

2017

Multisite Evaluation of APEX for Water Quality: II. Regional Parameterization

Nathan O. Nelson
Kansas State University

Claire Baffaut
United States Department of Agriculture

John A. Lory
University of Missouri

G.M.M.M. Anomaa Senaviratne
University of Missouri

Ammar B. Bhandari
Fort Hays State University

Follow this and additional works at: https://lib.dr.iastate.edu/abe_eng_pubs

Part of the [Agriculture Commons](#), [Bioresource and Agricultural Engineering Commons](#), [Environmental Indicators and Impact Assessment Commons](#), [Environmental Monitoring Commons](#), [Natural Resources Management and Policy Commons](#), [Sustainability Commons](#), and the [Water Resource Management Commons](#)

The complete bibliographic information for this item can be found at https://lib.dr.iastate.edu/abe_eng_pubs/879. For information on how to cite this item, please visit <http://lib.dr.iastate.edu/howtocite.html>.

Multisite Evaluation of APEX for Water Quality: II. Regional Parameterization

Abstract

Phosphorus (P) Index assessment requires independent estimates of long-term average annual P loss from fields, representing multiple climatic scenarios, management practices, and landscape positions. Because currently available measured data are insufficient to evaluate P Index performance, calibrated and validated process-based models have been proposed as tools to generate the required data. The objectives of this research were to develop a regional parameterization for the Agricultural Policy Environmental eXtender (APEX) model to estimate edge-of-field runoff, sediment, and P losses in restricted-layer soils of Missouri and Kansas and to assess the performance of this parameterization using monitoring data from multiple sites in this region. Five site-specific calibrated models (SSCM) from within the region were used to develop a regionally calibrated model (RCM), which was further calibrated and validated with measured data. Performance of the RCM was similar to that of the SSCMs for runoff simulation and had Nash–Sutcliffe efficiency (NSE) > 0.72 and absolute percent bias (|PBIAS|) < 18% for both calibration and validation. The RCM could not simulate sediment loss (NSE < 0, |PBIAS| > 90%) and was particularly ineffective at simulating sediment loss from locations with small sediment loads. The RCM had acceptable performance for simulation of total P loss (NSE > 0.74, |PBIAS| < 30%) but underperformed the SSCMs. Total P-loss estimates should be used with caution due to poor simulation of sediment loss. Although we did not attain our goal of a robust regional parameterization of APEX for estimating sediment and total P losses, runoff estimates with the RCM were acceptable for P Index evaluation.

Disciplines

Agriculture | Bioresource and Agricultural Engineering | Environmental Indicators and Impact Assessment | Environmental Monitoring | Natural Resources Management and Policy | Sustainability | Water Resource Management

Comments

This article is published as Nelson, Nathan O., Claire Baffaut, John A. Lory, G. M. M. M. Anomaa Senaviratne, Ammar B. Bhandari, Ranjith P. Udawatta, Daniel W. Sweeney et al. "Multisite evaluation of APEX for water quality: II. Regional parameterization." *Journal of Environmental Quality* 46, no. 6 (2017): 1349-1356. DOI: [10.2134/jeq2016.07.0254](https://doi.org/10.2134/jeq2016.07.0254). Posted with permission.

Rights

Works produced by employees of the U.S. Government as part of their official duties are not copyrighted within the U.S. The content of this document is not copyrighted.

Authors

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

Multisite Evaluation of APEX for Water Quality: II. Regional Parameterization

Nathan O. Nelson,* Claire Baffaut, 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

Phosphorus (P) Index assessment requires independent estimates of long-term average annual P loss from fields, representing multiple climatic scenarios, management practices, and landscape positions. Because currently available measured data are insufficient to evaluate P Index performance, calibrated and validated process-based models have been proposed as tools to generate the required data. The objectives of this research were to develop a regional parameterization for the Agricultural Policy Environmental eXtender (APEX) model to estimate edge-of-field runoff, sediment, and P losses in restricted-layer soils of Missouri and Kansas and to assess the performance of this parameterization using monitoring data from multiple sites in this region. Five site-specific calibrated models (SSCM) from within the region were used to develop a regionally calibrated model (RCM), which was further calibrated and validated with measured data. Performance of the RCM was similar to that of the SSCMs for runoff simulation and had Nash–Sutcliffe efficiency (NSE) > 0.72 and absolute percent bias (|PBIAS|) < 18% for both calibration and validation. The RCM could not simulate sediment loss (NSE < 0, |PBIAS| > 90%) and was particularly ineffective at simulating sediment loss from locations with small sediment loads. The RCM had acceptable performance for simulation of total P loss (NSE > 0.74, |PBIAS| < 30%) but underperformed the SSCMs. Total P-loss estimates should be used with caution due to poor simulation of sediment loss. Although we did not attain our goal of a robust regional parameterization of APEX for estimating sediment and total P losses, runoff estimates with the RCM were acceptable for P Index evaluation.

Core Ideas

- Regionally calibrated APEX produced very good estimates of site-specific runoff.
- Regionally calibrated APEX failed to adequately estimate sediment loss.
- Regionally calibrated APEX P-loss estimates were worse than site-specific models.
- APEX runoff estimates are adequate for rigorous evaluation of P Index runoff components.
- APEX sediment loss estimates are unsuitable for evaluation of P Index.

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. 46:1349–1356 (2017)
doi:10.2134/jeq2016.07.0254

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 7 July 2016.

Accepted 2 Oct. 2016.

*Corresponding author (nonelson@ksu.edu).

THE PHOSPHORUS (P) INDEX was developed as a tool to assess the risk of P loss from agricultural fields. Although this tool has been used to encourage the adoption of conservation practices and develop nutrient management plans, excess P losses from agricultural fields and associated water quality degradation persist (Jarvie et al., 2013; Sharpley et al., 2015; USEPA, 2016). Due to the lack of water quality improvement and the disparity among state P Indices (Osmond et al., 2006), it has been proposed that the P Indices undergo evaluation to ensure accuracy in P-loss risk assessment (Sharpley et al., 2012).

