IDENTIFY METHODS TO REFINE PHOSPHORUS INDICES AND SYNTHESIZE AND EXTEND LESSONS AND OUTCOMES FROM THREE REGIONAL INDEXING EFFORTS

Andrew Sharpley [1], Peter Kleinman [2], Claire Baffaut [3], Doug Beegle [4], Carl Bolster [5], Zachary Easton [6], John Lory [7], Nathan Nelson [8], Deanna Osmond [9], David Radcliffe [10], Tamie Veith [2], and Jennifer Weld [11]

[1] *Crop, Soil and Environmental Sciences, Division of Agriculture, University of Arkansas, Fayetteville, AR.*

[2] *USDA-ARS, Pasture Systems and Watershed Management Research Unit, University Park, PA.*

[3] *USDA-ARS, Cropping Systems and Water Quality Research Unit, Columbia, MO.*

[4] *Plant Sciences, The Pennsylvania State University, University Park, PA.*

[5] *USDA-ARS, Food Animal Environmental Systems Research Unit, Bowling Green, KY.*

[6] *Biological Systems Engineering, Virginia Tech, Blacksburg, VA.*

[7] *Plant Science Division, University of Missouri, Columbia, MO.*

[8] *Agronomy, Kansas State University, Manhattan, KS.*

[9] *Crop and Soil Sciences, North Carolina State University, Raleigh, NC.*

[10] *Crop and Soil Sciences, University of Georgia, Athens, GA.*

[11] *Ecosystem and Science Management, The Pennsylvania State University, University Park, PA.*

SYNOPSIS

In the early 2000's, USDA-NRCS incorporated the P Index risk assessment tool into the 590 Nutrient Management Conservation Practice Standard to identify critical areas of P loss from agriculture and to target mitigation practices. In 2012, because of continued P impairment of surface waters and a disparity among Indices, NRCS funded three Regional Conservation Innovation Grants (CIGs) to assess and review P Indices across 22 states, along with a National CIG tasked with synthesizing and extending the findings of the Regional CIGs. Here we document the outcomes of these projects and identify options to refine and improve Indices. Because of the plethora of field data needed to cover all physical characteristics, management combinations, and spatial and temporal nuances, insufficient field data was available to assess the reliability of Indices to identify site risk as a function of land and P management. Thus, all projects evaluated the capability of nonpoint source models (i.e., APEX, APLE, SWAT and TBET) to provide surrogate estimates of P loss. These comparisons show promise for advancing the weighting and formulation of P Index components, but require careful vetting of simulation models. Also, differences among regional conclusions highlight model strengths and weaknesses. For example, the southern states region found that while fate and transport models could simulate the effects of nutrient management on P runoff, hydrology is often more accurately predicted than total P loads. Elsewhere, particulate P was over-predicted and dissolved P under-predicted, resulting in correct total P predictions but for the wrong reasons. Experience in the U.S. supports expanded regional approaches to P site assessment, assuming closely coordinated efforts that engage science, policy, and implementation communities, but poor precedent exists for uniform national P site assessment tools.

ABBREVIATIONS

APLE, Annual P Loss Estimator; APEX, Agricultural Policy / Environmental eXtender; CIG, Conservation Innovation Grant; NRCS, Natural Resources Conservation Service; P, Phosphorus; SWAT, Soil and Water Assessment Tool; TBET, TX Best management practice Evaluation Tool;
USDA, U.S. Department of Agriculture; USEPA, U.S. Environmental Protection Agency.

INTRODUCTION

Despite concerted efforts to improve phosphorus (P) management practices, P continues to be a major contributor to the impairment of a large proportion of surface waters in the U.S. (Dubrovsky et al. 2010; U.S. Environmental Protection Agency, 2015a, 2015b) and globally (Haygarth et al., 2014; Jarvie et al., 2015; McDowell and Nash, 2012; Schindler, 2012). Harmful algal blooms have been linked to excess P in Western Lake Erie (Michalak et al., 2013; Scavia et al., 2014) and Florida (Reddy et al., 2011; U.S. Environmental Protection Agency - Scientific Advisory Board, 2011), and hypoxia in the Northern Gulf of Mexico (Alexander et al., 2008; Dale et al., 2010; Rebich et al., 2011). These concerns, along with an inability to meet Chesapeake Bay Watershed P-reduction goals (Chesapeake Bay Program, 2013; U.S. Environmental Protection Agency, 2010), have heightened attention on the need to improve P management strategies, as well as identify and target more effective conservation practices.

In 2010, the U.S. Department of Agriculture – Natural Resources Conservation Service (NRCS) recognized the need to revise P Index risk assessment tools, which are a critical component of the National 590 Nutrient Management Conservation Practice Standard (U.S. Department of Agriculture – Natural Resources Conservation Service, 2011). This decision was based to a large extent on an in-depth review and assessment under the auspices of the Southern Extension and Research Activity – 17 (i.e., SERA-17; https://sera17.org/) of P Indices across the U.S. (Sharpley et al., 2012). Variations in outcomes among P Indices in many states (Osmond at al., 2012) and a lack of change in P management and P runoff, led NRCS to fund three regional Conservation Innovation Grants (CIGs) across 22 States. The goals of these regional projects were to assess and review P Indices in the Chesapeake Bay (Delaware, Maryland, New York, Pennsylvania, Virginia, and West Virginia), Heartland (Kansas, Iowa, Missouri, and Nebraska), and Southern U.S. (Alabama, Arkansas, Florida, Georgia, Kentucky, Louisiana, Mississippi, North Carolina, Oklahoma, South Carolina, Tennessee, and Texas) regions (fig. 1). Along with these, a national project synthesized outcomes across the regional efforts.

This paper documents the outcomes of the national project, in terms of the lessons learned from the three regional efforts to assess the accuracy of P Indices to determine the risk of P runoff, to identify shortcomings, and to provide options to refine and improve Indices.

USE OF MODELS

In the early stages of each regional project, it became apparent that there were limited edge-of-field data to verify P Indices in any given State or region. For example, the Heartland region team documented access to over 200 site-years of water quality data from field-scale watersheds within the region. However, these data represented only 27 soil-cropping-management scenarios, of which only three sites had greater than five years of data under constant management. All existing P Indices in the region evaluated long-term risk of P loss. However, the existing data set could not be used to directly test long-term risk assessment, as calculated by a P Index, for the majority of common soil types and management scenarios in the region. Similar limitations in availability of measured edge-of-field data were apparent for the two other regional projects.

Consequently, using nonpoint source models to provide sufficient and appropriate P loss estimates against which P Index values could be compared became a cornerstone technique for all three regions. Models used were calibrated and uncalibrated versions of the Agricultural Policy Environmental eXtender (APEX; ; Steglich and Williams, 2008; Wang et al., 2012) in the Heartland and Southern Regions; the Agricultural Phosphorus Loss Estimator (APLE; Vadas et al., 2009) in the Chesapeake Bay and Southern Regions; and the Soil and Water Assessment Tool (SWAT; Arnold et al., 2012) or Texas Best management practice Evaluation Tool (TBET; White et al., 2012) in the Chesapeake Bay and Southern Regions, respectively. Evaluation of the capability of these models to provide data of sufficient quality to assess P Indices within these regions became a major focus of all three projects.

