Assessing water quality changes in agricultural drainages: Examples from oxbow lake tributaries in northwestern Mississippi, United States, and simulation-based power analyses

J.C. Murphy, M. Hicks, and S. Stocks

Abstract: Hydrology (streamflow and stage) and water quality (suspended sediment, total nitrogen [N], ammonia [NH_3], total Kjeldahl N, nitrate plus nitrite [$NO_3 + NO_2$], and total phosphorus [TP]) were monitored in two small agricultural drainages in northwestern Mississippi to document changes in water quality that coincided with the implementation of best management practices (BMPs) in upstream drainages. Using an event-based data set and bootstrapping techniques, we tested for difference and equivalence in median event concentration and differences in concentration-streamflow (C–Q) relationships between an early and late period at each site, where most of the major BMP implementation occurred during the early period. Results for Bee Lake Tributary were inconclusive. Using 95% confidence intervals, none of the constituents were statistically different or equivalent for median event concentrations between the periods, indicating a lack of evidence to determine whether water quality had changed or stayed the same, and only TP had a significantly higher C–Q intercept during the late period. At Lake Washington Tributary, more than half the constituents had a significantly different median, slope, or intercept between periods, indicating a 35% or more decrease in event concentration following a period of intense BMP implementation. These mixed results could be due to a variety of differences between the sites including the type and timing of BMP implementation, production practices, and crop types. We also used the monitoring data to generate synthetic data and performed a simulation-based power analysis to explore the ability to detect change under 25 scenarios of sampled event counts and hypothetical percentage changes. The simulation-based power analysis indicated that natural variability in event concentration and flow hindered our ability to detect change. The results from this study can be used to guide future decisions pertaining to monitoring efforts in small agricultural drainage basins in the Mississippi Alluvial Plain.

Key words: agriculture—conservation practices—hydrology—Mississippi Alluvial Plain—nutrient pollution—sediment pollution

The efficacy of agricultural best management practices (BMPs) to reduce nutrient and sediment runoff has been documented across a range of scales; however, studies indicate apparently diminishing returns as the basin sizes increase from farm plots to regional basins (Tomer and Locke 2011; Sharpley et al. 2015; Mulla et al. 2008; USDA NRCS 2017). Small agricultural drainage ditches (drainage area <10 km²) provide a middle ground for evaluating water quality changes in response to BMPs and production practices. These smaller drainages are often runoff driven and typically have few additional land uses apart from agriculture. Without the added complexity of groundwater discharge and fewer land uses in the basin, the hope is that water quality changes are generally easier to detect without the cumbersome monitoring of edge-of-field setups at the local scale. However, even at this scale, water quality monitoring presents challenges, especially in some areas of the country.

For example, the Mississippi Alluvial Plain (MAP) ecoregion is a topographically low-relief area that stretches along the lower Mississippi River from below the confluence with the Ohio River to the Gulf of Mexico (Chapman et al. 2004). Historically, this area had extensive wetlands with over 10 million ha of bottomland hardwood forests (King and Keeland 1999). However, in the early to mid-1900s, deforestation due to agriculture, flood control, and rural development reduced these forests to less than 3 million ha, and extensive levee construction and channelization reshaped the landscape (King et al. 2006; King and Keeland 1999; Faulkner et al. 2011). This extensive and intensive alteration led to decreased ecosystem functions and the loss of natural processes equipped to mitigate excess nutrients and sediment (Faulkner et al. 2011). The highly altered landscape coupled with low-gradient streams make the MAP a complex hydrologic setting.

Like most agricultural regions of the country, producers in the MAP, often with financial assistance from federal and state governments, have implemented a variety of BMPs on the landscape to reduce the amount of nutrients and sediment leaving the fields and hopefully mitigate the deleterious effects of these constituents on downstream waters (Dabney et al. 2001). A general review of BMPs in the MAP found that during the early 2000s, about 1.27 million ha with 68 different types of BMPs were active in the region (Faulkner et al. 2011). The implementation of BMPs in the MAP, and across the entire Mississippi River basin, is an important part of efforts to minimize the size of the summer hypoxic zone in the Gulf of Mexico (Hypoxia Task Force 2017). Yet, little monitoring has been associated with pre- and postimplementation of BMPs, making it difficult to characterize their effectiveness individually or collectively (Faulkner et al. 2011). Furthermore, explicit information about which BMPs are applied and the exact location is typically not available as per Section 1619 of the Food, Conservation, and

Jennifer C. Murphy is a hydrologist, Matthew Hicks is an ecologist, and Shane Stocks is a hydrologic technician at the US Geological Survey, Lower Mississippi-Gulf Water Science Center, Nashville, Tennessee, and Jackson, Mississippi.

Energy Act of 2008 (US Congress 2008), which presents a challenge when trying to establish a link between water quality changes and BMP implementation.

In the late 2000s, the US Geological Survey (USGS) began monitoring streamflow, stage, and water quality in two agricultural ditches located in the MAP that drain row crop fields and have a variety of BMPs in place. These sites discharge into separate oxbow lakes (Bee Lake and Lake Washington) in the northwestern portion of the state of Mississippi (figure 1). Elevated concentrations of nutrients and sediment entering the two lakes from surrounding drainages are thought to contribute to lake infill, increased noxious aquatic plants, low dissolved oxygen (O) concentrations, algal blooms, and fish kills (MDEQ 2006, 2007). In an effort to remediate these effects, multiple BMPs have been installed in the drainages of both oxbow lakes.

We used these monitoring data to detect changes in water quality that could be potentially attributed to BMPs in the drainages; however, we did not have a control basin with which to assess these changes. Specifically, the objectives of our study were threefold. Our first objective was to test for differences in sediment and nutrient event concentrations between two periods that capture different intensities of BMP activities in the drainages. Our second objective was to explore the ability to detect changes in sediment and nutrients in small agricultural drainages under various monitoring scenarios using a power analysis based on our monitoring results. Our final objective was to use the information from this monitoring effort, data analysis, and power analysis to inform future monitoring strategies of small agricultural basins in the MAP.

Materials and Methods

Location Description. Bee Lake (33.05139°, −90.35056°) and Lake Washington (33.06083°, −91.03889°) are oxbow lakes located in the predominately agricultural region of northwestern Mississippi (figure 1). This area has low topographical relief. Bee Lake is a former meander of the Yazoo River, and Lake Washington is a former meander of the Mississippi River. Bee Lake drains approximately 124 km², and Lake Washington drains approximately 120 km² of mostly row crop agricultural land.

