Evaluation of the TBET Model for Potential Improvement of Southern P Indices

Adam Forsberg, David E. Radcliffe,* Carl H. Bolster, Aaron Mittelstet, Daniel E. Storm, and Deanna Osmond

Abstract

Due to a shortage of available phosphorus (P)-loss datasets, simulated data from an accurate quantitative P transport model could be used to evaluate a P Index. The objective of this study was to compare predictions from the Texas Best Management Practice Evaluation Tool (TBET) against measured P-loss data to determine whether the model could be used to improve P Indices in the southern region. Measured P-loss data from field-scale study sites in Arkansas, Georgia, and North Carolina were used to assess the accuracy of TBET for predicting field-scale loss of P. We found that event-based predictions using an uncalibrated model were generally poor. Calibration improved runoff predictions and produced scatterplot regression lines that had slopes near one and intercepts near zero. However, TBET predictions of runoff met the performance criteria (Nash–Sutcliffe efficiency $\geq 0.3$, percent bias $\leq 35\%$, and mean absolute error $\leq 10\%$) in only one out of six comparisons: North Carolina during calibration. Sediment predictions were imprecise, and dissolved P predictions underestimated measured losses. In North Carolina, total P-loss predictions were reasonably accurate because TBET did a slightly better job of predicting sediment losses from cultivated land. In Arkansas and Georgia, where the experimental sites were in forage production, the underprediction of dissolved P led directly to the underpredictions of total P. We conclude that TBET cannot be used to improve southern P Indices, but a curve number approach could be incorporated into P Indices to improve runoff predictions.

Core Ideas

- Predictions from an accurate P transport model could be used to evaluate a P Index.
- Predictions from an uncalibrated TBET model were generally poor.
- A calibrated TBET model was better but did not meet the performance criteria.
- A curve number approach for runoff could be incorporated into P Indices.

Given the limited amount of data available for evaluating and updating Phosphorus (P) Indices, the Southern Extension-Research Activity Group 17 (SERA-17) recommended that simulated output from P transport models be used as a substitute, provided the models have been shown to provide accurate estimates of P loss at the field scale (Sharpley et al., 2012). This approach has been successfully applied in multiple studies (Schoumans et al., 2002; Veith et al., 2005; Leone et al., 2006; Bolster et al., 2012). Bolster (2011) compared simulated P-loss data from the Annual Phosphorus Loss Estimator (APLE) against output from the Kentucky P Index. This analysis resulted in significant changes in the Kentucky P Index. Similarly, Bolster et al. (2012) used APLE to evaluate the Pennsylvania P Index. They demonstrated that APLE simulations could be used to derive more accurate P Index weighting factors and noted that correlating P Index ratings with quantitative P-loss model output can provide valuable estimates of uncertainty in the P Index.

The Texas Best Management Practice Evaluation Tool (TBET) was developed by White et al. (2010, 2012, 2014) for conservation planners that needed a simple yet accurate tool to predict sediment and nutrient losses from agricultural fields. It is a simplified, field-scale version of the Soil and Water Assessment Tool (SWAT; Arnold et al., 2012).

The objective of this study was to evaluate the quantitative performance of TBET against event-based measured P-loss data to evaluate the accuracy of the model and thus determine whether the model could be used to improve P Indices for states in the southern region of the United States. This evaluation of TBET included assessment of the uncalibrated model, how calibration improved model performance, and how well the calibrated model performed on a separate validation dataset.

Abbreviations: ADJ_PKR, peak-rate adjustment factor; APLE, Annual Phosphorus Loss Estimator; $C_{\text{m}}$, minimum cropping factor; CN, moisture condition value; HRU, hydrologic response unit; K, erosion factor value; KTR, Kendall–Theil robust line method; MAE, mean absolute error; NSE, Nash–Sutcliffe efficiency; PBIAS, percent bias; PHOS5K, phosphorus partitioning coefficient; PPERCO, phosphorus percolation coefficient; PPM Plus, Pasture Phosphorus Management Plus; SLSBBSN, subbasin slope length; SWAT, Soil and Water Assessment Tool; TBET, Texas Best Management Practice Evaluation Tool; USLE, Universal Soil Loss Equation.
Materials and Methods

Study Sites

The study sites were in a humid, temperate climate, and average annual rainfall among the sites ranged from 80 to 157 cm (Supplemental Table S1). Each study site consisted of a unique crop system and soil. Supplemental Table S2 provides complete site descriptions and sample collection and analysis methods at each site. The climate, soils, management schedules, and measured field data from Georgia and North Carolina were obtained through publications and their authors (Pierson et al., 2001a, 2001b; Larsen et al., 2014; Edgell et al., 2015). The Arkansas data were supplied directly from Andrew Sharpley (sharpley@uark.edu).