REGIONAL PROJECT OBJECTIVES

Chesapeake Bay Region

This regional project coordinated the revision, testing, and implementation of P management tools within the states encompassing the Chesapeake Bay watershed, with general objectives of harmonizing site assessment and nutrient management recommendations with the NRCS 590 standard and promoting consistency within each of the Bay's four major physiographic provinces (Coastal Plain, Piedmont, Ridge and Valley, Allegheny Highlands). There were four specific objectives;

1. For each physiographic province, identify site conditions and practices of priority concern and corresponding remedial practices of greatest efficacy and adoption.
2. Evaluate P site assessment tools by comparing their output with water quality monitoring data and fate-and-transport models.
3. Use water quality data (monitored edge-of-field or predicted by model) to refine P Indices, improving their prediction of P loss potential, ensuring consistency across state boundaries and within physiographic provinces, and promoting effective recommendations for P management.
4. Predict the management impact of P Indices (existing and refined) on nutrient management practices and water quality.

Heartland Region

This project's goals were to advance P management by developing and demonstrating procedures that ensure P indices are appropriately tested in accordance with the 2011 NRCS 590 Standard by achieving the following four main objectives;

1. Identify the most effective strategies for using APEX and existing data from watershed and plot studies to evaluate P Indices.
2. Use APEX to extend edge-of-field runoff P loss data to expand periods of time, alternative management practices, and alternative landscapes relative to measured data.
3. Evaluate and improve current P Index formulations in Iowa, Kansas, Missouri and Nebraska.
4. Engage farmers, technical service providers, stakeholder groups, state and regional regulators and state NRCS staff to facilitate acceptance of recommendations in each state, facilitate more

consistency across state borders, and demonstrate the utility of calibrated and validated P Indices for reducing P loss and protecting water quality.

Southern Region

The major goal of the project was to coordinate and advance P management in the Southern region by comparing southern P indices with edge-of-field water quality data, in order to produce consistent results across physiographic regions and thus, promote greater similarity in ratings and recommendations among P Indices from the Southern region. There were four specific objectives:

1. Collect pre-existing water quality and land treatment data from field- or plot-scale (29 sites) sites/treatments where nutrient management site assessment tools can be reliably evaluated for accuracy in predicting P loss potential and in generating nutrient management recommendations that will improve water quality.
2. Compare predictions of P-Index assessment tools to water quality data from benchmark sites.
3. Compare results from fate and transport models (APEX, APLE, and TBET), both calibrated and uncalibrated, against the water quality data. Then compare predictions of P-Index assessment tools against those of the calibrated and uncalibrated fate and transport models.
4. Collaborate with similar projects in the Chesapeake Bay and Heartland regions, and the National overarching CIG project to facilitate application of results to humid regions of the US.

National Synthesis

The objective of this study was to synthesize the recent work on investigating the accuracy of process based models and P indices in identifying the magnitude and extent of P loss risk and their utility for improving water quality and provide suggestions to refine or improve existing models and Indices.

REGIONAL OUTCOMES

Chesapeake Bay Region

- Stakeholder groups were established for each regional physiographic province, and stakeholder surveys were distributed in several Chesapeake Bay states including Delaware, New York, Pennsylvania, and West Virginia. Feedback from stakeholder group meetings and stakeholder survey results assisted in identifying site conditions and practices of priority concern. Specifically, in New York, stakeholder surveys, along with analysis of nutrient management planner supplied database of over 33,000 fields, led to a revision of the structure of the New York P Index that more accurately reflected the risk of P transport from inherent landscape properties (Kettering et al., 2017). Stakeholders recommended that where landscape properties made the risk of runoff high, irrespective of nutrient source management, the application of manure should be discouraged and adoption of risk reducing BMPs encouraged (e.g., cover crops, manure incorporation / injection, application setbacks, and vegetative buffers) (Cela et al., 2016).
- The effects of basic 4R management principles (method, rate, timing, and source; see: International Fertilizer Industry Association, 2009; International Plant Nutrition Institute, 2014) for manure management on P runoff were not accurately represented by current nonpoint source modeling frameworks, such as APEX, SWAT, and TBET (Collick et al., 2016). While Vadas et al. (2007) developed a model that more accurately represents the fate in soil of P added in manure and its release to rainfall-runoff water, the model has not yet been incorporated into publically available versions of these models. When the refined manure P routines were included in SWAT runs, estimates of P runoff as a function of the rate and timing of manure application were improved (Collick et al., 2016). Collick et al. (2016) found these improvements

to hold for simulations of runoff in Pennsylvania. Thus, inclusion of process equations that simulate P fate and transport after manure application is needed to accurately represent current nutrient and land management activities and to provide enhanced prediction of critical P source areas.

- Understanding how to initialize SWAT soil P pools and related simulated soil P pool values with actual soil test values is not a straightforward process. The interactions and synergy throughout this project led to great progress in clarifying how to correctly label and compare SWAT-modeled soil P pools with measured soil P data.
- Fuka et al. (2016) showed that accounting for the influence of field scale topography on soil properties in SWAT can result in more accurate field characterization and improve model predictions of P loss (Collick et al., 2016; Veith et al., 2017; Weld et al., 2017).

Heartland Region

- We were unable to attain our goal of using APEX to provide robust estimates of sediment and total P losses from a range of field-scale agricultural watersheds to test P Indices.
- Uncalibrated models failed to simulate total P and sediment loss. Models such as APEX require calibration with water quality data to reliably estimate total P loss from agricultural systems or conservation management effects on total P loss (Baffaut et al., 2017; Bhandari et al., 2017; Senaviratne et al., 2017).
- APEX reliably estimated P loss when calibrated with locally relevant measured field runoff, P and soil loss data (Baffaut et al., 2017; Bhandari et al., 2017; Senaviratne et al., 2017). However, APEX frequently failed to simulate field-scale sediment loss. Possible reasons included challenges to appropriately calibrating a model under low-sediment conditions (Bhandari et al., 2017; Nelson

et al., 2017; Senaviratne et al., 2017), inability of APEX to simulate erosion processes beyond sheet and rill erosion, and the effects of field delineation on APEX erosion estimates (Senaviratne et al., 2017).

- Use caution when applying calibrated models outside the systems (e.g., soils, geography, climate, and management) used during the calibration and validation (Bhandari et al., 2017; Senaviratne et al., 2017).
- A regional (multi-location) calibration of APEX satisfactorily simulated runoff and total P loss, but was not satisfactory for sediment loss simulation. The applicability of a regionally calibrated model outside the management systems and the characteristics of the locations used for calibration remains uncertain (Nelson et al., 2017).
- Models that failed to reliably simulate runoff, sediment, and P loss were also not reliable at estimating management practice effectiveness.

Southern Region

- Models were usually able to predict runoff accurately (calibrated or uncalibrated) but not P or sediment loss. Prior versions of the Revised Universal Soil Loss Equation – version 2 (RUSLE2; Foster, 2013; Foster et al., 2001) can dramatically over estimate soil loss estimates, especially from pastures, for both models and Indices which use uncorrected versions of RUSLE2. This overestimation of sediment was due to low biomass estimates in RUSLE2 crop management routines (Dabney and Yoder, 2012).
- None of the models used (APEX, APLE, and TBET), were deemed to be appropriate for use as a field-based tool, especially if uncalibrated. Specifically, parameterization, calibration, and use of TBET and APEX were time consuming with high uncertainty, while APLE was less time-consuming (Bolster et al., 2017; Forsberg et al., 2017; Ramirez-Avila et al., 2017).