Figure 1

(a) Lake Washington Tributary at Stein Road (LWSR) site (dot) and drainage (dotted outline), (b) Bee Lake Tributary No. 1 (BLT1) site and drainage, and (c) location of both study areas in relation to the state of Mississippi and the Gulf of Mexico (hatched area is the Yazoo River basin).
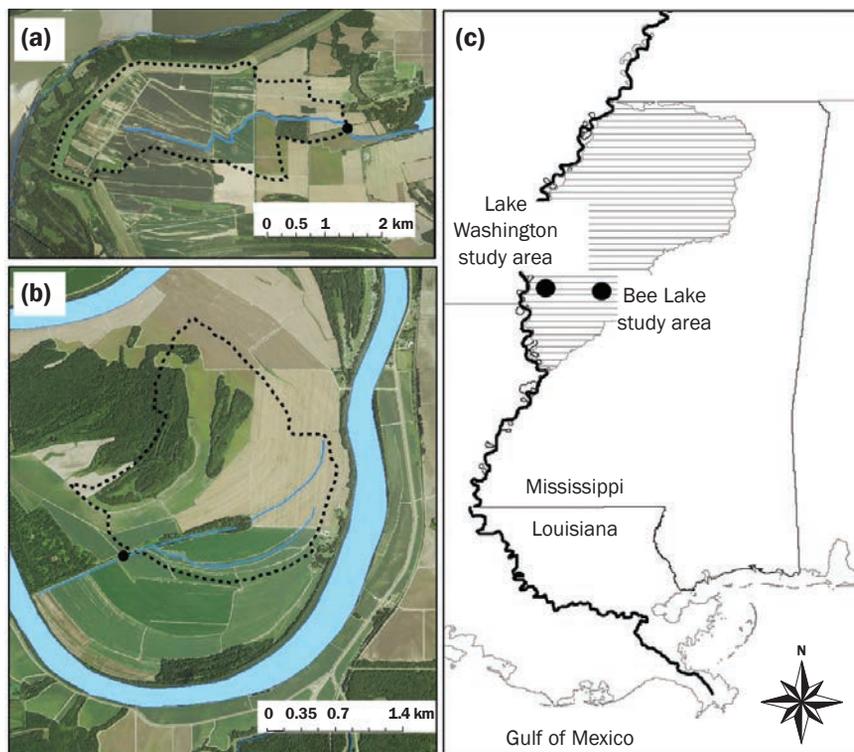
Base from US Geological Survey and Mississippi Automated Resource Information System, GCS NAD 1983

One tributary to each oxbow lake was monitored for water quality and hydrology for an eight- or nine-year period to assess changes in water quality related to the installation of BMPs in their drainages. Both tributaries are relatively large ditches draining row crop fields (figure 1). Bee Lake Tributary No. 1 (BLT1; USGS site number 330304090210100) near Thornton, Mississippi, drains a 3.4 km² area near the center of Bee Lake and was monitored for nine years. Lake Washington Tributary at Stein Road (LWSR; USGS site number 330548091055100) near Chatham, Mississippi, drains a 5.8 km² area to the northwest of Lake Washington and was monitored for eight years.

Land use in the MAP is overwhelmingly agricultural with row crop farming of largely corn (*Zea mays*), soybean (*Glycine max*), and cotton (*Gossypium*), and both drainages were used exclusively for row crop agriculture. The Cropland Data Layer produced by the USDA was used to provide hectarage estimates of various crop types in both drainages by year (USDA NASS 2007-2016). For both sites, two or more crops were planted within the drainage each year. In the BLT1 drainage, corn and cotton were the predominant crops planted between 2008 and 2016. Other crops planted in the drainage included winter wheat (usually *Triticum aestivum*), soybeans, and peanuts (*Arachis hypogaea*). Woody wetlands also comprise about a quarter of the BLT1 drainage. In the LWSR drainage, soybean was the predominate crop. Corn was also planted every year and occasionally winter wheat, but these two crops combined rarely exceeded the soybean hectarage.

Hydrologic Monitoring and Data Processing. Water-surface elevation (stage) was continuously monitored at 15-minute intervals at both study sites using a nonsubmersible pressure sensor attached to a steel pipe near the center of the drainage ditch. The sensor measures water pressure at the bottom of the steel pipe, contrasts this to local atmospheric pressure, and then converts the pressure to stage. Streamflow was periodically measured at each site across a range of flow conditions according to standard USGS guidelines as described by Rantz et al. (1982).

At each station, relations between stage and measured streamflow were developed (Buchanan and Somers 1969), and these relations were used to compute continuous streamflow at 15-minute intervals using the monitored stage data (Kennedy 1983).

Since both drainage ditches go dry during portions of the year, and the precision of the flow measurements was poor below 0.03 $m^3\ s^{-1}$, no observations below this threshold were retained. Thus, the flow data set can be interpreted as left censored at 0.03 $m^3\ s^{-1}$; missing values in the 15-minute data set indicate either flows below 0.03 $m^3\ s^{-1}$, no flow (a dry channel), or a submerged gage (lake surface is higher than gage orifice). We aggregated the 15-minute flow information into an event-based data set by isolating periods of continuous flow, separating these periods into individual flow events (i.e., a rise and fall of the hydrograph), and finally calculating several event-based statistics for each event (Murphy et al. [2019] for processing details). Hereafter, the event mean flow is referred to as "event flow." Additionally, only events that also had recorded rainfall at the closest National Oceanic and Air Administration (NOAA) National Weather Service (NWS) stations (located within 48 km of each site) on the day of or the day before the event were included in the analysis. Thus, the event-based data set used in this analysis only considers runoff events due to storms, not groundwater flow or irrigation return flow.

Water Quality Sampling and Data Processing. Water samples were collected for nutrient and suspended-sediment (SS) analysis at BLT1 between July 9, 2007, and August 17, 2016, and at LWSR between July 10, 2008, and June 14, 2016. Unfiltered water samples were collected over event hydrographs using one or two automatic samplers (ISCO 6712 or 3700; Teledyne ISCO, Lincoln, Nebraska) at each site and were programmed to sample streamflow above a specified stage height. When flow was above the specified stage, samples were collected in 1 L plastic bottles at set time intervals. Sampling ceased when the flow dropped below the specified stage or once all 24 bottles in the sampler were filled. The specified stage height and sampling intervals varied seasonally and over the monitoring period in an effort to capture the most events, sample as much of the event hydrograph as possible (i.e., rising and falling

limbs), and to find the most optimal settings during different hydrologic conditions. At the beginning of the monitoring effort, two automated water samplers were used at each site (one for nutrient samples and another for sediment), though later only one sampler was used, and multiple bottles were filled for each sample. Water samples for nutrient analysis were preserved with 4.5-N sulfuric acid and chilled on ice immediately after sample collection (MDEQ 2013). Within 48 hours of collection, one set of water samples was delivered to the Mississippi Department of Environmental Quality (MDEQ) laboratory in Pearl, Mississippi, for determination of nutrient constituent concentrations (MDEQ 2013). Analyses were performed in accordance with US Environmental Protection Agency (USEPA) Methods for Chemical Analysis of Water and Wastes (USEPA 1983), Test Methods for Evaluating Solid Waste SW-846 (USEPA 1986), and the Standard Methods for Examination of Water and Wastewater (APHA 2005). The second set of water samples was delivered to the USGS sediment laboratory in Baton Rouge, Louisiana, for determination of SS concentration and processed according to Fishman and Friedman (1989). Quality-control procedures were followed according to the MDEQ Quality Assurance Project Plan (MDEQ 2013) and the 2006 Quality Assurance Plan for Water Quality Activities in the USGS Mississippi Water Science Center (R.A. Rebich, US Geological Survey, personal communication, 2006). All streamflow, nutrient, and SS data are available from the USGS National Water Information System (NWIS) web interface (USGS 2018).

This analysis explores changes in SS and nutrient concentrations, including total nitrogen (TN), ammonia (NH_3), total Kjeldahl nitrogen (TKN), nitrate plus nitrite (NO_3^- + NO_2^-), and total phosphorus (TP) (table 1). At times, concentration determinations were censored at a given detection level due to conditions at the laboratory. In these cases, one half the detection limit was substituted for the left-censored value. The water quality observations were aggregated to a flow-weighted mean concentration for each event, hereafter referred to as an "event concentration." This process was accomplished by first interpolating concentrations at 15-minute intervals between the observed concentrations, calculating 15-minute (unit) loads, summing the

unit loads over each event, and dividing the event load by the total flow volume of the event (Murphy et al. 2019).

