County Level Wind Erosion Estimation Using National Resources Inventory Survey

Taps Maiti

Department of Statistics, Iowa State University

E.mail: taps@iastate.edu

Joint work with W. Fuller and P. Mukhopadhyay

- **NRI Outline**

- **Variable of Interest**

- **Main Objective**

- **Methodology**

- **Result**

- **Conclusion**

The National Resources Inventory Survey (NRI):

Longitudinal survey, conducted by U.S. Natural Resources Conservation Services (NRCS) assisted by the Iowa State University.

Objective: Assess conditions and trends for land cover, soil, water, and related natural resources on non-federal lands.

Design

- NRI was conducted every 5 years, 1982-1997, rotational panel survey, 2000-.

- Two-stage/two-phase stratified sample

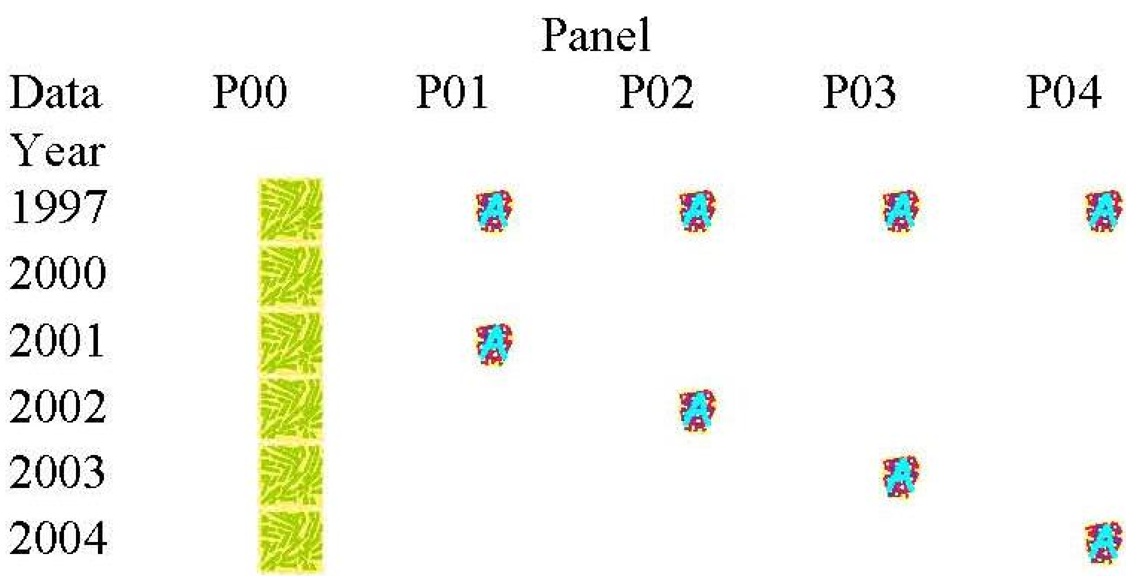
- PSU/segments are areas of land 40-640 acres; 1997 NRI contains about 300,000 PSU.

- Approx. 3 points/PSU; 800,000 points in 1997 NRI

- Data on Urban land, small water etc. are collected at the PSU level. Detail data on soil properties and land use are collected at the point level.

  (Nusser and Goebel, 1997)

- Since 2000 changed to panel surveys; approx. 42,000 PSU repeated every year with approx. 30,000 nonoverlapping (mostly) PSU's from supplemental panel.


Panel
Data Year
P00
P01
P02
P03
P04
1997
2000
2001
2002
2003
2004

Variable of Interest

Wind erosion is a serious problem;

Worse in arid or semiarid regions;

In north America, susceptibility of agricultural land to wind erosion is a serious concern.

In 1930, a prolonged dry spell culminated in dust storms and soil destruction of disastrous proportion. Even after 70 years of this Dust Bowl, wind erosion continues to threaten the sustainability of US natural resources.

Some facts from Wind Erosion Research Institute (WERE)

- In Spring, 1996, wind erosion severely damaged agricultural land throughout the Great Plains.