- Statistical analysis of the relationship between measured and calculated P-loss ratings were similar or better for most P Indices than the three models used (i.e., APEX, APLE, and TBET; Osmond et al., 2017). However, further work is needed to establish a framework that details how to assign P loss to a potential risk category, as receiving waters have differing sensitivities to P inputs, which is generally not considered in P Indices.

OVERARCHING OUTCOMES

Lessons Learned for Model Developers

Updating nonpoint source models has proceeded rapidly for some processes, such as the representation of landscape hydrology (Fuka et al., 2016) and use of GIS-based layers of pertinent soil characteristics. Some processes, however, have not been adequately updated to reflect improved understandings of the processes involved. This is particularly true for the fate of manure-derived P in models such as APEX, SWAT, and TBET which are based on the Erosion-Productivity Impact Calculator (EPIC). The original EPIC routines (Jones et al., 1984) were based on mineral fertilizer additions of P, which are immediately plant available for the most part, and thus may not accurately reflect P cycling when added to soil in the form of animal manure, which can slowly release P. Recent models developed by Vadas et al. (2007), more accurately describe the breakdown, fate, and transport of land applied P in manures; however, they have not yet been widely incorporated into nonpoint source models.

The findings of Collick et al. (2016) clearly show that use of SWAT without updated P cycling routines may produce correct results for the wrong reason. The comparisons in Figure 2 are for event-based output, which are not commonly reported in verification efforts of fate-and-transport models. With simulations spanning multiple seasons or years, analyses over longer time periods show that manure management effects on average P loss trends can be derived from the original P routines of Jones et al. (1984). That is, gradual changes in soil P and soil P release eventually converge with the dynamic, short-term effects of recently applied manure wash off and soil pool imbalances. Other process-based models will likely benefit from similar improvements to the P cycling routines. For example, Forsberg et al. (2017) found that total P loss in runoff was accurately simulated but there were clear indications it may have been for the wrong reasons. Here, dissolved P losses were under-predicted by TBET, while sediment losses and thus, sediment bound P, were over-predicted resulting in reasonable predictions of total P loss (Forsberg et al., 2017).

If fundamental processes governing P loss to soil or water, such as erosion and/or P cycling from applied sources are inadequately represented, the calibrated models will likely lack portability across landscapes and management systems, which agrees with conclusions of multiple studies (Baffaut et al., 2017; Bhandari et al., 2017; Bolster et al., 2017; Nelson et al., 2017; Ramirez-Avila et al., 2017; Senaviratne et al. 2017).

Attaining the goal of using models to extend measured data with robust estimates of sediment and total P for expanded periods of time, alternative management practices, and alternative landscapes will require continued investment in model development.

Lessons Learned for Potential Model Users

The potential of fate and transport models to provide accurate estimates of P loss for multiple locations, climates, and scenarios has led to widespread interest to evaluate water quality policy and program outcomes at a watershed scale (e.g., U.S. Environmental Protection Agency, 2010; Whittaker et al., 2015). Indeed, experience exists with “quantitative” P Indices that leverage fate and transport to predict edge-of-field runoff under different management and site conditions (Good et al., 2012; Osmond et al., 2017; White et al., 2010). A common theme is to develop stripped-down versions of fate and transport models that can be exported to sites (or fields) where they have not previously been applied for use by nutrient management planners who lack expertise in fate and transport modeling.

In every region (Chesapeake Bay, Heartland and South), fundamental problems were identified in applying fate and transport models to areas where they had neither been carefully calibrated or corroborated. These concerns would be amplified substantially if end users lacked modeling expertise. Specifically, parameterization, calibration, and use of SWAT, TBET and APEX were time consuming with high uncertainty, while APLE was less time-consuming (Bolster et al., 2017; Forsberg et al., 2017; Ramirez-Avila et al., 2017). Circumventing or avoiding the calibration process, or relying on expert opinion to set model parameters, resulted in poor model estimates for P and sediment in several applications (Baffaut et al., 2017; Forsberg et al., 2017; Ramirez-Avila et al., 2017; White et al., 2010). Clearly, caution is required when applying calibrated models outside the systems (e.g., soils, geography, climate, and management) used during the calibration and validation (Fig. 3; Bhandari et al., 2017; Nelson et al., 2017; Senaviratne et al., 2017).

Lessons Learned About Using Models and P Indices

Most P Indices focus on assessments of long-term risk. For example, they use 30-year records of weather data and use RUSLE2 estimates of long-term annual average soil loss for a specific year in a rotation or across all years of a rotation. Most water-quality data sets capture events during un-frozen periods of the year and frequently have only a few years of data representing a specific management practice (Harmel et al., 2006; Harmel et al., 2008). Even in regions with comparatively extensive management and water quality data (e.g., Veith et al. 2015), variability in time and space prevents clear, controlled, cause and effect conclusions. Consequently, measured water quality and weather data for a given location typically do not match the length of time required to appropriately test P Indices. However, research supported by this project and others has documented that calibrated models work well when used within the soils, nutrient, and land management conditions for which they were developed (Baffaut et al., 2017; Bhandari et al., 2017; Forsberg et al., 2017; Senaviratne et al., 2017; Sommerlot et al., 2013).

A major objective of model use has been to estimate losses from alternative management practices for the calibrated location. Generally, APEX did a good job estimating runoff volume for alternate scenarios (Bhandari et al., 2017; Nelson et al., 2017; Ramirez-Alvia et al., 2017; Senaviratne et al., 2017). Similarly, Forsberg et al. (2017) found that the calibrated TBET model did a good job of predicting runoff. However, the ability of models to estimate the P and sediment loss was limited, but was more likely to be successful if several management types were present in the calibration data set (Bhandari et al., 2017; Forsberg et al. 2017; Ramirez-Alvia et al., 2017; Senaviratne et al., 2017).

Another issue confronted during the comparison of models and P Indices is that it can be very difficult to translate some common management parameters into appropriate model parameters, to ensure that comparisons between model output and P Index calculations are directionally and magnitudinally correct. For example, there is no direct way in SWAT to relate a soil test P concentration, which is a common management parameter used in P Indices, to the pools of P in the model, so as to be sure that the same soil P level is being represented in both the model and the P Index.

Osmond et al. (2017) showed differences in ratings produced by southern P Indices. To a certain extent, this assessment affirms earlier findings by Osmond et al. (2006) and Osmond et al. (2012), and begs the question of how much progress has been made in refining and ensuring that Indices adequately reflect site vulnerability to P loss in runoff. Several states have implemented newly revised Indices in the last five years. For example, Maryland (Maryland Department of Agriculture, 2016), Arkansas (DeLaune et al., 2006; Sharpley et al., 2010b), Kentucky (Bolster et al., 2014), and Texas (White et al., 2012) all recently revised their Indices. All the revised Indices have provided more restrictive nutrient management for similar site conditions than their preceding versions.

