12-1-2017

Calibration of the APEX Model to Simulate Management Practice Effects on Runoff, Sediment, and Phosphorus Loss

Ammar B. Bhandari Ph.D.
Fort Hays State University

Nathan O. Nelson
Kansas State University

Daniel W. Sweeney
Kansas State University

Claire Baffaut
University of Missouri

John A. Lory
University of Missouri

See next page for additional authors

Recommended Citation

Bhandari, A.B., Nelson, N.O., Sweeney, D.W., Baffaut, C., Lory, J.A., Senaviratne, A., Pierzynski, G.M., Janssen, K.A., Barnes, P.L. (2016). Calibration of the APEX Model to Simulate Management Practice Effects on Runoff, Sediment, and Phosphorus Loss. Journal of Environmental Quality, 46(6), 1332-1340. 10.2134/jeq2016.07.0272

Authors
Ammar B. Bhandari Ph.D., Nathan O. Nelson, Daniel W. Sweeney, Claire Baffaut, John A. Lory, Anomaa Senaviratne, Gary M. Pierzynski, and Phillip L. Barnes

Published December 15, 2016

Journal of Environmental Quality
SPECIAL SECTION
THE EVOLVING SCIENCE OF PHOSPHORUS SITE ASSESSMENT

Calibration of the APEX Model to Simulate Management Practice Effects on Runoff, Sediment, and Phosphorus Loss

Ammar B. Bhandari,* Nathan O. Nelson, Daniel W. Sweeney, Claire Baffaut, John A. Lory, Anomaa Senaviratne, Gary M. Pierzynski, Keith A. Janssen, and Philip L. Barnes

Abstract

Process-based computer models have been proposed as a tool to generate data for Phosphorus (P) Index assessment and development. Although models are commonly used to simulate P loss from agriculture using managements that are different from the calibration data, this use of models has not been fully tested. The objective of this study is to determine if the Agricultural Policy Environmental eXtender (APEX) model can accurately simulate runoff, sediment, total P, and dissolved P loss from 0.4 to 1.5 ha of agricultural fields with managements that are different from the calibration data. The APEX model was calibrated with field-scale data from eight different managements at two locations (management-specific models). The calibrated models were then validated, either with the same management used for calibration or with different managements. Location models were also developed by calibrating APEX with data from all managements. The management-specific models resulted in satisfactory performance when used to simulate runoff, total P, and dissolved P within their respective systems, with $r^2 > 0.50$, Nash–Sutcliffe efficiency > 0.30, and percent bias within ±35% for runoff and ±70% for total and dissolved P. When applied outside the calibration management, the management-specific models only met the minimum performance criteria in one-third of the tests. The location models had better model performance when applied across all managements compared with management-specific models. Our results suggest that models only be applied within the managements used for calibration and that data be included from multiple management systems for calibration when using models to assess management effects on P loss or evaluate P Indices.

Core Ideas

- The APEX model has limited ability to simulate effects of changing management.
- If available, multiple management should be used to calibrate and validate the model.
- Policy makers must exercise caution in using model-estimated P losses to evaluate PIs.

J. Environ. Qual. 46:1332–1340 (2017)
doi:10.2134/jeq2016.07.0272

Supplemental material is available online for this article.
Received 21 June 2016.
Accepted 16 Oct. 2016.
*Corresponding author (abbhandari@fhsu.edu).

Agricultural watersheds export substantial amounts of phosphorus (P) to water resources that can accelerate biological productivity, promote algal growth and eutrophication, and lead to general water quality degradation (Carpenter et al., 1998; Sharpley et al., 2003; Sharpley and Wang 2014). Land managers need accurate information on the effects of management practices on P loss so they can choose practices that reduce P loss and protect water quality. However, there is a general lack of data on P loss from the various and complex management practices.

Phosphorus Indices (PIs) are used to assess the vulnerability of agricultural fields for P loss and make recommendations to producers (Sharpley et al., 2012; Bolster et al., 2012). Ideally, a PI should accurately predict P loss risk due to changes in management practices. However, concerns have been raised about the ability of PIs to accurately rank the impacts of management practices on P loss (Benning and Wortmann, 2005; Osmond et al., 2006; Drewry et al., 2011; Nelson and Shober, 2012; Sharpley et al., 2012). In response to these concerns, the NRCS mandated that PIs be calibrated to standardize the P loss risk categories across regional, state, and watershed boundaries (USDA–NRCS, 2012).

Field studies provide valuable data on the water quality impacts of agricultural management systems, but there are drawbacks to the use of field study data for the evaluation of PIs. Field studies generally have a limited number of treatment comparisons. Because results are highly influenced by the weather patterns that occur during the study, edge-of-field monitoring data provide a snapshot of nutrient losses for a given weather scenario at the event or seasonal timescale. In contrast, PIs are intended to rank management impacts on long-term average annual losses. Furthermore,

A.B. Bhandari, Dep. of Agriculture, 204 B Albertson Hall, Fort Hays State Univ., Hays, KS, 67601; N.O. Nelson, K.A. Janssen, and G.M. Pierzynski, Dep. of Agronomy, Kansas State Univ., Manhattan, KS, 66506; D.W. Sweeney, Southeast Agricultural Research Center, Kansas State Univ., Parsons, KS, 67357; C. Baffaut, USDA–ARS Cropping Systems and Water Quality Research Unit, 241 Agricultural Engineering Building, Univ. of Missouri, Columbia, MO 65211; J.A. Lory and A. Senaviratne, Division of Plant Sciences, 108 Waters Hall, Univ. of Missouri, MO, 65211; P.L. Barnes, Dep. of Biological & Agricultural Engineering, Kansas State Univ., Manhattan, KS, 66502. Assigned to Associate Editor Deanna Osmond.

Abbreviations: APEX, Agricultural Policy Environmental eXtender model; CONV-T, fertilizer incorporated with chisel-disk-field cultivate; DP, dissolved phosphorus; FERTC, conventional tillage with commercial nitrogen and phosphorus fertilizers; NSE, Nash–Sutcliffe efficiency; NTSA, no tillage with surface-applied fertilizer; NTDB, no tillage with deep band fertilizer application; PBIAS, percentage bias; PEC, model performance evaluation criteria; PI, Phosphorus Index; TLNC, conventional tillage with nitrogen-based turkey litter application; TLPC, conventional tillage with phosphorus-based turkey litter and commercial nitrogen fertilizer application; TP, total phosphorus.

the available measured edge-of-field P loss data represent a limited number of locations and management systems.