A wide variety of methods have been used to assess P Indices (Nelson and Shober, 2012); however, the ideal assessment would include comparison of P Index results to independently obtained quantitative estimates of long-term average annual P loss across a wide range of soils, topography, and management practices (Sharpley et al., 2013). This type of quantitative independent assessment requires long-term average annual estimates of P loss because the P Index is a generalized assessment of the average risk of P loss across an extended period, as opposed to an assessment of P loss for a specific year or weather sequence. Because measured edge-of-field P-loss data are highly dependent on the weather patterns during the years of data collection, and because data collection generally occurs for a brief time period (<10 yr) and limited number of treatments, these data are generally unsuitable for use as an independent assessment of the P Index (Sharpley et al., 2013). Alternatively, process-based models could be used to generate independent assessments of long-term (>30 yr) P loss for multiple locations and management practices, provided that models are adequately validated.

Baffaut et al. (2017) found that an uncalibrated Agricultural Policy Environmental eXtender (APEX) model failed to adequately

N.O. Nelson, Dep. of Agronomy, 2004 Throckmorton Plant Sciences Center, Kansas State Univ., Manhattan, KS, 66506; 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 G.M.M.M. Anomaa Senaviratne, Division of Plant Sciences, 108 Waters Hall, Univ. of Missouri, MO, 65211; A.B. Bhandari, Dep. of Agriculture, 204 B Albertson Hall, Fort Hays State Univ., Hays, KS, 67601; R.P. Udawatta, Soil, Environmental and Atmospheric Sciences, 203 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, formerly, Biological Systems Engineering Dep., 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 6830726 Contribution no. 17-008-J from the Kansas Agricultural Experiment Station. Assigned to Associate Editor Deanna Osmond.

Abbreviations: APEX, Agricultural Policy Environmental eXtender; BPJ, best professional judgment; NSE, Nash–Sutcliffe efficiency; PBIAS, percent bias; PEC, performance evaluation criteria; RCM, regionally calibrated model parameterization; SSCM, site-specific calibrated model parameterization.

simulate edge-of-field P loss. However, it was found that, with calibration, the APEX model could simulate P loss resulting from different management practices at nearby locations with similar soils (Senaviratne et al., unpublished data, 2016; Bhandari et al., 2017). If APEX could be calibrated at a regional level, then it could be used to generate the required P-loss estimates for P Index evaluation.

Process-based models are commonly only calibrated for a single location or dataset, which potentially limits the applicability of the model across a wide range of soils, climates, and landscape positions. Calibration over multiple locations with different soils, management practices, and watershed characteristics could increase the area where the model can be applied. The Soil and Water Assessment Tool (SWAT) was calibrated for the Oklahoma and Texas region for runoff, sediment, and P loss by White et al. (2012), who calibrated hydrology at the basin (>500 km²) scale, followed with sediment and P calibration at the field scale. The resulting model had good to very good calibration and validation performance statistics for edge-of-field runoff and total P loss. Performance for sediment loss was worse, but still satisfactory.

Because the APEX model is well suited for simulating water quality impacts of management practices at the field scale (Wang et al., 2012) and has been promoted as a regional- and national-scale assessment tool (Saleh et al., 2011), it would be beneficial to develop a regional-scale calibrated parameterization. The objectives of this research were to develop a regional model parameterization of APEX 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

A regional model parameterization was developed based on parameterizations from previously calibrated and validated models from within the region and was evaluated against measured edge-of-field datasets. The process required model selection, systematic parameter comparison to identify differences, sensitivity analysis for parameters with contrasting values, a multisite calibration of disparate parameters to maximize model performance across all datasets, model validation with independent datasets, and model evaluation by comparing simulation results from the regional parameterization with those from locally calibrated parameterizations.

The APEX model was used to simulate runoff, sediment, and total P losses from agricultural systems. The APEX model is a process-based daily time step combined hydrologic and crop growth model that includes processes for chemical and nutrient transport and transformation (Gassman et al., 2010; Steglich and Williams, 2013). Hydrology, crop growth, chemical transformation, and sediment and chemical loss are simulated for subareas with uniform soil, topography, vegetation, and management. Losses are routed through subareas to the watershed outlet. Therefore, it is well suited for simulation of nutrient and sediment losses at the field to small watershed scale (Wang et al., 2012). This study used the APEX 0806 version with modifications as described by Baffaut et al. (2017). The APEX code and executable used to obtain the results presented herein are available on request from the corresponding author.

Model Development

Five site-specific calibrated models (SSCM) from four locations (one location with two distinct management periods) within the Heartland region (Supplemental Table S1; Supplemental Fig. S1) were selected as the basis for developing the regionally calibrated model (RCM). These are the same five models used by Baffaut et al. (2017) to evaluate an uncalibrated best professional judgement (BPJ) model parameterization. The general watershed characteristics, data collection, and methods of calibration and validation are described in papers within this special issue (Baffaut et al., 2017; Bhandari et al., 2017) and in previously published works (Udawatta et al., 2002; Zeimen et al., 2006; Senaviratne et al., 2012, 2013, 2014; Sweeney et al., 2012). Calibration of each SSCM started with the BPJ control file options and parameters and resulted in five SSCMs that differed from the BPJ parameterization for three options in the control file (Supplemental Table S2) and 21 values in the parameter file (Supplemental Table S3).

In general, APEX control file options are used to select equations or methods for process simulation and parameter file values are used as constants in equations. Control file values for the RCM were set by selecting values that were most common among the majority of SSCMs. The SSCM parameter file values that differed between the models (Supplemental Table S3) were examined to determine if there were site characteristics that contributed to or explained the differences obtained through site-specific calibration, and if not, they were selected for regional optimization. The regional calibration process included conducting additional sensitivity analysis to identify site-specific differences in parameter sensitivity. Parameter values that were found to be equal among the majority of SSCMs and were nonsensitive for the remaining SSCMs were set in the RCM to the most common value used for the SSCMs. The RCM parameter values were set to the average of the SSCM parameter values for parameters whose values were close to each other and had uniformly low sensitivity within that range for all SSCMs. Parameters with nonuniform values in the SSCMs that were highly sensitive to changes within the range were selected for calibration through a regional calibration process.

