

Multisite Evaluation of APEX for Water Quality:

I. Best Professional Judgment Parameterization

Claire Baffaut,* Nathan O. Nelson, John A. Lory, G.M.M.M. Anomaa Senaviratne, Ammar B. Bhandari, Ranjith P. Udawatta, Daniel W. Sweeney, Matt J. Helmers, Mike W. Van Liew, Antonio P. Mallarino, and Charles S. Wortmann

Abstract

The Agricultural Policy Environmental eXtender (APEX) model is capable of estimating edge-of-field water, nutrient, and sediment transport and is used to assess the environmental impacts of management practices. The current practice is to fully calibrate the model for each site simulation, a task that requires resources and data not always available. The objective of this study was to compare model performance for flow, sediment, and phosphorus transport under two parameterization schemes: a best professional judgment (BPJ) parameterization based on readily available data and a fully calibrated parameterization based on site-specific soil, weather, event flow, and water quality data. The analysis was conducted using 12 datasets at four locations representing poorly drained soils and row-crop production under different tillage systems. Model performance was based on the Nash–Sutcliffe efficiency (NSE), the coefficient of determination (r^2) and the regression slope between simulated and measured annualized loads across all site years. Although the BPJ model performance for flow was acceptable (NSE = 0.7) at the annual time step, calibration improved it (NSE = 0.9). Acceptable simulation of sediment and total phosphorus transport (NSE = 0.5 and 0.9, respectively) was obtained only after full calibration at each site. Given the unacceptable performance of the BPJ approach, uncalibrated use of APEX for planning or management purposes may be misleading. Model calibration with water quality data prior to using APEX for simulating sediment and total phosphorus loss is essential.

Core Ideas

- Uncalibrated, APEX produced unacceptable site-specific sediment and TP estimates.
- Acceptable runoff estimates do not translate to acceptable water quality estimates.
- Distributions of successfully calibrated performance indicator values were normal.
- Calibration of the APEX model with water quality data remains an essential step.

EXCESSIVE PHOSPHORUS (P) and sediment loss from agricultural fields continues to degrade fresh water quality despite decades of efforts to understand loss processes and implement management practices to reduce nonpoint pollution (e.g., Sharpley et al., 1994, 2015; Sims and Kleinman, 2005; Jarvie et al., 2013). Failure of conservation practices to produce expected improvements in water quality has renewed appreciation for the complexity of P movement within landscapes and to waterbodies (Jarvie et al., 2013; Sharpley et al., 2013). Process-based watershed- and field-scale models offer the prospect of integrating knowledge to assess and quantify impacts of conservation and management practices.

The P Index was developed in the mid-1990s to assess risk of P loss from agricultural land (Lemunyon and Gilbert, 1993) and is now an integral part of the NRCS 590 Nutrient Management Standard and other state and federal programs (Sharpley et al., 2003, 2017). However, the diversity in P Index ratings and P management recommendations for similar conditions led Sharpley et al. (2012) and Osmond et al. (2006) to emphasize the need for science-based assessment and improvement of existing P Indices.

Extensive testing of P Indices requires water quality data from field-scale watersheds from a broad range of soils and management scenarios with a sufficient number of monitoring years to estimate long-term average annual losses. Such extensive datasets are rare. Hence, computer models must be a core component of any strategy to assess and improve P Indices (Bolster et al., 2012; Sharpley et al., 2012) by extending limited datasets to simulated edge-of-field losses over a wide range of conditions and to long-term average estimates. Phosphorus Index usefulness is assessed by ensuring

C. Baffaut, USDA–ARS Cropping Systems and Water Quality Research Unit, 241 Agricultural Engineering Building, Univ. of Missouri, Columbia, MO 65211; N.O. Nelson, Dep. of Agronomy, 2004 Throckmorton Plant Sciences Center, Kansas State Univ., Manhattan, KS 66506; J.A. Lory, Division of Plant Sciences, 108 Waters Hall, Univ. of Missouri, MO 65211; G.M.M.M. Senaviratne, The Center for Agroforestry, Anheuser-Busch Natural Resource Building, Univ. of Missouri, Columbia, MO 65211; A.B. Bhandari, Dep. of Agriculture, 204 B Albertson Hall, Fort Hays State Univ., Hays, KS 67601; R.P. Udawatta, Dep. of Soil, Environmental and Atmospheric Sciences, 302 Anheuser–Busch Natural Resource Building, Univ. of Missouri, Columbia, MO 65211; D.W. Sweeney, Southeast Agricultural Research Center, Kansas State Univ., Parsons, KS 67357; M.J. Helmers, Agricultural and Biosystems Engineering, 4354 Elings Hall, Iowa State Univ., Ames, IA 50011; M.W. Van Liew, former address, Biological Systems Engineering Department, Univ. of Nebraska, Lincoln, NE 68583; A.P. Mallarino, Dep. of Agronomy, Iowa State Univ., Ames, IA 50011; C.S. Wortmann, Dep. of Agronomy and Horticulture, 369 Keim Hall, Univ. of Nebraska, Lincoln, NE 68583. Assigned to Associate Editor Andrew Sharpley.

Abbreviations: APEX, Agricultural Policy Environmental eXtender; BPJ, best professional judgment; CEAP, Conservation Effects Assessment Project; DP, dissolved phosphorus; MUSLE, Modified Universal Soil Loss Equation; NSE, Nash–Sutcliffe efficiency; PBIAS, percent bias; PEC, performance evaluation criteria; PSP, phosphorus sorption coefficient; TP, total phosphorus.

Copyright © American Society of Agronomy, Crop Science Society of America, and Soil Science Society of America. 5585 Guilford Rd., Madison, WI 53711 USA. All rights reserved.

J. Environ. Qual. 46:1323–1331 (2017)

doi:10.2134/jeq2016.06.0226

This is an open access article distributed under the terms of the CC BY-NC-ND license (<http://creativecommons.org/licenses/by-nc-nd/4.0/>)

Supplemental material is available online for this article.

Received 27 June 2016.

Accepted 22 Nov. 2016.

*Corresponding author (Claire.baffaut@ars.usda.gov).

that P Index ratings and ensuing P management recommendations are directionally consistent with model results. Alternatively, models can be used to assess specific components of a P Index, e.g., the runoff volume, which serves to assess the accuracy of the component and the weights assigned to each component (Bolster et al., 2012). These strategies, however, require that the model used is appropriately parameterized over the range of conditions relevant for P Index validation, i.e., that one has confidence in the model results. The benefits of models can only be achieved if a model parameterization can be successfully used for management scenarios, weather scenarios and/or physiographic areas beyond the calibration space of the model. The greatest benefits of models could be achieved if it was possible to successfully parameterize a model without calibration using water quality data.