TBET Model Description

The TBET model is an extensively simplified graphical user interface for a modified version of SWAT. Unlike SWAT, which allows a subbasin to contain many unique hydrologic response units (HRUs), TBET only permits simulation of seven predefined HRUs. Thus, TBET does not require a geographic information system interface. The minimum inputs required to run TBET include field area, slope, and distance to stream, soil series (maximum of three), daily weather (precipitation and temperature), and soil-test P in terms of Mehlich-3 extraction. Unique combinations of soil, slope, and land use become HRUs. The standard output produced by TBET includes annual totals of runoff, sediment, total nitrogen (N), total P, and dissolved P. Complete descriptions of TBET can be found in White et al. (2010, 2012).

White et al. (2012) calibrated and validated TBET for Texas and Oklahoma. Part of the process included calibrating basin-scale parameters. These parameters included the soil evaporation compensation factor, soil available water in the surface layer, deep aquifer percolation fraction, threshold depth of water in the surface layer, deep aquifer percolation fraction, threshold depth of water in the shallow aquifer for return flow, effective hydraulic conductivity of the main channel bed, and the coefficient for water movement from the shallow aquifer to the unsaturated zone (White et al., 2012). This was done by simulating flow in 20 basins in Texas, ranging in size from 550 to 6500 km². The calibrated basin-scale parameters were then fixed to reduce the number of parameters that needed to be adjusted to only field-scale parameters.

Currently, the soils, climate, and crop system databases for TBET mainly include information for the Texas State Soil and Water Conservation Board regions. Thus, most soils, weather, and management input files for this study were prepared manually by accessing and modifying the text input files in the TBET file directory.

Model Evaluation Process

The process of evaluating TBET as a quantitative P transport model included uncalibrated, calibration, and validation runs. We used event-based data as the best way to determine if the model processes were correct. The time periods used in the calibration datasets were the first 2 yr in Arkansas and North Carolina, where there were three total years of data available (Table 1). The third year was used for validation, which was a year of higher rainfall so that the runoff was about the same in the calibration and validation periods. In Georgia, where there were four total years available, years were chosen so that there was a year each of low and high rainfall in both the calibration and validation datasets. Model simulations were run for a single calendar year, and each run incorporated a 2-yr model warmup to initialize soil-moisture and nutrient levels. The decision to use a 2-yr warmup period was determined by making 10-yr, annual time-step model runs and examining the model output to see in which year the output stabilized.

Statistical Analyses

Model performance was evaluated by comparing measured runoff, sediment yield, and P load to event-based model predictions. Scatterplots and regression supplied an initial visual comparison of simulated and measured data, while the 95% confidence interval and prediction interval of the regression slope provided a visual measure of uncertainty. The confidence interval shows the likelihood that the true population mean is within the model limits. The prediction interval shows the likelihood that a single data point is within the model limits and is more relevant for this study (Helsel and Hirsch, 1992). For the sake of completeness, we show both. Regression was used in conjunction with statistical measures of goodness-of-fit for model performance evaluation.

Given that the regression residuals did not meet the requirements for normality and homoscedasticity (even with transformations), nonparametric regression in the form of the Kendall–Theil robust line method (KTR) was used (Helsel and Hirsch, 1992). The KTR slope estimate is calculated as the median slope of all possible pairwise slopes for each pair of points in the dataset. The intercept is calculated as the median of all possible intercepts computed by solving the KTR using the median slope and each data point. The KTR estimates for this study were calculated using the R package “mbml” version 0.12 (Komsta, 2013).