In Arkansas for instance, information gathered during the nutrient management planning process between 2004 and 2015 was used to compare risks assigned by the original and revised P Indices (Sharpley et al., 2012). Revision of the Index involved increasing the sensitivity of runoff P concentration to increasing soil test P based on information from plot-scale, simulated rainfall – runoff research studies, adding the mineralization of manure to provide additional source P following land application, and increasing the risk assigned to P applications made during the rainy spring and early summer months (e.g., March to June). Between 2004 and 2015, 18,278 fields were assessed, with the revised Index being on average 1.11 times greater than the original Index (Fig. 4). The average risk assigned over the 12-year period was 66 for the original and 83 for the revised Index, which transitions average site interpretation from a "medium" to "high" risk category, along with an associated reduction in manure application rate from N- to P-based (i.e., typically from 7 to 3 Mg broiler litter ha^{-1}).

Lessons Learned About Regional Research Collaboration

The grant structure for this work was unique in that it included three regional projects and a synthesis project funded, with similar but not identical objectives. Additionally, each regional group pursued a somewhat different approach to modeling and used different models. This structure had the advantage of providing information from a wide geographical area. Interestingly, each group had similar modeling results regardless of model or region; runoff and total P (the sum of dissolved and particulate P in runoff) were reliably predicted with model calibration, and sediment was often under predicted, even with calibration. The synthesis component provided constant communication between the regional projects and provided a synthetic analysis of the lessons learned. We believe that consideration of funding structures that involve similar regional transdisciplinary collaboration among researchers, extension, NRCS, and other farm advisory agencies would help better address agricultural questions and provide applicable outcomes.

Implications to National P Site Assessment Approaches

Given the diversity of P site assessment tools used by U.S. states, there is persistent interest in developing a single, national approach to P site assessment. There is a large body of research, conducted over the last twenty years, which has assessed the impacts of nutrient and land management on P runoff at a field scale, addressing all physical characteristics, management combinations, and spatial and temporal scales. Despite this, there were insufficient data across the Southern, Chesapeake and Heartland regions to adequately verify Indices with sufficient technical rigor using edge-of-field data alone.

One of the most consistent findings of P site assessment efforts was that, at present, there is no scientific justification to implement a single, national P Index in the US. The differences in regional and state-wide nutrient and land management priorities, landscape properties, climatic regimes, and dominant hydrologic process were so great as to render any attempt at a single, national P Index exceeding difficult. Fate and transport models remain research tools that are not yet capable of providing accurate estimates of P loss under the diverse set of management scenarios and locations necessary to test and/or replace the current state-by-state system of P Indices. For instance, results from this assessment suggest that southern P Indices are just as robust as the harder to use fate and transport models (Osmond et al., 2017).

Another consistent conclusion is that there are still limitations to the predictive capability of models regarding the effects of management practices on P runoff. The use of use fate and transport models to estimate the impacts of agricultural management on P and sediment loss in runoff, including accessing data, parameterizing models, and verifying estimates, is a labor intensive and a complex process. While it is possible to successfully run a model with a limited understanding of the input data and without calibration, this research confirmed that such practices result in poor water quality predictions and undermine the validity of the resulting model outcomes and subsequent recommendations.

Our assessment of the models should not, however, be taken as a lack of confidence in the potential of these models to contribute to a better understanding of P loss from agricultural systems. Instead, our results emphasize that progress in understanding water quality from small agricultural fields remains a three-legged stool approach. Future success requires continued support for:

1. collection of soil, water quality, and land management data through field-scale watershed studies;
2. application of models to extend the conclusions of measured data, particularly through variations in climate and management, and to provide broader understandings of individual and combined uncertainties in the natural system;
3. Continued updating of existing process-based models based on experimental knowledge, particularly for sediment loss and P transport processes at a field-scale.

Finally, there remains a critical disconnect between assigned P loss risk and biological response of any given receiving water resource (Hirt, 2016). Including site-specific variables that account for the sensitivity or biological response of a water body to P inputs into the Indexing framework adds a complexity that few states have been willing to adopt. A first step to overcome these challenges would be to base the P Indexing framework on an ecoregional delineation. However, because P Indices are established, approved, and legislated at a state level, it will be both technically and politically problematic to move ownership from state boundaries to an ecoregional or watershed delineation.

ACKNOWLEDGEMENTS

This research was supported by four USDA-NRCS Conservation Innovation Grants to assess the accuracy of P Indices to determine the risk of P runoff, to identify shortcomings, and to provide options to refine and improve Indices.

REFERENCES

Alexander, R.B., R.A. Smith, G.E. Schwarz, E.W. Boyer, J.V. Nolan, and J.W. Brakebill. 2008. Differences in phosphorus and nitrogen delivery to the Gulf of Mexico from the Mississippi River. Environ. Sci. Technol. 42:822-830.

Arnold, J.G., D. Moriasi, P.W. Gassman, K.C. Abbaspour, M.J. White, M.S. Srinivasan, et al. 2012. SWAT: Model use, calibration, and validation. Trans. Am. Soc. Agric. Biol. Eng. 55: 1491-1508.

Baffaut, C., N.O. Nelson, J.A. Lory, G.M.M.M.A. Senaviratne, A.B. Bhandari, R.P. Udawatta, D.W. Sweeney, M.J. Helmers, M.W. Van Liew, A.P. Mallarino, and C.S. Wortmann. 2017. Multi-site evaluation of APEX for water quality: 1. Best professional judgement parameterization. J. Environ. Qual. *In press*.

Bhandari A.B., N.O. Nelson, D.W. Sweeney, C. Baffaut, J.A. Lory, G.M.M.M.A. Senaviratne, G.M. Pierzynski, K.A. Janssen, and P.L. Barnes. 2017. Calibration of the APEX model to simulate management practice effects on runoff, sediment, and phosphorus loss. J. Environ. Qual. 46: *In press.*

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.

Bolster, C.H., A. Forsberg, A. Mittelstet, D.E. Radcliffe. D. Storm, J. Ramirez-Avila, A.N. Sharpley, and D. Osmond. 2017. Comparing an annual and daily time step model for predicting field-scale phosphorus loss. J. Environ. Qual. *In press.*

Cela, S., Q.M. Ketterings, K.J. Czymmek, J. Weld, D.B. Beegle, and P.J.A. Kleinman. 2016. Nutrient management planners' feedback on New York and Pennsylvania phosphorus indices. J. Soil Water Conserv. 71:281-288.

Chesapeake Bay Program. 2013. Bay barometer 2011 – 2012. Spotlight on health and restoration in the Chesapeake Bay and its watershed. Chesapeake Bay Program, Annapolis, MD. 7 pages. Available at http://www.chesapeakebay.net/documents/10-Pg_CBP_Barometer_for_web.pdf.

Collick, A.S., T.L. Veith, D.R. Fuka, P.J.A. Kleinman, A.R. Buda, J.L. Weld, R.B. Bryant, P.A. Vadas, M.J. White, D. Harmel, and Z.M. Easton. 2016. Improved simulation of edaphic and manure phosphorus loss in SWAT. J. Environ. Qual. *In press*. doi:10.2134/jeq2015.03.0135.

Dabney, S.M., and D.C. Yoder. 2012. Improved description of herbaceous perennial growth and residue creation for RUSLE2. Agron. J. 104 (3):771-784.