Process-based computer models have been used to estimate management practice impacts on P loss (Yin et al., 2009; Gassman et al., 2010; Wang et al., 2012). Models are advantageous because they are not restricted to a limited number of treatment comparisons, durations, or weather scenarios. Although processed-based models may be too complex for routine use by nutrient management planners, modeling results could be used to evaluate, revise, and calibrate PIs (Nelson and Shober, 2012).

Computer models have been widely used to assess influences of management practices on P loss and to guide water resource policy, management, and regulations (Plotkin et al., 2013; Francesconi et al., 2014; Ford et al., 2015). The Agricultural Policy Environmental eXtender (APEX) model was used in the Conservation Effects Assessment Project (CEAP) to assess the benefits of the USDA conservation program (Mausbach and Dedrick, 2004; Wang et al., 2009). However, models must be tested (calibrated and validated) to ensure they are capable of simulating loss accurately over a variety of management practices. Model calibration decreases margins of error and minimizes uncertainties related to model parameters (Wang et al., 2009; Winchell et al., 2011; Baffaut et al., 2016).

Model use for PI evaluation is contingent on the ability of a model to accurately simulate the effects of many management practices on P loss. Most models are validated with the same management practices that were used during calibration (Ramanarayanan et al., 1997; Gassman et al., 2002; Williams and Izaurralde, 2006; Wang et al., 2008; Yin et al., 2009; Gassman et al., 2010; Mudgal et al., 2010, Kumar et al., 2011; Wang et al., 2012; Senaviratne et al., 2013; Francesconi et al., 2014). This could result in management-specific parameterizations that are not valid if applied to management practice scenarios that are not included in the calibration dataset. Some management-specific and sensitive parameters may be ignored during model testing if the model is calibrated and validated with a single set of management practices and then used to simulate water quality impacts of a different set of management practices. Previous studies have concluded that APEX can simulate effects of changing tillage (Wang et al., 2008) and implementation of best management practices (Wang et al., 2009) on runoff and sediment loss. However, no studies have evaluated the ability of APEX to simulate the impact of multiple management scenarios on P loss.

The objectives of this study were to determine if APEX can accurately simulate runoff, sediment, total P (TP) and dissolved P (DP) losses from management practices that are (i) similar to the calibration data (i.e., constant management practices) or (ii) different from the calibration data (i.e., constant management practices).

Materials and Methods

Measured runoff and water quality data from two field-scale watershed studies were used to calibrate and validate the APEX model. The Franklin site was located in Franklin County, KS (38°25′ N, 95°7′ W). All soils on the site were in the Summit soil series (fine, smectitic, thermic Oxyaquic Vertic Argiudolls) in NRCS hydrologic soil group C (D. Gastineau, unpublished data, 2013) with an average slope of 4 to 7%. The study site was terraced, creating six drainage areas ranging from 0.4 to 1.5 ha. The study included three tillage and fertilizer treatments applied in replicate to a soybean [*Glycine max* (L.) Merr.] and grain sorghum [*Sorghum bicolor* (L.) Moench] rotation, with water quality data collected from 2001 to 2004 (Table 1). The nitrogen (N) source was urea-ammonium nitrate applied at 78 kg N ha^{-1} and the P source was ammonium polyphosphate applied at 16 kg P ha^{-1} to grain sorghum. Additional site characterization and data collection details are described elsewhere (Zeimen et al., 2006; Mankin et al., 2010).

The Crawford runoff study was located in Crawford County, KS (37°30′ N, 94°59′ W). The site has claypan soils with 1% slope, mapped as Parsons silt loam (fine, mixed thermic Mollic Albaqualf) in NRCS hydrologic group D (D. Gastineau, unpublished data, 2013). There were 10 adjacent small watersheds, 133 by 31 m (0.40 ha) in size and separated on all sides by a soil berm to isolate runoff, with berms on the downslope end of the watershed angled toward the outlet. The study included four tillage and poultry litter treatments applied in replicate to a continuous grain sorghum cropping system with runoff collected from 2011 to 2013 (Table 1). Prior management was sorghum–soybean (2001–2004), continuous grain sorghum with poultry litter application (2005–2007), and continuous soybean (2008–2010). Additional details on site characteristics are available in Sweeney et al. (2012) and Zeimen et al. (2006).

For each location, runoff volume was measured from April to November at each watershed outlet with a 90° V-notch weir, instrumented with ISCO 6700 samplers (ISCO, Lincoln, NE). Water quality data included runoff volume, sediment loss, total N loss, TP loss, and DP loss based on flow-weighted composite samples for each runoff event, with some events including multiple days. Because detailed hydrograph data were not available, event durations were defined based on the onsite precipitation records, where consecutive days with precipitation were regarded as a single event. Measured data were reviewed for quality control and events were omitted from the analysis if the runoff:rainfall ratio was >0.9 or if equipment errors were confirmed by personal communication with location managers.

Model Development

The APEX model is a farm- to small watershed-scale, daily time-step, process-based model that simulates daily water flux, plant growth, nutrient cycling, soil erosion, and nutrient loss (Williams and Izaurralde, 2006; Wang et al., 2009; Gassman et al., 2010; Wang et al., 2012; Williams and Steglich, 2012). The APEX model is written in Fortran with an open source code and the version 0806, compiled in August 2015, was used for this study.

Both sites had onsite, daily precipitation data collected during the months where runoff was monitored. Precipitation data from nearby weather stations (National Climatic Data Center) were used to fill in winter precipitation, missing precipitation data during the growing season, and daily temperature data (Supplemental Section S1, Supplemental Fig. S1).

Soil chemical data were obtained from archived soil samples collected at 0- to 5-cm and 5- to 15-cm depths from each watershed at the beginning of the monitoring period, which were analyzed for total carbon (LECO, 1995), total N, and TP (Bremner and Mulvaney, 1982) and Bray-1 P (Brown 1998). Anion exchangeable P, which was used as a measure of labile P, was estimated from Bray-1 P using the regression equation suggested by

Mallarino and Atia (2005). The soil P sorption coefficient (PSP) was calculated as PSP = 1/{[(TP − organic P)/(5 × labile P)] + 4/5) (Nelson and Parsons, 2006). Soil physical characteristics were obtained from onsite soil investigations conducted in 2012 and 2013, with details described in Supplemental Section S2.

Site-specific management data such as tillage, fertilization rates and times, poultry litter rates, application methods, date of planting, date of harvesting, etc. were obtained from site managers and used to develop APEX management files. They are summarized in Table 1.