The regional calibration process was an event-based, two-step, multisite model calibration to optimize calibration parameters. Precipitation and runoff that occurred over multiple days were summed as an event, where the end of the event was defined as a day without any rainfall. First, the PAROPT tool was used to identify the multiple parameter sets that met model performance criteria for each dataset. The PAROPT tool is a step-wise, multi-objective, multivariable automatic parameter optimization tool that runs the APEX model for all possible combinations of selected parameter values and computes performance statistics with three objective functions (Senaviratne et al., 2014). The PAROPT tool was used to run APEX for each dataset using all possible combinations of the selected parameter values. Parameter combinations were compared to identify a parameter set that met model performance criteria for the majority of datasets. Following identification of the most commonly acceptable parameterization with PAROPT, manual calibration was used to maximize model performance. The manual calibration consisted of running model simulations for a limited set of parameter values and computing the performance statistics for event-based model output combined

across all calibration datasets. For example, measured event runoff from the calibration datasets for all locations was plotted against the model simulated runoff for the same set of events ($n = 158$). Following calibration, the RCM was validated by simulating runoff, sediment loss, and total P loss for the validation datasets and comparing simulated results with measured data.

Model Evaluation

The SSCMs had been calibrated and validated with 12 event-based datasets ranging from 2 to 8 yr in length. Five datasets had been used for calibration, one for each SSCM (Supplemental Table S1). These same five datasets were used for calibration of the RCM, as described above. Seven datasets had been used for validation of the SSCMs and were therefore also used for validation of the RCM (Supplemental Table S1). Model performance for simulation of runoff, sediment loss, and total P loss was assessed by computing the coefficient of determination (r^2), Nash–Sutcliffe model efficiency (NSE), and percent bias (PBIAS) for event-based comparison of measured data with model simulation results, computed as described by Moriasi et al. (2007). The statistics were also computed for an annualized comparison of measured versus simulated losses, where the annualized sums were calculated by summing verified events by year. Performance evaluation criteria (PEC) used to indicate acceptable model simulation were $r^2 > 0.5$, $NSE > 0.3$, and $|PBIAS| < 35, 60, \text{ and } 70$ for runoff, sediment loss, and total P loss, respectively. The justification for PEC is provided in Supplemental Section S1.

The performance statistics (PBIAS, r^2 , and NSE) were computed on either a combined dataset (including all datasets) or on single datasets, depending on the objectives of the analysis. The performance statistics were computed based on all datasets together for performance assessment across the region. They were computed for each of the 12 datasets to evaluate how well the SSCM or RCM parameterizations performed for a specific dataset (specific location, watershed, time period, and management system). Performance statistics computed for each dataset were further summarized based on mean, median, minimum, and maximum values obtained for each of the performance statistics. The distributions of performance statistic values obtained with the RCM and SSCMs were compared with each other and with the normal distribution using normal probability plots. The Shapiro–Wilk test for normality ($\alpha = 0.05$) was applied for performance statistics that passed the PEC obtained from the 12 datasets to determine if model performance statistics meeting PEC were normally distributed.

Results

Three control options were different from the BPJ parameterization: NVCN, ISW, and DRV (defined in Supplemental Table S2). The options NVCN and DRV were the same values for all SSCMs (NVCN = 4 and DRV = 3); therefore, these values were selected for the RCM. The value for ISW was three for the Knox B SSCM, but zero for all other SSCMs. Thus, ISW was set to zero for the RCM. Twenty global parameters in the SSCMs were different from the BPJ parameterization (Supplemental Table S3). Because Crawford was the only location with manure application, the Crawford SSCM was the only model that

could be calibrated for the manure-related parameters (62, 68, and 71). Therefore, the values obtained from the Crawford SSCM were used for the RCM. Because measured sediment loss was very low for the Knox B, Franklin, and Crawford datasets (Senaviratne et al., unpublished data, 2016; Bhandari et al., 2017), these datasets did not have enough information to inform the models relative to calibration for sediment. Therefore, sediment-related parameter 47 in the RCM was set based on the value obtained from the Knox and Chariton SSCMs. Parameter 85 had a general low level of sensitivity, and its value for the RCM was set based on theoretical descriptions of P subroutines and an evaluation of APEX model processes (Jones et al., 1984; Sharpley et al., 1984; Baffaut et al., 2013). Parameters 3, 15, 59, 76, and 90 were similar among the majority of SSCMs and nonsensitive for the remaining SSCMs and were therefore set to the most common value used for the SSCMs. Model results were relatively insensitive to changes for parameters 17, 19, and 46 within the range of values found for SSCMs; therefore, the RCM values were set as the average of SSCM values. Model results were found to be highly sensitive to parameters 8, 42, 46, 69, 70, 84, and 96. Consequently, these seven were included for regional calibration. Although parameters 29 and 31 were uniform for all SSCMs, they were very sensitive. Therefore, parameters 29 and 31 were also included in the regional calibration process to identify potential interactions between their optimal values and the other seven parameters included in the regional calibration process (Supplemental Table S4). The final RCM parameter set was unique relative to the SSCMs on which it was based (Supplemental Tables S2 and S3).

