Published online July 20, 2017

Journal of Environmental Quality | SPECIAL SECTION

THE EVOLVING SCIENCE OF PHOSPHORUS SITE ASSESSMENT

Southern Phosphorus Indices, Water Quality Data, and Modeling (APEX, APLE, and TBET) Results: A Comparison

Deanna Osmond,* Carl Bolster, Andrew Sharpley, Miguel Cabrera, Sam Feagley, Adam Forsberg, Charles Mitchell, Rao Mylavarapu, J. Larry Oldham, David E. Radcliffe, John J. Ramirez-Avila, Dan E. Storm, Forbes Walker, and Hailin Zhang

Abstract

Phosphorus (P) Indices in the southern United States frequently produce different recommendations for similar conditions. We compared risk ratings from 12 southern states (Alabama, Arkansas, Florida, Georgia, Kentucky, Louisiana, Mississippi, North Carolina, Oklahoma, South Carolina, Tennessee, and Texas) using data collected from benchmark sites in the South (Arkansas, Georgia, Mississippi, North Carolina, Oklahoma, and Texas). Phosphorus Index ratings were developed using both measured erosion losses from each benchmark site and Revised Universal Soil Loss Equation 2 predictions; mostly, there was no difference in P Index outcome. The derived loss ratings were then compared with measured P loads at the benchmark sites by using equivalent USDA–NRCS P Index ratings and three water quality models (Annual P Loss Estimator [APLE], Agricultural Policy Environmental eXtender [APEX], and Texas Best Management Practice Evaluation Tool [TBET]). Phosphorus indices were finally compared against each other using USDA–NRCS loss ratings model estimate correspondence with USDA–NRCS loss ratings. Correspondence was 61% for APEX, 48% for APLE, and 52% for TBET, with overall P index correspondence at 55%. Additive P Indices (Alabama and Texas) had the lowest USDA–NRCS loss rating correspondence (31%), while the multiplicative (Arkansas, Florida, Louisiana, Mississippi, South Carolina, and Tennessee) and component (Georgia, Kentucky, and North Carolina) indices had similar USDA–NRCS loss rating correspondence—60 and 64%, respectively. Analysis using Kendall's modified Tau suggested that correlations between measured and calculated P-loss ratings were similar or better for most P Indices than the models.

Core Ideas

- Southern region P Indices estimate P losses as well as water quality models.
- APLE and TBET P-loss predictions were more similar than were results from APEX.
- Assigning potential P-loss risk from P Indices to any given water resource is challenging.

J. Environ. Qual.
doi:10.2134/jeq2016.05.0200

Received 29 May 2016.
Accepted 16 Oct. 2016.
*Corresponding author (deanna_osmond@ncsu.edu).

Water quality impairment caused by nutrient enrichment remains a major concern (Dubrovsky and Hamilton, 2010). Recent harmful algal blooms in Lake Erie caused Toledo to shut down its drinking water supply for several days, refocusing the link between nutrient enrichment (particularly phosphorus [P]) and water quality impairment (Stow et al., 2015), with many of these nutrients being agriculturally derived. To control agricultural nutrient loading to surface waters, multiple control strategies are necessary at the source and during transport into the receiving water resources. The USDA–NRCS refers to this as "avoid, control, and trap."

Since the late 1990s, the USDA and USEPA jointly required all states to adopt a unified nutrient management policy through the NRCS Code 590 Standard (USDA and USEPA, 1999). States were required to establish a soil-test P threshold based on crop requirements (above which P applications were restricted), to establish an alternative soil test P threshold using water quality criteria, or to develop a P Index to identify fields at risk for P losses. Forty-eight states and some territories, including Puerto Rico, chose to use P Indices (Sharpley et al., 2003), a concept originally developed by USDA–NRCS for assigning relative risk of P loss to agricultural fields (Lemunyon and Gilbert, 1993). California and Connecticut use soil-test P crop response (Sharpley et al., 2003).

To address local hydrologic, soils, landscapes, crops, and nutrient sources, each state developed its own P Index. Some states modified the original P Index (Lemunyon and Gilbert, 1993), while other states, such as Arkansas, Georgia, and North Carolina, used different strategies to develop their P Index (Osmond et al., 2006). Not surprisingly, state P Index recommendations vary widely (Osmond et al., 2006, 2012; Sharpley et al., 2003). Osmond et al. (2006) compared P Indices from 12 southern US states and found that there were diverse P Index ratings between states for the

D. Osmond, Crop and Soil Sciences, North Carolina State Univ., Raleigh, NC; C. Bolster, USDA–ARS, Food Animal Environmental Systems Research Unit, Bowling Green, KY; A. Sharpley, Crop, Soil and Environmental Sciences, Division of Agriculture, Univ. of Arkansas, Fayetteville, AR; M. Cabrera, A. Forsberg, and D.E. Radcliffe, Crop and Soil Sciences, Univ. of Georgia, Athens, GA; S. Feagley, Soil and Crop Sciences, Texas A&M Univ., College Station, TX; C. Mitchell, Agronomy and Soils, Auburn Univ., Auburn, AL; R. Mylavarapu, Soil and Water Science, Univ. of Florida, Gainesville, FL; J.L Oldham, Plant and Soil Sciences, Mississippi State Univ., Starkville, MS; J.J. Ramirez-Avila, Civil and Environmental Engineering, Mississippi State Univ., Starkville, MS; D.E. Storm, Biosystems and Agricultural Engineering, Oklahoma State Univ., Stillwater, OK; F. Walker, Biosystems Engineering and Soil Science, Univ. of Tennessee, Knoxville, TN; H. Zhang, Plant and Soil Sciences, Oklahoma State Univ., Stillwater, OK. Assigned to Associate Editor Peter Kleinman.

Abbreviations: APLE, Annual Phosphorus Loss Estimator; APEX, Agricultural Policy Environmental eXtender; DP, dissolved phosphorus; HSS, Heidke skill score; RUSLE2, Revised Universal Soil Loss Equation 2; TBET, Texas Best Management Practice Evaluation Tool; TP, total phosphorus.

same conditions, which led to differences in P management and the amount of animal manures, effluents, and fertilizer allowed.

In 2011, the USDA–NRCS revised the 590 standard (USDA NRCS, 2011b), in part to address the stark differences in P Index ratings and recommendations across state boundaries, and to address the concern that there had been no change in elevated soil-test P levels and runoff P (Osmond et al., 2006; USEPA, 2010; Sharpley et al., 2011). Since the USDA–NRCS did not generally provide resources to test P Indices (Sharpley et al., 2012), the ability of states to verify their P Indices resided at the state level. Many states did not fund either development or validation of their P Indices, and southern states were no exception.

In one of the few P Index verifications from the South, Harmel et al. (2005) compared measured P runoff from pasture and cropped watersheds of the Texas Blackland Prairies with three indices from Arkansas, Iowa, and Texas. Even though the three indices were fundamentally different, the Iowa and Texas indices both provided reasonable estimates of P-loss potential ($p < 0.01$; Harmel et al., 2005). This was the case for Arkansas, even though this index was developed for pastures with low erosion potential (DeLaune et al., 2004). In another southern P Index validation, Butler et al. (2010) showed good agreement between the risk of P loss and measured total P (TP) loss in runoff from Georgia. Lately, Bolster et al. (2014) evaluated the recently modified Kentucky P Index with data used to evaluate the Annual P Loss Estimator (APLE) model (Vadas et al., 2009); they reported relatively good correlations between measured P loss and predicted risk of P loss.