Hydrologic and Water Quality Data Coverage. Both sites had intermittent flow through the channel, with most of the flow occurring in the winter (December to February) and spring (March to May). At BLT1, due to equipment malfunctions, the gage was not operational for approximately 3.2% of the record. The gage was continuously operational at LWSR (table 2). Across the entire study period, flow events represented 66% of the total time and 97% of the total flow volume at LWSR, and 41% of the total time and 82% of the total flow volume at BLT1. In total, there were 194 and 177 flow events at BLT1 and LWSR, respectively. At both sites, the hydrologic characteristics of the flow events were highly variable with some flow statistics varying by two or more orders of magnitude (table 3). The hydrologic characteristics and variability found in BLT1 and LWSR are consistent with findings of other research on agricultural ditches in Mississippi (Baker et al. 2018; Littlejohn et al. 2014).

The monitoring of water quality at BLT1 and LWSR resulted in many individual samples being collected that were then aggregated to event concentrations (table 1). Due to equipment and laboratory issues, holding time logistics, and adaptation of study design at BLT1 to include nutrient sampling later in the study, we had fewer samples for nutrients than sediment. At BLT1, 550 and 195 individual samples were collected for SS and nutrients, respectively, which resulted in 56 and 20 flow events sampled for SS and nutrients, respectively (table 1). Sampling for SS began approximately two years prior to sampling for nutrients at BLT1 (table 2). At LWSR, 367 and 234 individual samples were collected for SS and nutrients, respectively, which resulted in 40 and 27 flow events sampled for SS and nutrients, respectively (table 1). Sampling for SS and nutrients began concurrently at LWSR (table 2).

Conservation Practices. Due to concerns about sediment pollution in Bee Lake, Clean Water Act Section 319(h) funds were used to install 14 BMPs in the drainage of BLT1. These BMPs included the construction of a buffer zone, water-control structures, grade-stabilization structures, a vegetated ditch, and a two-stage ditch. Some BMPs were installed in July of 2007, and a second

Table 1

Water quality sample information and event-based descriptive statistics, by site.

Constituent	USGS parameter code	Number of samples (total)	Number of censored samples	Number of flow events	Minimum event concentration (mg L⁻¹)	Median event concentration (mg L⁻¹)	Mean event concentration (mg L⁻¹)	Maximum event concentration (mg L⁻¹)
Bee Lake Tributary No. 1 (BLT1)								
NH_3	00610	206	30	20	0.02	0.23	0.73	6.20
$NO_3^- + NO_2^-$	00630	206	6	20	0.04	1.00	1.40	3.80
SS	80154	550	0	56	52.00	210.00	340.00	1,200.00
TKN	00625	206	0	20	0.42	1.30	1.80	11.00
TN	00600	206	0	20	1.00	2.60	3.10	14.00
TP	00665	206	0	20	0.34	0.57	1.00	9.00
Lake Washington Tributary at Stein Road (LWSR)								
NH_3	00610	234	77	27	0.016	0.11	0.18	0.70
$NO_3^- + NO_2^-$	00630	235	0	27	0.075	0.81	0.95	3.90
SS	80154	367	0	40	82.00	1,100.00	1,100.00	4,000.00
TKN	00625	235	0	27	0.32	0.95	1.55	5.80
TN	00600	235	0	27	0.55	1.80	2.50	7.60
TP	00665	235	1	27	0.26	0.75	0.88	2.30

Notes: NH_3 = ammonia. $NO_3^- + NO_2^-$ = nitrate plus nitrite. SS = suspended sediment. TKN = total Kjeldahl nitrogen. TN = total nitrogen. TP = total phosphorus. USGS = US Geological Survey.

Table 2

Streamflow and water quality monitoring information by site.

Site	Streamflow period of record (number of events)	Dates of missing streamflow record	Percentage of flow record missing (%)	Water quality sampling date range (number of sampled events)	Flow events sampled (%)	End of early sampling period	Number of sampled events in early period	Number of sampled events in late period
Bee Lake Tributary No. 1 (BLT1)	July 9, 2007, to Aug. 17, 2016 (194 events)	Oct. 23 2008, to Oct. 29, 2008 (6 days) Oct. 18, 2010, to Nov. 1, 2010 (15 days) Sept. 1, 2011, to Sept. 6, 2011 (6 days) Feb. 2, 2015, to Feb. 3, 2015 (2 days) Feb. 7, 2015, to Feb. 26, 2015 (20 days) Aug. 28, 2015, to Sept. 30, 2015 (34 days) Oct. 1, 2016, to Oct. 24, 2016 (24 days)	3.2	SS*: Oct. 18, 2007, to Aug. 17, 2016 (56) Nutrients: May 2, 2010, to Aug. 17, 2016 (20)	SS: 29 Nutrients: 10	Aug. 15, 2013	SS: 39 Nutrients: 8	SS: 17 Nutrients: 12
Lake Washington Tributary at Stein Road (LWSR)	Aug. 10, 2008, to June 14, 2016 (177 events)	None	0.0	SS: Aug. 12, 2008, to June 14, 2016 (40) Nutrients: same as SS (27)	SS: 23 Nutrients: 15	Dec. 31, 2011	SS: 24 Nutrients: 15	SS: 16 Nutrients: 12

*SS = suspended sediment.

Site	Statistic	Period of record	Early period	Late period
Bee Lake Tributary No. 1 (BLT1)	Number of events	194	120	74
	Event mean flow ($m^3 s^{-1}$)	0.31 (0.05 to 1.42)	0.42 (0.05 to 1.42)	0.28 (0.08 to 1.20)
	Event peak flow ($m^3 s^{-1}$)	0.71 (0.11 to 3.37)	0.76 (0.11 to 3.37)	0.59 (0.11 to 2.75)
	Event total flow volume ($10^3 m^3 s^{-1}$)	30 (2 to 150)	34 (2 to 150)	26 (2 to 133)
	Event duration (h)	25 (8 to 153)	25 (8 to 153)	24 (9 to 84)
	Total flow volume ($10^3 m^3 s^{-1}$)	6,034	4,087	1,946
Lake Washington Tributary at Stein Road (LWSR)	Number of events	177	70	107
	Event mean flow ($m^3 s^{-1}$)	0.85 (0.06 to 5.38)	0.68 (0.06 to 4.39)	0.96 (0.08 to 5.38)
	Event peak flow ($m^3 s^{-1}$)	2.07 (0.11 to 13.68)	2.07 (0.11 to 13.68)	2.07 (0.11 to 9.49)
	Event total flow volume ($10^3 m^3 s^{-1}$)	84 (1 to 1,409)	63 (1 to 1,409)	85 (2 to 785)
	Event duration (h)	22 (2 to 89)	22 (2 to 89)	22 (6 to 55)
	Total flow volume ($10^3 m^3 s^{-1}$)	14,907	5,748	9,160

round of BMP installation was completed in August of 2012. The second round of installations included the construction of two-stage ditches to address nutrient pollution. For the BLT1 site, the "early" sampling period was defined as data collected on or before August 15, 2013. Hydrology and water quality data collected after this date were considered part of the "late" period (table 2). This break in the data captures both phases of BMP implementation in the early period.