The watersheds in the Franklin site were delineated using the ArcAPEX interface using 2-m digital elevation data (Kansas Geological Survey, 2012). The average upland slope, average upland slope length, mainstream channel slope, and channel slope of routing reach were adjusted based on site characteristics and measured field data. Watersheds at the Crawford site were defined using the WinAPEX interface with inputs based on measured field data and observed site characteristics. Because these watersheds were less <1.5 ha with relatively uniform management and soils, they were each modeled as a single subarea.

Data Analysis and Model Evaluation

Model estimates of runoff, sediment, TP, and DP from the daily watershed outlet (.DWS) file were compared with measured data for each event. The coefficient of determination (r^2), Nash–Sutcliff model efficiency (NSE) (Nash and Sutcliffe, 1970), and percentage bias (PBIAS) (Gupta et al., 1999) were used to evaluate model performance during calibration and validation. Considering the objectives of the current study, the acceptable model performance evaluation criteria (PEC) were r^2 > 0.50 and NSE > 0.30 for runoff, sediment, and TP loss and |PBIAS| < 35, 60, 70% for runoff, sediment, and TP respectively (Nelson et al., 2016; Baffaut et al., 2016).

Management-Specific Model Development

The APEX control and parameter files define equations used for specific model processes, process threshold values, and equation coefficients. The initial control and parameter file inputs were based on best professional judgment values described by Baffaut et al. (2016). The sensitivity analysis and calibration processes were a combination of both manual and automated calibration. Initially, sensitive parameters were identified by manually adjusting each input in the control and parameter files within reasonable ranges, as suggested in the APEX user manual. Sensitive parameters were identified as those for which relatively small changes resulted in substantial differences in model output. Second, automated calibration of 10 to 15 of the most sensitive parameters was conducted with the APEX-PROPOT program (Senaviratne et al., 2014), which uses a stepwise multiobjective, multivariable optimization algorithm. Finally, manual calibration was used to refine sensitive parameter values. The final results were confirmed as the best calibration by comparison with results from the automated calibration. Management-specific APEX models were calibrated for each management system by changing options in the control file and by adjusting sensitive parameter values identified in Supplemental Table S1 to optimize model performance (maximize r^2 and NSE, minimize |PBIAS|). Inputs in the soil, management, or watershed files were not considered for calibration, as those were determined based on measured data or field descriptions during model setup. All model simulations included a 4-yr warmup period to reduce the potential for adverse impacts of unknown initial values (i.e., soil moisture, residue cover, etc.) on model results.

To test objective (i), management-specific model parameterizations were validated by comparing model output with independently measured data from a watershed with management that was identical to the calibration data. Validation was completed without changing any model parameters. To test objective (ii), the management-specific model parameterizations were used to simulate runoff, sediment, and nutrient loss from watersheds where the management practices were different from the management of the calibration dataset.

The overall model performance was also assessed by combining event-based measured datasets from all calibration watersheds and computing the r^2, NSE, and PBIAS for the entire dataset. A similar process was followed for the validation datasets. Model-simulated crop yields were also assessed by aggregating data for all watersheds in each location for computation of PBIAS.

Location Model Development

A location model parameterization was developed for the Franklin and Crawford locations based on the

Table 1. Summary of the management practices used for calibration and validation at the Franklin (2001–2004) and Crawford (2011–2013) locations.

Location	Management abbreviation	Management practices	Watershed ID†	
			Calibration	Validation
Franklin	NTDB	No tillage; deep band P fertilizer application‡, 7.6–12.7 cm deep, 16 kg P ha^{-1} in spring prior to planting sorghum	7 (32)	2 (36)
Franklin	NTSA	No tillage; surface-applied P fertilizer‡, 16 kg P ha^{-1} in spring prior to planting sorghum	8 (33)	4 (33)
Franklin	CONV-T	Fertilizer incorporated‡ with chisel-disk and field cultivator, 16 kg P ha^{-1} in spring prior to planting sorghum	6 (36)	5 (34)
Crawford	Control	No tillage; No P fertilizer or turkey litter application, 0 kg P ha^{-1} yr^{-1}	103 (27)	205 (26)
Crawford	FERTC	Conventional tillage§; urea ammonium nitrate with ammonium polyphosphate; 135 kg N ha^{-1} yr^{-1} and 24 kg P ha^{-1} yr^{-1} applied in spring prior to planting.	102 (27)	203 (27)
Crawford	TLPC	Conventional tillage§; P-based turkey litter (1 Mg ha^{-1} yr^{-1}) with urea ammonium nitrate; 135 kg N ha^{-1} and 24 kg P ha^{-1} applied in spring prior to planting	104 (27)	204 (27)
Crawford	TLNC	Conventional tillage§; N-based turkey litter (7.5 Mg ha^{-1} yr^{-1}), 135 kg N ha^{-1} yr^{-1}, and 180 P kg ha^{-1} yr^{-1} applied in the spring prior to planting	105 (27)	201 (27)

† Identification number. Numbers in the parenthesis are the number of events used for calibration or validation. The total numbers of events collected were 36 at the Franklin site and 33 at the Crawford site.

‡ At Franklin, the fertilizers were only applied during the grain sorghum years.

§ Conventional tillage is Chisel (15-cm depth) followed by disk (5–10-cm depth) and field cultivate (5-cm depth).

management-specific models at the respective locations. Parameters that were equal for all management-specific models were set to that value. Parameter values that differed from one management-specific model to another were used for location calibration. During the location calibration, selected model parameters were manually adjusted to maximize PEC for calibration datasets from all management systems at the location (Table 1, Supplemental Table S1). The location models were then validated for all management practices present at the location using the validation datasets (Table 1).

Results

Management-Specific Model Calibration and Validation

The model-simulated crop yields were in good agreement with the measured data at both sites. The PBIAS for grain sorghum yield simulated by the calibrated model at Crawford was −22 and −7% for the calibration and validation datasets, respectively. Likewise, the PBIAS for grain yield simulated with the calibrated model at Franklin was −5 and −2% for calibration and validation datasets.

Simulated runoff with the calibration and validation datasets exceeded the model PEC for runoff in all three management practices at the Franklin site. For sediment loss, the CONV-T (fertilizer incorporated with chisel-disk-field cultivate) management met PEC after calibration and validation. The NTSA (no tillage with surface-applied fertilizer) management did not meet PEC for r^2, and the NTDB (no tillage with deep band fertilizer application) management practice did not meet PEC for r^2 and NSE during calibration. Both NTSA and NTDB failed to pass PEC for validation of sediment loss (Supplemental Table S2).