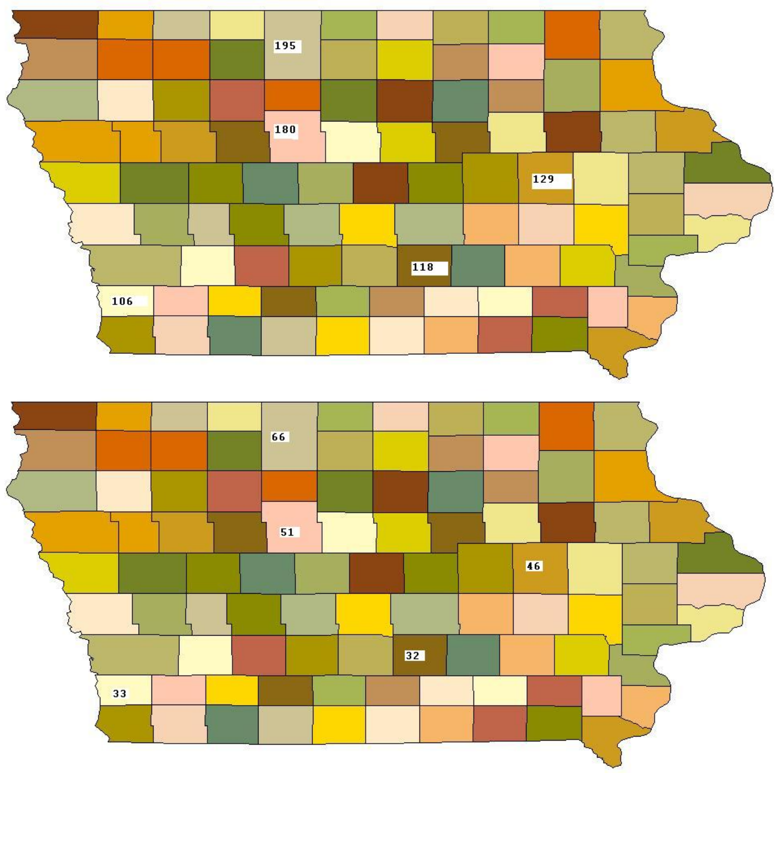
- On cropland, about 70 million hectares are eroded by wind and water at rates that exceed twice the tolerance level of sustainable production.

- On average, wind erosion is responsible for about 40% of this loss.

- Wind erosion can increase markedly in drought years.

County level Estimation

Small sample produce estimates with very large standard errors.



Ingredient needed

- a probability sample to construct the direct estimate
- auxiliary information available for all the counties
- a model specifying the relationship appropriately.

Popular SAE Models

$$\begin{aligned} \widehat{\theta}_i &= \theta_i + e_i; \\ & \quad E(e_i) = 0; Var(e_i) = D_i \\ \theta_i &= \boldsymbol{x}_i^T \boldsymbol{\beta} + u_i; \\ & \quad E(u_i) = 0; Var(u_i) = \sigma^2 \end{aligned}$$

Fay-Herriot (1979)

$$\widehat{\theta}_i^{EB} = \frac{\widehat{\sigma}^2}{\widehat{\sigma}^2 + D_i}\widehat{\theta}_i + (1 - \frac{\widehat{\sigma}^2}{\widehat{\sigma}^2 + D_i})\boldsymbol{x}_i^T \widehat{\boldsymbol{\beta}}$$

Unit level model:

$$y_{ij} = \boldsymbol{x}_{ij}^T \boldsymbol{\beta} + u_i + e_{ij}$$

Battese, Harter, Fuller (1987), Rao (2003).

Model searching for Wind Erosion

In NRI, wind erosion $\equiv WEQ$ and is not directly obtained from the field; a complicated function of several factors (Fuller, 2002).