The Agricultural Policy Environmental eXtender (APEX; Gassman et al., 2010) model is a daily time-step model that simulates plant growth, water movement, and fate and movement of sediment, nutrients, and pesticides in a catchment. It is especially useful for field-size catchments (30–300 ha) because fields can be described as linked subareas that are simulated separately, thus providing options to simulate features such as filter strips, buffers, and grass waterways. Surface and subsurface flow, and associated nutrient and pesticide fluxes, are routed from each subarea to the outlet of the catchment through a network of channels. Model inputs include the physical characterization of each subarea and of the channels (e.g., slope, Manning coefficients), management information, physical and chemical soil characterization, and daily weather data. Inputs also include control parameters that select methods used to simulate specific processes (e.g., equation used to simulate soil erosion by water), as well as global parameters that define thresholds and rate coefficients for selected processes. A complete description of the inputs is provided in the APEX user's manual (Steglich and Williams, 2013). Choice of simulation methods and global parameter values can significantly affect results of the model. However, there is little guidance on how to set these parameters for any given situation, and they are usually defined through the process of model calibration and validation.

The APEX model has successfully simulated edge-of-field runoff, sediment and P loss for specific management scenarios after calibration of the model using measured data from that location (Wang et al., 2008; Yin et al., 2009; Kumar et al., 2011; Mudgal et al., 2012; Plotkin et al., 2013; Senaviratne et al., 2013). There are published studies using APEX with limited or no calibration with water quality data to predict water quality outcomes (Harman et al., 2004; Williams et al., 2006; Tuppad et al., 2010). The most extensive use of APEX to date is the Conservation Effects Assessment Project (CEAP; Duriancik et al., 2008). The CEAP project used a detailed farmer survey conducted by the National Agricultural Statistical Service (NASS) and data from the USDA National Resource Inventory to develop regional parameterizations of APEX. A sensitivity analysis identified sensitive parameters at the field scale (Wang et al., 2006). Integration of APEX and the Soil and Water Assessment Tool (SWAT) to simulate the Upper Mississippi River Basin (Wang et al., 2011) provided some basis for the calibration of surface water and sediment transport at the eight-digit watershed level. However, sediment and P-loss estimates generated from these APEX parameterizations were not compared back with measured field-scale data. Thus, there is a need for formal APEX performance evaluation for event or

annual runoff and water quality constituents when APEX is not calibrated with edge-of-field data.

The objectives of this paper were to develop a best professional judgment (BPJ) parameterization of APEX0806 for estimating edge-of-field runoff, sediment, and P losses in restricted-layer soils common in Missouri and Kansas and to assess the performance of this parameterization using monitoring data from multiple sites in this region.

Materials and Methods

Processing of Monitoring Data

Four sites including 12 datasets were selected across the Heartland region (Supplemental Material, Section S1; Supplemental Fig. S1) based on the availability of edge-of-field runoff monitoring data (Supplemental Table S1), as well as site-specific soil, management, and weather (Supplemental Table S2). For each dataset, subdaily and daily rainfall and runoff data were first aggregated to event values. For these small (<35 ha) field-scale watersheds, there was no baseflow; events were defined as a day or consecutive days with runoff bracketed by days without runoff. Hourly rainfall records were examined to review rainfall on the day prior to a runoff event to determine whether that rain contributed directly to the runoff event. Rainfall and runoff data were further reviewed to exclude events where equipment malfunction was suspected, including instances of runoff greater than rainfall or runoff that was significantly less than or greater than expected, given other data from that location. Annualized sums were calculated by aggregating verified events by year. Event mean, standard deviation, and ranges of annualized sums are tabulated in Supplemental Table S3 for each dataset.

Model Used and Generic Implementation

We acquired the executable and associated code for APEX (version APEX0806; Wang et al., 2012) in October 2012. Subsequently, we modified it to correct identified errors or to increase modeling capability to fit our needs. These changes were made by, or in consultation with, the APEX developers, but we incorporated them in our version so that we would not be affected by other changes in the version maintained by the APEX developers. Most notable were changes in the simulation of contour filter strips and the change in P leaching from the surface (1-cm) layer of the soil profile. These changes are described in Supplemental Material, Section S2.

Best Professional Judgment and Site-Specific Parameterizations

The BPJ parameterization was defined based on our understanding of the hydrological and biological processes that take place in a field, our understanding of APEX, and the parameterization outlined for CEAP, as described by Wang et al. (2011). The BPJ parameterization is intended to be implemented with no field measurement, relying instead on regional and national data for soils, weather, and topography. It did assume that management records were available to define crops and associated field operations.

Delineation Process

For any site, the modeling process starts with subarea delineation based on the knowledge of the site. In APEX, each subarea has homogeneous soil, slope, and management properties. Each

site was delineated as simply as possible while maintaining the spatial variability of soils and management features. As a result, the larger sites had distinct subareas for each soil type, while the smaller sites had only one subarea. Features such as filters, buffers, terraces, and grass waterways were represented as individual subareas to represent different land covers in these areas. Main channels and drainage ways were also considered for the definition of the drainage network and subareas. Delineation was conducted either manually or with the aid of ArcGIS tools (ESRI, Redlands, CA) to process elevation and soil maps, and input files were built with the WinAPEX interface. Characteristics of channel lengths, routing reach lengths, channel top and bottom widths, slopes, slope lengths, Manning coefficients, flood plain width for buffer areas, and fraction of buffer that is bypassed were entered based on observations (channels, channel conditions, buffer conditions) made during site visits. Detailed maps of the subarea delineations for the complex watersheds can be found in Senaviratne et al. (2013) and in Supplemental Material, Section S1.