Three quantitative goodness-of-fit indicators were used: Nash–Sutcliffe efficiency (NSE), mean absolute error (MAE), and percent bias (PBIAS) (Abbaspour, 2013). The calculation of these statistics was performed using the R package “hydroGOF” version 0.3-8 (Zambrano-Bigiarini, 2014). The equations for each statistic are shown in the Supplemental Material [Eq. S1–S3]. The statistical criteria used to assess whether model runs were satisfactory were similar to those used by Nelson et al. (2017) to evaluate the Agricultural Policy Environmental eXtender (APEX) model: NSE ≥ 0.30 and |PBIAS| ≤ 35, 60, and 70% for runoff, sediment, and P. Nelson et al. (2017) included $R^2 ≥ 0.50$ as criteria, but we used MAE ≤ 10 mm, 0.5 Mg ha⁻¹, and 0.5 kg ha⁻¹ for runoff, sediment, and P instead. The values for NSE and PBIAS are more lenient than those recommended by Moriasi et al. (2015), but they indicate that the model would be directionally correct for estimating the losses required for a P Index. We also used the scatterplots as an essential, visual guide of the model performance.

Uncalibrated Performance

Performance of TBET on sites in Arkansas, Georgia, and North Carolina was first evaluated using the TBET parameter values chosen by White et al. (2012) for sites in Texas and Oklahoma. The uncalibrated dataset is shown in Table 1. An analysis of the contribution of P versus depth of runoff generated during an event showed that both measured and predicted
storms resulting in <1 mm of runoff contributed only 0.3 to 1.4% of total P from all storms. Subsequently, only simulated or observed events with runoff ≥1 mm were used in the uncalibrated and calibrated datasets.

**Model Calibration and Validation**

Each state was calibrated separately. A manual calibration using event-based measured and simulated data was performed by individually calibrating each constituent in the following order: runoff, sediment, total P, and dissolved P. Each model run for calibration followed the same procedure as the uncalibrated process with single-year runs, 2-yr warmup, and event-based comparisons.

The calibration parameters and ranges are listed in the Supplemental Table S3. Calibration of runoff used nine curve number moisture condition II values (CNwere ±16 points from the default). Sediment was calibrated using seven erosion factor values (K, ±60% of the default), seven minimum cropping factors (Cmin, ±60% of the default), seven peak-rate adjustment factors (ADJ_PKR, range 0.25–1.75), and seven slope length values (SLSBBSN, ±60% of the default). A two-step process was used to calibrate the sediment and sediment-P parameters, again due to the use of a manual calibration. First, Cmin factor and SLSBBSN were calibrated in a full factorial 7 × 7 analysis. The best-fit parameter values for Cmin and SLSBBSN were chosen and were thereafter used in the 7 × 7 factorial calibration of Universal Soil Loss Equation (USLE) K and ADJ_PKR. The calibration of dissolved P analyzed 5 P percolation coefficients (PPERCO, ranging between 5 and 15) and 10 P partitioning coefficients (PHOSKD, ranging between 100 and 200) values. The calibration resulted in a total of 132 parameter alterations (9 runoff, 98 sediment or sediment P, and 25 dissolved P), for a total of 4752 model simulations combined.

The calibration and validation datasets are shown in Table 1.

**Results and Discussion**

**Predictions for Uncalibrated Model**

Predictions using the uncalibrated model with the parameter set for TBET chosen by White et al. (2012) for Texas and Oklahoma were unsuccessful in most cases, meeting all three performance criteria only in Georgia for total P (Table 2). Supplemental Fig. S1, S2, S3, and S4 contain the scatterplots and KTR regression statistics for the uncalibrated performance for each site, organized by constituent. The TBET model was more accurate in predicting runoff with NSE = −7.00, 0.57, and 0.23 for Arkansas, Georgia, and North Carolina, respectively, compared with sediment (NSE = −59 to −88), soluble P (NSE = −1.6 to 0.19), and total P (NSE = −78 to 0.34) (Table 2). Despite the low NSE, Supplemental Fig. S1 shows that the regression line for runoff in North Carolina was close to the 1:1 line (slope = 1.01, intercept = 2.6).

Given that sediment losses at the Georgia site (pastures) were low, sediment losses were not measured. Measurements on the same plots in 2005 through 2007 averaged 0.084 Mg ha−1. The TBET model predicted low sediment loss for the Georgia site with annual predictions between 0.05 and 0.15 Mg ha−1 yr−1 and for the Arkansas site (pastures and hay fields) with annual predictions between 0.02 and 1.26 Mg ha−1 yr−1. The tool predicted much higher sediment losses for North Carolina (row crops), with annual predictions between 2.29 and 105.31 Mg ha−1 yr−1.