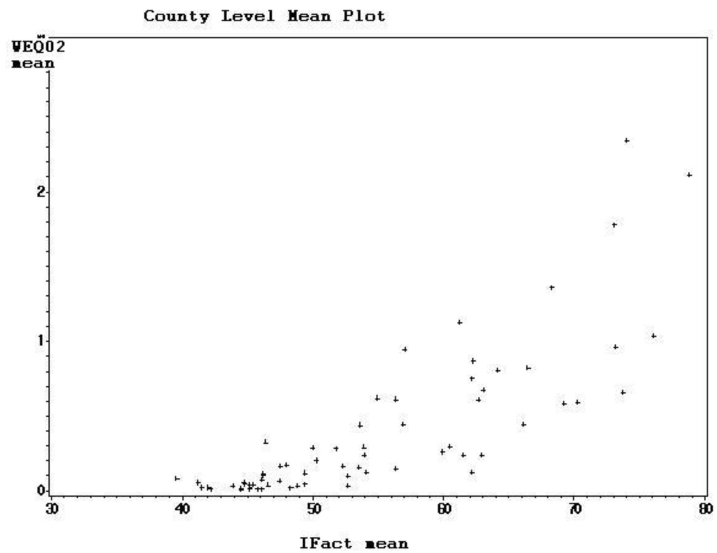
Our study area is Iowa.

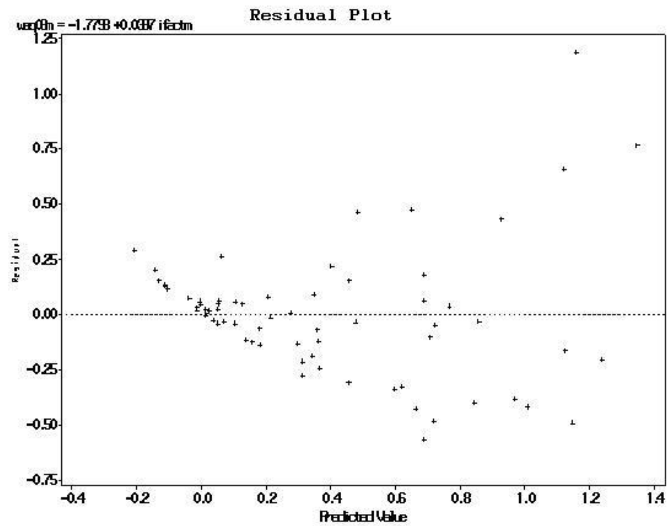
Response variable= WEQ02

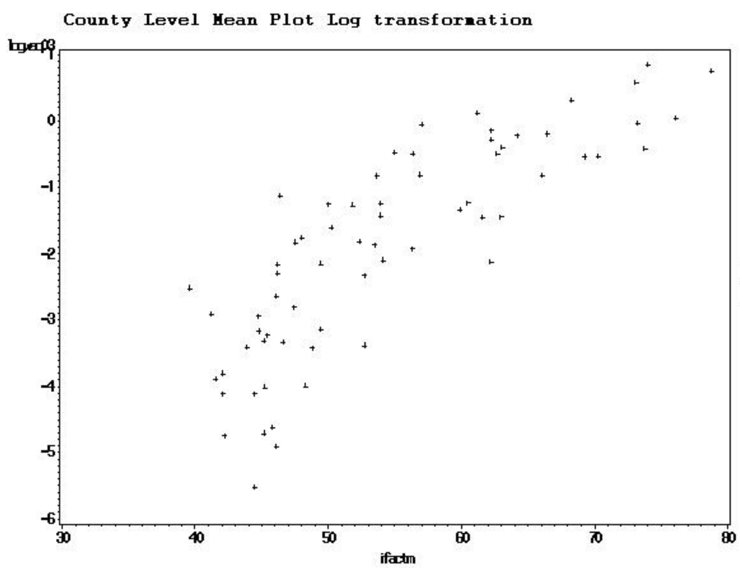
Predictor variable= soil erodibility index (I Fact)