To reduce P Index rating variability and management interpretation, the 2011 590 USDA–NRCS standard suggested that the Agricultural Policy Environmental eXtender (APEX) model replace each state P Index (USDA–NRCS, 2011a; Williams et al., 2012). Concerned that APEX could not adequately capture P losses, members of the USDA Southern Extension and Research Activity 17 (SERA-17) developed a white paper stating the need to compare indices and water quality model performance using edge-of-field-based P runoff data (Sharpley et al., 2011). To test the veracity of APEX and other commonly used water quality models, the Southern, Chesapeake, and Heartland project teams were funded by USDA–NRCS and coordinated through SERA-17, with the goals of supporting the refinement of state P Indices and demonstrating their accuracy in identifying the magnitude and extent of P-loss risk and their utility to improve water quality.

Specific objectives of the southern group were (i) to compare measured P runoff losses from prior edge-of-field studies (Arkansas, Georgia, Mississippi, North Carolina, Oklahoma, and Texas) with each of the 12 southern P Indices (Alabama, Arkansas, Florida, Georgia, Kentucky, Louisiana, Mississippi, North Carolina, Oklahoma, South Carolina, Tennessee, and Texas); (ii) to compare measured P runoff losses from the aforementioned datasets to estimates from APEX (Williams et al., 2012), APLE (Vadas et al., 2009), and Texas Best management Evaluation Tool (TBET) (White et al., 2012); and (iii) to compare southern P Indices against each other and model predictions.

Materials and Methods

Water Quality and Land Use Data

Six current and/or published edge-of-field water quality datasets from northwest Arkansas, central Georgia, the Mississippi Delta, western North Carolina, Oklahoma, and Texas were assembled and used to evaluate the southern P Indices (Pierson et al., 2001; White et al., 2012; Yuan et al., 2013; Edgell et al., 2015; Sharpley, unpublished data, 2016). Multiple water quality datasets existed from Oklahoma and Texas, but eight were selected to represent varied cropping systems that included small grains and row crops, as well as pasture and rangeland (White et al., 2012). Data from Georgia, Arkansas, and North Carolina represented different treatments at the same location, while data from the remaining locations were unique fields. These sites represented a range of agroecological areas, cropping systems, nutrient application rates, and tillage. Soil characteristics (Table 1), as well as land use information (Table 2), were assembled for these benchmark locations. A detailed description of these six benchmark sites can be found in Bolster et al. (2017).

Southern Phosphorus Indices

A description of the 12 southern P Indices can be found in Osmond et al. (2012). Indices, however, can be generally defined as additive, multiplicative, and component. Texas and Alabama P Indices are additive, with the weighted transport and source factors summed (Lemunyon and Gilbert, 1993). The multiplicative P Indices (Arkansas, Florida, Louisiana, Mississippi, South Carolina, and Tennessee) combine all source and transport factors into two separate factors, which are then multiplied to obtain the final P Index value (Gburek et al., 2000). Finally, component indices (Georgia, Kentucky, and North Carolina) sum P loss from each individual process contributing to P loads; each component is calculated as the product of transport and source factors (Bolster et al., 2012). Only the Oklahoma P Index cannot be categorized, as it is strictly qualitative. Since the last analysis of southern P Indices (Osmond et al., 2012), modifications have occurred in four states: Kentucky (Bolster et al., 2014), Mississippi, Tennessee, and Texas.

Thirty-four input variables were assembled from each site to serve as inputs to the various P Indices (Osmond et al., 2006; Osmond et al., 2012). Some factors, however, were not available and had to be assumed: water resource impairment, buffer (since all collected data were edge of field), irrigation, rock fragments > 25.4 cm, and infrequent flooding. Counties are required to process two state P Indices; selected counties were Putnam (Georgia) and Buncombe, Cabarrus, Chatham, Durham, Person, or Union (North Carolina), depending on the soil series and rainfall of the benchmark locations.

Ratings for individual state P Indices were determined for each of the six datasets using sediment losses measured from each location (Table 3). Phosphorus Indices were then recalculated using erosion losses using Revised Universal Soil Loss Equation 2 (RUSLE2) predictions for the benchmark locations of Arkansas, Georgia, North Carolina, and Mississippi (Table 3).

Water Quality Models

Three water quality models—APEX, APLE and TBET—were selected for this analysis. The APLE model is a field-scale model that operates on an annual time step, predicts annual P loss where runoff is the dominant transport process, and has been shown to provide good predictions of P loss at the field scale for a wide range of climatic and land use conditions (Vadas et al., 2009). The APEX model predicts surface runoff, erosion, sediment deposition and degradation, nutrient and pesticide

Table 1. Soil factors from six benchmark edge-of-field research sites: Arkansas (Sharpley, unpublished data, 2016), Georgia (Pierson et al., 2001), Mississippi (Yuan et al., 2013), North Carolina (Edgell et al., 2015), Oklahoma, and Texas (White et al., 2012).

Soil properties	Arkansas	Georgia	Mississippi	North Carolina	Oklahoma				Texas			
					Chickasaw	Cyril	Demo	El Reno	Goosebranch	Melde	Patton	Reisel
Predominant soil map unit	Captina	Cecil/ Sedgefield	Tensas	Delanco	McLain	Norge	Clarksville	Bethany	Duffau	Topsey	Duff	Houston Black
Texture	Silt loam	Silt loam	Silty clay loam	Silt loam	Silty clay loam	Silt loam	Silt loam	Silt loam	Fine sandy loam	Clay loam	Gravelly loam	Clay
Hydrologic group	C	C/D	D	C	C	B	B	–	B	C	C	D
Curve number	71	70/80	90	88	82	69	66	–	55	83	80	90
Slope, %	2.0	6.0–8.0	1.0	3.5–4.3	0.5	2.0	16.0	3.6	2.3	2.1	2.4	2.3
Slope length, m	30	30.5	–	18.3	124	106	25	105	76	76	69	114
Runoff class	–	Medium–rapid	High	Slow–medium	Low	Low	Medium–very high	Low	Low	Medium	Rapid	Very high
Soil drainage class	Moderately well drained	Well to somewhat poorly drained	Poorly drained	Moderately well drained	Moderately well drained	Well drained	Well drained	Well drained	Well drained	Well drained	Well drained	Moderately well drained
Depth to water table, m	10	5	–	0.8	–	–	–	–	–	>2	–	–
Rainfall, mm	1215	1003–1123	1300	1023–1194	840	864	1092	865	790	737	400	889

transport, and subsurface flow (Williams et al., 2012), while TBET is a field-scale application of the Soil Water Assessment Tool (White et al., 2012). All models estimate edge-of-field P and sediment loss. More detailed descriptions of APLE can be found in Bolster et al. (2017), of TBET in Forsberg et al. (2017), and of APEX in Ramirez-Avila et al. (unpublished data, 2016).

Water quality modeling was conducted with TBET for Arkansas, Georgia, and North Carolina datasets (Forsberg et al., 2017) and for Oklahoma and Texas datasets (White et al., 2012). The APLE model was applied to all data sites (Bolster et al., 2017), and APEX was employed with datasets from Arkansas, Georgia, Mississippi, and North Carolina (Ramirez-Avila et al. , unpublished data, 2016). Model type was matched to benchmark datasets based on a combination of prior work, such as the case of TBET use with the Oklahoma and Texas data, or time step differences (annual vs. daily) between the collected data and model requirements of APEX, which precluded the use of Oklahoma and Texas as benchmark sites, as the data did not match model requirements.