Management and Soil Properties

Availability of detailed management information was a selection requirement of the sites. Thus, information on crops grown and details on field operations such as tillage, fertilization, and harvest were known for each site and manually entered. Supplemental Table S2 provides information on the sources of data at each site. For the BPJ parameterization, we identified soil map units and used the corresponding soil properties from the SSURGO database (Soil Survey Staff, 2014). Since none of these sites were artificially drained and most soils were poorly drained, most soils were classified in the hydrologic group D. Using the representative value from the SSURGO database, the following soil properties were specified for each soil layer: soil albedo (first layer), soil layer depth, bulk density, sand and silt content, pH, organic carbon content, cation exchange capacity, and saturated conductivity. The number of years of cultivation, which affects mineralization of organic nitrogen and thus crop yields, was set to 100, a value in line with history of agriculture in the region. The P sorption coefficient (PSP), a critical input that affects the balance between active and labile P pools and thereby influences the initialization of total P (TP), P availability for loss in runoff, and general P cycling, fate, and transport was set to 0.5, which is the default value for noncalcareous soils when soil weathering is unknown. All other values were set to zero, which caused APEX to either use a default value or calculate the parameter values using algorithms internal to the model.

Control and Model Parameters

Control parameters define the type of output desired and equations used for specific processes when several are available. Model global parameters include S-curve parameters, process threshold values, and equation coefficients. Control and model parameters BPJ typically were set to what was used for the CEAP cropland study (Wang et al., 2011). Departures from this parameterization are shown in supplemental information: Supplemental Table S4 for the control parameters, Supplemental Table S5 for S-curve parameters, and Supplemental Table S6 for global parameters.

Location-Specific Calibration and Validation

Detailed descriptions of the APEX parameterization, calibration, and validation procedure at each of these sites are

given in Bhandari et al. (2017) for the Franklin sites, Bhandari (2016) for the Crawford sites, and in previously published work (Senaviratne et al. 2013, 2014). Summaries of the soil parameterization approach using site-specific data (soil texture, TP levels, and PSP) and of the calibration and validation strategy are provided in Supplemental Material, Section S4.

Evaluation of BPJ versus Fully Calibrated Parameterizations

Our first objective was to compare model performance for the BPJ with the fully calibrated parameterizations based on runoff, sediment, and TP event data for the 12 monitoring datasets. Model performance for event data was based on percent bias (PBIAS), coefficient of determination (r^2), Nash–Sutcliffe Efficiency (NSE; Nash and Sutcliffe, 1970), and in some cases, regression slope (Moriassi et al., 2007, 2015). First, performance indicator values for both the BPJ and the fully calibrated parameterizations were tabulated for each site and described in terms of their range (minimum, median, and maximum values) and the number of sites meeting the performance evaluation criteria (PEC). Acceptable PEC for event data at a site were defined as $r^2 \geq 0.5$; $\text{NSE} \geq 0.3$ for runoff, sediment, and TP; and $|\text{PBIAS}| \leq 35\%$ for runoff, $\leq 60\%$ for sediment, and $\leq 70\%$ for TP.

In addition, we evaluated the normality of the distribution of performance indicator values obtained with the BPJ and fully calibrated parameter sets using normal probability plots, a visual technique to explore the relationship between ranked observations and the expected values of a standard normal distribution. In normal probability plots, the y -axis values are the ranked observations and the x -axis values are the normal standardized value that corresponds to the probability of occurrence of the ranked observation. If the observations follow a normal distribution, they approximately follow a straight line with a slope equal to their standard deviation and an intercept equal to the mean. Normality of the distribution of PEC values implies that the observed values are from the same population where the average is an appropriate estimate of the population mean and that values of performance indicators can be related to the probability that a model is incorrectly rejected. The Shapiro–Wilk test (Shapiro and Wilk, 1965; Shapiro et al., 1968) for normality ($\alpha = 0.05$) was applied for each set of 12 performance indicators values for the BPJ and calibrated models to confirm normality.

Finally, all simulated and measured data from the 12 datasets were considered together for both the BPJ and the fully calibrated parameter datasets. The comparisons were based on all events ($n = 437$ for runoff, $n = 384$ for sediment, and $n = 358$ for TP) and annualized estimates ($n = 61$), with performance of each parameterization strategy assessed using the previously described PEC values. This collective assessment of event data tested if the parameterization was successful regionally, across all events, despite potential poor performance at specific locations. The collective assessment of annualized site–year data across all datasets assesses if model performance is likely to be acceptable when using the model to generate annual data typically used for P Index assessment.

Results

Table 1 summarizes performance indicator values from the 12 sets of monitoring data used to evaluate the BPJ and fully calibrated parameterizations of APEX for runoff, sediment, and TP. Performance of the BPJ parameterization was

acceptable for runoff for two-thirds of the sites and is not acceptable for sediment and TP at most sites.

With calibration, all indicators showed improvement in model performance (Table 1). Calibration resulted in a reduced range of values for each indicator, as indicated by a smaller standard deviation, and increased the number of sites where PEC were met. Sites for which sediment r^2 and NSE did not meet the PEC were characterized by small sediment loads (average annualized load < 0.8 Mg ha⁻¹; Bhandari et al., 2017; Nelson et al., 2017).

Normal probability plots of the 12 estimates of PBIAS, r^2 , and NSE for the BPJ models imply that these performance evaluation coefficients were not normally distributed (Fig. 1) and had skewed distributions, as indicated by the poor fit to a linear trend line. Small deviations from the linear trend line are indicative of small, random, and independent errors such as small measurement errors, small misrepresentations of the factors and processes that control the variable, proximity of the weather station, expertise of the modeler, etc. Large deviations can indicate misrepresentation of the processes simulated by the model.

Table 1. Characteristics of performance indicators for event runoff, sediment, and total phosphorus simulation at the 12 sites.

	PBIAS		r^2		NSE§	
	BPJ‡	Calibrated	BPJ	Calibrated	BPJ	Calibrated
	%					
Runoff						
Mean	30	18	0.64	0.81	0.29	0.73
Median	24	14	0.69	0.80	0.48	0.72
Min.	2	3	0.24	0.69	-1.62	0.63
Max.	71	36	0.87	0.91	0.81	0.87
Standard deviation	19	11	0.21	0.06	0.67	0.08
No. sites within PEC¶	8	11	9	12	9	12
Sediment						
Mean	1256	36	0.30	0.43	-1494	0.24
Median	757	35	0.26	0.35	-58	0.31
Min.	48	9	0.10	0.25	-17031	-0.26
Max.	8205	85	0.81	0.80	0.68	0.51
Standard deviation	2248	22	0.19	0.18	4894	0.23
No. sites within PEC	1	10	1	3	1	7
Total phosphorus						
Mean	112	22	0.59	0.80	-3.31	0.68
Median	62	15	0.60	0.79	-0.13	0.66
Min.	16	1	0.22	0.61	-28.58	0.48
Max.	552	59	0.97	0.99	0.82	0.96
Standard deviation	156	20	0.23	0.13	8.58	0.15
No. sites within PEC	6	12	8	12	3	12

† PBIAS, percent bias.