In the Georgia pastures, where most of the measured P loss was dissolved P, TBET underpredicted dissolved P but met the criteria for total P (Table 2). In Arkansas and North Carolina, TBET overpredicted total P and dissolved P. The overprediction in Arkansas, where most of the measured P was also dissolved P, was due in part to the overprediction of runoff. The overprediction of total P in the North Carolina row crops was due in part to the overprediction of sediment (Supplemental Fig. S1 and S3).

**Runoff Predictions for Calibration and Validation Periods**

Using the calibrated model, results for daily runoff were satisfactory in terms of NSE during the calibration and validation periods, except for Arkansas during the calibration period (Table 2). The scatterplots for runoff (Fig. 1) show regression lines with slopes near one and intercepts near zero. The prediction intervals were narrow (<50 mm) and the MAE values were usually <10 mm (Table 2).

Although the calibrated runoff results in Arkansas were unsatisfactory (MAE = 13.82 mm, NSE = −2, PBIAS = −112%) for the calibration period, they were an improvement over the uncalibrated model (MAE = 20.54 mm, NSE = −7, PBIAS = −179%) (Table 2). The poorer performance of runoff in Arkansas during the calibration period may have been due to the lack of site-specific rainfall input data. As noted in the Supplemental Material, the Arkansas dataset lacked complete weather data for the period of investigation; thus, existing weather data were supplemented with observations from three nearby weather stations.

The overprediction of runoff in North Carolina was due in part to the inability of the model to predict runoff in the no-till treatments (in two of the five fields). Anand et al. (2007) found that SWAT predicted field-scale runoff poorly on no-till plots with default parameters. Bonta and Shipitalo (2013) and Endale et al. (2011) both found CN values representing

### Table 1. Site-years and number of observations for the uncalibrated, calibrated, and validated model datasets.

<table>
<thead>
<tr>
<th>Site</th>
<th>No. fields</th>
<th>Years</th>
<th>Site-years</th>
<th>Runoff</th>
<th>Sediment</th>
<th>Total P</th>
<th>Dissolved P</th>
</tr>
</thead>
<tbody>
<tr>
<td>Uncalibrated, calibration</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Washington Co., AR</td>
<td>7</td>
<td>2009, 2010</td>
<td>14</td>
<td>42</td>
<td>41</td>
<td>41</td>
<td>41</td>
</tr>
<tr>
<td>Henderson Co., NC</td>
<td>5</td>
<td>2011, 2012</td>
<td>10</td>
<td>139</td>
<td>127</td>
<td>131</td>
<td>107</td>
</tr>
<tr>
<td>Validation</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Washington Co., AR</td>
<td>7</td>
<td>2011</td>
<td>7</td>
<td>48</td>
<td>48</td>
<td>48</td>
<td>48</td>
</tr>
<tr>
<td>Putnam Co., GA</td>
<td>6</td>
<td>1996, 1997</td>
<td>12</td>
<td>187</td>
<td>0</td>
<td>165</td>
<td>187</td>
</tr>
<tr>
<td>Henderson Co., NC</td>
<td>5</td>
<td>2013</td>
<td>5</td>
<td>245</td>
<td>134</td>
<td>160</td>
<td>152</td>
</tr>
</tbody>
</table>
Table 2. Uncalibrated, calibrated, and validated goodness-of-fit statistics for the Texas Best Management Practice Evaluation Tool (TBET) model predictions on field sites in Arkansas, Georgia, and North Carolina. Model predictions that met the performance criteria for a satisfactory prediction are shown in bold.