Analysis and Statistics

The NRCS Title 190, National Instruction, Part 302 of the revised 590 Nutrient Management Standard suggested that edge-of-field P losses be categorical for potential risk (USDA–NRCS, 2012), using equivalencies of low (<2.2 kg ha^{-1} yr^{-1}), moderate (2.2–5.5 kg ha^{-1} yr^{-1}), and high (>5.5 kg ha^{-1} yr^{-1}). Water quality data from the benchmark locations included water volume, TP, and dissolved P (DP) concentrations and TP and DP loads (Table 3). Using NRCS equivalencies, TP loads were transformed to P ratings of low, moderate, and high (Table 3), which allowed a comparison between measured water quality losses and P Index ratings.

Kendall's modified Tau (τ_b) for ordinal data (Helsel and Hirsch, 2002) was calculated to test whether there was a significant correlation between how each P Index categorized P-loss risk and the actual risk associated with each field, based on our assigned thresholds for low, moderate, and high risk. To test for significance, the test statistic was computed to get a one-sided *p*-value from a normal distribution table. Correlations were considered significant at $\alpha < 0.05$.

The accuracy of each model and P Index in assigning the correct risk category for each field was calculated using the Heidke skill score (HSS), a metric commonly used for evaluating accuracy of weather forecasts. The HSS measures the fraction of correct forecasts after accounting for correct forecasts due to chance by (Wilkes, 2011):

$$\mathrm{HSS} = \frac{\sum_{i=1}^{n} p(y_i, o_i) - \sum_{i=1}^{n} p(y_i)p(o_i)}{1 - \sum p(y_i)p(o_i)}$$

where the first term in the numerator is the proportion of correct forecasts and the second term in the denominator is the fraction of correct forecasts resulting from chance. A score of one means perfect forecast, whereas a value of zero means that all correct forecasts are due to random chance.

Results and Discussion

All southern P Indices were used to assess P loss using land-use data from the six benchmark sites, except the Arkansas and Oklahoma index tools. The Arkansas P Index is only used for pasture or hay conditions, and the Oklahoma P Index could not be used with the Georgia dataset (Pierson et al., 2001), as more litter was applied than would be allowed by Oklahoma state law.

Measured and Predicted Phosphorus Loads

Measured TP and DP loads from the six benchmark edge-of-field research sites ranged from 0.02 to 19.8 kg ha^{-1} yr^{-1} (Table 3), with 65% of the sites measuring losses <2.2 kg ha^{-1} yr^{-1}. Smith et al. (2014) reported similar losses for the majority of the fields and years monitored in Ohio. Modeled average loads to Lake Erie suggest that the 2.5 kg ha^{-1} contributed yearly may be sufficient for algal production (USDA–NRCS, 2011a). In Ireland, however, measured TP losses were an order of magnitude lower, averaging around 0.10 kg ha^{-1} yr^{-1} or less (Shore et al., 2016). Losses of TP were elevated at the Georgia location ($5.1–19.8$ kg ha^{-1} yr^{-1}) due to the extremely high annual litter application ($\geq$215 kg P ha^{-1} yr^{-1}) (Pierson et al., 2001). The range of DP losses ($0.02–13.4$ kg ha^{-1} yr^{-1}) was almost as great as TP losses (Table 3). The ratio of DP to TP was much greater when animal manures and effluents were applied (Arkansas, Georgia, Oklahoma Demo North, and Texas Goosebranch), with ratios of ~70 to 100% of total losses as DP (Table 3). The ratio of DP to TP loss was greatest (1.0) for the Demo North site in Oklahoma, where manure was applied but measured losses were extremely low (0.02 kg ha^{-1} yr^{-1}). Crop field DP to TP ratios ranged from almost no DP to about 30%, even when poultry litter pellets were surface applied to some fields (North Carolina, organic production).

Phosphorus loads predicted by APEX, APLE, and TBET for TP and DP varied from a low of 0.03 to a high of 15.7 kg ha^{-1} yr^{-1},

Table 2. Factors for the southern Phosphorus (P) Indices derived from six benchmark edge-of-field research sites: Arkansas (Sharpley, unpublished data, 2016), Georgia (Pierson et al., 2001), Mississippi (Yuan et al., 2013), North Carolina (Edgell et al., 2015), Oklahoma, and Texas (White et al., 2012).

Benchmark field sites	Crop	Field management	Timing	Mehlich 3	Measured erosion	RUSLE2† erosion	Applied P
				mg P kg^{-1}	—— Mg ha^{-1} ——		kg P ha^{-1}
		Arkansas					
Check	Fescue hay (*Festuca* L.)	No nutrients applied	April	91	0.03	0.2	0
Broad litter		Litter broadcast		131	0.04	0.2	40
Injected litter		Litter Injected		112	0.03	0.2	40
Injected litter ×2		2× litter rate injected		183	0.03	0.2	80
Continuous grazing	Fescue pasture	Continuous grazing + litter applied		160	0.03	0.2	50
Rotational grazing		Rotational grazing + litter applied		150	0.05	0.2	50
Hay	Fescue hay	Litter applied		135	0.05	0.2	50
		Georgia					
Field 2,1	Bermudagrass [*Cynodon dactylon* (L.) Pers.]/fescue	None	March and September or October	35	0.5	1.6	215
Field 2,2		None		53	0.5	1.6	230
Field 4,1		None		22	0.5	1.7	222
Field 4,2		None		50	0.5	1.7	304
Field 6,1		None		25	0.5	1.7	243
Field 6,2		None		53	0.5	1.7	327
		Mississippi					
Yuan field 1	Cotton (*Gossypium hirsutum* L.)/winter wheat (*Triticum aestivum* L.) or soybeans [*Glycine max* (L.) Merr.]/winter wheat	Reduced tillage	October	38	2.3	6.5	9.5
Yuan field 2		Reduced tillage		50	2.3	6.7	9.5
		North Carolina					
CT,C 2011	Sweet corn (*Zea mays* L.)	Conventional tillage and management	May	27	5.0	5.7	0
CT,C 2012				33	1.0	5.7	0
NT,C 2011		No tillage; conventional management		58	0.1	1.2	0
NT,C 2012				51	0.05	1.2	0
CT,O 2011		Conventional tillage; organic management		54	0.8	5.7	100
CT,O 2012				52	0.3	5.7	69
NT,O 2011		No tillage; organic management		81	0.1	1.2	114
NT,O 2012				69	0.05	1.2	62
		Oklahoma					
Chicksaw	Cotton	Cotton	–	20	3.9	–	67
Cyril	Wheat	Reduced tillage	Fall	35	1.4	–	12
Demo North	Pasture	Pasture	June	50	0	–	44
El Reno	Native grass	0.051 animal units ha^{-1} for 91 d	–	15	–	–	–
		Texas					
Goosebranch	Hay	Broadcast manure	June	435	0.09	–	131
Melde	Sorghum [*Sorghum bicolor* (L.) Moench)/ oats (*Avena strigosa* Schreb.)	Broadcast manure	March and September	34	1.1	–	35/45‡
Patton	Rangeland	0.027 animal units ha^{-1} for 180 d	August	10	0.4	–	29
Reisel	Corn/corn/wheat		April	52	2.9	–	–

† RUSLE2, Revised Universal Soil Loss Equation 2.