‡ BPJ, best professional judgment parameterization.

§ NSE, Nash–Sutcliffe efficiency.

¶ PEC, performance evaluation criteria

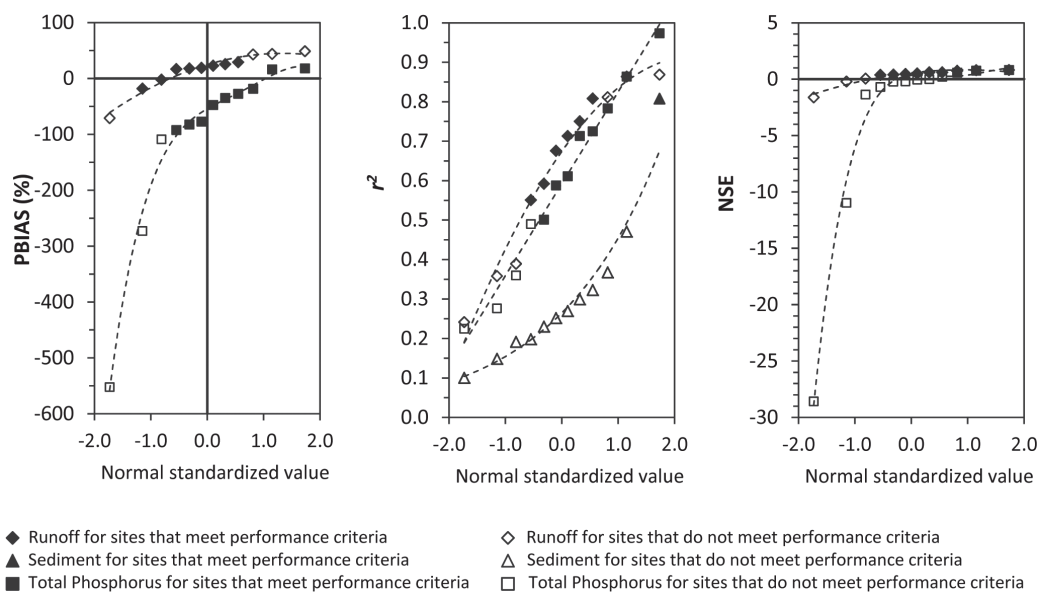


Fig. 1. Normal probability plots of percent bias (PBIAS), coefficient of determination (r^2), and Nash–Sutcliffe efficiency (NSE) for the best professional judgment model for runoff, sediment, and total phosphorus for 12 watersheds. Sediment PBIAS and NSE are not shown because their range prevents readability of the plots.

Deviations were particularly marked for NSE and PBIAS; distributions for r^2 were closer to a normal distribution, especially for TP, as indicated by the good fit to the linear trend line. On the other hand, all values obtained with the calibrated models were approximately normally distributed (Fig. 2). The Shapiro–Wilks statistics confirmed that, among the BPJ models, NSE and PBIAS distributions did not follow a normal distribution. Only r^2 values for BPJ runoff and TP met the normality assumption. In contrast, for the calibrated models, all PEC except r^2 values for sediment met the normality assumption.

Regional event-based evaluation confirmed these results (Table 2), as shown by $|PBIAS| > 100\%$ and negative NSE for sediment and TP under BPJ. Negative NSE values indicate that the mean of measured loads is a better predictor than results simulated with BPJ. For event data at the regional scale, all PEC for the calibrated models were met (Table 2), except that r^2 was marginally >0.5 and the regression slope for event sediment loss was >0.5 due to underestimation of high sediment loads. Results based on annualized values confirmed the event-based results (Fig. 3): BPJ runoff estimates were acceptable but not sediment or TP. Annualized values obtained

