Processing Error in Erosion Estimations Based on NRI Longitudinal Survey

Center for CSSM Survey Statistics and Methodology

Yang Li, Emily Berg, Wayne A. Fuller

Department of Statistics, Iowa State University, Ames, IA

Center for Survey Statistics and Methodology, Iowa State University, Ames, IA

Introduction

Erosion analysis is an important component in the National Resources Inventory (NRI) conducted by U.S. Department of Agriculture's Natural Resources Conservation Service (NRCS). The Universal Soil Loss Equation (USLE) is a model that predicts the long term average annual rate of erosion on a field, which was released in 1960s. It is replaced by the newly released Revised Universal Soil Loss Equation model. USLE is collected in the NRI through the year 2007, data collection for RUSLE2 starts in 2004. To estimate long term soil erosion trends and maintain consistency as an input to other models, it is necessary to impute USLE erosion after 2007[1].

The USLE is based on the climate, soil characteristics, topography, cropping systems, and conservation practices at a given location. The USLE model is

$$A = R \times K \times LS \times C \times P$$

- ✓ A: long-term average annual soil loss
- ✓ R: rainfall erosivity factor
- ✓ K: soil erodibility factor
- ✓ LS: slope length-gradient factor
- ✓ C: crop/vegetation and management factor
- ✓ P: support practice factor

RUSLE2 is a more nuanced approximation to erosion. Its algorithms are more complicated than USLE and do not have closed form expressions. RUSLE2 also produces numerical summaries that are intended to be equivalent to the USLE R, K, LS, P, and C factors.

USLE and RUSLE2 have a four-year overlap between 2004—2007, i.e., a calibration period, which allows us to build models and predict at unobserved years.

	…	2002	2003	2004	2005	2006	2007	2008	…
USLE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	No	No
RUSLE2	No	No	No	Yes	Yes	Yes	Yes	Yes	Yes
Others	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes

Goals

To build a predictive model between USLE and auxiliary information from the calibration period as an effort to predict USLE soil erosion between the years 2008 and 2012 when USLE is not collected anymore.

Predictor Variables

1) Broaduses:

An NRI point is classified into one of 18 broaduse (BU) categories. Erosion is measured for four of them:

- ➢ BU = 1: cultivated cropland
- ➢ BU = 2: noncultivated cropland
- ➢ BU = 3: pasture
- ➢ BU = 15: Conservation Reserve Program (CRP)

Different broaduses have different erosion characteristics.

2) Permanent USLE factors:

R, K, LS factors in USLE do not change with time (at least approximately). Define a new variable $\text{USLEfac} = R \times K \times LS$.

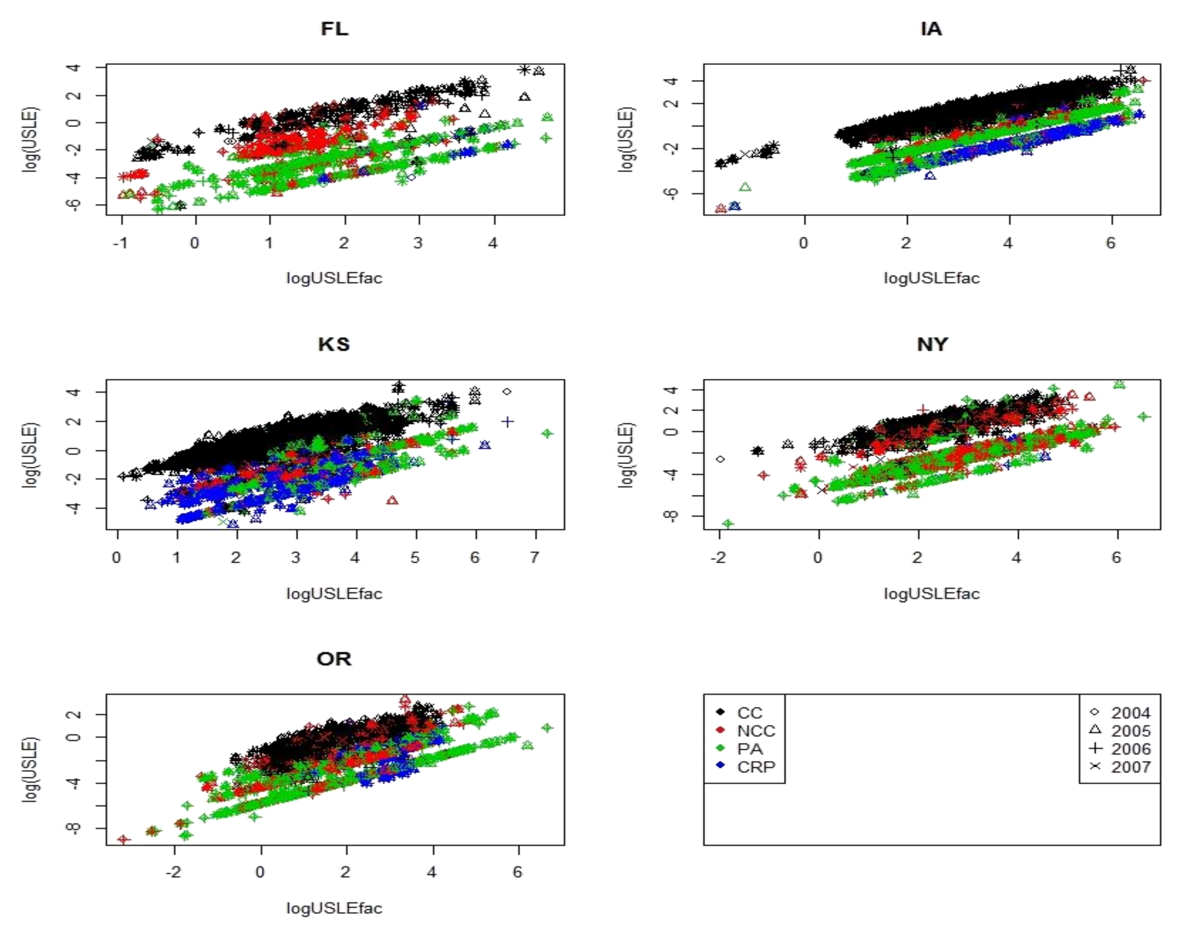
3) RUSLE2:

RUSLE2 are collected for all years after 2004. RUSLE2 contains information on practice and management.

4) Historical USLE data

We are doing prediction in a sequential way. Predicted USLE in past years are treated as true values and can be used for prediction in subsequent years.

Exploratory Analysis

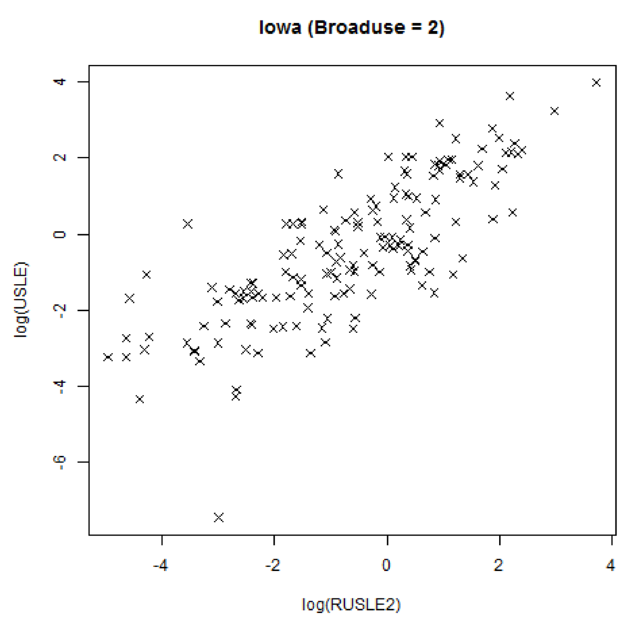


The multiplicative nature of USLE inspires us to work in the log scale. Log transformation also serves to better satisfy the assumptions of regression models.

Regression-type methods:

- ➢ Simple linear regression (e.g., log(USLE) on log(RUSLE2))
- ➢ Multiple linear regression
- ➢ Multivariate Adaptive Regression Splines (MARS)[2]

Conclusions are there are evidence of nonlinear relationships between log(USLE) and log(RUSLE2). We need information besides RUSLE2 to get a good prediction of USLE. Outliers exist in some states.



We notice many observations are stable across years. Most core points have constant soil losses and broaduses across consecutive years.

No change in RUSLE2 or broaduse typically indicates no change in USLE.

Kansas between 2006 and 2007		(RUSLE2 + Broaduse) Change	(RUSLE2 + Broaduse) No Change
(USLE)	Change	205	41
(USLE)	No Change	142	**1207**

Methods and Procedures

Different states have different erosion patterns.

We fit a separate model for each broaduse in each state

Proposed model: $Y_{i,t} = \beta_0 + \beta_1 X_{i,t} + \beta_2 W_{i,t} + \beta_3 Z_{i,t} + \epsilon_{i,t}$

- ❖ $\epsilon_{i,t} \sim (0, \sigma^2)$
- ❖ $(Y_{i,t}, X_{i,t}) = (\log(\text{USLE}), \log(\text{RUSLE2}))$
- ❖ $W_{i,t} = \log(R \times K \times LS)$
- ❖ $Z_{i,t} = (X_{i,t} - \bar{X}_{..}) I[X_{i,t} \geq \bar{X}_{..}]$

Variable selection implemented to ensure that coefficients are reasonable. $Z_{i,t}$ incorporates possible nonlinearity in log(RUSLE2).

Back transformation:

After getting the predicted log(USLE), a back transformation is used to get predicted USLE on the original scale. Some adjustments are made to remove the transformation bias and control the variance[3].

Carry-forward rule: Carry forward the USLE value for the most recent year if there is no change in the broaduse or the RUSLE2 between the most recent observed year and the current year.