The three management-specific models at Franklin met calibration and validation PEC for TP loss (Supplemental Table S2). Measured TP losses over a 4-yr rotation with CONV-T, NTDB, and NTSA management practices were 3.44, 2.75, and 5.23 kg ha^{-1}, respectively. The model-simulated TP loss followed a similar trend (CONV-T = 1.92 kg ha^{-1}; NTDB = 1.49 kg ha^{-1}; NTSA = 4.19 kg ha^{-1}). Model performance criteria for DP loss simulated by the calibrated NTSA model exceeded PEC for both calibration and validation datasets. For the NTDB management, the calibrated model NSE (0.34) and PBIAS (+16%) passed PEC, and r^2 (0.48) was only slightly less than acceptable. However, the DP simulation for the NTDB model did not meet PEC for validation. Furthermore, the CONV-T management model did not meet PEC for DP loss in either calibration or validation (Supplemental Table S2).

The simulated runoff with the calibrated model at the Crawford site exceeded PEC for both calibration and validation datasets. For the sediment loss, only the model for FERTC (conventional tillage with commercial N and P fertilizers) management practice met PEC. After calibration, the PBIAS for sediment loss was within the acceptable criteria for all management practices. However, control (no tillage, fertilizer, or turkey litter application), TLPC (conventional tillage with P-based turkey litter and commercial N fertilizer application), and TLNC (conventional tillage with N-based turkey litter application) management practices did not meet the criteria for either r^2 or NSE. Overall, the sediment loss was underpredicted during the calibration by 6 to 35%. During validation, the model overpredicted sediment loss by 48 to 200% and none of the managements met PEC for sediment simulation in the validation datasets (Supplemental Table S2).

Management-specific models for all management systems at Crawford site met PEC for TP loss during calibration, except for the control model, which had a low r^2. All management-specific models met PEC for TP loss with validation datasets. The management-specific calibrated models exceeded PEC for calibration and validation of DP, except for a low r^2 for control and low NSE for TLNC during calibration (Supplemental Table S2).

Performance of the management-specific models across management practices and locations was assessed by combining observed and simulated event-based data from all management systems at both locations for computation of statistics (Fig. 1). The results showed that the runoff and TP loss simulated by the calibrated management-specific models exceeded the threshold criteria from both locations for calibration and validation datasets (Fig. 1a, 1b, 1e, and 1f). Across all managements and locations, the PBIAS for both calibration and validation of sediment simulation met PEC, but r^2 and NSE did not (Fig. 1c and 1d). Combined runoff and TP loss simulated for validation datasets exceeded PEC (Fig. 1b, 1f). The NSE and PBIAS for DP loss were above PEC for calibration and validation of DP loss; however, the r^2 was slightly lower than PEC for the calibration dataset (Fig. 1h and 1i).

Using APEX to Simulate Management Effects on Water Quality

Objective (ii) was tested by using the previously calibrated and validated management-specific models to simulate runoff, sediment, and P loss for validation datasets with contrasting management practices.

The ability of the APEX model to simulate changes in P fertilizer placement was determined by using a model calibrated and validated for subsurface-applied P fertilizer (NTDB) to simulate P loss when the fertilizer is surface applied (NTSA). The reverse was also tested. The model was successful at simulating a change from subsurface to surface placement of P fertilizer, but not the reverse (Table 2). Using the NTSA model to simulate P loss from NTDB managements resulted in a low r^2 for TP loss and overpredicted DP loss with generally poor NSE and r^2. The ability of APEX to simulate the effects of tillage and P placement was tested using the model calibrated and validated for CONV-T to simulate P loss from no-tillage management systems (NTDB and NTSA). The reverse was also tested. In both cases the simulated runoff, sediment, TP, and DP loss did not pass PEC (Table 2).

The ability of the APEX model to simulate effects of different nutrient sources on P loss in a conventionally tilled system was tested by using a model calibrated with data from the TLPC system to simulate the P loss from FERTC management. The runoff, TP, and DP loss all met PEC, indicating that the fully calibrated and validated APEX model is capable of simulating the change in fertilizer source with the same tillage (Table 2).

The ability of the APEX model to simulate the effect of P application rate on P loss was tested. Models calibrated and validated with data from conventionally tilled systems for which either poultry litter (TLPC) or fertilizer (FERTC) were applied at 24 kg P ha^{-1} were used to simulate P loss from a conventional tillage system with poultry litter applied at 180 kg P ha^{-1} (TLNC). The results