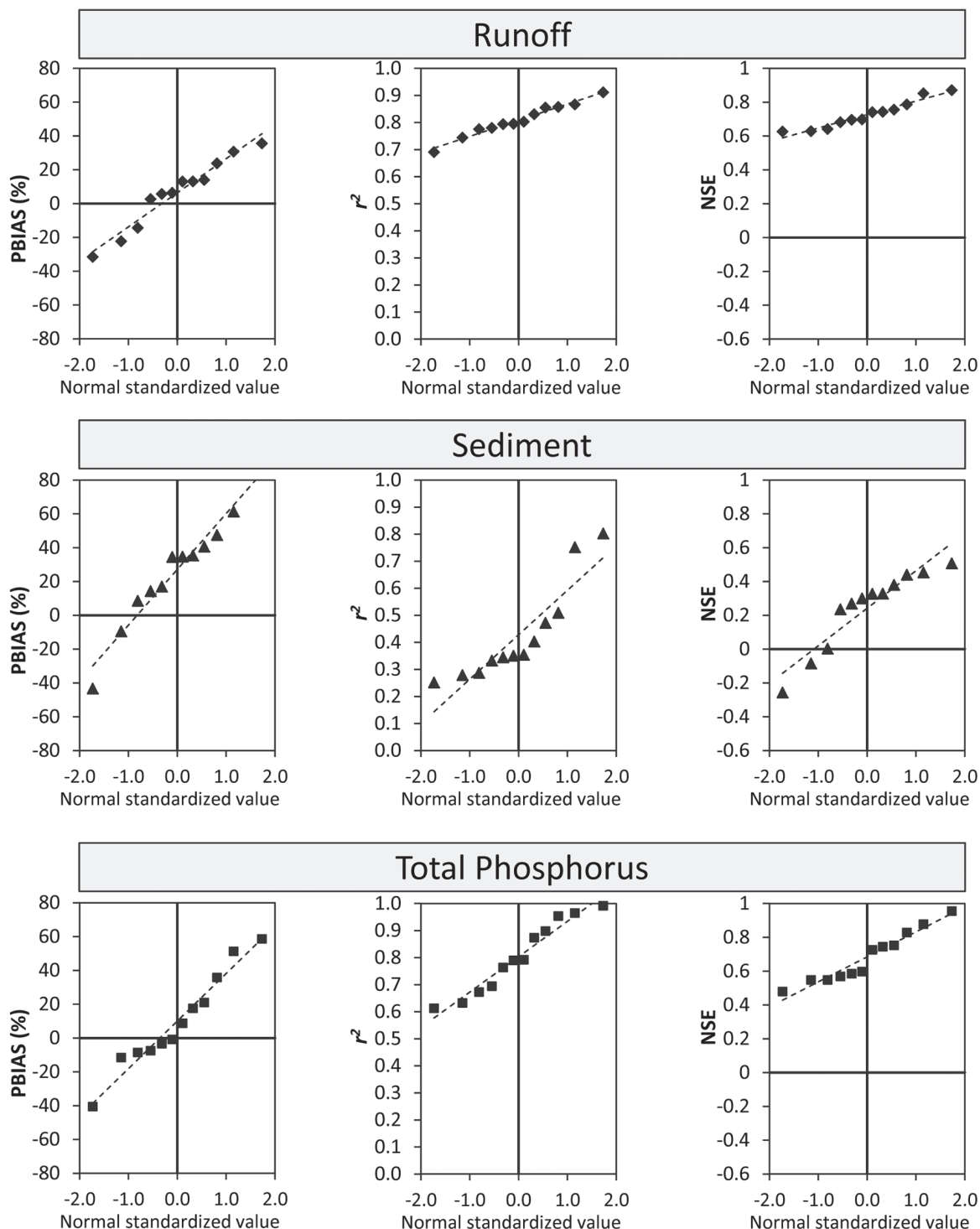


Fig. 2. Normal probability plots of percent bias (PBIAS), coefficient of determination (r^2), and Nash–Sutcliffe Efficiency (NSE) obtained with the fully calibrated models for runoff, sediment, and total phosphorus for 12 watersheds.

Table 2. Regional performance indicators at the event temporal scale for runoff, sediment, and total phosphorus (TP).

	BPJ†			Calibrated model		
	Runoff	Sediment	TP	Runoff	Sediment	TP
PBIAS‡ (%)	28	−653	−103	14	24	12
r^2	0.59	0.05	0.77	0.74	0.47	0.92
NSE§	0.53	−44	−0.14	0.72	0.43	0.92
Regression slope	0.58	1.42	1.57	0.69	0.35	0.89

† BPJ, best professional judgment parameterization.

‡ PBIAS, percent bias.

§ NSE, Nash–Sutcliffe efficiency.

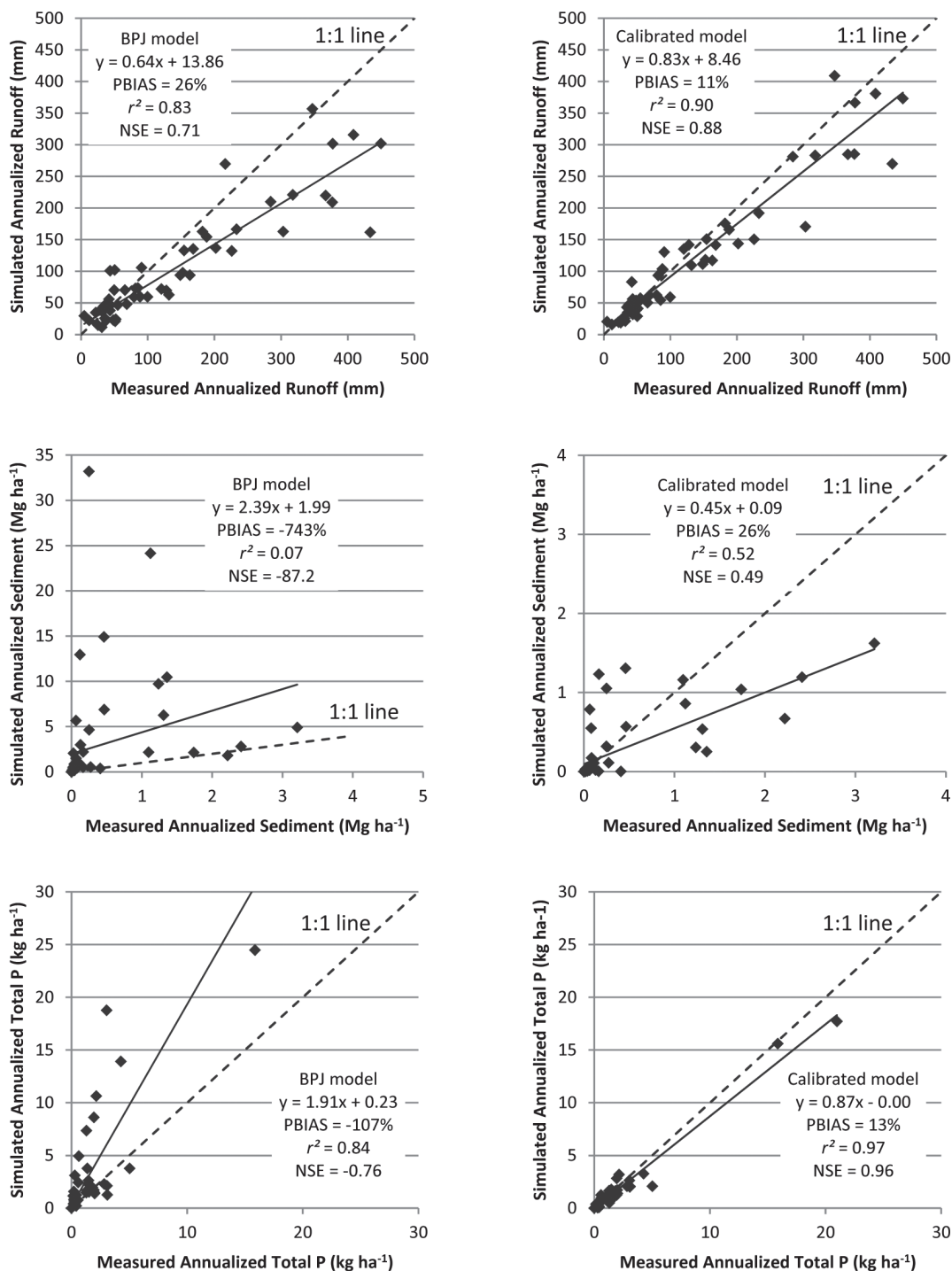


Fig. 3. Simulated versus measured annualized runoff, sediment, and total phosphorus (P) for 61 site-years across 12 watersheds for the best professional judgment (BPJ) and fully calibrated models, along with values of percent bias (PBIAS), coefficient of determination (r^2), and Nash–Sutcliffe efficiency (NSE).

with the calibrated models all passed PEC; performance was very good for runoff and TP and less so for sediment, as indicated by the smaller values of r^2 , NSE and regression slope.