‡ Denotes spring-applied/fall-applied P from litter

which was a similar range to the measured data. For this study, neither APLE nor TBET performed better than the other (Table 3). Sometimes TBET predicted greater TP or DP losses than APLE, while at other times APLE gave the greater predictions. The one instance where TP losses were almost always greater for TBET than APLE was in the cropped North Carolina fields; this has implications that, if sediment-attached P is the primary loss pathway, TBET results may be more restrictive to P application than APLE. Overall, APLE, the simpler model, performed about as well as the more complex and calibrated TBET model, although both models had limited accuracy for predicting field-scale P losses in the South (Table 3). Bolster et al. (2017) provided a detailed comparison analysis of the benchmark data and modeled P loads from APLE and TBET. Using a slightly more expanded dataset for the fields evaluated in our study, predictions of DP loss by APLE (model efficiency [E] = 0.52, percent bias [PBIAS] = −9.8%) were slightly better than those with TBET (E = 0.42; PBIAS = 40%) (Bolster et al., 2017). For predictions of TP, model efficiencies for both models were negative, indicating that the models provided worse predictions of TP loss than simply taking the average of the measured values. Percent bias values, however, were low for both models, with values of −8.3 and 1.5% for APLE and TBET, respectively. Bolster et al. (2017) showed that the relatively poor predictions with the APLE model were due, in part, to poor predictions of runoff and erosion. When using measured runoff and

Table 3. Measured total phosphorus (TP) and dissolved phosphorus (DP) loads from the six benchmark edge-of-field research sites (Pierson et al., 2001; White et al., 2012; Yuan et al., 2013; Edgell et al., 2015; Sharpley, unpublished data, 2016), TP and DP loads predicted by the Agricultural Policy Environmental eXtender (APEX), Annual P Loss Estimator (APLE), and Texas Best Management Practice Evaluation Tool (TBET) models, and USDA–NRCS P Index ratings.

Benchmark field sites	Measured			APEX		APLE		TBET		NRCS P Index rating†
	TP	DP	DP:TP ratio	TP	DP	TP	DP	TP	DP	
	kg ha^{-1} yr^{-1}			kg ha^{-1} yr^{-1}						
Arkansas										
Check	0.10	0.07	0.70	1.15	0.93	0.63	0.49	0.58	0.42	Low
Broad litter	0.71	0.60	0.85	0.39	0.33	2.89	2.51	1.26	0.42	Low
Injected litter	0.84	0.76	0.90	0.52	0.49	1.40	0.85	1.67	0.51	Low
Injected litter ×2	0.87	0.79	0.91	0.94	0.91	1.38	1.06	0.99	0.45	Low
Rotational grazing + litter	2.12	1.72	0.81	0.69	0.63	4.71	4.56	1.44	1.24	Low
Continuous grazing + litter	1.52	1.41	0.93	0.42	0.39	5.66	5.47	2.70	2.39	Low
Hay + litter	1.50	1.29	0.86	0.07	0.01	5.44	4.73	2.39	0.96	Low
Georgia										
Field 2,1	5.77	4.09	0.71	0.28	0.20	7.39	7.38	6.78	6.41	High
Field 2,2	5.05	3.20	0.63	0.36	0.32	8.27	8.08	5.38	5.02	Moderate
Field 4,1	9.32	7.20	0.77	0.87	0.32	4.70	4.74	4.14	3.92	High
Field 4,2	13.7	11.7	0.86	0.33	0.27	8.89	14.20	5.88	8.90	High
Field 6,1	10.8	7.53	0.70	0.77	0.55	9.59	9.61	8.60	8.11	High
Field 6,2	19.8	13.4	0.68	0.51	0.50	9.70	15.71	5.97	8.85	High
Mississippi										
Field 1	2.26	0.36	0.16	2.64	0.97	3.23	1.62	1.58	0.07	Moderate
Field 2	2.23	0.60	0.27	2.25	0.88	2.17	0.83	2.08	0.19	Low
North Carolina										
CT,C 2011	7.88	0.18	0.02	0.84	0.33	2.87	0.24	5.3	0.10	High
CT,C 2012	1.62	0.07	0.04	0.84	0.33	4.01	0.19	10.3	0.15	Low
NT,C 2011	0.69	0.15	0.22	0.09	0.03	4.33	0.47	5.8	0.17	Low
NT,C 2012	0.27	0.06	0.22	0.09	0.03	5.37	0.23	9.8	0.16	Low
CT,O 2011	3.08	0.63	0.20	0.23	0.06	11.9	6.29	9.2	0.26	Moderate
CT,O 2012	1.79	0.26	0.15	0.23	0.06	10.1	2.66	15.6	0.23	Low
NT,O 2011	1.05	0.30	0.29	0.05	0.01	9.83	4.43	9.3	0.76	Low
NT,O 2012	0.47	0.14	0.30	0.05	0.01	7.80	1.37	11.7	0.28	Low
Oklahoma										
Chickasha	6.69	1.09	0.16	–	–	1.32	0.25	3.15	0.06	High
Cyril	1.08	0.15	0.14	–	–	0.74	0.06	1.25	0.04	Low
Demo North	0.02	0.02	1.00	–	–	0.05	0.05	0.06	0.06	Low
El Reno	0.28	–	1.00	–	–	0.16	–	0.33	–	Low
Texas										
Goosebranch	1.62	1.11	0.69	–	–	0.16	0.13	0.10	0.09	Low
Melde	1.05	0.06	0.06	–	–	0.53	0.29	0.90	0.07	Low
Patton	0.51	0.20	0.39	–	–	1.34	1.33	0.50	0.16	Low
Riesel	3.07	0.86	0.28	–	–	1.28	0.75	0.76	0.23	Moderate

† NRCS P Index ratings: low (<2.2 kg ha^{-1} yr^{-1}), moderate (2.2–5.5 kg ha^{-1} yr^{-1}), and high (>5.5 kg ha^{-1} yr^{-1}).

erosion data, model efficiencies for APLE increased to 0.62 for DP and from −0.13 to 0.43 for TP.

APEX always predicted lower TP compared with measured losses, except for Mississippi fields, and often these P loss differences were more than an order of magnitude (Table 3). Overall, compared with measured P loads, APEX calibrated predictions were highly variable among the different soil types but compared more favorably for row crops than pasture and better for inorganic fertilizer than animal manures (Ramirez-Avila et al., unpublished data, 2016). The linear regression analysis between measured and APEX-predicted P loads showed no significant relationship at $\alpha = 0.05$. Finally, APEX results were almost always lower than loads predicted by APLE or TBET (Table 3).

Comparison of NRCS and Southern Phosphorus Index Ratings

The USDA–NRCS P Index ratings were associated with measured TP loads using the suggested equivalencies of low (<2.2 kg ha^{-1} yr^{-1}), moderate (2.2–5.5 kg ha^{-1} yr^{-1}) and high (>5.5 kg ha^{-1} yr^{-1}) (USDA–NRCS, 2012). Using this metric, P losses would be low for all benchmark fields except the Georgia site, one of two fields in Mississippi, two of eight site years in North Carolina, and one location each in Oklahoma and Texas (Table 4). High P losses in the Georgia fields were most certainly associated with the very high P applied as poultry litter. The Mississippi fields were adjacent and TP losses were similar, but the TP loss in field 1 was just barely above the 2.2-kg ha^{-1} yr^{-1} threshold. In North Carolina, the conventionally tilled and managed field had large P losses in 2011 due to high erosion rates, hence the high loss rating; TP losses from the conventionally tilled, organically managed field were classified as moderate, probably due to the applied chicken litter pellets. The Chickasaw, OK, location was rated as moderate, while the Riesel, TX, location was rated as high.