Dale, V.H., C.L. Kling, J.L. Meyer, J. Sanders, H. Stallworth, T. Armitage, D. Wangsness, T.S. Bianchi, A. Blumberg, W. Boynton, D.J. Conley, W. Crumpton, M.B. David, D. Gilbert, R.W. Howarth, R. Lowrance, K.R. Mankin, J. Opaluch, H.W. Paerl, K. Reckhow, A.N. Sharpley, T.W. Simpson, C. Snyder, and D. Wright. 2010. Hypoxia in the Northern Gulf of Mexico. Springer Series on Environmental Management. New York, NY: Springer Science.

DeLaune, P.B., Haggard, B.E., Daniel, T.C., Chaubey, I., and Cochran, M.J. 2006. The Eucha/Spavinaw Phosphorus Index: A court mandated index for litter management. J. Soil Water Conserv. 61:96-105.

Dubrovsky, N.M., K.R. Burow, G.M. Clark, J.M. Gronberg, P.A. Hamilton, K.J. Hitt, D.K. Mueller, and M.D. Munn. 2010. The quality of our Nation's waters—Nutrients in the Nation's streams and groundwater, 1992–2004: U.S. Geological Survey Circular 1350. 174 pages. Available at: http://water.usgs.gov/nawqa/nutrients/pubs/circ1350.

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. *In press*.

Foster, G.R. 2013. Science Documentation User's Guide Version 2 RUSLE2. USDA – Agricultural Research Service, Washington, D.C. Available at http://www.ars.usda.gov/sp2UserFiles/Place/60600505/RUSLE/RUSLE2_Science_Doc.pdf.

Foster, G.R., D.C. Yoder, G.A. Weeseis, and T.J. Toy. 2001. The design philosophy behind RUSLE2: Evolution of an empirical model. p. 95-98. In J.C. Ascough II and D.C. Flanagan (eds.), Soil Erosion Research for the 21st Century. Proc. Int. Symposium, January 2001, Honolulu, HI. Am. Soc. Am. Eng. 701P0007. St. Joseph, MI:

Fuka, D.R., A.S. Collick, P.J.A. Kleinman, D. Auerbach, D, Harmel, and Z.M. Easton. 2016. Improving the spatial representation of soil properties and hydrology using topographically derived initialization processes in the SWAT model. Hydrol. Proc. doi: 10.1002/hyp.10899.

Garcia, A.M., R.A. Alexander, J.G. Arnold, L. Norfleet, M.J. White, D.M. Robertson, and G. Schwarz. 2016. Regional effects of Agricultural conservation practices on nutrient transport in the Upper Mississippi River Basin. Environ. Sci. Technol. 50(13):6991-7000.

Good, L.W., P.A. Vadas, J.C. Panuska, C.A. Bonilla, and W.E. Jokela. 2012. Testing the Wisconsin phosphorus index with year-round, field-scale runoff monitoring. J. Environ. Qual. 41:1730-1740.

Harmel, R.D., R.J. Cooper, R.M. Slade, R.L. Haney, J.G. Arnold. 2006. Cumulative uncertainty in measured streamflow and water quality data for small watersheds. Trans. Am. Soc. Agric. Biol. Eng. 49(3): 689-701.

Harmel, D., S. Qian, K. Reckhow, and P. Casebolt. 2008. The MANAGE database: Nutrient load and site characteristic updates and runoff concentration data. J. Environ. Qual. 37:2403-2406.

Haygarth, P.M., H.P. Jarvie, S.M. Powers, A.N. Sharpley, J. Elser, J. Shen, H.M. Peterson, N.-I. Chan, N.J.K. Howden, T. Burt, F. Worrall, F. Zhang, and X. Liu. 2014. Sustainable phosphorus management and the need for a long-term perspective: The legacy hypothesis. Environ. Sci. Technol. 48:8417-8419.

Hirt, Claire Regan. 2016. Stream condition and nutrient runoff: Linking the Soil and Water Assessment Tool (SWAT) model with empirical ecological measures in an agricultural watershed in central Pennsylvania. M.S. thesis, Geography. The Pennsylvania State University. https://etda.libraries.psu.edu/catalog/28815.

International Fertilizer Industry Association. 2009. The Global "4R" nutrient stewardship framework: developing fertilizer best management practices for delivering economic, social and environmental benefits. IFA, Paris, France. Available at: https://www.ipni.net/ipniweb/portal/4r.nsf/0/BAB4157B488871A385257DF100739D94/$FILE/The%20Global%204R%20Nutrient%20Stewardship%20Framework.pdf.

International Plant Nutrition Institute. 2014. 4R Nutrient stewardship portal. Available at: http://www.ipni.net/4R.

Jarvie, H.P., A.N. Sharpley, D. Flaten, P.J.A. Kleinman, A. Jenkins, and T. Simmons. 2015. The pivotal role of phosphorus in a resilient water-energy-food security nexus. J Environ. Qual. 44(5):1308-1326.

Jones, C. A., C.V. Cole, and A.N. Sharpley. 1984. A simplified soil and plant phosphorus model. I. Documentation. Soil Sci. Soc. Am. J. 48:800-805.

Kettering, Q.M., S. Cela, A. Collick, S. Crittenden, and K.J. Czymmek. 2017. Restructuring the P Index to better address P management in New York. J. Environ. Qual. *In press*.

Maryland Department of Agriculture. 2016. The agricultural phosphorus initiative. Available at http://mda.maryland.gov/Pages/PMT.aspx.

McDowell, R.W., and D. Nash. 2012. A review of the cost-effectiveness and suitability of mitigation strategies to prevent phosphorus loss from dairy farms in New Zealand and Australia. J. Environ. Qual. 41:680–693.

McGrath, J.M., F.J. Coale, and N.M. Fiorellino. 2015. University of Maryland Phosphorus Management Tool: Technical users guide. University of Maryland Extension and The Department of

Environmental Science and Technology publication, Extension Bulleting EB-405. Available at http://extension.umd.edu/sites/default/files/_docs/articles/EB-405%20UMD%20Phosphorus%20Management%20Tool-Technical%20Users%20Guide.pdf. 16 pages.

Michalak. A.M. E. J. Anderson, D. Beletsky, S. Boland, N.S. Bosch, T.B. Bridgeman, J.D. Chaffin, K. Cho, R. Confesor, I. Daloglu, J. V. DePinto, M A. Evans, G. L. Fahnenstiel, L. He, J. C. Ho, L. Jenkins, T.H. Johengen, KC Kuo, E. Laporte, X Liu, MR.McWilliams, M.R. Moore, D. J. Posselt, R. Peter Richards, D. Scavia, A. L. Steiner, E. Verhamme, D. M. Wright and M.A. Zagorski. 2013. Record-setting algal bloom in Lake Erie caused by agricultural and meteorological trends consistent with expected future conditions. PNAS 110 6448 – 6452.

Nelson, N.O., C. Baffaut, J.A. Lory, G.M.M.M.A. Senaviratne, A.B. Bhandari, R.P. Udawatta, D.W. Sweeney, M.J. Helmers, M.W. Van Liew, A.P. Mallarino, and C.S. Wortmann. 2017. Multi-site evaluation of APEX for water quality: II. Regional parameterization. J. Environ. Qual. *In press*.

Osmond, D., M. Cabrera, S. Feagley, G. Hardee, C. Mitchell, P. Moore, R. Mylavarapu, J. Oldham, J. Stevens, W. Thom, F. Walker, and H. Zhang. 2006. Comparing Southern P Indices. J. Soil Water Conserv. 61:325-337.