Discussion

Using APEX with BPJ Parameterization for Event Data at a Specific Site

Individual site results indicated that using APEX with BPJ parameterization may be appropriate for the purpose of estimating surface runoff volume from cropland underlain by claypan soils but can lead to erroneous results for sediment and TP components. In our dataset, runoff performance indicators using the BPJ parameterization strategy were acceptable for two-thirds of the sites (Table 1). Obviously, it is difficult to set an acceptable fail/pass ratio when using a model without calibration and two-thirds of sites that meet the PEC may be sufficient in some cases; objectives of the specific study would dictate what an acceptable number should be. Nonetheless, in cases where water quality simulation is an objective, correct simulation of runoff is critical (e.g., Arnold et al., 2015).

In this study, in spite of correct runoff simulation for two-thirds of the sites, APEX did not simulate correct sediment loads in all but one case with the BPJ parameterization. At 7 of 12 sites, sediment loss was overestimated by >500%, in large part because of the choice of erosion equation. A comparison of BPJ and calibrated sediment-related parameters revealed significant differences that explain sediment overestimation with BPJ (Supplemental Tables S7 and S8). The Modified Universal Soil Loss Equation (MUSLE) was selected for BPJ because that option was used in the CEAP study (Wang et al., 2011). None of our datasets could be calibrated with the MUSLE, instead requiring the MUSS equation, which is an adaptation of MUSLE to small watersheds with no channel erosion (Williams et al., 2012). While MUSS and MUSLE both calculate sediment loss as a function of runoff volume, runoff peak rate, and drainage area, MUSLE typically estimates greater sediment loss. In addition, sediment routing coefficients were greater with BPJ, which also increases sediment loss. The choice of MUSS for erosion calculation seems appropriate for these watersheds, given that they are small (<35 ha), many of them are in a no-till system, and some have grassed waterways. However, there are no guidelines in the APEX manual or documentation to guide the choice of erosion equation.

Sediment results after calibration were still not satisfactory for some of the Knox, Franklin, and Crawford datasets. Possible factors include the large measurement errors for edge-of-field sediment losses (Harmel et al., 2006), sediment deposition caused by backwater induced by the flume or the weir, and the use of a USLE-type equation, which does not consider the true rainfall distribution during an event and does not account for gully erosion. Given the amount of work required to complete these calibrations, results were disappointing and point out the need for improvement of the algorithms used for estimating edge-of-field sediment loss with APEX.

The BPJ results for TP were also not acceptable: the BPJ NSE values were acceptable for only 25% of sites in spite of acceptable r^2 values for two-thirds of the sites. Percent bias was acceptable for half of the sites, and for sites with unacceptable PBIAS, the BPJ overestimated TP loss by an average of 200%. Total P

estimates are the sum of dissolved P (DP) and sediment-bound P estimates. Because simulated sediment losses were grossly overestimated, the sediment-bound fractions of TP were overestimated as well. The accuracy of model estimates for particulate P and DP can be evaluated by comparing measured and simulated DP/TP ratios, which was only possible at the Franklin and Crawford sites. At the Franklin sites, where inorganic fertilizer was used, the mean measured ratio of DP/TP was 45%, compared with 16% (or 30% points less) with the BPJ and 63% with the calibrated models. At the Crawford site fertilized with turkey litter, the mean measured ratio was 88%, compared with 36% with the BPJ and 99% with the calibrated models.

While inaccurate initialization of TP concentrations in the soil could explain the overestimation of sediment-bound P, this was not the case here. The APEX model initializes total inorganic P pools as a function of the initial labile P and the PSP, where increasing the PSP decreases the initial amount of total inorganic P in the soil. Calculated values of PSP used for the calibrated models were all <0.5, which was the default PSP for the BPJ model. Therefore, total soil P values in the BPJ models were less than both measured data and that simulated by the calibrated models. Thus, overestimation of sediment loss by the BPJ model, not inaccurate initial soil P concentrations, caused the overestimation of sediment-bound P and thus TP. Although the calibrated models overestimated this DP/TP ratio, likely by underestimating the sediment transport, as indicated by the positive sediment PBIAS for the calibrated models (Table 2), the ratios were closer to measured data.

Overall, when APEX was used with the BPJ parameterization, sediment and TP losses from most locations were overestimated by large amounts, indicating that we cannot use BPJ to generate data and assess the P Index at specific locations. Furthermore, the fact that the BPJ model provided acceptable estimates of runoff yet unacceptable results for sediment and TP loss illustrates that acceptable estimates of runoff do not necessarily translate into acceptable estimates of sediment or TP loss.

Distribution of Performance Indicators Values

Performance indicators measure the goodness of fit between the simulated values and the data. Indicator values that meet the PEC reflect the adequacy of the model, including the algorithms and equations embedded in the code and the parameterization used for each site, with the data collected at that site. The fact that distributions of performance indicators (r^2 , NSE, and PBIAS) for calibrated models approximately fit a normal distribution indicated that variability in performance was caused by small errors in the model, the model parameters, or the data used. One exception to this was the distribution of sediment r^2 obtained with the calibrated models, as the assumption of normality had to be rejected ($\alpha = 0.05$). In contrast, most distributions of the same performance indicators obtained with the BPJ parameterization did not follow a normal distribution, indicating a fundamental flaw between simulated values and measured values. Given that it was possible to calibrate the model at each of the sites, we conclude that the model and the data were essentially good but that the BPJ parameterization was not appropriate.

In their development of PEC, Moriasi et al. (2015) found that r^2 values in their database of published studies were approximately normally distributed. Nash–Sutcliffe efficiency and

PBIAS, which they also considered, were not. Similarly to these authors, we found that r^2 values for runoff and TP were approximately normally distributed with the fully calibrated parameterization among these 12 watersheds. We could not confirm this for sediment, perhaps indicative that several models (5 out of 12) were not well calibrated for sediment. In contrast with the results from Moriasi et al. (2015), both NSE and PBIAS values were normally distributed. Thus, in this study, when the model was performing acceptably, performance indicators values followed a normal distribution. If this can be confirmed with a larger dataset of calibrated models, PEC values identified as outliers relative to the distributions of performance measures may indicate that the model is not simulating processes accurately or that the data used for evaluation are not representative.