The RCM was well calibrated for runoff simulation with a PBIAS close to zero and high r^2 and NSE (Fig. 1a). It also passed the PEC for simulated runoff with the validation datasets (Fig. 1d). The RCM did a poor job of simulating sediment loss and did not pass any of the PEC for sediment loss with either the calibration or validation datasets (Fig. 1b and 1e). Despite the poor ability of the RCM to simulate sediment loss, the RCM had very high performance statistics when simulating total P loss for the calibration dataset (Fig. 1c). The calibration dataset contained one event with very high total P loss that may have a disproportionate influence on the performance statistics. The PBIAS, r^2 , and NSE were 1.1%, 0.55, and 0.39, respectively, when this extreme event was omitted from the computation, indicating that the RCM still met PEC for events with lower P loss. The total P loss simulated with the RCM also met PEC for the validation datasets (Fig. 1f).

Table 1 summarizes the three performance indicators used to evaluate model simulations with the RCM and SSCM parameterizations of APEX for each of the 12 datasets for runoff, sediment, and total P, respectively. For runoff, the RCM performed similarly to SSCMs. In contrast, the RCM did not perform as well as the SSCMs for sediment (Table 1). The SSCMs passed all PEC on 2 of 12 datasets, with a third only marginally rejected for high PBIAS (PBIAS = 61). Results obtained with the regional parameterization for the same three datasets were the only ones to pass performance criteria for sediment. The very low performance statistics for sediment simulation with the regional model were associated primarily with datasets that had low sediment loss and did not have successful model simulation as part of the calibration process. For total P, the regional model was capable of

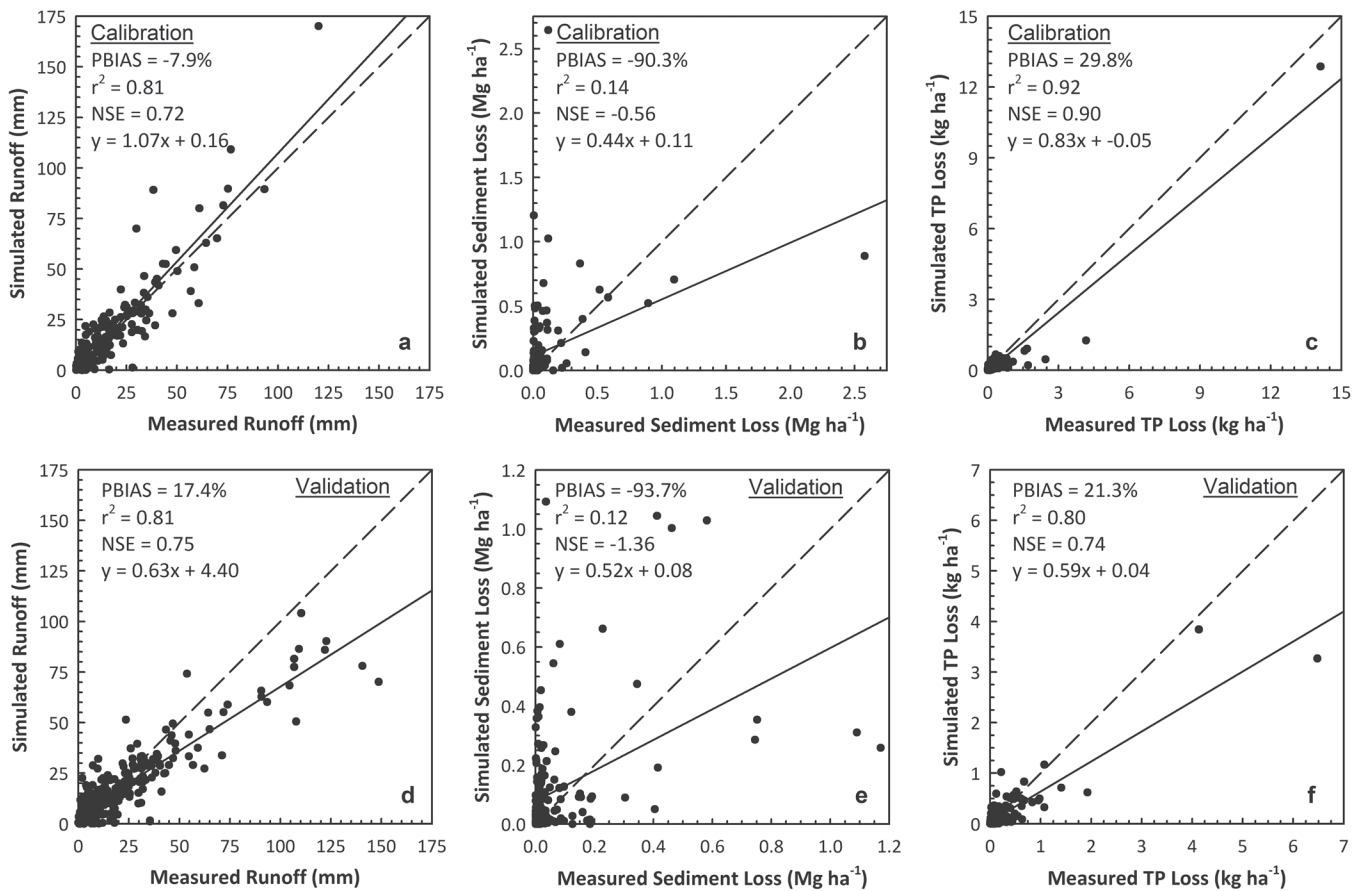


Fig. 1. Event-based runoff, sediment loss, and total phosphorus (TP) loss simulated with the regionally calibrated model (RCM) compared with measured data for the (a, b, and c) calibration and (d, e, and f) validation datasets with the percent bias (PBIAS), coefficient of determination (r^2), Nash–Sutcliffe Model efficiency (NSE), and regression equation for each dataset (dashed lines represent the 1:1 line and solid lines are the linear regression lines).

Table 1. Characteristics of model performance indicators for simulated event runoff, sediment loss, and total P loss with the Agricultural Policy Environmental eXtender (APEX) model using a regionally calibrated model (RCM) and site-specific calibrated models (SSCM) for 12 datasets.