Due to concerns about nutrient pollution in Lake Washington, Clean Water Act Section 319(h) funds were also used to install numerous BMPs throughout the drainage basin between early 2008 and late 2010. In the LWSR drainage, BMPs included land leveling, construction of pads, pipes, and filter strips, and planting of cover crops. During this time and continuing beyond 2010, additional BMPs were installed or implemented throughout the basin with funds provided by the USDA Natural Resources Conservation Service (NRCS)'s Environmental Quality Incentives Program and Conservation Reserve Program. These BMPs included more land-leveling and construction of pads and pipes, in addition to nutrient management, construction of an underground irrigation line, and tree planting. The precise timing of BMP construction and implementation is unknown. The BMPs implemented in the LWSR drainage, and throughout the Lake Washington drainage basin in general, were originally designed for sediment control even though nutrient pollution was the primary concern. For LWSR, the "early" sampling period was defined as data collected on or before December 31, 2011, and data collected after this date were considered part

of the "late" period (table 2). This break in the data was intended to capture the main BMP implementations and most intensive periods of construction in the early period. Water quality samples were not collected prior to the early period at either site.

Difference and Equivalence Testing. For each constituent, a nonparametric bootstrap method was used to estimate 90% confidence intervals for the difference in median event concentration between the early and late sampling periods. The difference in median event concentration was calculated according to equation 1:

$$M_{late} - M_{early} = M_{diff},\qquad(1)$$

where M_{late} is the median of event concentrations during the late period, M_{early} is the median of event concentrations during the early period, and M_{diff} is the difference in medians between the early and late periods.

All analyses were completed using the R statistical software program (R Core Team 2017), and bootstrapping was implemented using the boot R package (Canty and Ripley 2017; Davison and Hinkley 1997). A replicate consisted of sampling the observed event concentration data, with replacement, according to the strata (early and late periods). For each replicate, the difference in the median event concentration (M_{diff}, equation 1) was calculated. A total of 10,000 bootstrap replicates were generated, and the adjusted bootstrap percentile interval (Davison and Hinkley 1997) was used to estimate 90% confidence interval for M_{diff}.

The 90% confidence interval of M_{diff} was used to assess statistically significant differences and equivalences. Statistically

significant differences were defined as the 90% confidence interval not encompassing zero (figure 2). Statistically significant equivalences were assessed using ±25% equivalence intervals. For each constituent, the lower and upper bounds of the equivalence interval were calculated by multiplying M_{early} by −0.25 and 0.25, respectively. If the 90% confidence interval of M_{diff} was wholly contained within the equivalence interval, then M_{early} and M_{late} were considered to be significantly equivalent within ±25% (figure 2). Using a 90% confidence interval equates to a significance level (α) of 0.10 for the difference test and 0.05 for the equivalence test. Combining these two tests allowed us to determine if median event concentration differed between the early and late periods or if it remained essentially unchanged. If neither test was statistically significant, then the results were inconclusive, indicating there was too much uncertainty to detect if event concentrations were higher, lower, or similar between early and late periods (figure 2).

Difference in Concentration–Streamflow (C–Q) Relationships. We also tested for differences in the C–Q relationship between early and late periods for all constituents in the BLT1 and LWSR drainages. Logarithmically (base *e*) transformed event concentrations were regressed against logarithmically transformed event mean flows using ordinary least squares regression, and a dummy variable was used to identify whether data were collected in the early or late period. First, we tested for significant differences in slope by fitting concentration and flow data to equation 2:

$$log(C_i) = \alpha + \beta log(Q_i) + \gamma P_i + \delta(Q_i P_i),\ (2)$$

Figure 2

Explanation of possible result combinations and interpretations from tests for difference and equivalence in median event concentrations between the early and late periods. Horizontal error bars are representative of 90% confidence intervals.
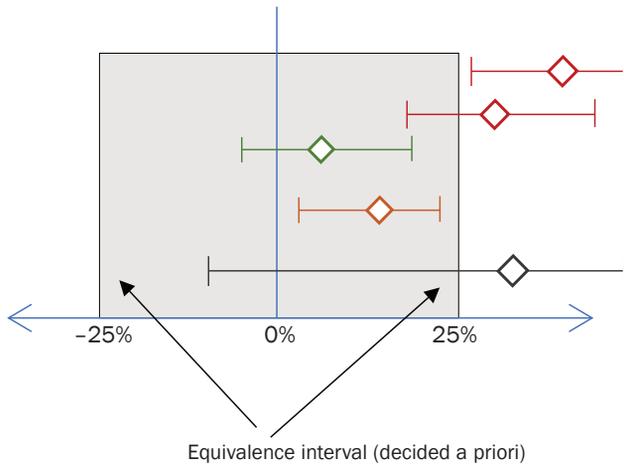
Result of difference test	Result of equivalence test	Interpretation of combined results
Significant	Nonsignificant	Medians differ
Nonsignificant	Significant	Medians equivalent
Significant	Significant	Medians equivalent; no meaningful difference
Nonsignificant	Nonsignificant	Inconclusive; not enough evidence

Equivalence interval (decided a priori)

where C is the event concentration; Q is event flow; P is a dummy variable coded as 1 or 0 for the early and late periods, respectively; i indicates a given flow event; and α, β, γ, and δ are fitted coefficients.

Using the boot R package (Canty and Ripley 2017; Davison and Hinkley 1997), 10,000 bootstrap replicates were used to estimate 90% confidence intervals using the adjusted bootstrap percentile interval (Davison and Hinkley 1997) for all coefficients in equation 2. If the 90% confidence interval for δ (the difference in slope between the early and late periods) did not include zero, then the slope was considered to be significantly different between the early and late periods.

For constituents that did not show significant differences in slope, we then tested for significant differences in intercept of the C-Q relationships. In these cases, equation 3 was fit to the data (variables previously defined):

$$log(C_i) = \alpha + \beta log(Q_i) + \gamma P_i. \quad (3)$$

Similarly, a nonparametric bootstrap (10,000 replicates) was used to estimate 90% confidence intervals for all coefficients. If the 90% confidence interval for α (the difference in intercept between the early and late periods) did not include zero, then the intercept was considered to be significantly different between the early and late periods. For constituents with significantly different intercepts, the intercept coefficient was backtransformed by exponentiating the coefficient, subtracting by 1 and multiplying by 100. This calculated value represented the percentage change in event concentrations

between the early and late sampling periods if event flow was held constant.

Power Analysis. In addition to exploring differences in water quality between the early and late sampling periods at BLT1 and LWSR, we also completed a power analysis to explore the sensitivity of three statistical tests of difference under varying scenarios of percentage change (effect size) and number of sampled events. Using selected distributional parameters and covariate effect sizes (i.e., influence of event flow on event concentrations; table 4), we generated synthetic data to simulate SS, TN, and TP event concentrations for control and treatment groups for various combinations of percentage

change in event concentration (i.e., the effect size) and number of sampled events. The distributional parameters and covariate effect sizes (table 4) were derived from the monitoring data (Murphy et al. 2019), and the control and treatment groups were meant to simulate the influence of limited to no influence of BMPs on water quality compared to decreased event concentrations due to BMP implementations. The number of total sampled events were set to 20, 50, 100, 150, or 200, and percentage changes were set to −10%, −25%, −50%, −75%, or −90% between the control and treatment group, resulting in 25 scenarios. For each scenario and constituent, 1,000 simulations were run and

Table 4

Parameters used in power analysis and derived from the observed data at each site.