Using APEX with BPJ Parameterization for Regional Analyses

Results obtained with the results from the 12 sites together, either at the event temporal scale or by annualizing the results, showed that APEX could be used with the BPJ parameterization strategy for the purpose of evaluating runoff from cropland underlain by claypan soils (Table 2, Fig. 3). Thus, it would be appropriate to use APEX with the BPJ parameterization to evaluate the variability of runoff within this region, as affected by topography or management, for example. Variability of soils was less well represented within this dataset, since all soils were considered claypan soils and were classified in the hydrological group D. In the context of the P Index, APEX using the BPJ parameterization could be used to evaluate the runoff volume component of the index—e.g., by generating runoff values across a large range of slopes, weather scenarios, and management scenarios on claypan soils and other soils with a restrictive layer.

Sediment loss estimates obtained with the BPJ parameter set were still not accurate when considered collectively across the 12 sites (Table 2, Fig. 3). Minimal sediment losses from the Crawford and Franklin sites (four no-till datasets) were so overestimated that estimates were greater than soil loss estimates from any other site, thus making inappropriate the use of this model for regional sediment loss analyses. Consequently, TP losses were not accurate as well because of the overestimation of sediment-bound P. As a result, APEX and the BPJ parameterization could not be used to test whether a P Index, or any other tool, provides directionally correct information because the BPJ parameterization did not provide that information itself when tested across these 12 sites. These results do not necessarily mean that APEX never can be used for locations that have no calibration data. A regional parameterization, based on the calibrated parameter sets obtained for each dataset, was investigated by Nelson et al. (2017) and showed encouraging results for runoff and TP on restricted-layer soils.

Conclusion

Performance of the APEX model was assessed for 12 datasets from four sites in Missouri and Kansas using two parameterization strategies: a BPJ parameterization common to all sites and a site-specific parameterization derived through systematic calibration and validation of the model. Each parameterization was assessed for model performance using runoff, sediment, and P-loss monitoring data at each site. The BPJ parameterization resulted

in a large range of performance indicators values, some indicating acceptable performance but the majority showing unacceptable performance. The model performance was greatly improved through calibration and the range of indicator values was reduced.

Model performance was also assessed with an annualized time step across the 12 sites. While BPJ performance for runoff simulation was acceptable, performance for sediment and TP was not. Thus, APEX with a BPJ parameterization could not be used as the basis for sediment or TP evaluation or for the evaluation of P Indices. It potentially could be used to evaluate the runoff component of these indices, provided it is successfully tested on soils with different characteristics. Other papers in this special section include analyses of the calibrated parameterizations within and beyond the calibration conditions (Bhandari et al., 2017) and development of a regional parameterization that could provide acceptable performance (Nelson et al., 2017).

The results of this analysis emphasize the essential role of calibration of APEX with water quality data to ensure accurate model estimates of TP in edge-of-field runoff from small agricultural watersheds.

Acknowledgments

The authors are grateful for partial funding for this project through a USDA Conservation Innovation Grant (Agreement Number 69-3A75-12-193).

References

- Arnold, J.G., M.A. Youssef, H. Yen, M.J. White, A.Y. Sheshukov, A.M. Sadeghi et al. 2015. Hydrological processes and model representation: Impact of soft data on calibration. *Trans. ASABE* 58:1637–1660. doi:10.13031/trans.58.10726
- Bhandari, A.B. 2016. Improving phosphorus loss assessment with the APEX model and phosphorus index. Ph.D. diss., Kansas State Univ., Manhattan, KS.
- Bhandari, A.B., N.O. Nelson, D.S. Sweeney, C. Baffaut, J.A. Lory, A. Senaviratne et al. 2017. Calibration of the APEX model to simulate management practice effects on runoff, sediment, and phosphorus loss. *J. Environ. Qual.* doi:10.2134/jeq2016.07.0272 (in press).
- Bolster, C.H., P.A. Vadas, A.N. Sharpley, and J.A. Lory. 2012. Using a P loss model to evaluate and improve P Indexes. *J. Environ. Qual.* 41:1758–1766. doi:10.2134/jeq2011.0457
- Duricank, L.F., D. Bucks, J.P. Dobrowolski, T. Drewes, S.D. Eckles, L. Jolley et al. 2008. The first five years of the Conservation Effects Assessment Project. *J. Soil Water Conserv.* 63:185A–197A. doi:10.2489/jswc.63.6.185A
- Gassman, P.W., J.R. Williams, X. Wang, A. Saleh, E. Osci, L.M. Hauck et al. 2010. Invited review article: The Agricultural Policy/Environmental eXtender (APEX) model: An emerging tool for landscape and watershed environmental analyses. *Trans. ASABE* 53:711–740. doi:10.13031/2013.30078
- Harman, W.L., E. Wang, and J.R. Williams. 2004. Reducing atrazine losses: Water quality implications of alternative runoff control practices. *J. Environ. Qual.* 33:7–12. doi:10.2134/jeq2004.7000
- Harmel, R.D., R.J. Cooper, R.M. Slade, R.L. Haney, and J.G. Arnold. 2006. Cumulative uncertainty in measured streamflow and water quality data for small watersheds. *Trans. ASABE* 49:689–701. doi:10.13031/2013.20488
- Jarvie, H.P., A.N. Sharpley, P.J.A. Withers, J.T. Scott, B.E. Haggard, and C. Neal. 2013. Phosphorus mitigation to control river eutrophication: Murky waters, inconvenient truths, and “postnormal” science. *J. Environ. Qual.* 42:295–304. doi:10.2134/jeq2012.0085
- Kumar, S., R.P. Udawatta, S.H. Anderson, and A. Mudgal. 2011. APEX model simulation of runoff and sediment losses for grazed pasture watersheds with agroforestry buffers. *Agrofor. Syst.* 83:51–62. doi:10.1007/s10457-010-9350-7
- 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
- Moriassi, D.N., J.G. Arnold, M.W. Van Liew, R.L. Bingner, R.D. Harmel, and T.L. Veith. 2007. Model evaluation guidelines for systematic quantification of accuracy in watershed simulations. *Trans. ASABE* 50:885–900. doi:10.13031/2013.23153