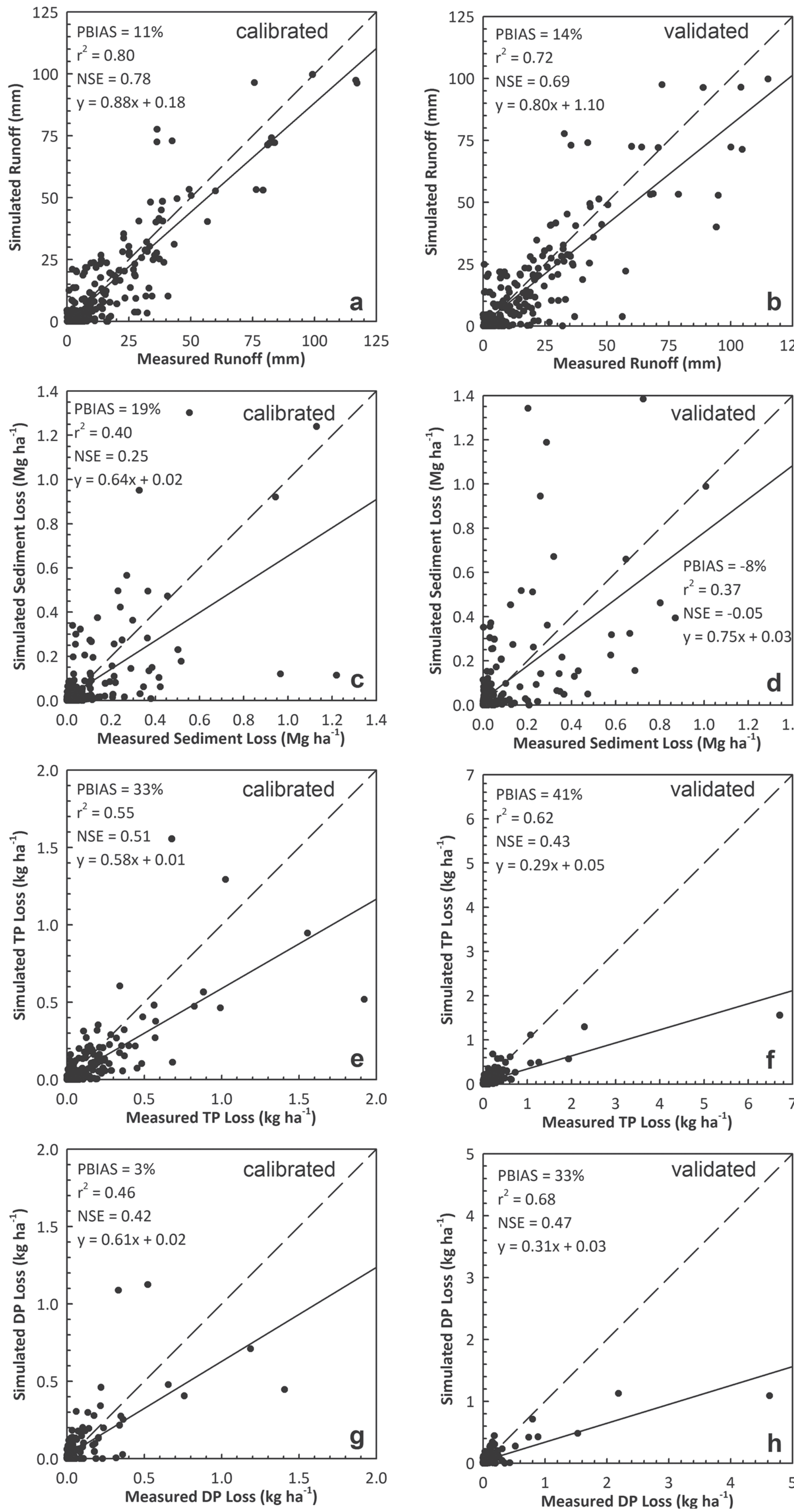

Fig. 1. Event-based linear regression of measured versus simulated values with model performance statistics for Agricultural Policy Environmental eXtender Model (APEX)-simulated runoff (a, b), sediment loss (c, d), total phosphorus (TP) loss (e, f), and dissolved phosphorus (DP) loss (g, h) for models that were calibrated (a, c, e, and g) for specific management systems and validated (b, d, f, and h) within the management systems used for calibration (data combined from Franklin and Crawford locations).

indicated that, while PEC were met for runoff, this was not the case for sediment, TP, or DP loss (Table 2).

Likewise, the ability of the APEX model to simulate effects of changing nutrient source, rate, and tillage was tested using a model calibrated for no-tillage management without any P application (control) to simulate P loss from FERTC, TLPC, or TLNC. Although the model did not simulate sediment loss well, it did pass PEC for runoff, TP, and DP when the model was calibrated with no-tillage simulated P loss in conventional tillage at low P application rates. The model did not pass PEC when simulating P loss at high P application rates (for TLNC). When the models calibrated with conventional tillage were used to simulate P loss for no tillage, they did not meet PEC for sediment, TP, or DP. Management-specific models only passed PEC for both runoff and TP loss in 5 of the 18 tests when applied outside the management system used for calibration (Table 2). In each of these cases, the models also passed PEC for simulation of DP loss.

Location Models

Because management-specific calibrated and validated models failed to provide consistently accurate estimates of P loss when applied to different management systems, a location model was developed by calibrating with data from all management systems at each location (Franklin and Crawford). The final parameter values for the location models are listed in Supplemental Table S1.

Full calibration of location models was not possible at either location. Both models exceeded the calibration PEC for runoff, but generally failed to meet PEC for sediment (Table 3). The Franklin model met PEC for TP calibration for the two no-tillage management systems but could not simultaneously be calibrated for P loss in the conventional-tillage system. The Crawford model was calibrated for TP loss in the three systems with lower P inputs (<25 kg P ha^{-1} yr^{-1}) but could not be simultaneously calibrated for TP loss from the system with high P inputs (180 kg P ha^{-1} yr^{-1}). The location models only passed PEC for calibration of DP loss in one of the management systems at their respective locations.

The Franklin model exceeded PEC for simulation of runoff and TP loss in only one of the three validation managements (NTSA); however, it was very close to passing PEC in the NTDB management also (Table 3). The Crawford model exceeded PEC for runoff and TP loss in three of the four management systems used for validation. Therefore, the location models passed four of seven validation tests, which is nearly twice the passing rate compared with when management specific models were applied outside the calibration management.

Discussion

The APEX estimates of runoff were very good when calibrated and validated within a single management system (Supplemental Table S2) and for nearly all cases when APEX was used to simulate runoff in a contrasting management system (Table 2). Other studies reported good results for runoff calibration and validation with APEX (Gassman et al., 2010; Kumar

Table 2. Model performance statistics for models used to simulate runoff, sediment, and phosphorus (P) loss for management systems that are different from the management practices for data used to calibrated the model at Franklin and Crawford runoff study sites (bolded values indicate model performance that did not meet threshold criteria).