Osmond, D.L., A.N. Sharpley, C. Bolster, M. Cabrera, S. Feagley, B. Lee, C. Mitchell, R. Mylavarapu, L. Oldham, F. Walker, and H. Zhang. 2012. Comparing phosphorus indices from twelve southern USA states against monitored phosphorus loads from six prior southern studies. J. Environ. Qual. 41:1741-1750.

Osmond, D., C. Bolster, A.N. Sharpley, M. Cabrera, S. Feagley, A. Forsberg, C. Mitchell, R. Mylavarapu, J. L. Oldham, D. E. Radcliffe, J. J. Ramirez-Avila, D.E. Storm1, F. Walker1, and H. Zhang. 2017. Southern P Indices, water quality data and modeling (APPEC, APLE and TBET) results: A comparison. J. Environ. Qual. *In press*.

Ramirez-Aliva, J.J., Osmond, D. Radcliffe, C. Bolster, S.L. Ortega-Achury, A. Forsberg, and A. Sharpley, J.L. Oldham. 2017. Evaluation of the APEX model to simulate runoff quality from agricultural fields in the southern region of the U.S. J. Environ. Qual. *In press*.

Rebich, R.A., N.A. Houston, S.V. Mize, D.K. Pearson, P.B. Ging, C.E. Hornig. 2011. Sources and delivery of nutrients to the northwestern Gulf of Mexico from streams in the south-central United States. J. Am. Water Res. Assoc. 47(5):1062-1086.

Reddy, K.R., S. Newman, T.Z. Osborne, J.R. White, and H.C. Fitz. 2011. Phosphorus cycling in the Greater Everglades Ecosystem: Legacy Phosphorus implications for management and restoration. Crit. Reviews Environ. Sci. Technol. 41(6):149-186.

Scavia, D., J.D. Allan, K.K. Arend, S. Bartell, D. Beletsky, N.S. Bosch, S.B. Brandt, R.D. Briland, I. Daloglu, J.V. DePinto, D.M. Dolan, M.A. Evans, T.M Farmer, D. Goto, H. Han, T.O. Hoeoek, R. Knight, S.A. Ludsin, D. Mason, A.M. Michalak, R.P. Richards, J.J. Roberts, D.K. Rucinski, E. Rutherford, D.J. Schwab, T.M. Sesterhenn, H. Zhang, Y. Zhou.. 2014. Assessing and addressing the re-eutrophication of Lake Erie: Central basin hypoxia. J. Great Lakes Res. 40: 226-246.

Schindler, D.W. 2012. The dilemma of controlling cultural eutrophication of lakes. Proc. Royal Soc. Biol. Sci. 12 pages. DOI: 10.1098/rspb.2012.1032. Available at http://rspb.royalsocietypublishing.org/content/early/2012/08/14/rspb.2012.1032.

Senaviratne, G.M.M.M.A., C. Baffaut, J.A. Lory, R.P. Udawatta, N.O. Nelson, and A.B. Bhandari. 2017. Evaluation of four parameterization strategies for the APEX model. J. Environ. Qual. 46: *In press*.

Sharpley, A.N., D.G. Beegle, C. Bolster, L.W. Good, B. Joern, Q. Ketterings, J. Lory, R. Mikkelsen, D. Osmond, and P.A. Vadas. 2012. Phosphorus indices: Why we need to take stock of how we are doing. J. Environ. Qual. 41:1711-1718.

Sharpley, A.N., M. Daniels, K. VanDevender, B. Haggard, N. Slaton, and C. West. 2010a. Using the 2010 Arkansas Phosphorus Index. Cooperative Extension Service, Division of Agriculture, University of

Arkansas. Miscellaneous Publication MP487. 17 pages. Available at http://www.uaex.edu/Other_Areas/publications/PDF/MP487.pdf.

Sharpley, A.N., P.A. Moore, Jr., K. VanDevender, M. Daniels, W. Delp, B. Haggard, T.C. Daniel, and A. Baber. 2010b. Arkansas Phosphorus Index. Cooperative Extension Service, Division of Agriculture, University of Arkansas. Fact Sheet FSA 9531. 8 pages. Available at http://www.uaex.edu/Other_Areas/publications/PDF/FSA-9531.pdf.

Sommerlot, A.R., A.P. Nejadhashemi, and S.A. Woznicki. 2013. Evaluating the impact of field-scale management strategies on sediment transport to the watershed outlet. J. Environ. Managt. 128:735-748.

Steglich, E., and J.R. Willliams. 2008. Agricultural Policy Environmental Extender: User's manual version 0604. Blackland Research and Extension Center. Available at http://epicapex.brc.tamus.edu/media/25398/the%20apex%20user%20manual%207-8-10.pdf.

U.S. Department of Agriculture - Natural Resources Conservation Service. 2011. Conservation Practice Standard, Nutrient Management 590. Available at http://www.nrcs.usda.gov/Internet/FSE_DOCUMENTS/stelprdb1046177.pdf.

U.S. Department of Agriculture – Natural Conservation Resources Service. 2015. Mississippi River Basin Healthy Watershed Initiative: Conservation beyond boundaries. 2014 Progress Report. Fact Sheet available at http://www.nrcs.usda.gov/wps/portal/nrcs/detailfull/national/home/?cid=stelprdb1048200.

U.S. Environmental Protection Agency. 2010. Draft Chesapeake Bay Phase 5.3; Community Watershed Model. EPA 903S10002 - CBP/TRS-303-10. U.S. Environmental Protection Agency, Chesapeake Bay Program Office, Annapolis MD. Available at http://ches.communitymodeling.org/models/CBPhase5/documentation.php.

U.S. Environmental Protection Agency. 2015a. Water quality assessment and total maximum daily loads information. Available at http://www.epa.gov/waters/ir/.

U.S. Environmental Protection Agency. 2015b. Watershed assessment, tracking & environmental results. National Probable Sources Contribution to Impairments. Available at http://iaspub.epa.gov/waters10/attains_nation_cy.control#prob_source.

U.S. Environmental Protection Agency - Scientific Advisory Board. 2011. Review of EPA's draft Approaches for deriving numeric nutrient criteria for Florida's estuaries, coastal waters, and southern inland flowing waters. EPA Scientific Advisory Board, Washington, DC. 67 pages. Available at http://yosemite.epa.gov/sab/sabproduct.nsf/DCC3488B67473BDA852578D20058F3C9/$File/EPA-SAB-11-010-unsigned.pdf.

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.

Vadas, P.A., W.J. Gburek, A.N. Sharpley, P.J.A. Kleinman, P.A. Moore Jr., M.L. Cabrera, et al. 2007. A model for phosphorus transformation and runoff loss for surface-applied manures. J. Environ. Qual. 36: 324-332.

Veith, T.L., S.C. Goslee, D.B. Beegle, J.L. Weld, and P.J.A. Kleinman. 2017. Analyzing the distribution of hydrogeomorphic characteristics across Pennsylvania as a precursor to Phosphorus Index modifications. J. Environ. Qual. 46: *In press*.

Veith, T.L., J.E. Richards, S.C. Goslee, A.S. Collick, R.B. Bryant, D.A. Miller D.A. Miller, B.W. Bills, A.R. Buda, R.L. Sebring, and P.J.A. Kleinman. 2015. Navigating spatial and temporal complexity in developing a long-term land use database for an agricultural watershed. J. Soil Water Conserv. 70(5):288-96.