	PBIAS †		r^2		NSE†	
	RCM	SSCM	RCM	SSCM	RCM	SSCM
	%					
Performance indicators for runoff						
Mean	21	18	0.82	0.81	0.69	0.73
Median	16	14	0.82	0.80	0.72	0.72
Minimum	3	3	0.72	0.69	0.43	0.63
Maximum	63	36	0.90	0.91	0.85	0.87
Standard deviation	17	11	0.06	0.06	0.13	0.08
# Sites within PEC†	11	11	12	12	12	12
Performance indicators sediment loss						
Mean	257	36	0.39	0.43	-44	0.24
Median	115	35	0.34	0.35	-2	0.31
Minimum	7	9	0.14	0.25	-394	-0.26
Maximum	1094	85	0.80	0.80	0.57	0.51
Standard deviation	349	22	0.21	0.18	114	0.23
# Sites within PEC	5	10	4	3	5	7
Performance indicators total P loss						
Mean	11	22	0.72	0.80	0.50	0.68
Median	10	15	0.64	0.79	0.52	0.66
Minimum	-55	1	0.51	0.61	0.02	0.48
Maximum	66	59	0.99	0.99	0.94	0.96
Standard deviation	41	20	0.17	0.13	0.24	0.15
# Sites within PEC	12	12	12	12	10	12

† PBIAS, percent bias; NSE, Nash–Sutcliffe efficiency; PEC, performance evaluation criteria as $r^2 > 0.5$, $NSE > 0.3$, and $|PBIAS| < 35, 60, \text{ and } 70$ for runoff, sediment loss, and total P loss, respectively.

meeting PECs for 10 of 12 datasets but the model performance for those datasets declined relative to the fully calibrated models.

Normal probability plots for PBIAS, NSE, and r^2 confirmed that performance statistics for runoff prediction by the RCM and SSCM are likely from the same population, with the exception of NSE values from two datasets (Fig. 2). Model performances should not be expected to be similar for multiple datasets. Rather, model performance varies from dataset to dataset based on random variability for model inputs (i.e., distribution of soil properties or watershed characteristics) and variability in measured data. Therefore, we assumed that performance statistics obtained with datasets for which the model adequately simulated measured data would be normally distributed. Furthermore, model performance statistics obtained with datasets for which the model failed to simulate measured data would not be within the normal distributions of performance statistics. In contrast to normal probability plots for runoff, the normal probability plots for total P performance statistics for the RCM (Fig. 2) indicated a steeper slope for PBIAS and a lower median value for r^2 and NSE (Table 1) compared with SSCMs, all indicative that the RCM, while meeting performance criteria, performed more poorly than the SSCMs.

Figure 3 shows measured versus simulated annualized values of runoff, sediment, and total P using the RCM when all of the site-years are analyzed together. The PEC values were acceptable for runoff and total P but not for sediment, where NSE