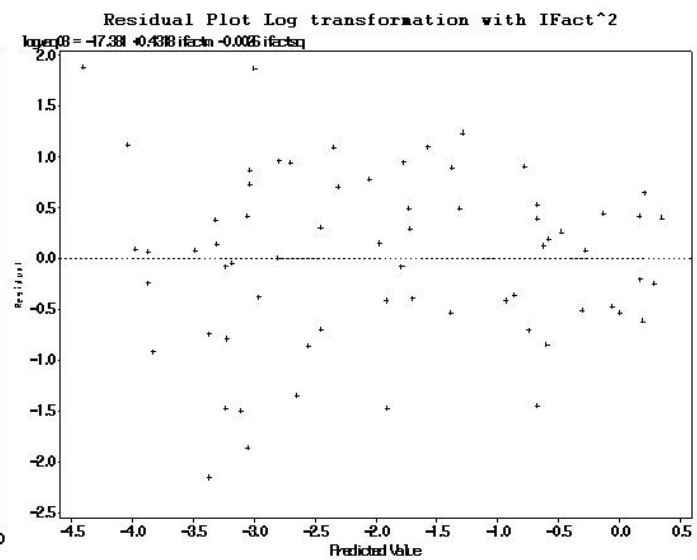
- Higher value of Ifact indicates high susceptibility to wind erosion (Alberts, 2002).

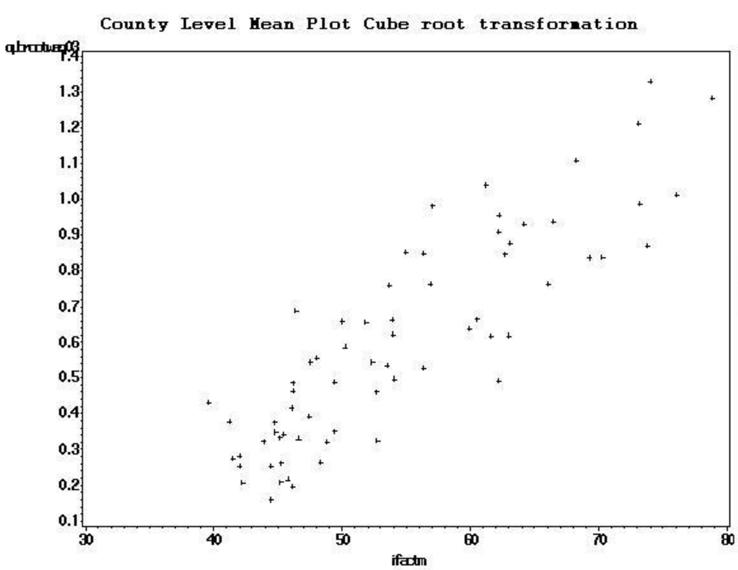
- I Fact can be obtained from NRCS soil database for each county.

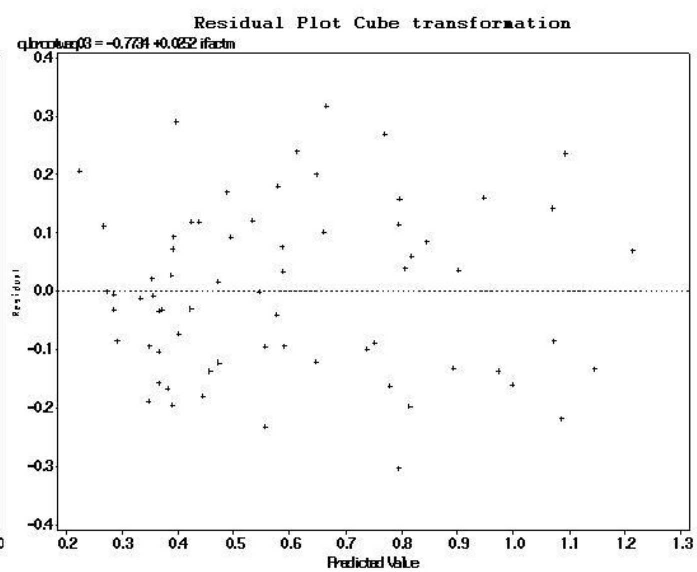

County Level Mean Plot
WEQ02
mean
IFact mean


Residual Plot
Residual
Predicted Value


County Level Mean Plot Log transformation
ifactm


Residual Plot Log transformation with IFact^2
Residual
Predicted Value


County Level Mean Plot Cube root transformation
ifactm


Residual Plot Cube transformation
Residual
Predicted Value

Transformation is not new : SAIPE use log transformation.

Maiti and Slud (2002) and Slud and Maiti (2004) studied the transformation effect on SAE.

Models

- Notations:

  i = county index, y_i = WEQ02, x_{1i} = IFact, u_i = Area specific random effect, e_i = Sampling error.