Wang, X., J.R. Williams, P.W. Gassmann, C. Baffaut, R.C. Izaurralde, J. Jeong, et al. 2012. EPIC and APEX: Model Use, Calibration, and Validation. Trans. Am. Soc. Agric. Biol. Eng. 55: 1447-1462.

Weld, J.L., T.L. Veith, A.S. Collick, D.B. Beegle, R. Day, P.J.A. Kleinman. 2017. Comparing phosphorus transport approaches in TopoSWAT and the Pennsylvania phosphorus index. J. Environ. Qual. 46: *In press*.

White, M.J., R.D. Harmel, and R.L. Haney. 2012. Development and validation of the Texas Best Management Practice Evaluation Tool (TBET). J. Soil Water Conserv. 67(6):525-535.

White, M.J., D.E. Storm, P.R. Busteed, M.D. Smolen, H. Zhang, and G.A. Fox. 2010. A quantitative phosphorus loss assessment tool for agricultural fields. Environ. Modelling Software 25:1121-1129.

Whittaker, G., B.L. Barnhart, R. Srinivasan, and J.G. Arnold. 2015. Cost of areal reduction of gulf hypoxia through agricultural practice. Sci. Total Environ. 505:149-153.

LIST OF FIGURES

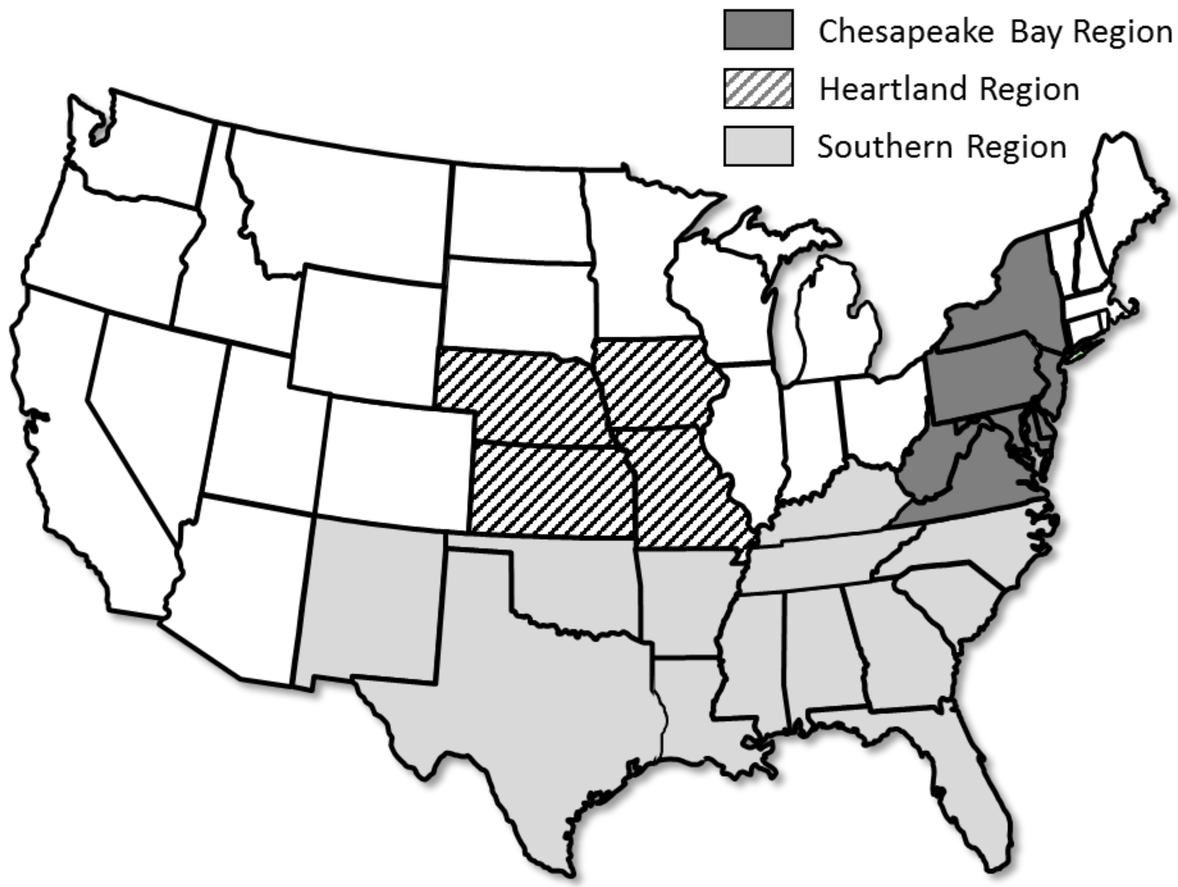

Figure 1. The National P Index project comprised assessments by the Chesapeake Bay, Heartland, and Southern regions of the U.S.

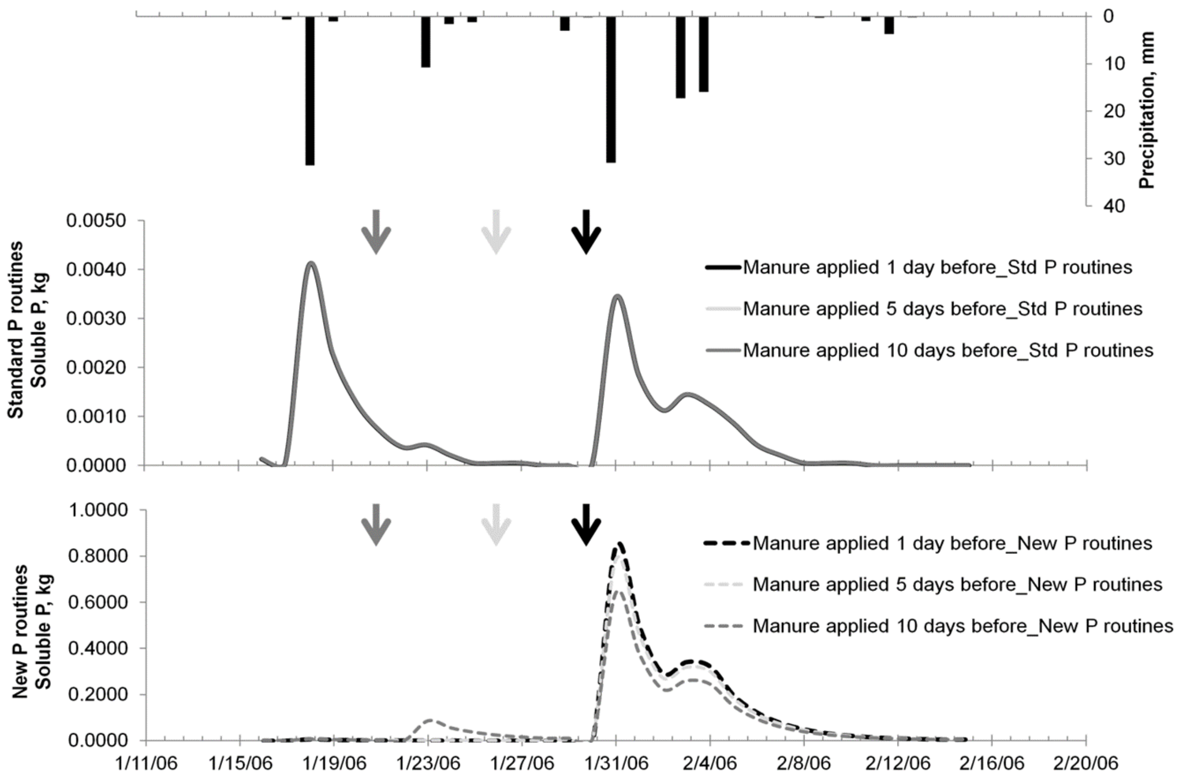

Figure 2. Simulation of P loss from dairy manure amended soils using conventional and new model routines for P, adapted with permission from Collick et al. (2016). The new routines were developed as part of U.S. efforts to improve model representation of the effects of manure application timing, rate, method and source on P fate and transport to enable better comparison with P Index.