Phosphorus Indices were calculated twice, first with erosion determined from measured data, and then using RUSLE2 predictions (Table 4). The RUSLE2 erosion predictions were always much greater than measured sediment losses, often by an order of magnitude or more (Table 2). Most P Indices, however, were insensitive to these differences, generally because predicted sediment losses were low. Since most state USDA–NRCS nutrient management conservation standards require the use of RUSLE2 (USDA–NRCS, 2011b), erosion may be overpredicted and, in some cases, may lead to P Index ratings greater than actual conditions. Researchers have noted a tenfold variation in soil-loss estimates, made using RUSLE2 on pastures, that were consistently greater than measured soil loss (Dabney et al., 2006; USDA–ARS, 2013) primarily due to an underestimation of biomass production for grazed and hayed pastures (Dabney and Yoder, 2012). Harmel et al. (2005) found that the correlation between P loss and index rating increased with measured, rather than estimated, erosion (using RUSLE2), giving r^2 of 0.09 and 0.32 for Arkansas, 0.31 and 0.90 for Iowa, and 0.31 and 0.51 for Texas. Thus, any future P Index refinements must ensure that accurate soil-loss estimates are being given by the most up-to-date, locally relevant version of RUSLE2 to increase the accuracy of any Index.

Not unlike prior analyses (Osmond et al., 2006, 2012), when each state's P Index was determined and then rated using each state's rating system (typically four or five ratings), southern P Index ratings varied for each location and field (Table 4). For most benchmark field sites, state P Index ratings ranged from low to very high or severe, the exceptions being the Mississippi fields and the Oklahoma and Texas locations, where the range was generally from low to high (Table 4).

Management of P is tied to index ratings, which allow nitrogen-based rates for low, P-based rates for medium, and no P with high ratings. Interpreting management outcomes, however, has additional complications, because most southern P Indices still use four or five categories, rather than the three that USDA–NRCS established in 2011 (USDA–NRCS, 2011b); only Kentucky uses three categories. Translation of state P Index ratings to three USDA–NRCS management categories was subjective and varied from state to state (Table 5). Many states (Arkansas, Georgia, Louisiana, Mississippi, North Carolina, Oklahoma, and South Carolina) allow nitrogen-based rates at low and medium, P-based rates at high, and no P applications with very high P Index ratings. The Tennessee P Index rating does not absolutely prevent application of P at very high ratings, but further application is strongly discouraged. Phosphorus management in Texas is a function of both the index value and the water resource of concern.

Overall, southern P Indices matched the USDA–NRCS loss rating 55% of the time. Individual state matches between southern P Indices and USDA–NRCS loss ratings were as follows: Alabama (39%), Arkansas (59%), Florida (52%), Georgia (71%), Kentucky (61%), Louisiana (61%), Mississippi (58%), North Carolina (61%), Oklahoma (44%), South Carolina (77%), Tennessee (55%), and Texas (23%). The additive P Indices (Alabama and Texas) had the lowest USDA–NRCS loss rating correspondence (31%), primarily because Alabama and Texas ratings often overestimated loss and thus were more conservative than other types of indices. Multiplicative (Arkansas, Florida, Louisiana, Mississippi, South Carolina, and Tennessee) and component (Georgia, Kentucky, and North Carolina) indices had similar USDA–NRCS loss rating correspondence—60 and 64%, respectively. Most southern P Index ratings matched USDA–NRCS loss ratings from the measured P losses of Arkansas, Oklahoma, and Texas (mostly rated low) benchmark sites (Table 6). Neither high P losses measured from the Georgia pastures nor mostly low P losses from the North Carolina site were accurately classified, suggesting that high P applications (Georgia) and high or low erosion losses and the application of poultry pellets (North Carolina) were inadequately captured in many Southern indices. Field-scale evaluation of P Indices is difficult, as most states do not have sufficient water quality data, and because rating determinations relative to P losses are subjective. Williams et al. (2017) were able to use TP and DP data from 112 site–years to evaluate the performance of the individual components within the Ohio P Index, separate from the rating thresholds; they concluded that local P loading thresholds are critical to make the Ohio P Index more sensitive to local conditions. Our datasets allowed us to compare the 12 southern P Indices to each other. The data, however, were insufficiently robust to allow us to determine specific state P Index strengths and weaknesses. Sensitivity analyses have been used to determine the importance of individual factors within a P Index (Johnson 2004), but this also requires significant state-specific data, which were unavailable for this analysis. Finally, the wide variation in measured P loss, used as ground-truthing information in the present research, highlights

Table 4. USDA–NRCS Phosphorus (P) Index ratings using measured P and sediment loss from the six benchmark sites and the corresponding 12 southern state P-loss rankings for each benchmark site.

Benchmark field sites	Southern state P Index ratings using measured sediment for erosion losses†												
	NRCS‡	AL	AR	FL	GA	KY	LA	MS	NC	OK	SC	TN	TX
Arkansas													
Check	Low	Low	Low	Low	Low	Low	Low	Low	Low	High	Low	Low	Low
Broad litter	Low	High	Mod	Mod	Low	Low	Low	Mod	Low	High	Low	Low	Mod
Injected litter	Low	High	Low	Mod	Low	Low	Low	Mod	Low	High	Low	Low	Mod
Injected litter ×2	Low	High	Mod	Mod	Low	Low	Low	Mod	Low	Severe	Mod	Low	Mod
Rotational grazing + litter	Low	High	Mod	Mod	Low	Low	Low	Mod	Low	High	Mod	Low	Mod
Continuous grazing + litter	Low	High	Mod	Mod	High	Mod	Low	Mod	Low	Severe	Mod	Low	Mod
Hay + litter	Low	High	Mod	Mod	Low	Low	Low	Mod	Low	High	Mod	Low	Mod
Georgia													
Field 2,1	High	Ex. high	Low	V. high	V. high	Low	Mod	High	High	–§	V. high	Mod	High
Field 2,2	Mod	Ex. high	Low	V. high	V. high	Low	Mod	High	Mod	–	V. high	Mod	High
Field 4,1	High	Ex. high	Low	High	V. high	Low	Mod	High	Low	–	V. high	Mod	Mod
Field 4,2	High	Ex. high	Low	V. high	V. high	Low	High	High	Mod	–	V. high	High	High
Field 6,1	High	Ex. high	Low	Mod	V. high	Mod	Mod	High	Mod	–	V. high	Mod	High
Field 6,2	High	Ex. high	Low	Mod	V. high	Mod	High	High	High	–	V. high	High	High
Mississippi													
Field 1	Mod	High	–	Low	Low	Low	Low	Mod	Low	Mod	Low	Mod	Mod
Field 2	Low	High	–	Low	Low	Mod	Low	Mod	Low	Mod	Low	Mod	Mod
North Carolina													
CT,C 2011	High	Low	–	Low	Low	Low	Low	Low	Low	Low	Low	Low	Low
CT,C 2012	Low	Low	–	Low	Low	Low	Low	Low	Low	Mod	Low	Low	Low
NT,C 2011	Low	Low	–	Low	Low	Low	Low	Low	Low	Mod	Low	Low	Mod
NT,C 2012	Low	Low	–	Low	Low	Low	Low	Low	Low	Mod	Low	Low	Mod
CT,O 2011	Mod	V. high	–	Mod	V. high	Mod	Low	Mod	Low	Mod	High	Mod	Mod
CT,O 2012	Low	V. high	–	Mod	High	Low	Low	Mod	Low	Mod	Mod	Mod	Mod
NT,O 2011	Low	V. high	–	Mod	V. high	Low	Low	High	High	Mod	High	High	Mod
NT,O 2012	Low	V. high	–	Mod	Mod	Low	Low	Mod	Mod	Mod	High	High	Mod
Oklahoma													
Chickasha	High	High	–	Mod	Low	Mod	Low	Low	Low	Low	Mod	High	Mod
Cyril	Low	Mod	–	Low	Low	Low	Low	Low	Low	Mod	Low	Low	Low
Demo North	Low	High	–	Mod	Low	Low	Low	Mod	Low	Mod	Low	Mod	High
El Reno	Low	Low	–	Low	Low	Low	Low	Low	Low	Low	Low	Low	Mod
Texas													
Goosebranch	Low	V. high	–	High	Low	Low	V. high	High	Mod	Severe	Low	Mod	Mod
Melde	Low	High	–	Low	Mod	Low	Low	Mod	Low	Mod	Low	Mod	High
Patton	Low	High	–	Mod	Low	Low	Low	Low	Low	Low	Mod	Low	Low
Riesel	Mod	Mod	–	Low	Low	Low	Mod	Mod	Low	Mod	Low	Low	High

