Evaluating the Impact of Nonsampling Errors on Erosion Estimates for the Conservation Effects Assessment Project

Andreea L. Erciulescu and Emily J. Berg
Center for Survey Statistics and Methodology, Iowa State University, Department of Statistics, Ames, IA 50011

Motivation

The Conservation Effects Assessment Project (CEAP) is a series of surveys intended to evaluate environmental outcomes associated with conservation practices. Erosion is one such environmental outcome of interest, because it affects soil quality on cropland and crop productivity, water quality and quantity, and air quality. Errors produce biased erosion estimators at different geographical levels, identified by Hydrologic Unit Codes (HUCs). We discuss possible ways to evaluate effects of two sources of nonsampling error on erosion estimates for the eight digit HUCs.

CEAP Background

NRI (National Resources Inventory)
Periodic survey of status and changing conditions of the soil, water, and related resources on private land in the US (here focus on soil erosion)

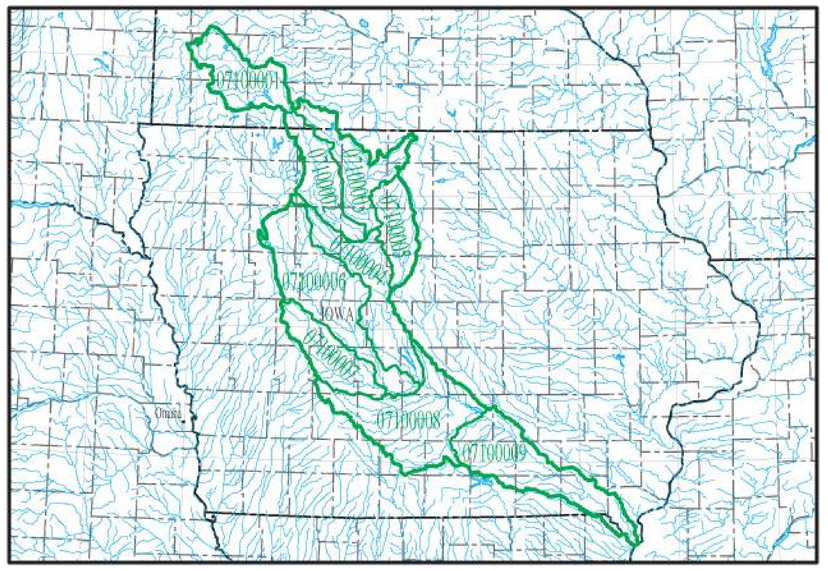
CEAP
1. Sample of NRI cultivated cropland (CC) points
2. Collect data: farmer interview and NRCS (National Resources Conservation Service) field office databases
3. APEX model (black box) $\mapsto$ erosion estimates

Objectives: Impacts of conservation practices

Users: Policy makers, farmers

Scope: HUC8 estimation

Des Moines River Watershed



Auxiliary information
1. Soil Survey
   - K-factor (soil erodibility index)
   - Slope
   - unit = map unit
2. 2007 NRI Pointgen
   - no nonresponse or frame errors
   - USLE (Universal Soil Loss Equation) erosion
   - unit = point

Nonsampling Errors

1. Nonresponse error

- occurs due to refusals
- the overall response rate is 67.44%
- highest response rate in HUC 7100005, 82.35%
- lowest response rates in HUCs 7100008 and 7100009, 62.16% and 57.14%, respectively

2. Frame problems

- only farming operations with CC are eligible
- current information on landuse is not available at the sample design stage

Number of eligible CEAP points, real, core, cc

03	CC 04	!CC 04	CC 05	!CC 05	CC 06	!CC 06
frame	531	2	528	5	525	8
!frame	4	0	8	0	16	0

- 16 points eligible as CC in 2006, but not in the 2003 frame
  - noncultivated cropland
  - most in HUCs 7100008 (9) and 7100009 (5)
  - misclassification counts increase over years

Acknowledgement

This research was partially supported by USDA NRCS CESU agreement 68-7482-11-534.

Results

Nonresponse error

Tests for covariate means respondents vs nonrespondents
(W test is Wilcoxon rank sum test)

	Avg USLE		Kfactor		Slope	
	stat	pval	stat	pval	stat	pval
T test	0.917	0.36	0.509	0.611	1.906	0.058
W test	24714	0.737	25471	0.364	25791	0.247

- no significant differences at HUC4 or HUC8 levels
- little evidence of nonresponse bias

Frame problem

Tests for points in the frame vs points not in the frame

	USLE		Kfactor		Slope	
	stat	pval	stat	pval	stat	pval
T test	-1.5504	0.1414	-5.1686	0.0001	-3.1210	0.0066
W test	3688.5	0.4067	1377.5	0.0000	1988.5	0.0002

- significant differences in Kfactor and Slope
- frame induces undercoverage bias in estimates

Response variables

- 16 continuous and 3 ordinal scores
- RUSLE2 (Revised Universal Soil Loss Equation)
  - continuous response variable

RUSLE2 HUC8 sample means ($\times 10$) and sample standard deviations ($\times 1000$)

	1	2	3	4	5	6	7	8	9
n_i	78	17	29	32	14	44	29	46	28
$\bar{y}_i$	11.7	25.4	29.5	21.1	14.2	20.2	38.4	55.5	49.9
$\frac{s_y}{\sqrt{n_i}}$	1.3	14.8	6.2	3.9	4.6	3.8	10.5	12.4	18.2

(HUC4 mean = 28.5, HUC4 sd = 1.0)

Estimated correlations between auxiliary variables and RUSLE2 with 95% CIs

	$\hat{\rho}$	95%
Kfactor	0.352	(0.261, 0.447)
Slope	0.532	(0.522, 0.608)
USLE	0.583	(0.504, 0.653)

Discussion / Future work

Calibration

- construct weights calibrated to the population mean of Slope and Kfactor
- incorporate adjustments in a small area model to estimate response mean at HUC8 level

Challenges

- definition of *unit* differs for the CEAP sample and for the Soil Survey
- unknown population mean for the NRI auxiliary variables

Other sources of nonsampling error

- processing error: imperfections in the APEX model
- location error: imperfections in data collection protocols and GPS instruments