Management used to calibrate†	Management validated‡	Criteria met§	Runoff¶			Sediment			Total P			Dissolved P		
			r^2	NSE	PBIAS	r^2	NSE	PBIAS	r^2	NSE	PBIAS	r^2	NSE	PBIAS
Franklin runoff study site														
NTDB	NTSA	Yes	0.81	0.76	22	**0.21**	**0.09**	20	0.68	0.60	35	0.72	0.54	26
NTSA	NTDB	No	0.63	0.55	24	**0.28**	**0.20**	17	**0.41**	0.37	21	**0.24**	**−1.84**	**−83**
CONV-T	NTSA	No	0.77	0.45	**49**	**0.06**	**−0.31**	**97**	0.64	**−0.17**	**92**	**0.44**	**−0.01**	**87**
CONV-T	NTDB	No	0.62	**0.18**	**−52**	**0.12**	**−0.30**	**91**	**0.36**	**−0.56**	**89**	**0.10**	**−0.39**	**−75**
NTSA	CONV-T	No	0.66	0.38	−24	0.62	**−1.27**	**−70**	0.50	**−0.99**	−53	**0.16**	**−232**	**−846**
NTDB	CONV-T	No	0.70	0.52	−13	0.63	**0.17**	−18	0.51	**0.13**	−7	**0.21**	**−72.0**	**−462**
Crawford runoff study site														
TLPC	FERTC	Yes	0.79	0.77	11	**0.36**	**0.04**	−17	0.70	0.53	45	0.64	0.56	33
FERTC	TLPC	Yes	0.70	0.67	12	**0.10**	**0.01**	**62**	0.56	0.33	55	0.50	0.42	33
TLPC	TLNC	No	0.81	0.81	−4	**0.37**	**−4.59**	**−96**	0.55	**0.11**	**−116**	0.53	**−0.58**	**−182**
FERTC	TLNC	No	0.80	0.80	−4	**0.36**	**−3.68**	**−80**	0.51	**−0.11**	**−135**	0.50	**−1.05**	**−212**
TLNC	TLPC	No	0.70	0.67	13	**0.15**	**−1.18**	−38	0.51	**0.07**	**78**	0.94	**−0.18**	**−108**
TLNC	FERTC	No	0.77	0.75	8	**0.32**	**0.28**	26	0.63	**0.05**	**80**	0.70	**0.00**	**82**
Control	TLPC	Yes	0.69	0.66	14	**0.05**	**−5.71**	**−104**	0.71	0.42	51	0.54	0.37	48
Control	TLNC	No	0.79	0.79	−2	**0.48**	**−29.0**	**−333**	0.53	**0.03**	**−112**	0.53	**−0.36**	**−155**
Control	FERTC	Yes	0.78	0.76	13	**0.27**	**−3.12**	**−135**	0.75	0.62	39	0.66	0.52	41
TLPC	Control	No	0.75	0.75	11	**0.08**	**−0.04**	**67**	**0.40**	**0.27**	51	**0.40**	0.37	38
TLNC	Control	No	0.72	0.71	6	**0.05**	**−0.09**	**70**	**0.44**	**−0.21**	**88**	**0.38**	**−0.08**	**87**
FERTC	Control	No	0.75	0.74	10	**0.08**	**−0.05**	**67**	**0.36**	**0.25**	48	**0.38**	0.34	33

† CONV-T, fertilizer incorporated with chisel-disk-field cultivate; NTDB, no tillage, deep band fertilizer application; NTSA, no tillage, surface-applied fertilizer. FERTC, conventional tillage, commercial N and P fertilizers; control, no tillage, fertilizers, or poultry litter; TLPC, conventional tillage, P-based turkey litter, and commercial N fertilizer; TLNC, conventional tillage, N-based poultry litter.

‡ Includes data from both calibration and validation watersheds with the given management (see Table 1).

§ Yes means model threshold criteria were met for runoff and total P loss; no means model threshold criteria were not met for either runoff or total P loss.

¶ NSE, Nash–Sutcliffe efficiency; PBIAS, percent bias.

et al., 2011; Senaviratne et al., 2013; Francesconi et al., 2014). This indicates that, if calibrated and validated for runoff, APEX can reliably simulate effects of agricultural management on runoff and could be used to evaluate runoff components of PIs. This applies whether the model is calibrated with datasets that include one or multiple management systems.

Simulation of sediment loss was generally poor, regardless of the method used to calibrate the model. These results may have been influenced by the datasets. At the Franklin site, event-based sediment loss for no-tillage management was low, with a range of 0.00 to 0.52 Mg ha^{-1} and a median loss of 0.03 Mg ha^{-1}, compared with 0.00 to 0.94 Mg ha^{-1} and a median loss of 0.03 Mg ha^{-1} for conventional-tillage management. Similarly, the event-based sediment loss was low at Crawford due to low slope (1%), with a range of 0.00 to 0.38 Mg ha^{-1} (median = 0.01 Mg ha^{-1}) for no tillage and 0.00 to 1.22 Mg ha^{-1} (median = 0.02 Mg ha^{-1}) for conventional-tillage systems. Because the datasets had such low sediment loss, there was not enough information to calibrate sediment-related parameters in APEX. Several studies have indicated similar difficulty in calibration and validation of the APEX model for sediment loss, especially when the measured loss is very low (Kumar et al., 2011; Mudgal et al., 2012; Senaviratne et al., 2013, 2016). Poor model performance for sediment loss prevents the application of these parameter sets to situations where higher erosion loss would be expected.

The APEX model did a better job of simulating TP loss than it did for sediment. This may be because 50 to 60% of the P loss from these watersheds was DP. The model tended to do a better job of simulating DP loss for management systems with higher DP loss (i.e., NTSA). Aggregated model performance results indicate that, when calibrated for specific management practices, the APEX model is able to satisfactorily simulate runoff, TP, and DP loss across multiple locations and management practices (Fig. 1).

Model performance was mixed when APEX was calibrated for a specific management and then applied to another management. In general, when the changes in management were small (such as change in P source or a small change in P application rate), APEX provided good estimates of TP loss. However, APEX did not do well at simulating effects of changes in tillage or large changes in P application rate on TP loss (Table 2). Poor model performance for changing tillage could be related to the generally poor model calibration for sediment loss. Only one of the seven management-specific models passed calibration and validation for runoff, sediment, and TP loss (CONV-T, Supplemental Table S2). However, this model still overpredicted sediment loss, and thus P loss, when applied to no-tillage management systems (NTDB and NTSA, Table 2).

The NTDB model tended to overestimate DP loss during initial validation (PBIAS = −52). When this model was applied to a system with higher DP losses (NTSA), it still passed the validation, but under predicted DP loss (PBIAS = 26). By contrast, when the NTSA model was calibrated, model parameters were adjusted to maximize model performance for DP loss within that system, which resulted in very good NSE and r^2 and overprediction of DP loss (PBIAS = −44). Because the model overpredicted DP loss in a high DP loss system, it greatly overpredicted DP loss when applied to a system with lower DP loss (NTDB) and further failed to meet PEC for TP loss (Table 2). A key parameter for DP loss is P8, the soluble P sorption coefficient, which was set high for the NTDB model (to reduce DP loss) and set low for the NTSA model (to increase DP loss).

Neither of the location models passed PEC for simulating TP loss with all calibration and validation datasets. At Franklin, the location model did well at simulating TP loss for the no-tillage managements, but not for conventional tillage. At Crawford, the location model tended to underestimate P loss from the low P application systems and overestimate P loss from the system with high P application (TLNC). Wang et al. (2009) calibrated APEX for runoff and sediment loss with data from two different tillage systems (conventional and ridge). They produced a calibrated parameterization that was successful for simulating runoff and sediment loss with change in tillage systems. One difference is that they calibrated the model with tillage-dependent curve numbers. Although we may have been able to improve our

Table 3. Model performance statistics for runoff, sediment, and phosphorus (P) loss simulated with Agricultural Policy Environmental eXtender model (APEX) location-calibrated models at Franklin and Crawford runoff study sites (bolded values indicate that model performance did not meet the threshold criteria).