Constituent	Parameter	Selected parameters for power analysis	Bee Lake Tributary No. 1 (BLT1)	Lake Washington Tributary at Stein Road (LWSR)
SS	log(mean)	6.00	5.43	6.55
	log(standard deviation)	1.00	0.90	1.13
	slope of SS-flow	6.50	6.50	−4.70
TN	log(mean)	0.80	0.91	0.69
	log(standard deviation)	0.65	0.65	0.65
	slope of TN-flow	−0.05	−0.05	−0.01
TP	log(mean)	−0.37	−0.39	−0.34
	log(standard deviation)	0.68	0.69	0.68
	slope of TP-flow	−0.006	−0.006	−0.003
Q	log(mean)	2.50	2.18	2.92
	log(standard deviation)	0.85	0.71	1.06

Notes: SS = suspended sediment. TN = total nitrogen. TP = total phosphorus. Q = event flow.

significant differences for a given statistical test were recorded. The number of significant differences divided by the number of total simulations (1,000) gave the power of that statistical test for the given scenario.

Three statistical tests were assessed through these simulation-based power analyses: (1) difference in median event concentration (equation 1), (2) difference in the intercept of log-transformed C-Q relationships (equation 3), and (3) difference in mean of log-transformed event concentration, calculated according to equation 4:

$$U_{diff} = U_{treatment} - U_{control},\qquad(4)$$

where $U_{treatment}$ is the mean of log-transformed event concentrations from the treatment period, $U_{control}$ is the mean of log-transformed event concentrations from the control period, and U_{diff} is the difference.

This test was used to explore the power due to the addition of a covariate to the statistical analysis (as in equation 2), as compared to the use of a different measure of central tendency, specifically the mean of log-transformed values (i.e., the geometric mean) compared to the median. These three tests are referred to as the difference-in-median (equation 1), difference-in-intercept (equation 3), and difference-in-meanLog (equation 4), hereafter. For each of the 1,000 simulations, 2,000 bootstrap replicates were used to determine the 90% confidence interval of the given difference statistic using the adjusted bootstrap percentile interval. If the 90% confidence interval of the given difference statistic did not include zero, the difference was considered statistically significant.

Results and Discussion

Difference and Equivalence Testing. At both sites, median event concentrations for some constituents were lower during the late period; however, for many constituents at both sites, the results from the difference and equivalence tests were inconclusive (table 5; figure 2). None of the constituents at either site were significantly equivalent between the early and late periods (table 5). At BLT1, the difference tests indicate all constituents were not significantly different between periods even though five of the six constituents had median event concentrations in the late period that were at least 10% higher or lower than event concentrations from the early period. The insignificant results from

both tests were inconclusive, indicating there was not enough evidence to detect whether the median event concentration changed or stayed the same between the two periods at BLT1. At LWSR, the difference tests indicate half of the constituents (SS, TN, and TP) had significantly lower event concentrations in the late period compared to concentrations from the early period; medians differed by −52%, −36%, and −44%, respectively. The remaining constituents did not have different median event concentrations between periods even though the early median of one constituent (NH_3) was 60% higher than the late period median. Thus, at LWSR, the difference and equivalence tests were inconclusive for NH_3, TKN, and $NO_2^- + NO_3^-$ (table 5). The inclusion of the equivalence test along with the difference test (figure 2) provided complementary information so that we did not mistakenly assume that event concentrations had not changed between periods for all the constituents at BLT1 and for half the constituents at LWSR. Instead, the tests together indicated a lack of evidence and a need for additional sampled events.

Differences in Concentration-Streamflow Relationships. At BLT1, C-Q relationships were similar between early and late periods; only one statistical difference was detected in the slope or intercept between early and late periods across all six constituents (TP; figure 3). For N constituents, the C-Q relationships were negative or near-horizontal (figure 3), suggesting a slight dilution response of these constituents at higher flow events during both periods or no response at all. For NH_3 and $NO_3^- + NO_2^-$, a slight steepening of the C-Q slope was observed during the late period; however, the intercepts or slopes were not significantly different between periods. The C-Q relationships for SS and TP were positive, suggesting a high supply of available sediment and P for transport from the drainage basin. For SS, this relationship is nearly identical between periods and not significantly different. For TP, the intercept for the late period was significantly higher than the early period, indicating about 60% higher TP event concentrations during the late period (figure 3). This increase in intercept is largely attributed to a single event with extremely high concentration during the late period. When we investigated this seemingly anomalous concentration, we did not find any issues with the laboratory or

field methods and thus retained the event in the analysis. We suspect this event may have occurred when fertilizer was applied just prior to a runoff event.

At LWSR, the C-Q relationships for most constituents suggest improved water quality conditions between the early and late periods; however, only three constituents (SS, TKN, and TN) had statistical differences in either the slope or intercept (figure 3). During the early period, event concentrations were negatively related to event flows for all constituents, indicating a tendency for event concentrations to be lower for large compared to small events, possibly suggesting dilution of a finite amount of constituent (figure 3). Although still negative, NH_3 was the only constituent where the C-Q relationship for the late period lay entirely above the early period, though these C-Q relationships were not statistically different. Additionally, TKN and TN had negative C-Q relationships for the late period. These relationships had statistically different intercepts between sampling periods resulting in event concentrations being approximately 35% lower during the late period for both constituents. For the other constituents ($NO_3^- + NO_2^-$, TP, and SS), the C-Q relationship shifted to positive during the late period, though only SS had significantly different slopes between periods. Much of this shift in slope was driven by lower concentrations during events with smaller flows during the late period and not a large increase in concentrations during high flow events (figure 3), suggesting BMPs may have decreased concentrations during small or moderate events but had little effect on larger events. Our results are consistent with Sharpley et al. (2008), who found a significant increase in storm flow dissolved P concentration with storm size, suggesting that as events increase in size, P release from soils and/or the area of the watershed producing runoff also increases.

Power Analysis. As expected, the results of the power analysis for the difference-in-median and difference-in-intercept tests show that as the number of sampled events increases so does the power and our ability to detect change (figure 4). However, this positive relationship between power and the number of sampled events is not linear and varies depending on the statistical test, constituent, and percentage change. The difference-in-median test typically had greater power for the synthetic TN and TP data

Table 5
Summary of difference and equivalence test results for median event concentrations, by constituent and site.

Site	Constituent	Early sampling period median concentration (mg L⁻¹)	Difference between median concentration M_{diff} (mg L⁻¹)	Bootstrap 90% confidence intervals for M_{diff} (mg L⁻¹)	Percentage difference in median concentrations	Significant difference	Equivalence interval for M_{diff} (mg L⁻¹)	Significant equivalence using 25% equivalence interval
Bee Lake	SS	230.00	–55.00	(–204, 157)	–24	ns	(–56, 56)	ns
Tributary	TN	2.20	0.46	(–1.08, 1.52)	21	ns	(–0.56, 0.56)	ns
No. 1	NH_3	0.37	–0.15	(–0.69, 0.31)	–41	ns	(–0.09, 0.09)	ns
(BLT1)	TKN	1.20	0.05	(–0.96, 0.76)	4	ns	(–0.30, 0.30)	ns
	$NO_3^- + NO_2^-$	1.10	–0.14	(–0.96, 0.36)	–13	ns	(–0.28, 0.28)	ns
	TP	0.52	0.13	(–0.05, 0.31)	25	ns	(–0.13, 0.13)	ns
Lake	**SS**	**1,200.00**	**–650.00**	**(–1,170, –71)**	**–52**	**sig**	(–313, 313)	ns
Washington	**TN**	**2.50**	**–0.89**	**(–3.77, –0.24)**	**–36**	**sig**	(–0.62, 0.62)	ns
at Stein	NH_3	0.084	0.05	(–0.24, 0.13)	60	ns	(–0.02, 0.02)	ns
Road	TKN	0.97	–0.16	(–1.59, 0.21)	–16	ns	(–0.243, 0.24)	ns
(LWSR)	$NO_3^- + NO_2^-$	1.00	–0.30	(–0.87, 0.15)	–29	ns	(–0.26, 0.26)	ns
	TP	**1.10**	**–0.48**	**(–1.03, –0.005)**	**–44**	**sig**	(–0.27, 0.27)	ns