  μ_i^* = Small area estimated mean in the transformed scale, and μ_i = small area estimated mean in the original scale.

- Assumptions:

  $u_i \sim^{iid} N(0, \sigma_u^2)$, $e_i \sim^{ind} N(0, D_i)$, u_i and e_i are independent, and D_i are known.

- Model I:

$$y_i = \beta_0 + \beta_1 x_{1i} + u_i + e_i \qquad (1)$$

- Model II:

$$log(y_i) = \beta_0 + \beta_1 x_{1i} + \beta_2 x_{1i}^2 + u_i + e_i \qquad (2)$$

- Model III:

$$(y_i)^{(1/3)} = \beta_0 + \beta_1 x_{1i} + u_i + e_i \qquad (3)$$

Estimated mean

- Model I: $\widehat{\mu}_i = \widehat{\gamma}_i y_i + (1 - \widehat{\gamma}_i) x_i^T \widehat{\beta}$, where $\gamma_i = \frac{\sigma_u^2}{\sigma_u^2 + D_i}$

- Model II: $\widehat{\mu}_i = exp\{\widehat{\gamma}_i ln(y_i) + (1 - \widehat{\gamma}_i) x_i^T \widehat{\beta} + 0.5 * \widehat{\sigma}_u^2 (1 - \widehat{\gamma}_i)\}$

- Model III: $\widehat{\mu}_i = \{\widehat{\gamma}_i y_i^{(1/3)} + (1 - \widehat{\gamma}_i) x_i^T \widehat{\beta}\}^3 \{\frac{3\widehat{\sigma}_u^2 + (x_i^T \widehat{\beta})^2}{3\widehat{\sigma}_u^2 \widehat{\gamma}_i + (x_i^T \widehat{\beta})^2}\}$

Estimation of MSE for transformed models is calculated using delta method.

- Model 1:
  $mse(\hat{\mu}_i) = g_{1i}(\hat{\sigma}_u^2) + g_{2i}(\hat{\sigma}_u^2) + 2g_{3i}(\hat{\sigma}_u^2)$,
  Prasad & Rao (1990)

- Model 2:
  $mse(\hat{\mu}_i) = \{exp(\hat{\mu}_i^*)\}^2 mse(\hat{\mu}_i^*)$

- Model 3:
  $mse(\hat{\mu}_i) = \{3\hat{\mu}_i^{2*}\}^2 mse(\hat{\mu}_i^*)$

A fully calibrated model

Is the weighted sum of county estimates equals to the state level estimates?

Model IV:

$$\sum_i w_i(y_i - \widehat{y}_i) = 0$$

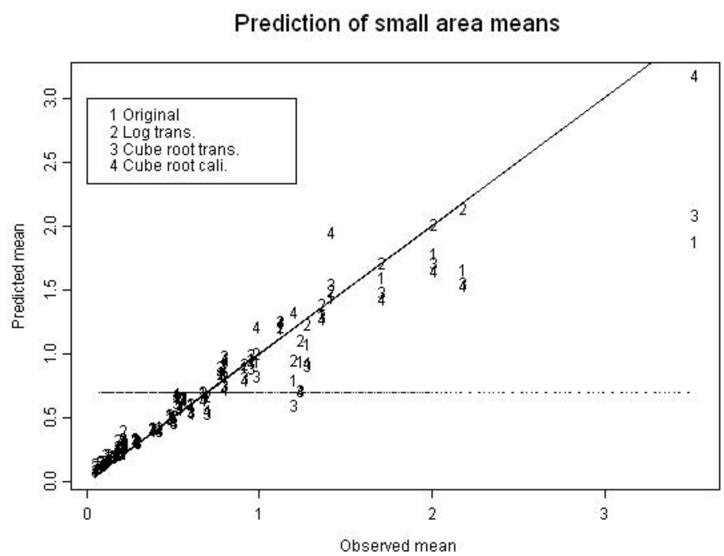
$$(y_i)^{(1/3)} = \beta_0 + \beta_1 x_{1i} + \beta_2 x_{2i} + u_i + e_i \quad (4)$$

$x_{2i} = D_i w_i \widehat{\zeta}_i$, and $\widehat{\zeta}_i$ are the fitted values for regressing $y_i^{(2/3)} + x_{1i}^T \widehat{\beta} y_i^{(1/3)} + (x_{1i}^T \widehat{\beta})^2$ on 1, x_{1i}, x_{1i}^2. $\widehat{\beta}$ is estimated from the fixed part of model 3

The idea is to include weights in the model in such a way so that the score function for the normal equation remains zero.