Model tested	Management (watershed no.)†	Criteria met	Runoff‡			Sediment			Total P			Dissolved P		
			r^2	NSE	P-bias	r^2	NSE	P-bias	r^2	NSE	P-bias	r^2	NSE	P-bias
Franklin runoff study site														
Calibration	CONV-T (6)	No	0.79	0.54	−15	0.66	0.32	−15	**0.41**	**−0.02**	−8	**0.09**	**−145**	**−724**
	NTDB (7)	Yes	0.78	0.59	34	**0.35**	**0.02**	68	0.66	0.37	52	0.50	**−0.21**	−20
	NTSA (8)	Yes	0.83	0.71	32	**0.38**	**−0.10**	**83**	0.80	0.60	48	0.74	0.67	5
Validation	CONV-T (5)	No	0.70	0.63	4	0.60	**0.17**	2	0.57	**0.07**	−8	**0.32**	**−175**	**−659**
	NTDB (2)	No	0.59	0.38	**43**	**0.43**	**0.04**	**81**	0.47	0.30	49	**0.36**	**−1.02**	**−98**
	NTSA (4)	Yes	0.82	0.70	30	**0.12**	**0.03**	58	0.71	0.68	28	0.75	0.74	−21
Crawford runoff study site														
Calibration	FERTC (102)	Yes	0.87	0.85	19	**0.42**	**−0.78**	−56	0.82	0.52	56	0.76	0.35	60
	Control (103)	Yes	0.78	0.66	−19	**0.15**	**−0.46**	−13	0.50	0.30	55	**0.33**	**0.23**	53
	TLPC (104)	Yes	0.78	0.76	18	**0.19**	**−4.61**	**−124**	0.52	0.31	56	0.52	**0.21**	64
	TLNC (105)	No	0.83	0.81	3	**0.35**	**−19.3**	**−259**	0.76	**−7.56**	**−178**	0.60	**−15.0**	**−262**
Validation	FERTC (203)	Yes	0.72	0.57	0	0.73	**−70.2**	**−661**	0.62	0.58	30	0.63	0.40	50
	Control (205)	No	0.76	0.71	25	**0.16**	**−6.7**	**−121**	0.63	**0.23**	66	**0.45**	**0.13**	**71**
	TLPC (204)	Yes	0.63	0.60	11	**0.18**	**−23.2**	**−299**	0.57	0.40	49	**0.47**	**0.21**	64
	TLNC (201)	Yes	0.78	0.77	−12	0.450	**−32.6**	**−402**	0.82	0.77	−1	0.84	0.80	−9

† CONV-T, fertilizer incorporated with chisel-disk-field cultivate; NTDB, no tillage, deep band fertilizer application; NTSA, no tillage, surface-applied fertilizer. FERTC, conventional tillage, commercial N and P fertilizers; control, no tillage, fertilizers, or poultry litter; TLPC, conventional tillage, P-based turkey litter, and commercial N fertilizer; TLNC, conventional tillage, N-based poultry litter.

‡ NSE, Nash–Sutcliffe efficiency; PBIAS, percent bias.

model results by calibrating management-dependent curve numbers, we choose to assign curve numbers based on land use tables to increase the applicability of this work specific to model use in testing PIs where selection of curve number would likely be based on land use information.

Accurately simulating the effects of management practices on P loss is essential if a model is to be used for PI evaluation. The management-specific models are not suitable for generating data to evaluate PIs because these models failed to accurately simulate effects of changing management on TP loss in over 70% of the tests. Overall, the model performance improved with location models, compared with the management specific models, and met PEC for runoff and TP loss for slightly over 50% of the validation datasets. This indicates that the model's ability to simulate changes in management practices improved when the calibration dataset included data from multiple management practices.

Although the location models developed in this study are an improvement over the management-specific models, they still fall short of generating results that would be required for use in extensive PI evaluation and revision. These models could be used to generate data within the limits of the management practices that passed calibration and validation. The resulting data could be used to evaluate whether a PI is directionally correct. Perhaps a model would do a better job at simulating effects of changing management practices on P loss if the parameters were successfully calibrated and validated for all four water quality measures (runoff, sediment loss, TP loss, and DP loss). However, we were unable to achieve such a calibration for these datasets using the APEX model.

Acknowledgments

This work was funded through the Kansas Water Resources Institute (KWRI) and a USDA–NRS Conservation Innovation Grant (CIG).

References

Baffaut, C., N.O. Nelson, J.A. Lory, G.M.M.M.A. Senaviratne, A.B. Bhandari, R.P. Udawatta et al. 2016. Multisite evaluation of APEX for water quality: I. Best professional judgement parameterization. J. Environ. Qual. doi:10.2134/jeq2016.06.0226 (in press).

Benning, J., and C. Wortmann. 2005. Phosphorus indexes in four Midwestern states: An evaluation of the differences and similarities. J. Soil Water Conserv. 60:221–227.

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

Bremner, J.M., and C.S. Mulvaney. 1982. Salicylic acid thiosulfate modification of the Kjeldhal method to include nitrate and nitrite. In: R.H. Miller and D.R. Keeney, editors, Methods of soil analysis. Part 2. ASA, Madison, WI. p. 621.

Brown, J.R., Editor. 1998. Recommended chemical soil test procedures for the North Central Region. SB 1001. Missouri Agric. Exp. Stn, Colombia, MO.

Carpenter S.R., N.F. Caraco, D.L. Correll, R.W. Howarth, A.N. Sharpley, and V.H. Smith. 1998. Nonpoint pollution of surface waters with phosphorus and nitrogen. Ecol. Appl. 8:559–568. doi:10.1890/1051-0761(1998)008[0559:NPOSWW]2.0.CO;2

Drewry, J.J., L.T.H. Newham, and R.S.B. Greene. 2011. Index models to evaluate the risk of phosphorus and nitrogen loss at catchment scales. J. Environ. Manage. 92:639–649. doi:10.1016/j.jenvman.2010.10.001

Ford W., M. King, J. Williams, and N. Fausey. 2015. Sensitivity analysis of the Agricultural Policy/Environmental eXtender (APEX) for phosphorus loads in tile-drained landscapes. J. Environ. Qual. 44:1099–1110. doi:10.2134/jeq2014.12.0527

Francesconi, W., D.R. Smith, G.C. Heathman, X. Wang, and C.O. Williams. 2014. Monitoring and apex modeling of no-till and reduced-till in tile-drained agricultural landscapes for water quality. Trans. ASABE 57:777–789.