† V., very; Ex., extremely; Mod, moderate.

‡ NRCS P Index ratings: low (<2.2 kg ha^{-1} yr^{-1}), moderate (2.2–5.5 kg ha^{-1} yr^{-1}), and high (>5.5 kg ha^{-1} yr^{-1}).

§ Oklahoma P Index would not allow these high rates of P application

Table 5. Translation of state Phosphorus (P) Index ratings into USDA–NRCS ranking categories of low (<2.2 kg ha^{-1} yr^{-1}), moderate (2.2–5.5 kg ha^{-1} yr^{-1}), and high (>5.5 kg ha^{-1} yr^{-1}) for P management decisions.

State	USDA–NRCS P Index rating		
	N-based (low)	P-based (moderate)	No P (high)
	State P Index rating		
Alabama	Low	Medium, high	Very high
Arkansas	Low, medium	High	Very high
Florida	Low	Medium	High, very high
Georgia	Low, medium	High	Very high
Kentucky	Low	Medium	High
Louisiana	Low, medium	High	Very high
Mississippi	Low, medium	High	Very high
North Carolina	Low, medium	High	Very high
Oklahoma	Low, medium	High	Very high, severe
South Carolina	Low, medium	High	Very high
Tennessee	Low	Medium, high	Very high
Texas†	Low	Medium, high, very high	High, very high with P impairment

† Rankings vary based on water quality impairment.

Table 6. USDA–NRCS Phosphorus (P) Index rating based on measured P loss from the six benchmark sites, the number of southern P Index ratings corresponding to different loss risks, and the percentage of southern index ratings corresponding to different loss risks.

Benchmark field sites	State	NRCS P Index rating†	Number of P Index ratings			P Index ratings		
			Low	Moderate	High	Low	Moderate	High
						%		
Check	AR	Low	11	1	0	92	8	0
Broad litter		Low	8	4	0	67	33	0
Injected litter		Low	8	4	0	67	33	0
Injected litter ×2		Low	8	3	1	67	25	8
Rotational grazing + litter		Low	8	4	0	67	33	0
Continuous grazing + litter		Low	6	5	1	42	50	8
Hay + litter		Low	8	4	0	67	33	0
Field 2,1	GA	High	4	4	4	33	33	34
Field 2,2		Moderate	5	3	4	42	25	34
Field 4,1		High	5	3	4	42	25	34
Field 4,2		High	4	4	4	33	33	34
Field 6,1		High	4	4	4	33	33	34
Field 6,2		High	2	6	4	17	50	33
Field 1	MS	Moderate	8	3	0	73	27	0
Field 2		Low	7	4	0	64	36	0
CT,C 2011	NC	High	11	0	0	100		0
CT,C 2012		Low	11	0	0	100		0
NT,C 2011		Low	10	1	0	91	9	0
NT,C 2012		Low	10	1	0	91	9	0
CT,O 2011		Moderate	5	4	2	46	36	18
CT,O 2012		Low	7	3	1	64	27	9
NT,O 2011		Low	4	5	2	36	46	18
NT,O 2012		Low	7	2	1	64	27	9
Chickasha	OK	High	7	4	0	64	36	0
Cyril		Low	10	1	0	91	9	0
Demo North		Low	8	3	0	73	27	0
El Reno		Low	10	1	0	91	9	0
Goosebranch	TX	Low	4	3	4	36	28	36
Melde		Low	8	3	0	73	27	0
Patton		Low	8	3	0	73	27	0
Riesel		Moderate	8	3	0	73	27	0

† NRCS P Index ratings: low (<2.2 kg ha^{-1} yr^{-1}), moderate (2.2–5.5 kg ha^{-1} yr^{-1}), and high (>5.5 kg ha^{-1} yr^{-1}).

the need for long-term data records and a large number of sites to provide reliable estimates of P runoff. Without this data, our results highlight the difficulties in accurately defining or quantifying loss categories for each state index at the present time.

After transforming APEX predicted loads to P-loss ratings, all ratings except the two Mississippi sites, or 94%, would be classified low (Table 3); data from Oklahoma and Texas benchmark sites could not be used. The APEX predictions matched rating losses 61% of the time, primarily because loss ratings and APEX ratings were low. In addition, APEX could only be run on 17 fields. Direct comparisons between overall APEX results with APLE and TBET, as well as the southern P Indices, must be made with care. The APLE-predicted P ratings designated 11 of the 31 benchmark sites as low (35%), nine as moderate (29%), and 11 as high (35%), with the correct frequency of 48%. The TBET loss ratings partitioned similarly to APLE: 14 low (45%); six moderate (19%); and 11 high (35%), with the correct frequency of 52%. Comparing model-predicted P ratings with southern P Index ratings, all were essentially similar, with a range in prediction frequency between 48 and 61%.

Based on Kendall's τ_b, the risk categorizations assigned to each field by each model and the majority of P Indices (Arkansas and Oklahoma being the exceptions) were positively correlated, with the risk categorizations determined by the measured P-loss data (Table 7). Of these correlations, only the APEX and Kentucky, Louisiana, and North Carolina P Index categorizations were not statistically significant at the 0.05 level. Overall, correlations for the majority of P Indices were similar or better than those of the models. This may be due to the use of measured erosion data for the P Indices and estimated erosion data for models, and/or the fact that the majority of fields have low P losses (20 of 31 fields classified as low risk based on assigned thresholds). The HSS scores were in general agreement with the correlation analysis (i.e., higher HSS values tended to coincide with higher τ_b values) (Table 7). When lower thresholds were constructed (low [<0.57 kg ha^{-1}], moderate [0.57–1.68 kg ha^{-1}], and high [>1.68 kg ha^{-1}]) and then compared with ratings from the water quality models and southern P Indices (Table 7), the greatest impact was on HSS for some of the indices, whereas impact on τ_b was limited. While reducing the threshold values for the categorizations did change (primarily, reduce) τ_b for

Table 7. Kendall modified Tau analysis (τ_b), Z-score, and Heidke skill score (HSS) testing the correct forecast for the models and the southern P Indices† for both USDA–NRCS threshold and reduced threshold conditions using measured erosion rates from the benchmark field sites. APEX, Agricultural Policy Environmental eXtender; APLE, Annual P Loss Estimator; TBET, Texas Best Management Practice Evaluation Tool. All models and state P Indices are significant at $\alpha < 0.05$ except APEX, Kentucky, Louisiana and North Carolina.