<table>
<thead>
<tr>
<th>Statistic†</th>
<th>Uncalibrated</th>
<th>Calibration</th>
<th>Validation</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>AR</td>
<td>GA</td>
<td>NC</td>
</tr>
<tr>
<td>Runoff</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>MAE, mm</td>
<td>20.54</td>
<td>9.74</td>
<td>6.72</td>
</tr>
<tr>
<td>NSE</td>
<td>−7</td>
<td>0.57</td>
<td>0.23</td>
</tr>
<tr>
<td>PBIAS, %</td>
<td>−179</td>
<td>−48</td>
<td>−39</td>
</tr>
<tr>
<td>Sediment</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>MAE, Mg ha$^{-1}$</td>
<td>0.02</td>
<td>–</td>
<td>0.53</td>
</tr>
<tr>
<td>NSE</td>
<td>−59</td>
<td>–</td>
<td>−88</td>
</tr>
<tr>
<td>PBIAS, %</td>
<td>−256</td>
<td>–</td>
<td>−963</td>
</tr>
<tr>
<td>Total P</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>MAE, kg P ha$^{-1}$</td>
<td>0.17</td>
<td>0.35</td>
<td>0.89</td>
</tr>
<tr>
<td>NSE</td>
<td>−1.5</td>
<td>0.34</td>
<td>−78</td>
</tr>
<tr>
<td>PBIAS, %</td>
<td>−55</td>
<td>20</td>
<td>−654</td>
</tr>
<tr>
<td>Dissolved P</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>MAE, kg P ha$^{-1}$</td>
<td>0.13</td>
<td>0.32</td>
<td>0.01</td>
</tr>
<tr>
<td>NSE</td>
<td>−1.6</td>
<td>0.19</td>
<td>−0.08</td>
</tr>
<tr>
<td>PBIAS, %</td>
<td>−35</td>
<td>11</td>
<td>−21</td>
</tr>
</tbody>
</table>

† MAE, mean absolute error; NSE, Nash–Sutcliffe efficiency; PBIAS, percent bias.

Fig. 1. Kendall–Theil robust line regression statistics and scatterplots for event-based measured and Texas Best Management Practice Evaluation Tool predicted runoff using the validation datasets and calibration best-fit parameters.
long-term, no-till management to be 16 units smaller than standard handbook table values.

Sediment Predictions for Calibration and Validation Periods

Sediment predictions met the performance criteria in only one of the six comparisons (Fig. 2, Table 2). Since the uncalibrated model overpredicted sediment losses by a large margin, most of the parameter changes during calibration reduced sediment losses (Supplemental Table S4). The USLE $C_{\min}$ and USLE $K$ were reduced by 60% in Arkansas and North Carolina in the best-fit calibration. The ADJ_PKR was reduced by 25% in Arkansas and by 75% in North Carolina. The default slope length was calculated using the area of the field, which is an input to the model. No changes were made in Georgia because sediment losses were so low that they were not measured (other than a few times at the beginning of the study) and the predicted values were low. These changes reduced sediment losses and improved the NSE, but the result was a trend to underpredict sediment losses with positive PBIAS values during the calibration and validation periods (Table 2).

The measured sediment losses from the Arkansas site (forage production) during calibration and validation periods were low (0.01–0.17 Mg ha$^{-1}$ yr$^{-1}$) and the MAE values were also low (0.01 Mg ha$^{-1}$). In North Carolina, where the plots were row crops, measured sediment losses were higher (0.05–2.42 Mg ha$^{-1}$ yr$^{-1}$) and the calibrated TBET predictions were better (NSE $= -0.67$ for the calibration period, 0.31 for validation period), with relatively low MAE values (0.07 for calibration, 0.13 for validation). Our results seemed to show that the calibrated TBET model could not accurately predict low sediment losses but did a better job when measured losses were high (cultivated land instead of forage production fields). The inaccuracy with low sediment loss predictions may not be as important as it would be with high sediment losses. Also, the fields in North Carolina were plots with a much smaller area (0.017 ha) than the Arkansas (0.4 ha) and Georgia (0.72–0.79 ha) fields. In a review article, Boix-Fayos et al. (2006) cite several studies that found that small-plot erosion rates extrapolated to larger scales resulted in overestimation of erosion.

Dissolved P Predictions for Calibration and Validation Periods

Calibration improved the predictions of dissolved P, with satisfactory values for PBIAS and MAE in all cases and satisfactory values for NSE in Georgia during the calibration period (Table 2). In Arkansas during the validation period, the NSE was 0.27, just below the critical goodness-of-fit value. The best-fit P parameter values for Arkansas and North Carolina (Supplemental Table S5) were a combination of a 33% increase in the PHOSKD and a 50% decrease in the PPERCO. After calibration, the scatterplots (Fig. 3) still showed a consistent underprediction at high measured values for all states with positive PBIAS values (Table 2). The calibrated PHOSKD value of 200 m$^3$ Mg$^{-1}$ was close to the value of 242 m$^3$ Mg$^{-1}$ derived by Radcliffe et al. (2009b) using the ratio of labile P in the soil to dissolved P in runoff for typical Piedmont soils.