Gassman, P., E. Osei, A. Saleh, and L. Hauck. 2002. Application of an environmental and economic modeling system for watershed assessments. J. Am. Water Resour. Assoc. 38:423–438. doi:10.1111/j.1752-1688.2002.tb04327.x

Gassman, P.W., J.R. Williams, X. Wang, A. Saleh, E. Osei, L.M. Hauck et al. 2010. 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

Gupta, H.V., S. Sorooshian, and P.O. Yapo. 1999. Status of automatic calibration for hydrologic models: Comparison with multilevel expert calibration. J. Hydrol. Eng. 4:135–143. doi:10.1061/(ASCE)1084-0699(1999)4:2(135)

Kansas Geological Survey. 2012. Data catalog. Kansas Data Access and Support Center (DASC). Kansas Geological Survey, Univ. of Kansas, Lawrence, KS. http://kansasgis.org/catalog/index.cfm?data_id=1921&show_cat=5 (accessed 25 May 2016).

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

LECO. 1995. CNS-200 instrument method–Carbon, nitrogen, sulfur testing in soil/plant tissue. LECO Corp., St. Joseph. MI.

Mallarino, A., and A. Atia. 2005. Correlation of a resin membrane soil phosphorus test with corn yield and routine soil tests. Soil Sci. Soc. Am. J. 69:266–272. doi:10.2136/sssaj2005.0266

Mankin, K.R., D. Maski, K.A. Janssen, P. Tuppad, and G.M. Pierzynski. 2010. Modeling nutrient runoff yields from combined in-field crop management practices using SWAT. Trans. ASABE 53:1557–1568. doi:10.13031/2013.34914

Mausbach, M., and A. Dedrick. 2004. The length we go—Measuring environmental benefits of conservation practices. J. Soil Water Conserv. 59:96A–103A.

Mudgal, A., C. Baffaut, S.H. Anderson, E.J. Sadler, N.R. Kitchen, K.A. Sudduth, and R.N. Lerch. 2012. Using the Agricultural Policy 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

Mudgal, A., C. Baffaut, S.H. Anderson, E.J. Sadler, and A.L. Thompson. 2010. Apex model assessment of variable landscapes on runoff and dissolved herbicides. Trans. ASABE 53:1047–1058. doi:10.13031/2013.32595

Nash, J.E., and J.V. Sutcliffe. 1970. River flow forecasting through conceptual models: Part I. A discussion 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. Udawatta et al. 2016. Multisite evaluation of APEX for water quality: II. Regional parameterization. J. Environ. Qual. doi:10.2134/jeq2016.07.0254 (in press).

Nelson, N.O., and J.E. Parsons. 2006. Modification and validation of GLEAMS for prediction of phosphorus leaching in waste-amended soils. Trans. ASABE 49:1395–1407. doi:10.13031/2013.22055

Nelson, N.O., and A.L. Shober. 2012. Evaluation of phosphorus indices after twenty years of science and development. J. Environ. Qual. 41:1703–1710. doi:10.2134/jeq2012.0342

Osmond, D.L., M.L. Cabrera, S.E. Feagley, G.E. Hardee, C.C. Mitchell, P.A. Moore, Jr. et al. 2006. Comparing ratings of the southern phosphorus indices. J. Soil Water Conserv. 61:325–337.

Plotkin, S., X. Wang, T.L. Potter, D.D. Bosch, J.R. Williams, E.S. Hesketh, and J.K. Bagdon. 2013. APEX calibration and validation of water and herbicide transport under US Southern Atlantic Coastal Plain conditions. Trans. ASABE 56:43–60. doi:10.13031/2013.42589

Ramanarayanan, T.S., J.R. Williams, W.A. Dugas, L.M. Hauck, and A.M.S. McFarland. 1997. Using APEX to identify alternative practices for animal waste management. Part I: Model description and validation. Paper No. 97-2209. Texas Institute of Applied Environmental Research, Tarleton State Univ., Stephenville, TX.

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

Sharpley, A.N., 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., P. Kleinman, R. McDowell, M. Gitau, and R. Bryant. 2003. Modeling phosphorous transport in agricultural watersheds: Processes and possibilities J. Soil Water Conserv. 57:425–439.

Sharpley, A.N., and X. Wang. 2014. Managing agricultural phosphorus for water quality: Lessons from the USA and China. J. Environ. Sci. (China) 26:1770–1782. doi:10.1016/j.jes.2014.06.024

Sweeney, D.W., G.M. Pierzynski, and P.L. Barnes. 2012. Nutrient losses in field-scale surface runoff from claypan soil receiving turkey litter and fertilizer. Agric. Ecosyst. Environ. 150:19–26. doi:10.1016/j.agee.2012.01.008

USDA–NRCS. 2012. Part 302—Nutrient management policy implementation. Title 190-National Instruction, amended January 2012. USDA–NRCS, Washington, DC. http://www.nrcs.usda.gov/Internet/FSE_DOCUMENTS/stelprdb1046435.pdf (accessed 20 Jan. 2016).

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., D.W. Hoffman, J.E. Wolfe, J.R. Williams, and W.E. Fox. 2009. Modeling the effectiveness of conservation practices at shoal creek watershed, Texas, using apex. Trans. ASABE 52:1181–1192. doi:10.13031/2013.27794

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.R.C.I., and E.M. Steglich. 2012. Agricultural Policy/Environmental eXtender model: Theoretical documentation. Texas A&M AgriLife Blackland Research and Extension Center, Temple, TX.

Williams, J.R., and R.C. Izaurralde. 2006 The APEX model. In: Singh, V.P., and D.K. Frevert, editors, Watershed models. CRC Press, Boca Raton, FL. p. 437–482.

Winchell, M.F., R.L. Jones, and T.L. Estes. 2011. Comparison of models for estimating the removal of pesticides by vegetated filter strips. In: K.S. Goh, B.L. Bret, T.L. Potter, and J. Gan, editors, Pesticide mitigation strategies for surface water quality ACS Symposium Ser. Vol. 1075. ACS Publications, Washington, DC. p. 273–286 doi:10.1021/bk-2011-1075.ch017

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

Zeimen, M.B., K.A. Janssen, D.W. Sweeney, G.M. Pierzynski, K.R. Mankin, D.L. Devlin et al. 2006. Combining management practices to reduce sediment, nutrients, and herbicides in runoff. J. Soil Water Conserv. 61:258–267.