Notes: NH_3 = ammonia. $NO_3^- + NO_2^-$ = nitrate plus nitrite. SS = suspended sediment. TKN = total Kjeldahl nitrogen. TN = total nitrogen. TP = total phosphorus. ns = nonsignificant result. sig = significant result.

sets, whereas the difference-in-intercept test typically had greater power for synthetic SS data. Across all scenarios and both tests, the power analysis indicated consistently lower power for synthetic TN (figure 4). More sampled events were particularly helpful when the percentage change was moderate (25% or less). When the percentage decrease was about 50%, increasing the number of sampled events rapidly increased the resulting power for most of the constituents and for both statistical tests. For some constituents, this pushed the power above 0.80, a commonly used threshold for power (Krzywinski and Altman 2013). For percentage decreases beyond this range (decrease of 75% or more), increasing the number of sampled events had little effect on power since the power was already high for both tests with as few as 20 sampled events (the exception being power <0.80 when synthetic TN data had a low number of sampled events). For scenarios with small percentage decreases (10%), increasing the number of sampled events resulted in only small increases in power across both tests and all constituents (figure 4).

Comparing the power of both the difference-in-median and difference-in-intercept tests to the power of the difference-in-meanLog test shows that the measure of central tendency being tested is important, and that accounting for event flow as a covariate does not appear to increase our ability to detect a change given the characteristics of the monitoring data. The difference-in-intercept test was anticipated to have greater power than the difference-in-median test because it accounts for event flow as a covariate; however, this proved to be true only for SS synthetic data and for synthetic TP and TN data during scenarios with only a few sampled events (figure 5). Furthermore, the difference-in-intercept test has almost the same power as the difference-in-meanLog test for all scenarios and constituents, indicating that accounting for flow as a covariate only marginally increased our ability to detect a change, given the synthetic data characteristics we used. Thus, the difference in power between the difference-in-median test and the difference-in-intercept test is primarily due to the use of different measure of central tendency. Use of the median as the measure of central tendency resulted in higher power for most of the scenarios using synthetic TN and TP data, likely due to the skewness of the synthetic (and thus the monitored) data (figure 5).

Potential Changes between Early and Late Periods. Our study demonstrated mixed results for detecting changes in sediment and nutrient concentrations from two small agricultural drainages in the MAP. At BLT1 there is a lack of evidence for a change or for equivalence in event concentrations between early and late periods, and only TP concentrations showed a significantly different intercept between early and late periods, likely driven by one event with an event concentration of 9 mg L⁻¹ (figure 3). Conversely, there was some evidence of significant decreases in event concentrations for about half of the constituents at LWSR. At least one of the statistical tests (i.e., difference in median, C-Q intercept, or C-Q slope) identified significant decreases for four of the six constituents (SS, TKN, TN, and TP), and these decreases ranged between –35% and –52%, depending on the constituent and test. The results were inconclusive for NH_3 and $NO_3^- + NO_2^-$. The mixed results presented here are consistent with other studies in agricultural drainages in the MAP, as well as other intensive agricultural areas with BMPs. Some studies have shown decreases in sediment (Lizotte et al. 2014; Corsi et al. 2005), P (Bishop et al. 2005; Baker et al. 2016; Littejohn et al. 2014), and N (Corsi et al. 2005; Littlejohn et al. 2014), while other studies have shown little to no reduction in sediment or nutrient (Lizotte et al. 2017; Baker et al. 2018; Corsi et al. 2012).

The changes in water quality, or lack thereof, identified in our study could be due to a variety of differences between early and late periods at BLT1 and LWSR. Many of these differences were not controlled for in this study and can be highly variable. Variability in crop type and soil type can greatly affect sediment and nutrient concentrations and loads (Aryal and Reba 2017). Variability in hydrology and crop production practices can make detecting changes in nutrients and sediment challenging (Baker et al. 2018). Rainfall frequency, duration, and intensity vary within and between years, also
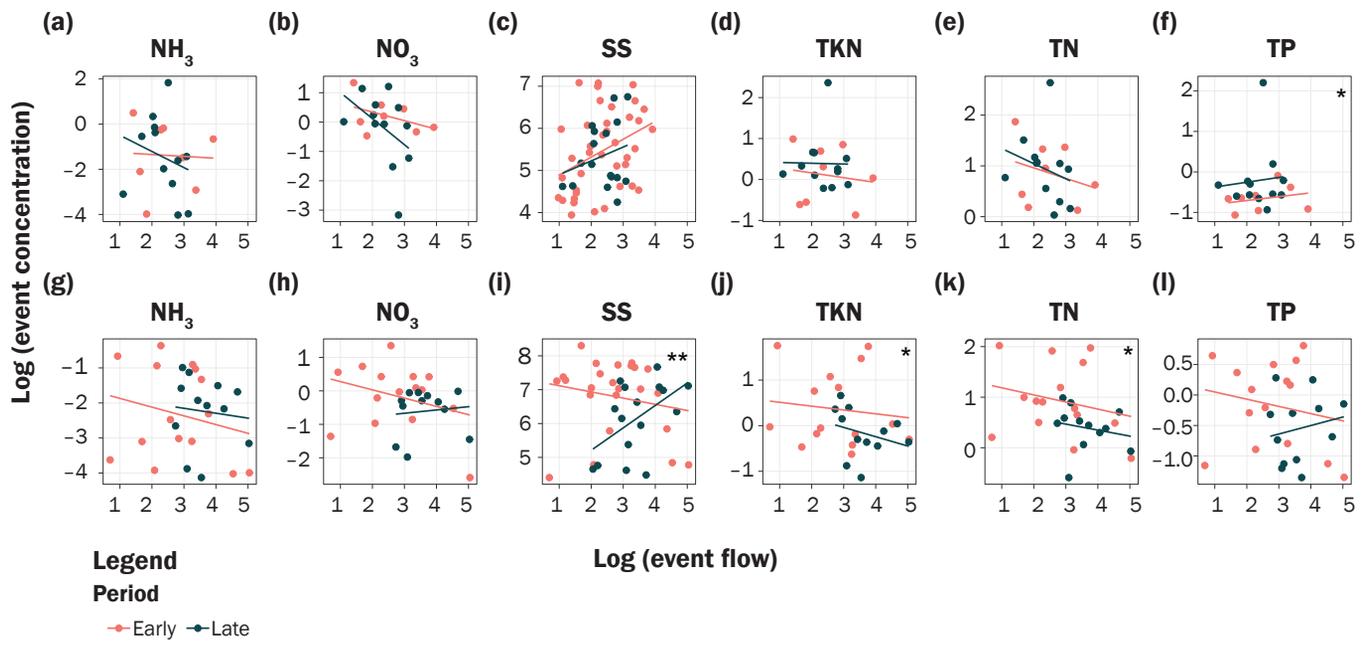
contributing to seasonal and annual variability in concentrations and loads (Sharpley et al. 2015). Attributing a specific cause to a change, or lack of change, in water quality or quantity is a challenge recognized across the hydrologic literature (Harrigan et al. 2014; Clark et al. 2011; Merz et al. 2012; Ryberg 2017). Specifically, our study design of two periods that capture differences in the intensity and density of BMP implementation, without an associated control basin, negates our ability to differentiate changes in water quality due exclusively to BMPs versus other influences (Kibler et al. 2011). Nevertheless, we can hypothesize and draw tentative conclusions about the role BMPs, and production practices, may have had on changes, or lack thereof, in water quality in these two agricultural drainages.