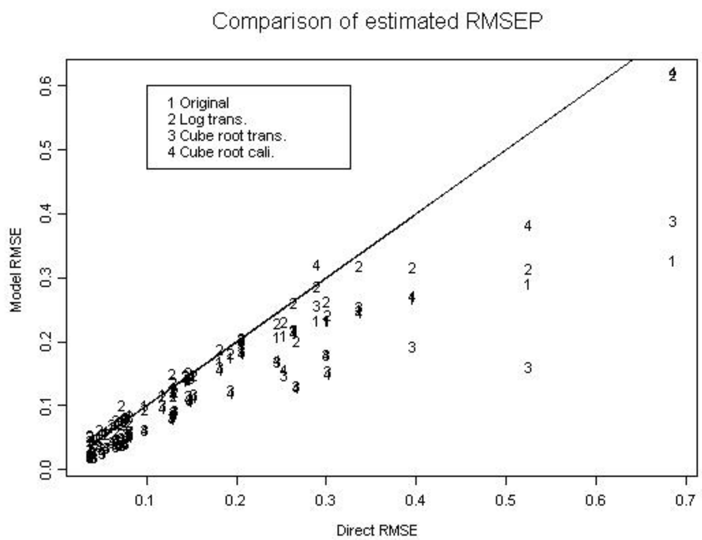
$\widehat{\mu}_i$ and its mse is calculated as in model 3.

Estimated small area means:



Model	Min.	1Q	Median	Mean	3Q	Max.
DE	0.05	0.20	0.51	0.70	0.96	3.53
M1	0.05	0.21	0.54	0.62	0.90	1.85
M2	0.09	0.25	0.54	0.70	0.93	3.31
M3	0.06	0.23	0.49	0.60	0.81	2.06
M4	0.06	0.23	0.48	0.65	0.90	3.15

Estimated MSEP: $\frac{M_i}{D_i}$



Summary of ratio of MSEP:

Model	Min.	1Q	Median	Mean	3Q	Max.
M1	0.22	0.74	0.88	0.82	0.97	0.99
M2	0.35	0.80	0.88	0.93	0.99	1.76
M3	0.09	0.24	0.38	0.41	0.51	0.97
M4	0.14	0.27	0.37	0.43	0.51	1.19

Details of a few selected counties

County	n_i	DE	MI	MII	MIII	MIV
93	31	0.53	0.58	0.59	0.55	0.54
		0.038	0.029	0.031	0.014	0.013
197	34	0.22	0.24	0.28	0.27	0.26
		0.002	0.002	0.003	0.001	0.001
73	35	0.61	0.59	0.58	0.52	0.50
		0.018	0.016	0.014	0.008	0.007
75	43	0.21	0.21	0.20	0.20	0.20
		0.002	0.002	0.001	5e-4	5e-4
145	47	0.05	0.05	0.09	0.06	0.06
		0.002	0.002	0.002	2e-4	3e-4
161	50	0.96	0.94	0.96	0.86	0.90
		0.060	0.041	0.050	0.027	0.028
109	55	1.71	1.57	1.68	1.46	1.40
		0.021	0.018	0.020	0.020	0.019
167	122	2.01	1.76	1.99	1.68	1.63
		0.043	0.032	0.040	0.040	0.038

Comparison between Model III and Model IV

- Relative Absolute Calibration $\frac{|\sum_i w_i(\widehat{y}_i - y_i)|}{\sum_i w_i y_i}$

  Model III: 13% Model IV: 4%

- Difference (MIV -MIII) in estimated MSEP

  Mean 0.009 and Max. 0.231

Conclusion

- Proposed a small area model for wind erosion

- Required non-linear regression

- Ifact. is very useful

- Estimates can be fully calibrated

- Proposed method significantly improves over direct estimates.

Thank You