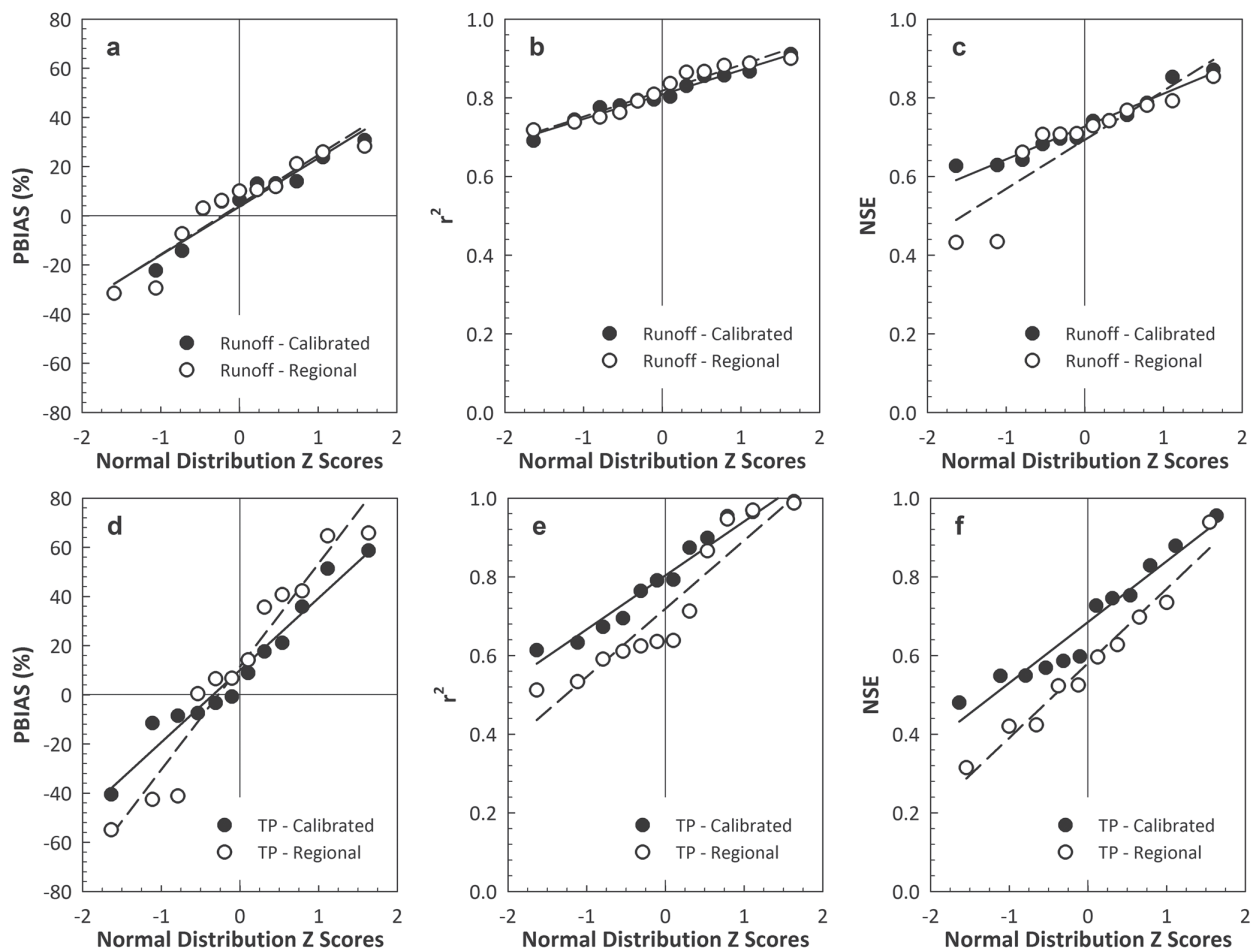


Fig. 2. Normal probability plots of (a and d) percent bias (PBIAS), (b and e) coefficient of determination (r^2), and (c and f) Nash-Sutcliffe Model efficiency (NSE) for (a, b, and c) runoff and (d, e, and f) total P (TP) simulation with site-specific calibrated models (SSCM) and the regionally calibrated model (RCM) for 12 datasets.

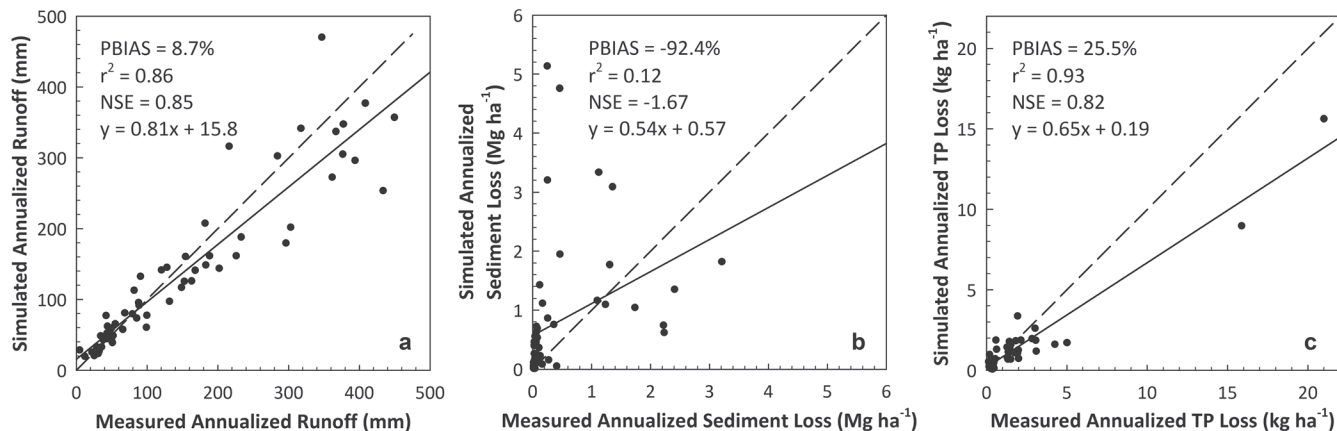


Fig. 3. Simulated versus measured annualized (a) runoff, (b) sediment, and (c) total P (TP) loss for 61 site years across 12 watersheds for the regionally calibrated model (RCM). Reported model performance evaluation criteria include percent bias (PBIAS), coefficient of determination (r^2), and Nash-Sutcliffe efficiency coefficient (NSE). The dashed line represents the 1:1 relationship and solid lines are linear regression lines.

was negative, indicating that the mean was a better estimate than the RCM estimates.

Discussion

Success of the RCM to estimate event runoff from multiple watersheds with nearly equal results as SSCM parameterizations (Table 1) supports the use of the RCM to estimate runoff from diverse managements and weather scenarios on restricted-layer

soils represented by the RCM. The RCM was developed and tested on datasets that included four locations and management practices that spanned a wide range of tillage, conservation practices, and fertilizer management strategies in grain production systems (Supplemental Table S1). The RCM performed significantly better than the uncalibrated BPJ parameterization developed for and tested on these same datasets (Baffaut et al., 2017). The RCM parameterization successfully simulated runoff for

more datasets (11 vs. 8) than the BPJ. Furthermore, the median NSE and r^2 values obtained with the RCM were greater than those obtained with the BPJ model for the datasets where the BPJ model met PEC, indicating improved performance of the RCM over the BPJ model. This outcome endorses the concept that the RCM can be used to estimate runoff volumes for new scenarios and emphasizes the importance of model calibration with water quality data from similar soils, although the extent of the applicability of the RCM beyond restricted-layer soils used for its development is unclear.

Model parameters sensitive for runoff included P15, P42, and P44 (see Supplemental Table S3 for definitions). Development of the RCM ensured that these sensitive parameters were adjusted within an appropriate range for restricted-layer soils. For example, the value of P42 selected based on best professional judgement was 1.0 (Baffaut et al., 2017), which is less than the value of all SSCMs, whereas the value of P42 was 2.3 for the RCM, which was well within the range of values selected during calibration of the SSCMs.

The RCM was particularly poor at simulation of sediment loss for the Knox B and Crawford datasets (as indicated by performance statistics outside the PEC), where conservation practices and site characteristics resulted in small sediment loads for the event-based calibration datasets (Fig. 4). Small sediment loss values relative to the variability in measured data provide little information for calibrating parameters that affect sediment loss for these datasets and can inflate statistics, like PBIAS, that are proportional to the range of the data (Bhandari et al., 2017). While model performance statistics for sediment were generally poor for the SSCMs (Table 1), they generally captured the scale of loss for each dataset (Fig. 4). In contrast, the RCM both performed poorly with performance statistics (Table 1) and frequently overestimated sediment loss (Fig. 4). The most dramatic example of this was simulation of the Crawford dataset, where the median value of the RCM was eight times greater than measured data and the RCM overestimated the maximum event by over two orders of magnitude. Clearly, information from the datasets where sediment loss was successfully calibrated did not provide the RCM the capacity to estimate losses from low sediment-loss scenarios. Model factors beyond calibration could contribute to the poor simulation of sediment loss, such as the use of assumed rainfall distributions that differ from actual rainfall intensity or the use of a Universal Soil Loss Equation-based approach for erosion estimates, as opposed to more process-based approaches, like that used by the Water Erosion Prediction Project (WEPP) model (Ascough et al., 1997).

