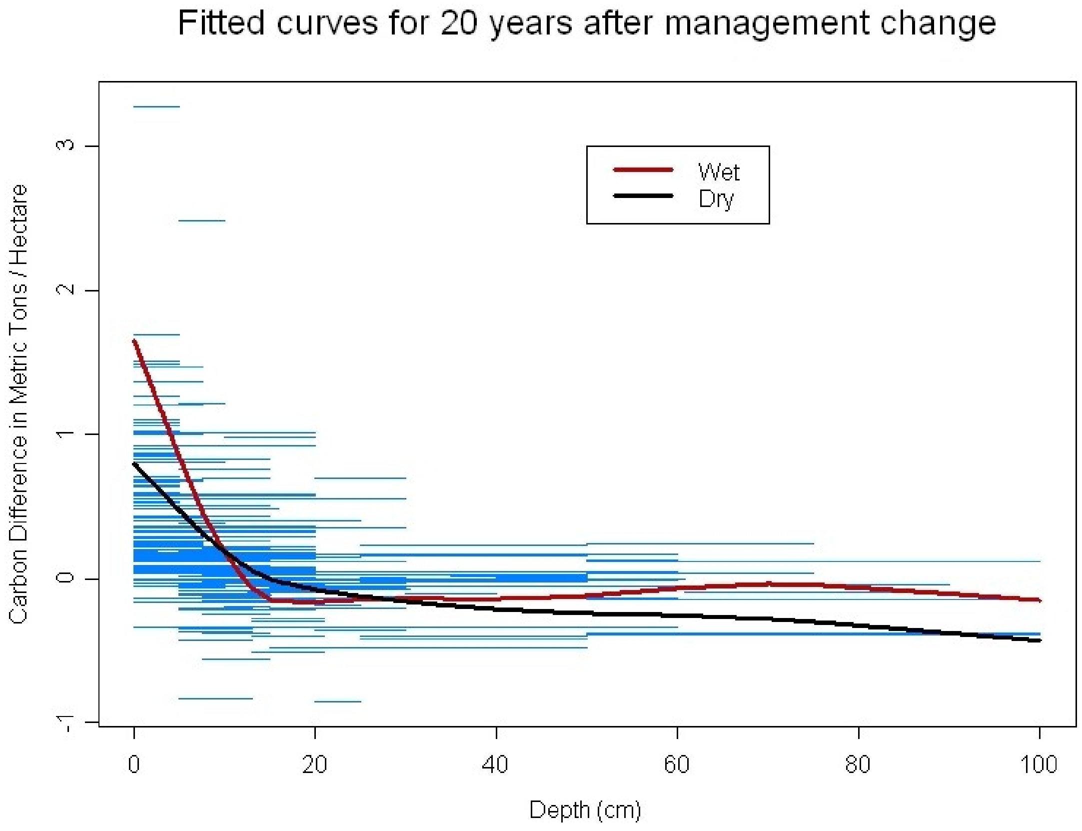
$\{a_{k\ell}\}$ iid $\mathsf{N}(0,\sigma^2_{a\ell})$, $\{b_i\}$ iid $\mathsf{N}(0,\sigma^2_b)$, $\{\epsilon_{ij}\}$ iid $\mathsf{N}(0,\sigma^2)$

S-Plus Code

```
#
# Following Ngo and Wand (2003), Smoothing with mixed model software.
# Set up linear splines, then integrate over increments.
#
Z1 <- outer(d1, knots, "-")
Z1 <- Z1 * (Z1 > 0)
Z2 <- outer(d2, knots, "-")
Z2 <- Z2 * (Z2 > 0)
Z <- 0.5 * diag(1/(d2 - d1)) %*% (Z2^2 - Z1^2)
#
# Create design matrices.
#
X <- cbind(wet, middepth*wet, years, middepth*years)
Z <- cbind(wet*Z, years*Z, study)
re.block.inds <- list(1:K, (K + 1):(2 * K), (2 * K + 1):(2 * K + nstudy))
Z.block <- list()
for(i in 1:length(re.block.inds))
    Z.block[[i]] <- as.formula(paste("~Z[,c(", paste(re.block.inds[[
        i]], collapse = ","), ")]-1"))
#
# Fit the mixed model using lme.
#
fit <- lme(y ~ -1 + X, random = pdBlocked(Z.block, pdClass = "pdIdent"),
    method = "REML")
```