- ➢ $\widetilde{USLE}_{i,t} = \exp\{\hat{\beta}_0 + \hat{\beta}_1 X_{i,t} + \hat{\beta}_2 W_{i,t} + \hat{\beta}_3 Z_{i,t}\}$
- ➢ $\widehat{USLE}_{i,t} = \hat{c}_{0b(i,t)} + \hat{c}_{1b(i,t)}(\widetilde{USLE}_{i,t} - \hat{c}_{2b(i,t)})$
- ➢ $(\hat{c}_{0b(i,t)}, \hat{c}_{2b(i,t)}), (s_{0b(i,t)}, s_{2b(i,t)})$: mean and standard deviation of $(USLE, \widetilde{USLE}_{i,t})$ across 2004—2007 for broaduse $b(i,t)$
- ➢ $\hat{c}_{1b(i,t)} = \min\{\frac{\hat{c}_{0b(i,t)}}{\hat{c}_{2b(i,t)}}, \frac{\hat{s}_{0b(i,t)}}{\hat{s}_{2b(i,t)}}\}$
- ➢ Final predictor guaranteed non-negative, adjustment not depend on a log-normal assumption.

Results and Discussions

➢ Variable selection:

State	R^2 (full)	Reduced Model	R^2 (reduced)
Florida	0.382	(W, Z)	0.380
Iowa	0.860	(X, W, Z)	0.860
Kansas	0.347	(X, W, Z)	0.347
New York	0.524	(X, W, Z)	0.524
Oregon	0.683	(W)	0.678

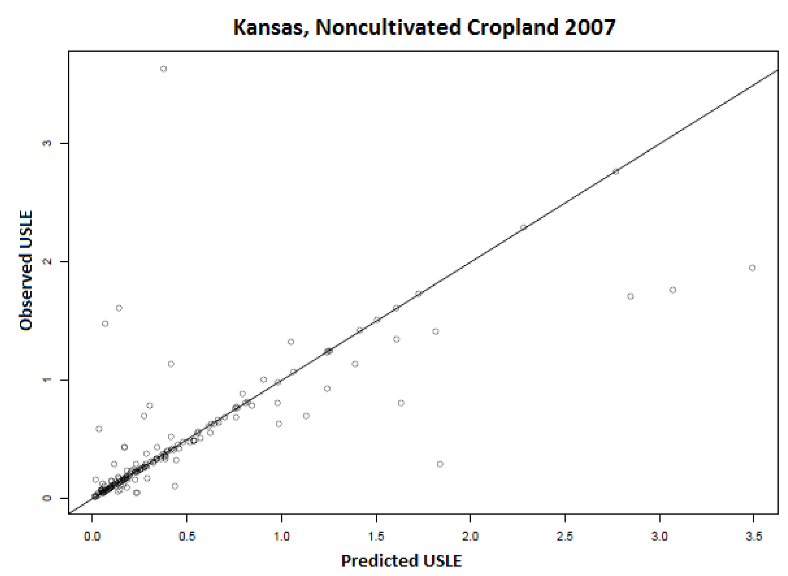
➢ Summary of prediction (Kansas):

BU	USLE Pred 07 Mean	USLE Pred 07 SE	USLE 07 Mean	USLE 07 SE	Diff (obs – pred) Mean	Diff (obs – pred) SE
1	2.130	0.051	2.137	0.053	0.007	0.016
2	0.463	0.041	0.477	0.042	0.014	0.012
3	0.824	0.083	0.853	0.087	0.030	0.025
15	0.207	0.011	0.214	0.012	0.007	0.003

- ➢ Some states have small numbers of observations for some broaduses. Data were combined across different broaduses in the same state for such situations.
- ➢ Observation points are nested in segments. Correlation also exists for the same point across time. Mixed effects models can better represent the structure of the data and avoid the subjectively chosen carry-forward rule.

$$y_{ijt} = \mathbf{z}'_{ijt}\boldsymbol{\beta} + \alpha_i + \delta_{ij} + \epsilon_{ijt}$$

- ❖ $\alpha_i \sim N(0, \sigma_\alpha^2), \delta_{ij} \sim N(0, \sigma_\delta^2), \epsilon_{ijt} \sim N(0, \sigma^2)$.
- ❖ $\alpha_i, \delta_{ij}, \epsilon_{ijt}$ are mutually independent.
- ❖ $\text{Cor}(\epsilon_{ijt}, \epsilon_{ijs}) = \varphi^{|t-s|}, \text{Cor}(\epsilon_{ijt}, \epsilon_{kls}) = 0$ for $i \neq k$ or $j \neq l$.



References

1. CSSM Report (2011), "Estimation for the 2005 NRI and 2007 NRI".
2. Friedman, J. H. (1991), "Multivariate Adaptive Regression Splines", Annals of Statistics, 19, 1-67.
3. Emily Berg *et al* (2013). "Small Area Estimation for County-Level Cash Rental Rates", manuscript.

Acknowledgements

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