With total P, the RCM typically was successful at capturing local watershed total P loss (Table 1), but unlike runoff, there was clear evidence that the RCM performed less effectively than SSCMs for those 12 datasets (Table 1, Fig. 2). The variation of performance statistics increased with the RCM. This implies that, in contrast to runoff, the total P loss simulated by the RCM is not functionally equivalent to that simulated by the SSCMs. This is not unexpected, given the inability of the RCM to appropriately estimate sediment loss (see discussion above). Sediment loss frequently plays an important role in driving total P loss from agricultural fields (Sharpley et al., 1994; Eghball and Gilley, 2001). In the measured event datasets, there was a weak correlation between sediment loss and total P ($r^2 = 0.10$, data not shown). This poor correlation was driven by total P losses observed in datasets such as Crawford, where sediment

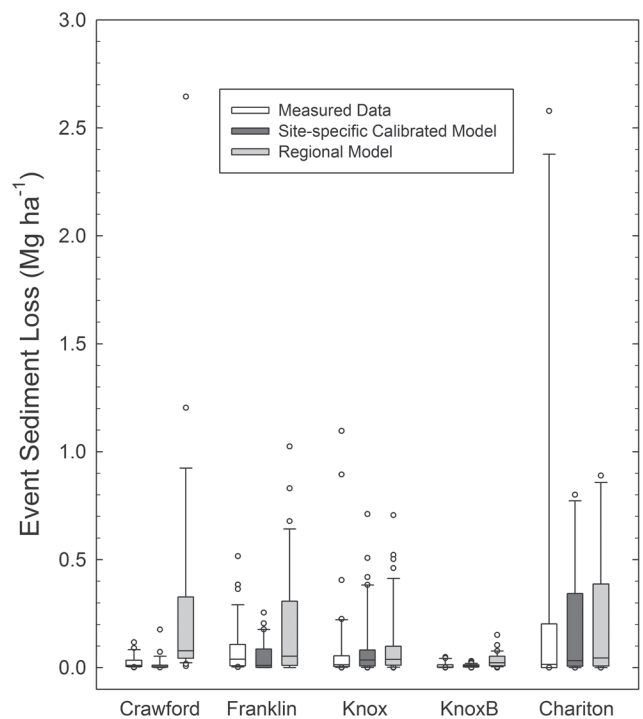


Fig. 4. Box plot of event sediment loss data comparing measured values and estimated values using the Agricultural Policy Environmental eXtender (APEX) model with site-specific calibrated models (SSCM) and the regionally calibrated model (RCM). Boxes represent the 25% quartile (bottom), the median (inner line), and the 75% quartile. Top and bottom whiskers are the 90th and 10th percentile, respectively, with outliers indicated as points above and below whiskers.

loss was controlled and management practices (e.g., applied poultry litter) contributed to high dissolved P loss. The SSCMs were able to account for the differences in the dissolved P:total P ratio resulting from different management practices used for the measured datasets and therefore maintained a weak correlation between sediment and total P loss ($r^2 = 0.04$), but the RCM resulted in a stronger relationship between sediment and total P for all event data ($r^2 = 0.52$). This highlights that the RCM is not appropriately capturing the mechanisms driving total P loss for these diverse scenarios. In agreement with this observation, the biggest drops in total P performance statistics, when comparing SSCMs and RCM for specific scenarios, were associated with simulations for datasets that had low sediment loss (Knox B and Crawford) and where the RCM consistently overestimated sediment loss. The RCM was likely successful at simulating total P loss for these datasets despite the poor simulation of sediment loss because dissolved P loss contributed to a high proportion of total P loss from these sites. For example, the measured dissolved P loss for the Franklin and Crawford datasets, the only two with measured dissolved P loss, was 44 and 92% of total P loss, respectively (Zeimen et al., 2006; Sweeney et al., 2012).

Annualized losses of total P met PEC (Fig. 3), but the RCM underestimated large total P losses. A similar assessment of results from the SSCM by Baffaut et al. (2017) had a slope of 0.87 and better performance indicators (PBIAS = 13%, $r^2 = 0.97$, NSE = 0.96). This acceptable performance of the RCM suggests it may be able to provide annualized assessments across multiple studies. This conclusion is tempered by concerns expressed about the functionality of the RCM to represent total P-loss processes.

Some of the weakness of the RCM in addressing sediment loss may be a consequence of how we developed the RCM. Our approach was to develop a “best” site-specific calibrated and validated model using distinct datasets. We then identified values for sensitive parameters. In this process, we emphasized sensitive parameters for sediment simulation from models that were successfully calibrated for sediment loss. The justification for this approach was that the models that failed calibration for sediment were calibrated with datasets that had very low sediment loss, which did not provide relevant information about settings for parameters sensitive to sediment loss and consequently had poor performance statistics for sediment loss. Looking to the future, there may be value in developing methods that would allow calibrating on multiple datasets simultaneously. This would offer the opportunity to benefit from datasets with an insufficient range to calibrate at that location. Wang et al. (2008) observed that simultaneously calibrating APEX for different tillage treatments resulted in a calibration that could handle both situations. Simultaneously calibrating a model for five datasets will require expanding APEX automated calibration tools and engaging enhanced computing resources to handle the numerous permutations needed in this computationally intensive approach.

Failure of the RCM to estimate sediment loss may be attributed to limitations of our capacity to calibrate the model for this critical water quality component. However, our results may also reflect weaknesses in the algorithms used to estimate sediment loss. One benefit of more robust calibration procedures that identify the “best” calibration or calibrations for multiple scenarios is that it will highlight which routines of the model are not capable of accounting for alternative locations, climate, and/or management scenarios.