The underprediction of dissolved P was consistent with the findings of White et al. (2014) using Pasture Phosphorus Management Plus (PPM Plus; White et al., 2010), the precursor model to TBET. They found that PPM Plus predicted dissolved P more accurately on cultivated land than pasture. The systematic underestimation of dissolved P may be due to the lack of a specific P pool for manure application in TBET (TBET uses the SWAT 2005 executable file). The SWAT model assumes that surface-applied manure P that is not incorporated immediately becomes a fraction of the soil-P pool; however, when manure is surface-applied.
applied and not incorporated, P is readily transported via surface runoff (Vadas et al., 2007; Radcliffe et al., 2009a). It may also be because SWAT is seldom used to predict losses from pastures or hay fields, and thus its default parameter values may be more representative of large P loss from sites with cultivated land.

Total P Predictions for Calibration and Validation Periods

Total P predictions were satisfactory in only two comparisons: Georgia (MAE = 0.30 kg ha⁻¹, NSE = 0.51, PBIAS = 10%) during the calibration period and North Carolina (MAE = 0.22 kg ha⁻¹, NSE = 0.35, PBIAS = −16.9%) during the validation period (Table 2). In Arkansas during the validation period, the NSE value (0.27) was just below the critical value, and the values for PBIAS and MAE were satisfactory. Figure 4 shows a wide scatter in the North Carolina data, but the slope of the regression line was near one and the intercept was near zero. Predictions of total P loss are directly dependent on particulate-P predictions associated with prediction of sediment loss. In North Carolina, since there was a lot of scatter in the sediment-loss predictions (Fig. 2), this affected the total P loss. However, the total P-loss predictions in North Carolina were reasonably accurate because TBET did a slightly better job of predicting sediment losses when cultivated lands were the treatment, as noted before. The scatterplots for Arkansas and Georgia (Fig. 4) showed a consistent underprediction of total P at high values. In these states, where the experimental sites were in forage production fields, the imprecision in the TBET predictions for sediment loss had little effect, since sediment loss predictions were low. Thus, the underprediction in dissolved P led directly to the underpredictions in total P.

Using TBET to Improve Southern Region P Indices

We conclude that our calibrated TBET models cannot be used to improve southern P Indices at this time. A more extensive calibration effort might produce a TBET model that could be used. It would be important to include calibration of the basin-scale parameters in each state, something we did not do but was done by White et al. (2012) for Texas using stream flow in 20 basins. Calibrating and then setting up TBET for use in each state would be very time consuming. For each major region within a state, weather, soils, and management files for each common farming practice would have to be added. We estimate that it would take a graduate student 2 to 3 yr to calibrate and set up files for a TBET model.

Our study showed that TBET predictions of runoff were reasonable with a calibrated model. Unlike measured P-loss datasets, there are many datasets with measured runoff, so regional or
state-by-state calibration is feasible. The TBET model uses a curve number approach to predict runoff, and this can be implemented in a spreadsheet using long-term precipitation data. It may be possible to incorporate this approach into P Indices to estimate long-term annual runoff for different soils and management. Several states have done this, including Georgia and Kentucky (Bolster et al., 2014).

**Conclusions**

Using an uncalibrated model, TBET predictions for runoff, sediment, and P losses were generally poor and did not meet the performance criteria in most cases. After calibrating TBET for each site, we found that runoff predictions improved, with scatterplots that had regression line slopes near one and intercepts near zero. However, all of the performance criteria were met in only one out of the six comparisons: North Carolina during the calibration period. Calibration also improved sediment predictions, but the predictions again met the performance criteria only in North Carolina during the validation period. There was a consistent underprediction of dissolved P at the sites with forage production (Arkansas and Georgia), where most of the observed P loss was in the form of dissolved P. The calibrated TBET model seemed to perform better where sediment and particulate P losses were high (the cultivated land in North Carolina) compared with situations where sediment losses were low and most of the P was lost as dissolved P (forage production fields in Arkansas and Georgia).

**Acknowledgments**

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

**References**


Fig. 4. Kendall–Theil robust line regression statistics and scatterplots for event-based measured and Texas Best Management Practice Evaluation Tool predicted total phosphorus (P) using the validation datasets and calibration best-fit parameters.