	APEX	APLE	TBET	AL	AR	FL	GA	KY	LA	MS	NC	OK	SC	TN	TX
NRCS thresholds and measured erosion rates‡															
τ_b	0.10	0.23	0.26	0.27	−0.18	0.30	0.41	0.23	0.17	0.37	0.18	−0.28	0.43	0.28	0.08
Z-score	0.62	1.99	2.20	2.28	−1.19	2.53	3.37	1.86	1.36	2.98	1.44	−2.38	3.61	2.23	0.61
HSS	0.17	0.19	0.29	0.17	−0.15	0.24	0.46	0.15	0.04	0.21	0.04	−0.28	0.54	0.27	0.06
Reduced thresholds and measured erosion rates§															
τ_b	0.16	0.26	0.30	0.30	−0.20	0.26	0.36	0.25	0.10	0.26	0.12	−0.16	0.27	0.30	0.17
Z-score	1.15	2.21	2.60	2.58	−1.31	2.24	2.99	2.09	0.84	2.11	0.92	−1.40	2.29	2.43	1.42
HSS	0.04	0.20	0.37	0.35	−0.10	0.25	0.21	1.34E−03	−0.02	0.03	0.02	0.02	0.17	−0.07	0.04

† $n = 31$, except for APEX ($n = 23$, not run on OK and TX) and AR P Index ($n = 17$).

‡ NRCS P Index ratings: low (<2.2 kg ha^{-1} yr^{-1}), moderate (2.2–5.5 kg ha^{-1} yr^{-1}), and high (>5.5 kg ha^{-1} yr^{-1}).

§ Reduced ranking equivalencies: low (<0.57 kg ha^{-1}), moderate (0.57–1.68 kg ha^{-1}), and high (>1.68 kg ha^{-1}).

some of the models and P Indices, it did not affect whether the correlation was considered statistically significant (Table 7). The drastic reduction in HSS with the Kentucky P Index is consistent with the fact that it was developed based on rating thresholds suggest by the NRCS, as we used in our analysis (Bolster et al., 2014). These data support the results of Table 6, with an overall pattern of P Indices working as well as, if not better than, the water quality models.

It is critically important, however, to recognize that establishing P Index rating equivalencies tied to management and eventually offsite P loss and water quality impact is difficult and very complex. A recent study from Smith et al. (2014) indicated that, in Ohio, TP losses were <2.2 kg ha^{-1} yr^{-1} and would be sufficient to trigger algal blooms in Lake Erie. If the rating equivalency was established as 1.0 kg ha^{-1}, almost no southern P Index would have matched the NRCS rating (Table 7), but neither would the water quality models. Overall this analysis suggests that P loss from agricultural fields is difficult to predict (Buda et al., 2009; Edgell et al., 2015), and that the simpler, field-based southern P Indices are as predictive as the more complex water quality models.

Conclusion

Comparisons among southern P Indices (Alabama, Arkansas, Florida, Georgia, Kentucky, Louisiana, Mississippi, North Carolina, Oklahoma, South Carolina, Tennessee ,and Texas) were derived using water quality and land treatment data from benchmark sites (Arkansas, Georgia, Mississippi, North Carolina, Oklahoma, and Texas). Measured P loads from the benchmark sites were converted to USDA–NRCS P Index rating equivalents [low (<2.2 kg ha^{-1} yr^{-1}), moderate (2.2–5.5 kg ha^{-1} yr^{-1}), and high (>5.5 kg ha^{-1} yr^{-1})] and compared with P Indices. Concomitantly, the benchmark data were used in three water quality models (APLE, APEX, and TBET) to predict TP losses. When state P Indices were compared against each other and USDA–NRCS loss ratings, there was 55% correspondence. Correspondence with model predictions to USDA–NRCS loss ratings was similar: 61% for APEX, 48% for APLE, and 52% for TBET. Analysis using Kendall's modified Tau indicated that correlations between measured and calculated P-loss ratings were similar or better for most P Indices than the models. These results suggest that southern P Indices are just as robust as the harder-to-use water quality models. However, a critical challenge to risk assessment tools is the difficulty of assigning potential P-loss risk to any given water resource.

Acknowledgments

The authors are grateful for funding supplied by USDA–NRCS Conservation Innovation Grants (Agreement No. 69-3A75-12-182).

References

Bolster, C.H., A. Forsberg, A. Mittelstet, D.E. Radcliffe, D. Storm, J. Ramirez-Avila et al. 2017. Comparing an annual and daily time-step model for predicting field-scale phosphorus loss. J. Environ. Qual. doi:10.2134/jeq2016.04.0159

Bolster, C.H., T. Horvath, B.D. Lee, S. Mehlhope, S. Higgins, and J.A. Delgado. 2014. Development and testing of a new Phosphorus Index for Kentucky. J. Soil Water Conserv. 69:183–196. doi:10.2489/jswc.69.3.183

Bolster, C.H., P.A. Vadas, A.N. Sharpley, and J.A. Lory. 2012. Using a P loss model to evaluate and improve P Indices. J. Environ. Qual. 41:1758–1766. doi:10.2134/jeq2011.0457

Buda, A.R., P.J.A. Kleinman, M.S. Srinivasan, R.B. Bryant, and G.W. Feyereisen. 2009. Effects of hydrology and field management on phosphorus transport in surface runoff. J. Environ. Qual. 38:2273–2284. doi:10.2134/jeq2008.0501

Butler, D.M., D.H. Franklin, M.L. Cabrera, L.M. Risse, D.E. Radcliffe, L.T. West, and J.W. Gaskin. 2010. Assessment of the GA Phosphorus Index on farm at the field scale for grassland management. J. Soil Water Conserv. 65:200–210. doi:10.2489/jswc.65.3.200

Dabney, S.M., and D.C. Yoder. 2012. Improved descriptions of herbaceous perennial growth and residue creation for RUSLE2. Agron. J. 104:771–784. doi:10.2134/agronj2011.0356,

Dabney, S.M., D.C. Yoder, G.R. Foster, and M.A. Nearing. 2006. Application of RUSLE 2 to pasturelands. In: Water management and soil conservation in semi-arid environments. Proceedings of the 14th International Soil Conservation Organization Conference, Marrakech, Morocco. 14–19 May 2006. International Soil Conservation Organization.

DeLaune, P.B., P.A. Moore, Jr., D.K. Carman, A.N. Sharpley, B.E. Haggard, and T.C. Daniel. 2004. Evaluation of the phosphorus source component in the Phosphorus Index for pastures. J. Environ. Qual. 33:2192–2200. doi:10.2134/jeq2004.2192

Dubrovsky, N.M., and P.A. Hamilton, 2010. Nutrients in the nation's streams and groundwater: National findings and implications. Fact sheet 2010–3078. USGS. http://pubs.usgs.gov/fs/2010/3078/pdf/fs20103078.pdf (accessed 20 Apr. 2016).