Other research reported in this special edition established the importance of meeting calibration and validation criteria of APEX as a critical step in having confidence in the model to estimate runoff, sediment, and total P losses for climate, management practices, and/or similar locations beyond the calibration and validation set (Bhandari et al., 2017). These studies implied that calibration at one location was most reliable for estimating other situations when the model had successfully been calibrated for multiple water quality components, e.g., runoff, sediment, and total P. In this project, we have successfully integrated calibration information from multiple locations to provide a regional calibration that provides robust estimates of runoff from multiple locations. This success is with a regional parameterization that is unsuccessful at estimating sediment loss and has clear limits on the capacity to estimate total P losses. Although the RCM met the performance criteria for simulation of total P loss, the application of this model for estimating total P loss should be limited to soils and management systems with similarly low sediment loss.

Conclusions

The goal of environmental mechanistic models such as APEX is to provide accurate estimates of water quality parameters for multiple locations, climates, and scenarios. Our hypothesis was that multilocation calibration could establish values for key sensitive parameters that allow model algorithms to account for diverse scenarios and estimate runoff, sediment, and total P losses needed to test tools such as a P Index.

In support of this concept, we successfully created a regional calibration of APEX for restricted layer soils that provided estimates of runoff from multiple locations under a wide range of scenarios. The RCM, while successful at meeting PECs for total P for most of the test watersheds, clearly was not accounting for sediment transport processes at some of the locations, a key component of total P loss. Consequently, we did not attain our goal of a robust regional parameterization of APEX for estimating sediment and total P losses.

Future work will focus on calibration techniques that allow simultaneous calibration of sensitive parameters across locations. These approaches will require developing more sophisticated calibration optimization software for APEX. Our results confirm the potential of models as a tool to expand the impact of measured water quality data to understand scenarios and climate situations beyond measured data. However, our results also emphasize that calibration with measured water quality remains a key element of model parameterization and that the capacity of the current construct of APEX to effectively capture sediment and total P-loss processes in small agricultural watersheds is still unclear.

References

- Ascough, J.C., C. Baffaut, M.A. Nearing, and B.Y. Liu. 1997. The WEPP watershed model: I. Hydrology and erosion. *Trans. ASAE* 40:921–933. doi:10.13031/2013.21343
- Baffaut, C., C. Bolster, N. Nelson, M.W. Van Liew, J. G. Arnold, and J.R. Williams. 2013. Session B2: Model development. SWATing your APEX model: A how-to from the trenches. In: *Proceedings of the 2013 International SWAT Conference*, Toulouse, France. 17–19 July 2013. Texas Water Resources Inst. Tech. Rep. TR-471. Texas A&M University, College Station, TX. p. 86–97.
- Baffaut, C., N.O. Nelson, J.A. Lory, G.M.M.A. Senaviratne, A.B. Bhandari, R.P. Udawatta et al. 2017. Multisite evaluation of APEX for water quality: I. Best professional judgement parameterization. *J. Environ. Qual.* doi:10.2134/jeq2016.06.0226 (in press).
- 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).
- Eghball, B., and J. Gilley. 2001. Phosphorus risk assessment index evaluation using runoff measurements. *J. Soil Water Conserv.* 56:202–206.
- 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
- 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
- Jones, C., C. Cole, A.N. Sharpley, and J. Williams. 1984. A simplified soil and plant phosphorus model: I. Documentation. *Soil Sci. Soc. Am. J.* 48:800–805. doi:10.2136/sssaj1984.03615995004800040020x
- 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
- 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 et al. 2006. Comparing ratings of the southern phosphorus indices. *J. Soil Water Conserv.* 61:325–337.
- Saleh, A., O. Gallego, E. Osei, H. Lal, C. Gross, S. McKinney, and H. Cover. 2011. Nutrient tracking tool—a user-friendly tool for calculating nutrient reductions for water quality trading. *J. Soil Water Conserv.* 66:400–410. doi:10.2489/jswc.66.6.400
- Senaviratne, G.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.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
- Senaviratne, G.M.M.A., R.P. Udawatta, K.A. Nelson, K. Shannon, and S. Jose. 2012. Temporal and spatial influence of perennial upland buffers on corn and soybean yields. *Agron. J.* 104:1356–1362. doi:10.2134/agronj2012.0081
- Sharpley, A.N., D.B. Beegle, C. Bolster, L. Good, B.C. Joern, Q.M. 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* 44:163–179. doi:10.1007/s13280-014-0612-x
- Sharpley, A.N., C. Bolster, C. Conover, E. Dayton, J. Davis, Z. Easton et al. 2013. Technical guidance for assessing phosphorus indices. Rep. Southern Coop. Ser. Bull. 417. Virginia Tech University, Blacksburg, VA.
- 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.N., C.A. Jones, C. Gray, and C.V. Cole. 1984. A simplified soil and plant phosphorus model: II. Prediction of labile, organic, and sorbed phosphorus. *Soil Sci. Soc. Am. J.* 48:805–809. doi:10.2136/sssaj1984.03615995004800040021x
- Steglich, E.M., and J.R. Williams. 2013. Agricultural policy/environmental eXtender model user's manual- version 0806. Blackland Research and Extension Center, Temple, TX.
- 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
- Udawatta, R.P., J.J. Krstansky, G.S. Henderson, and H.E. Garrett. 2002. Agroforestry practices, runoff, and nutrient loss: A paired watershed comparison. *J. Environ. Qual.* 31:1214–1225. doi:10.2134/jeq2002.1214
- USEPA. 2016. National summary of state information. USEPA, Office of Wetlands, Oceans, and Watersheds. https://ofmpub.epa.gov/waters10/attains_nation_cy.control (accessed 24 June 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., 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
- White, M.W., R.D. Harmel, and R.L. Haney. 2012. Development and validation of the Texas best management practice evaluation tool (TBET). *J. Soil Water Conserv.* 67:525–535. doi:10.2489/jswc.67.6.525
- 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.