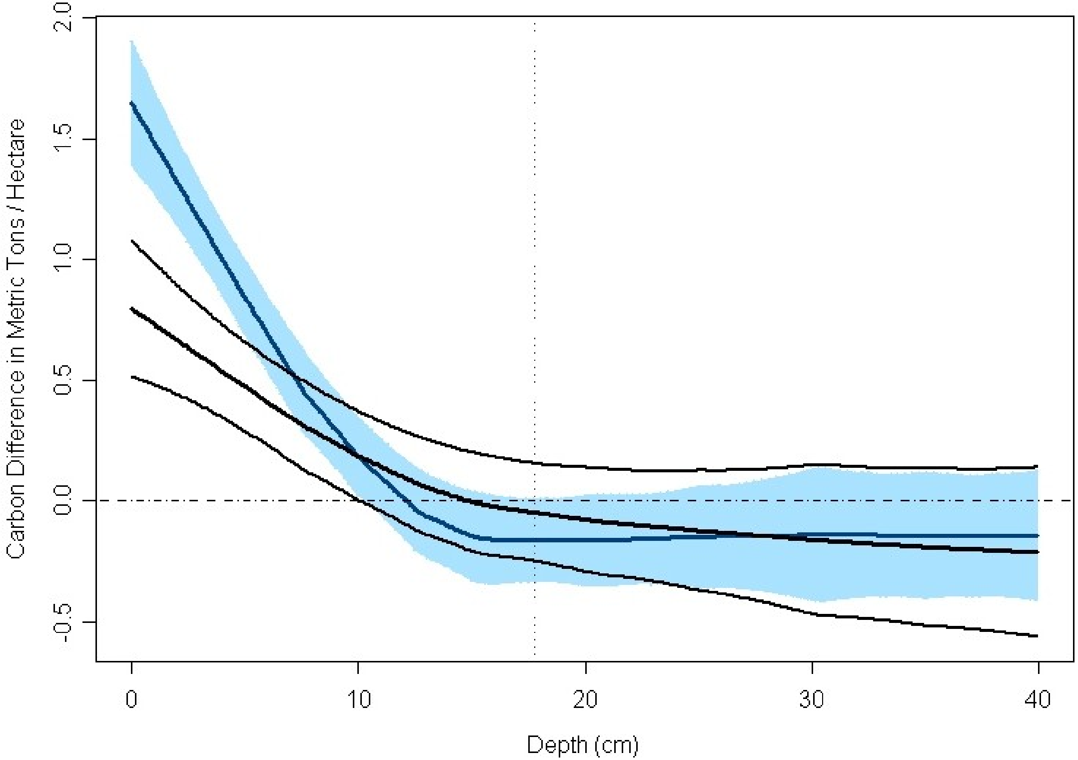
Results: Fitted Curves

- Varying coefficient model fit to no-till data



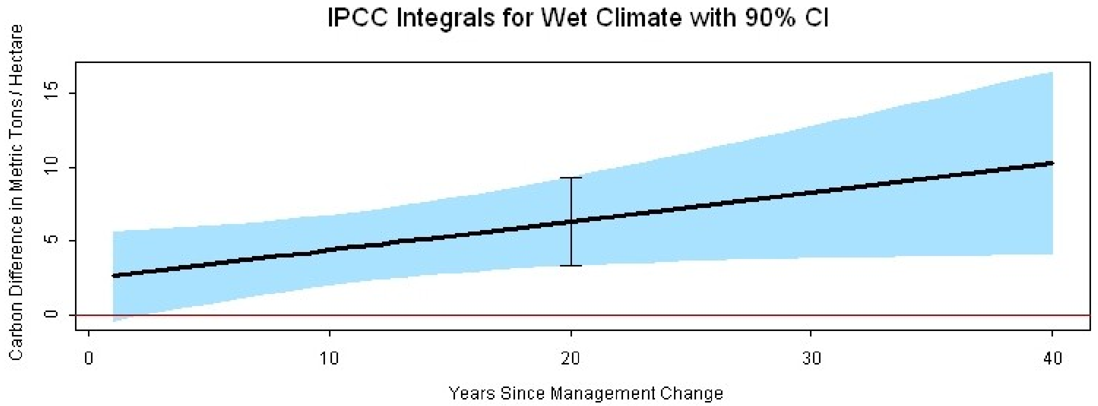
Results: Pointwise Confidence Intervals

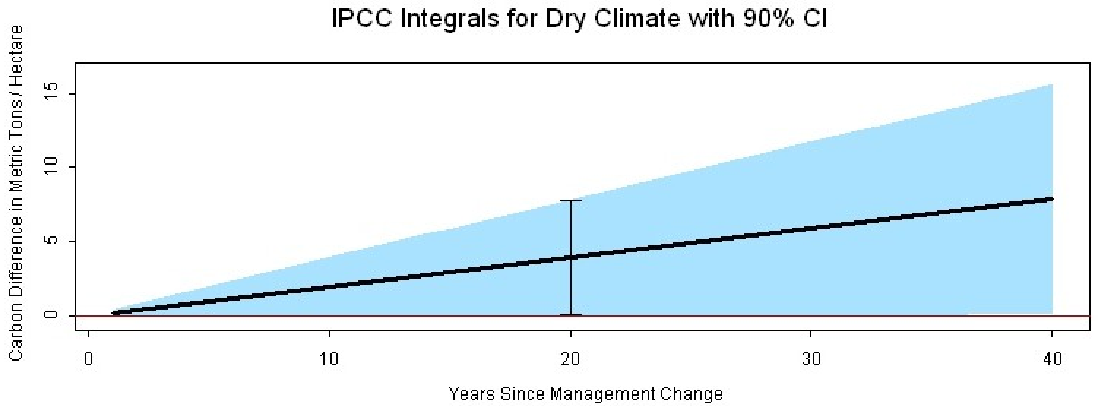
- Fitted curves and confidence intervals



Intergovernmental Panel on Climate Change Integrals

- Cumulative carbon storage versus year in top 30 cm





Summary

- Increment averaging cannot be ignored in core data
- Semiparametric mixed model
  - flexibly models increment averages
  - handles fixed and random effects
  - fits in standard linear mixed model framework
- Further work
  - investigate other dependence structures for noise
  - diagnostics complicated due to increment averaging