- Moriasi, D.N., M.W. Gitau, N. Pai, and P. Daggupati. 2015. Hydrologic and water quality models: Performance measures and evaluation criteria. *Trans. ASABE* 58:1763–1785. doi:10.13031/trans.58.10715
- Mudgal, A., C. Baffaut, S.H. Anderson, E.J. Sadler, N.R. Kitchen, K.A. Sudduth et al. 2012. Using the Agricultural Policy/Environmental eXtender to develop and validate physically based indices for the delineation of critical management areas. *J. Soil Water Conserv.* 67:284–299. doi:10.2489/jswc.67.4.284
- Nash, J.E., and J.V. Sutcliffe. 1970. River flow forecasting through conceptual models part 1—A discussion of principles. *J. Hydrol.* 10:282–290. doi:10.1016/0022-1694(70)90255-6
- Nelson, N.O., C. Baffaut, J.A. Lory, G.M.M.M.A. Senaviratne, A.B. Bhandari, R.U. Udawatta et al. 2017. Multisite evaluation of APEX for water quality: II. Regional parameterization. *J. Environ. Qual.* doi:10.2134/jeq2016.07.0254
- 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.
- Plotkin, S., X. Wang, T.L. Potter, D.D. Bosch, J.R. Williams, E.S. Hesketh et al. 2013. APEX calibration and validation of water and herbicide transport under U.S. Southern Atlantic coastal plain conditions. *Trans. ASABE* 56:43–60. doi:10.13031/2013.42589
- Senaviratne, G.M.M.M.A., R.P. Udawatta, C. Baffaut, and S.H. Anderson. 2013. Agricultural Policy Environmental eXtender simulation of three adjacent row-crop watersheds in the claypan region. *J. Environ. Qual.* 42:726–736. doi:10.2134/jeq2012.0241
- Senaviratne, G.M.M.M.A., R.P. Udawatta, C. Baffaut, and S.H. Anderson. 2014. Evaluation of a stepwise, multiobjective, multivariable parameter optimization method for the APEX model. *J. Environ. Qual.* 43:1381–1391. doi:10.2134/jeq2013.12.0509
- Shapiro, S.S., and M.B. Wilk. 1965. An analysis of variance test for normality (complete samples). *Biometrika* 52:591–611. doi:10.1093/biomet/52.3-4.591
- Shapiro, S.S., M.B. Wilk, and H.J. Chen. 1968. A comparative study of various tests for normality. *J. Am. Stat. Assoc.* 63:1343–1372. doi:10.1080/01621459.1968.10480932
- Sharpley, A., C. Baffaut, D. Beegle, C. Bolster, Z. Easton, and P. Kleinman et al. 2017. Verification of phosphorus site assessment tools: Lessons from the USA. *J. Environ. Qual.* doi:10.2134/jeq2016.11.0427 (in press).
- Sharpley, A., D. Beegle, C. Bolster, L. Good, B. Joern, Q. Ketterings et al. 2012. Phosphorus indices: 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., L. Bergstrom, H. Aronsson, M. Bechmann, C.H. Bolster, K. Borling et al. 2015. Future agriculture with minimized phosphorus losses to waters: Research needs and direction. *Ambio* 44:S163–S179. doi:10.1007/s13280-014-0612-x
- Sharpley, A.N., S.C. Chapra, R. Wedepohl, J.T. Sims, T.C. Daniel, and K.R. Reddy. 1994. Managing agricultural phosphorus for protection of surface waters: Issues and options. *J. Environ. Qual.* 23:437–451. doi:10.2134/jeq1994.00472425002300030006x
- Sharpley, A., H.P. Jarvie, A. Buda, L. May, B. Spears, and P. Kleinman. 2013. Phosphorus legacy: Overcoming the effects of past management practices to mitigate future water quality impairment. *J. Environ. Qual.* 42:1308–1326. doi:10.2134/jeq2013.03.0098
- Sharpley, A.N., J.L. Weld, D.B. Beegle, P.J.A. Kleinman, W.J. Gburek, P.A. Moore et al. 2003. Development of Phosphorus Indices for nutrient management planning strategies in the United States. *J. Soil Water Conserv.* 58:137–152.
- Sims, J.T., and P.J.A. Kleinman. 2005. Managing agricultural phosphorus for environmental protection. In: J.T. Sims and A.N. Sharpley, editors, *Phosphorus: Agriculture and the environment*. Agron. Monogr. 46. ASA, CSSA, and SSSA, Madison, WI. p. 1021–1068. doi:10.2134/agronmonogr46.c31
- Soil Survey Staff. 2014. Web soil survey. USDA–NRCS, Washington, DC. <http://websoilsurvey.nrcs.usda.gov/> (accessed 29 Oct. 2014).
- Steglich, E.M., and J.R. Williams. 2013. Agricultural Policy/Environmental eXtender model user's manual. Version 0806. Blackland Research and Extension Center, Temple, Texas.
- Tuppad, P., C. Santhi, X. Wang, J.R. Williams, R. Srinivasan, and P.H. Gowda. 2010. Simulation of conservation practices using the APEX model. *Appl. Eng. Agric.* 26:779–794. doi:10.13031/2013.34947
- Wang, X., P.W. Gassman, J.R. Williams, S. Potter, and A.R. Kemanian. 2008. Modeling the impacts of soil management practices on runoff, sediment yield, maize productivity, and soil organic carbon using APEX. *Soil Tillage Res.* 101:78–88. doi:10.1016/j.still.2008.07.014
- Wang, X., S.R. Potter, J.R. Williams, J.D. Atwood, and T. Pitts. 2006. Sensitivity analysis of APEX for national assessment. *Trans. ASABE* 49:679–688. doi:10.13031/2013.20487
- Wang, X., J.R. Williams, J.D. Atwood, M.L. Norfleet, and A.D. King. 2011. APEX model upgrades, data inputs, and parameter settings for use in CEAP cropland modeling. USDA–NRCS, Washington, DC. http://www.nrcs.usda.gov/wps/portal/nrcs/detail/national/technical/?cid=nrcs143_014165 (accessed 6 May 2016).
- Wang, X., J.R. Williams, P.W. Gassman, C. Baffaut, R.C. Izaurralde, J. Jeong, and J.R. Kiniry. 2012. EPIC and APEX: Model use, calibration, and validation. *Trans. ASABE* 55:1447–1462. doi:10.13031/2013.42253
- Williams, J.R., W.L. Harman, M. Magre, U. Kizil, J.A. Lindley, G. Padmanabhan, and E. Wang. 2006. APEX feedlot water quality simulation. *Trans. ASABE* 49:61–73. doi:10.13031/2013.20244
- 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, Texas.
- Yin, L., X. Wang, J. Pan, and P.W. Gassman. 2009. Evaluation of APEX for daily runoff and sediment yield from three plots in the Middle Huaihe River Watershed, China. *Trans. ASABE* 52:1833–1845. doi:10.13031/2013.29212