Our results suggest BMPs and production practices may have played a role in improving water quality at LWSR but had little effect at BLT1. Differences in improvement in water quality between BLT1 and LWSR may be related to differences in intensity and types of BMPs implemented and various production practices between the drainages. In the Bee Lake drainage, there were fewer BMPs installed, and most of the BMPs that were installed were targeted to reduce sediment and not necessarily nutrients; however, two-

stage ditches were installed near the end of the early period, which were specifically targeted to reduce nutrient runoff. LWSR had many BMPs installed in the drainage, primarily during the early period, and, similar to BLT1, these BMPs were mostly designed for sediment reduction. During the late period, nutrient management was ongoing in the LWSR drainage through funding mechanisms apart from Clean Water Act Section 319(h) funds. Nutrient management in the LWSR drainage likely resulted in lower fertilizer usage on the fields, thus, likely decreasing the amount of available N for transport.

The drainages differed in the type of crops planted between early and late periods. In the BLT1 drainage, corn was the primary crop for most years in the early period, and during the late period, corn and cotton alternated as the primary crop annually. Since fertilizer is used on both corn and cotton crops, it is possible that fertilizer inputs were not very different between early and late periods in the BLT1 drainage. Additionally, woody wetlands are a considerable portion of the BLT1 drainage, which may have helped to moderate any changes in water quality between early and late periods at this site. At LWSR, soybeans were the dominant crop type during the late period, while during the early period the
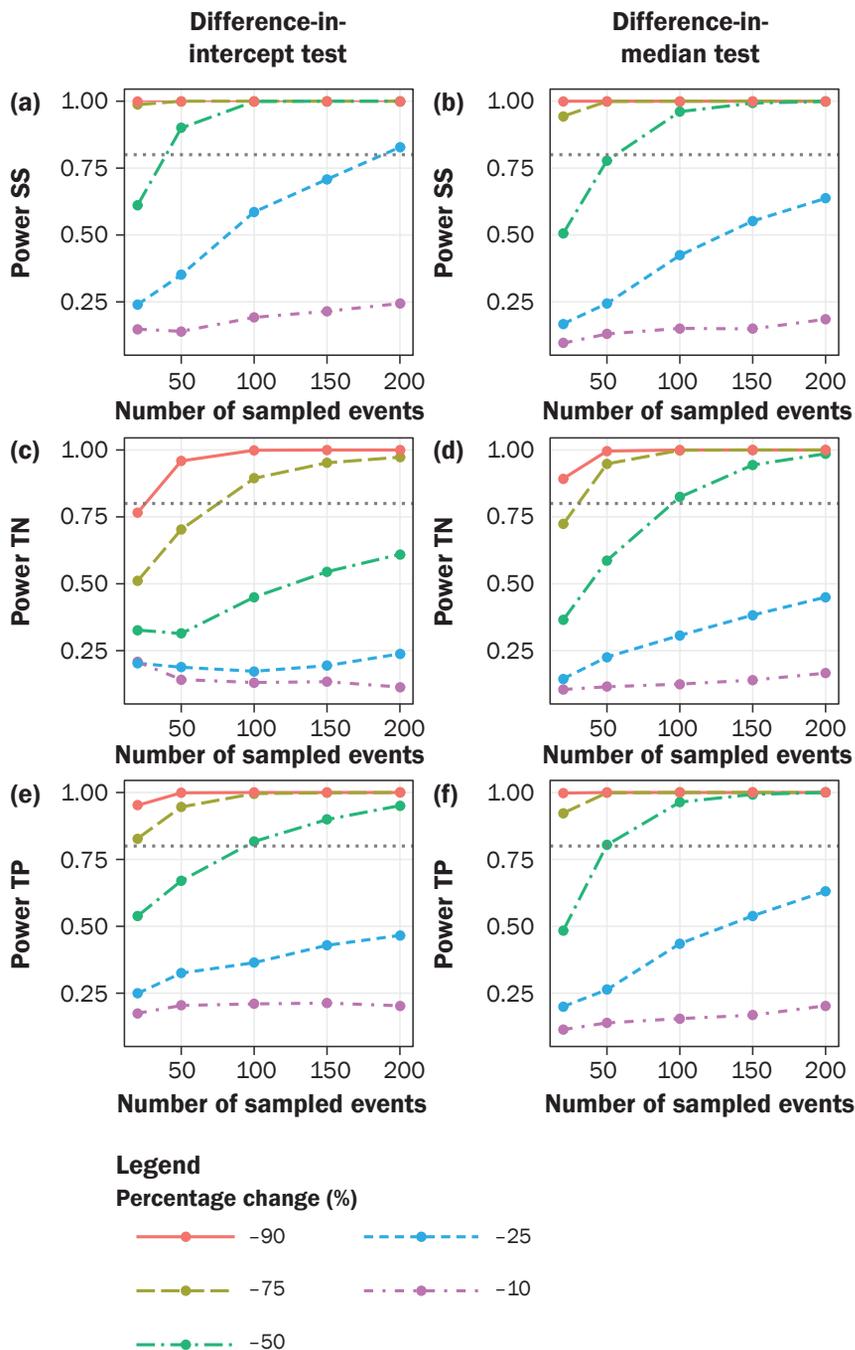
dominant crop type alternated between corn and soybean. The fertilizer requirements are typically quite different for these crops, such that corn is fertilized and soybeans generally are not. It is possible that this difference in crops between early and late periods may have led to lower fertilizer application and lower nutrient concentrations in field runoff in the late period compared to the early period.

Potential Considerations for Future Monitoring. Event concentrations and event flows were highly variable at both sites, which likely contributed to the low proportion of rejected null hypotheses across the four statistical tests (difference-in-median, equivalence-in-median, difference-in-slope, and difference-in-intercept). In this study, event concentrations for a given constituent can span two to three orders of magnitude (table 1), and event flows are similarly variable (table 3). The high variability of water quality and hydrology in the MAP has also been documented elsewhere (Littlejohn et al. 2014). A comparison of power across the various scenarios (percentage change versus number of sampled events) for both the difference-in-median and difference-in-intercept tests highlight why we may not have detected many changes—the number of sampled events was low given the natural variability of event con-

Figure 4
Comparison of statistical power for the (a, c, e) difference-in-intercept test and the (b, d, f) difference-in-median test for (a and b) suspended sediment (SS), (c and d) total nitrogen (TN), and (e and f) total phosphorus (TP) for each of the percentage change scenarios. Dotted line is at 0.80 power, a commonly used threshold for power (Krzywinski and Altman 2013).

analysis shows a large increase in power as the number of sampled events increases from 20 to 50. It would have been advantageous to assess the power of our data as an ongoing part of our monitoring activity. Thus, as other researchers have suggested (Levine et al. 2014), we support assessing the power of the collected data set as an ongoing activity during monitoring. This information can be used to better inform future monitoring and whether or not the appropriate number of events are being sampled to achieve the desired sensitivity given the variability of the observed data.