Edgell, J., D.L. Osmond, D.E. Line, G. Hoyt, J.M. Grossman, and E.M. Larsen. 2015. Comparison of surface water quality and yields from organically and conventionally produced sweet corn plots with conservation and conventional tillage. J. Environ. Qual. 44:1861–1870. doi:10.2134/jeq2015.02.0074

Forsberg, T.A., D.E. Radcliffe, C.H. Bolster, A. Mittelstet, D.E. Storm, and D.L. Osmond. 2017. Evaluation of the TBET model for potential improvement of Southern P Indices. J. Environ. Qual. doi:10.2134/jeq2016.06.0210

Gburek, W.L., A.N. Sharpley, A.L. Heathwaite, and G.J. Folmar. 2000. Phosphorus management at the watershed scale. J. Environ. Qual. 29:130–144. doi:10.2134/jeq2000.00472425002900010017x

Harmel, R.D., H.A. Torbert, P.B. DeLaune, B.E. Haggard, and R.L. Haney. 2005. Field evaluation of three Phosphorus Indices on new application sites in TX. J. Soil Water Conserv. 60:29–41.

Helsel, D.R., and R.M. Hirsch. 2002. Statistical methods in water resources techniques of water resources investigations. Book 4. Chapter A3. USGS. http://water.usgs.gov/pubs/twri/twri4a3/ (accessed 19 Oct. 2016).

Johnson, A.M. 2004. Phosphorus loss assessment in North Carolina. Master's thesis, North Carolina State Univ., Raleigh, NC.

Lemunyon, J.L., and R.G. Gilbert. 1993. Concept and need for a phosphorus assessment tool. J. Prod. Agric. 6:483–486. doi:10.2134/jpa1993.0483

Osmond, D., M. Cabrera, S. Feagley, G. Hardee, C. Mitchell, P. Moore, et al. 2006. Comparing southern P Indices. J. Soil Water Conserv. 61:325–337.

Osmond, D., M. Cabrera, A. Sharpley, C. Bolster, S. Feagley, B. Lee et al. 2012. Comparing Phosphorus Indices from twelve southern USA states against monitored phosphorus loads from six prior southern studies. J. Environ. Qual. 41:1741–1749. doi:10.2134/jeq2012.0013

Pierson, S.T., M.L. Cabrera, G.K. Evanylo, H.A. Kuykendall, C.S. Hoveland, M.A. McCann, and L.T. West. 2001. Phosphorus and ammonium concentrations in surface runoff from grasslands fertilized with broiler litter. J. Environ. Qual. 30:1784–1789. doi:10.2134/jeq2001.3051784x

Sharpley, A., C. Bolster, L. Good, B. Joern, Q. Kettering, J. Lory et al. 2011. Revision of the 590 nutrient management standard: SERA-17 recommendations. Southern Coop. Ser. Bull. 412. Virginia Tech Univ., Blacksburg, VA. http://saaesd.ncsu.edu/docs/SERS17%20Bulletin.pdf. (accessed 20 Apr. 2016).

Sharpley, A.N., D. Beegle, C. Bolster, L. Good, B. Joern, Q. Ketterings et al. 2012. Phosphorus Indices and the 590 revision: Why we need to take stock of how we are doing. J. Environ. Qual. 41:1711–1719. doi:10.2134/jeq2012.0040

Sharpley, A.N., J.L. Weld, D.B. Beegle, P.J.A. Kleinman, W.J. Gburek, P.A. Moore, Jr., and G. Mullins. 2003. Development of Phosphorus Indices for nutrient management planning strategies in the United States. J. Soil Water Conserv. 58:137–151.

Shore, M., P. Jordan, A.R. Melland. P.E. Mellander, N. McDonald, and G. Shortle. 2016. Incidental nutrient transfers: Assessing critical times in agricultural catchments using high-resolution data. Sci. Total Environ. 553:404–415. doi:10.1016/j.scitotenv.2016.02.085

Smith, D.R., K.W. King, L. Johnson, W. Francesconi, P. Richards, D. Baker, and A.N. Sharpley. 2014. Surface runoff and tile drainage transport of phosphorus in the Midwestern United States. J. Environ. Qual. 44:495–502 doi:10.2134/jeq2014.04.0176

Stow, C.A., Y. Kyung Cha, L.T. Johnson, R. Confesor, and R.P. Richards. 2015. Long-term and seasonal trend decomposition of Maumee River nutrient inputs to Western Lake Erie. Environ. Sci. Technol. 49:3392–3400. doi:10.1021/es5062648

USDA and USEPA. 1999. Unified national strategy for animal feeding operations: March 9, 1999. USDA, USEPA, Washington, DC. http://www.epa.gov/npdes/pubs/finafost.pdf (accessed 20 Apr. 2016).

USDA–ARS. 2013. Science documentation: Revised Universal Soil Loss Equation Version 2 (RUSLE2) (for the model with release date of May, 20, 2008). USDA–ARS, Washington, DC. http://www.ars.usda.gov/sp2UserFiles/Place/60600505/RUSLE/RUSLE2_Science_Doc.pdf (accessed 19 Oct. 2016).

USDA–NRCS. 2011a. Assessment of the effects of conservation practices on cultivated cropland in the Great Lakes region. USDA–NRCS, Washington, DC. http://www.nrcs.usda.gov/Internet/FSE_DOCUMENTS/stelprdb1045480.pdf (accessed 20 Apr. 2016).

USDA–NRCS. 2011b. Conservation practice standard: Nutrient management code 590. USDA–NRCS, Washington, DC. http://www.nrcs.usda.gov/Internet/FSE_DOCUMENTS/stelprdb1046177.pdf (accessed 20 Apr. 2016).

USDA–NRCS. 2012. National instructions title 190—Ecological sciences. NI_190_302. 2nd ed. USDA–NRCS, Washington, DC.

USEPA. 2010. Review of empirical approaches for nutrient criteria development. EPA Scientific Advisory Board, Washington, DC. http://yosemite.epa.gov/sab/sabproduct.nsf/95eac6037dbee075852573a00075f732/E09317EC14CB3F2B85257713004BED5F/$File/EPA-SAB-10-006-unsigned.pdf (accessed 20 April 2016).

Vadas, P.A., L.W. Good, P.A. Moore, Jr., and N. Widman. 2009. Estimating phosphorus loss in runoff from manure and fertilizer for a phosphorus loss quantification tool. J. Environ. Qual. 38:1645–1653. doi:10.2134/jeq2008.0337

White, M.W., R.D. Harmel, and R.L. Haney. 2012. Development and validation of the TX Best Management Practice Evaluation Tool (TBET). J. Soil Water Conserv. 67:525–535. doi:10.2489/jswc.67.6.525

Wilkes, D.S. 2011. Statistical methods in the atmospheric sciences. 3rd ed. Elsevier, Oxford, UK.

Williams, J.R., R.C. Izaurralde, and E.M. Steglich. 2012. Agricultural Policy/Environmental eXtender Model theoretical documentation. Version 0806. Blackland Research and Extension Center, Temple, TX.

Williams, M., K. King, G. LaBarge, R. Confesor, and N. Fausey. 2017. Edge-of-field evaluation of the Ohio Phosphorus Risk Index. J. Environ. Qual. doi:10.2134/jeq2016.05.0198

Yuan, Y., M.A. Locke, R.L. Bingner, and R.A. Rebich. 2013. Phosphorus losses from agricultural watersheds in the MS Delta. J. Environ. Manage. 115:14–20. doi:10.1016/j.jenvman.2012.10.028