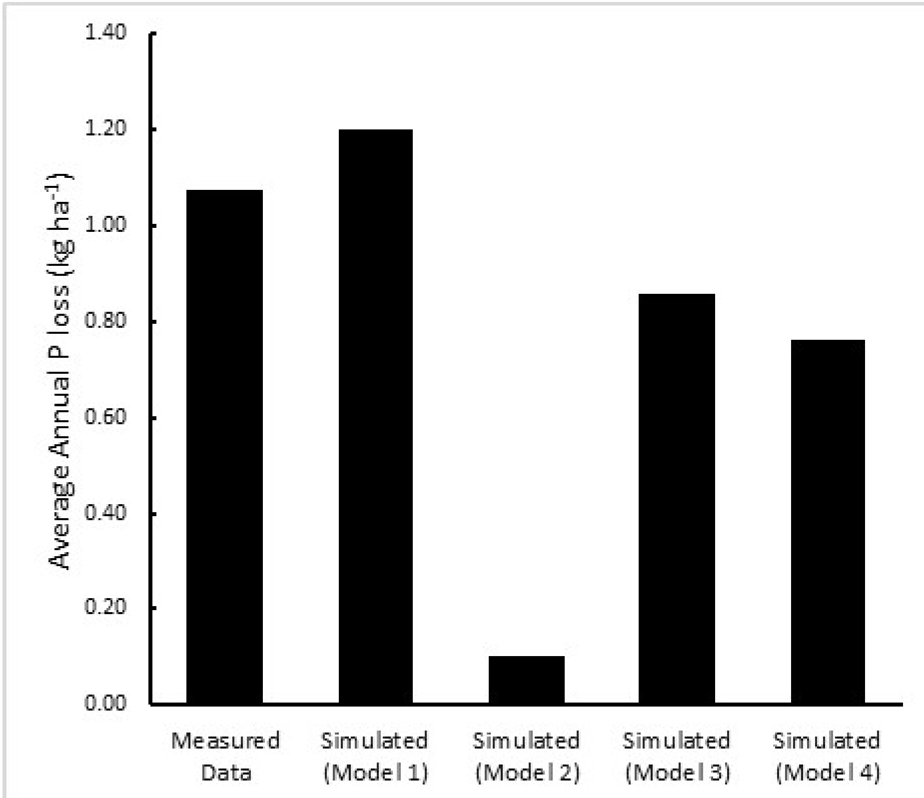

Figure 3. Average annual total P loss from a no-till sorghum-soybean rotation with surface-applied P fertilizer compared to APEX model simulation results from models calibrated using datasets with management similar or different from the measured data. Model 1: APEX calibrated with data from the same management system as measured data. Model 2: APEX calibrated with data from conventional-till sorghum soybean rotation with incorporated P fertilizer. Model 3: APEX calibrated with data from no-till sorghum-soybean rotation with subsurface-applied P fertilizer. Model 4: APEX calibrated with all data used for Models 1, 2, and 3 (Bhandari et al., 2017).

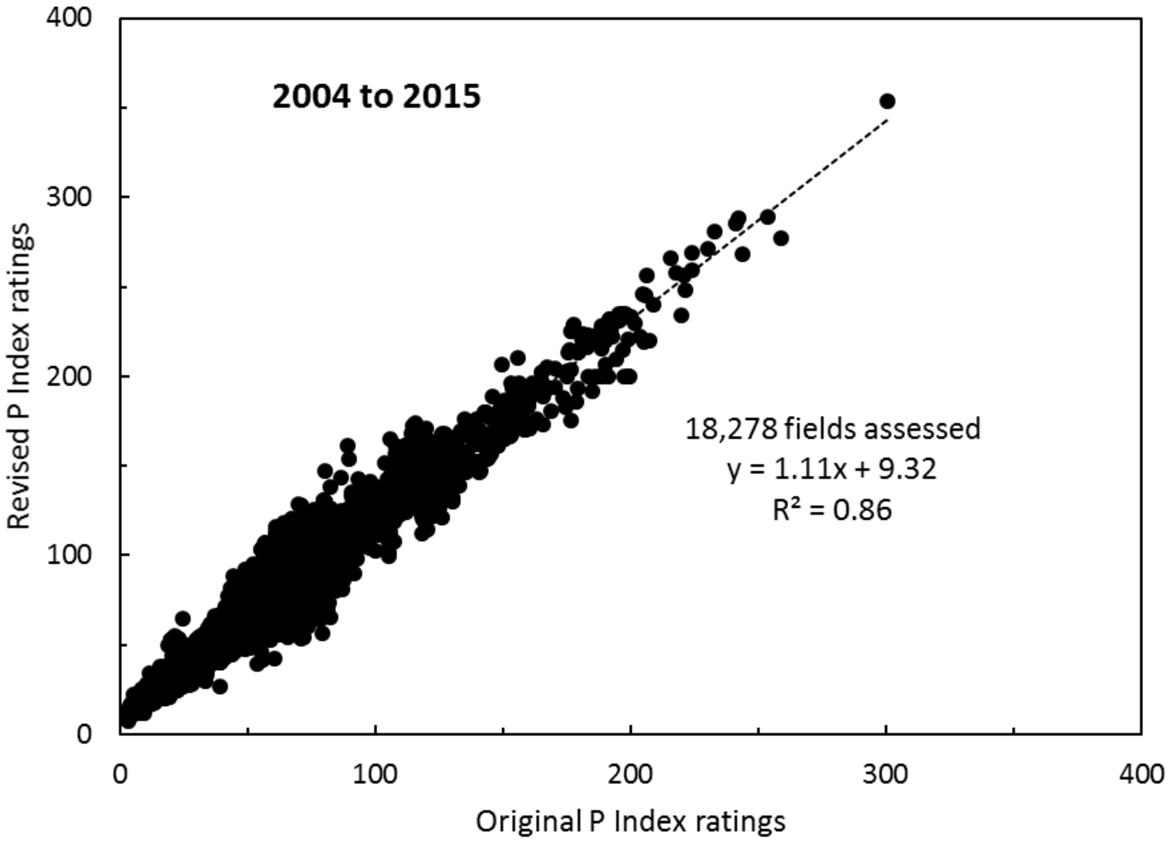

Figure 4. Relationship between the risk of P loss assigned by original and revised Indices for Arkansas, based on nutrient management planning information for 2004 to 2015.