Even if all the observed flow events that occurred at BLT1 and LWSR (194 and 177, respectively) were sampled, the power analysis suggests the difference-in-median and difference-in-intercept tests would still not be able to detect relatively small changes in event concentrations. The power analysis for the scenario with 200 sampled events, coupled with a power of 0.80, indicates only decreases of about 40% or more are likely to be detected using the difference-in-median test (figure 4). Typically, in such cases, a covariate can be incorporated into the analysis to help control some of the variability and increase the power of the statistical tests. The "gold standard" for studies evaluating changes in water quality due to the implementation of BMPs is to use a control basin as the covariate. Since a control basin was not monitored as part of this study, we used event mean flow as the covariate. The power analysis for the difference-in-intercept test indicates that the variability of event concentrations was too great and the synthetic covariate (event flow) was not that useful in controlling for this variability (figure 5). If resources had been available for our study, using a control drainage or monitoring another covariate with higher correlation to event concentrations would have improved our ability to detect change (Loftis et al. 2001). Other covariates to consider may have included soil moisture or precipitation. Thus, at the onset of future monitoring efforts, identifying and monitoring possible covariates may help to control some of the natural variability in the system, and increase the ability of future studies to detect small water quality changes.

An often overlooked aspect of testing for differences between two time periods is acknowledging that a nonsignificant result

centrations in these drainages. Given the characteristics of our data set, only large decreases are likely to be detected, and thus smaller decreases in event concentrations may have occurred but would have been largely undetectable. Interestingly, more sig-

nificant differences were identified at LWSR even though event concentrations and event flows appear to be more variable than BLT1 (table 1 and table 3). However, more events were sampled for nutrients at LWSR compared to BLT1 (table 2), and the power

Figure 5

Comparison of statistical power for difference tests according to percentage change for (a, b, c) suspended sediment (SS), (d, e, f) total nitrogen (TN), and (g, h, i) total phosphorus (TP), grouped by number of sampled events ([a, d, g] 20 sampled events, [b, e, h] 50 sampled events, and [c, f, i] 100 sampled events). Dotted line is at 0.80 power, a commonly used threshold for power (Krzywinski and Altman 2013).
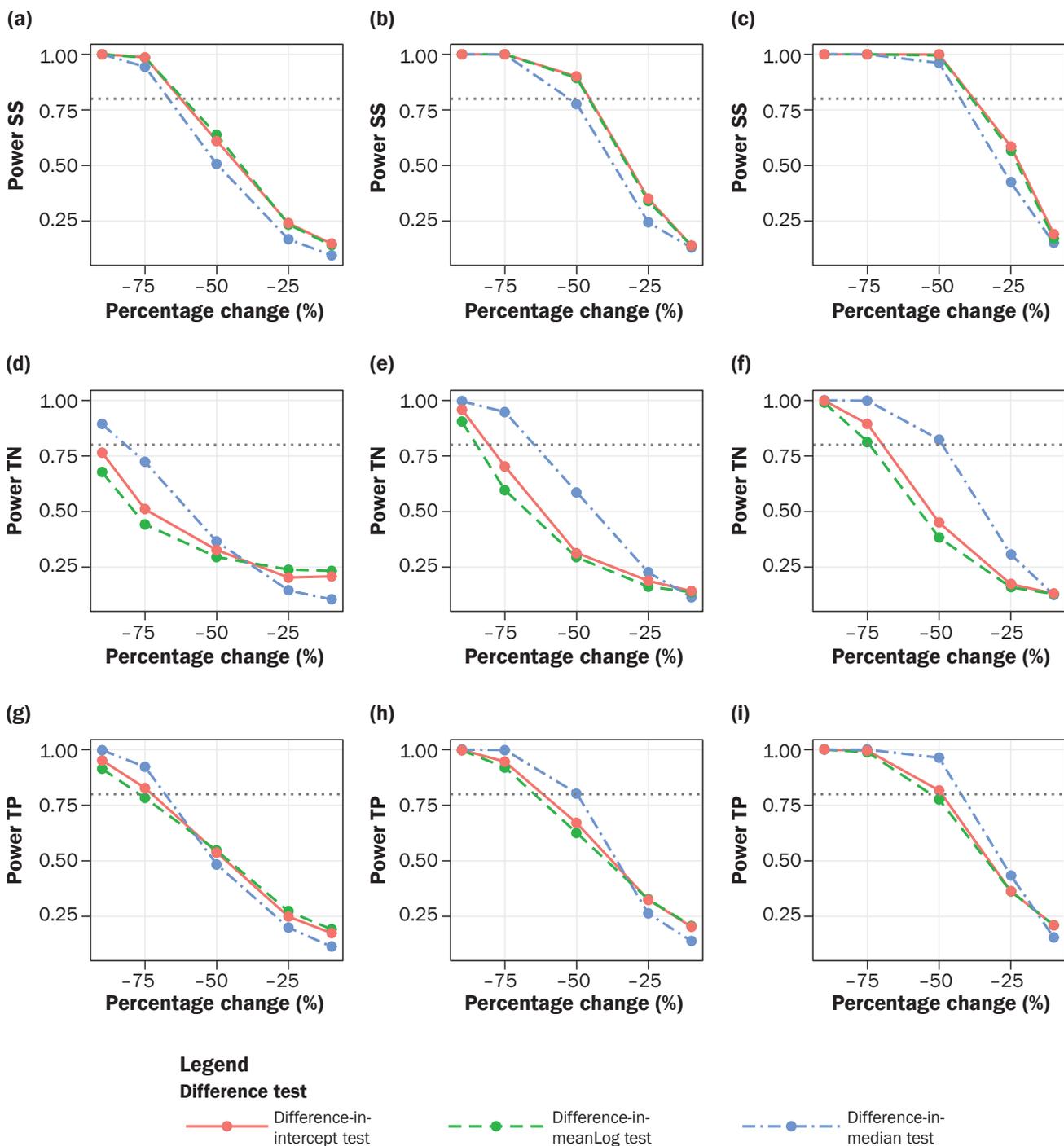
does not necessarily mean there was no effect or that conditions were the same between the two periods (McBride et al. 1993; Dixon and Pechmann 2005). This is why we tested for difference and equivalence of the median event concentrations in our analysis (table 5). While we found that several constituents had differences in median event concentra- tion between the early and late periods, none of the late period medians were equivalent within ±25% of the early period medians. Thus, for many constituents, it is impossible

to declare whether conditions were similar or different. Without this accompanying test for equivalence, we may have mistakenly assumed the nonsignificant results of the differences-in-median test indicated "no change" when in fact the data do not support this statement. We believe combining difference and equivalence testing in hydrologic studies can inform better understanding of how water quality conditions may or may not change over time and when there is a need for more or better data.

Summary and Conclusions
Monitoring and data analysis identified mixed changes in water quality coinciding with implementation of BMPs in two small agricultural drainages in the MAP. Statistical tests indicated no evidence of improvement in sediment and nutrient event concentrations at BLT1, but did suggest some improvement in water quality at LWSR, particularly during moderate flow events for several constituents and for TKN and TN across all flow event magnitudes. Nevertheless, equivalence testing, size of the confidence intervals, and the power analysis suggest a large amount of uncertainty in our results. BMPs and production practices may have played a role in the decreases in event concentrations observed at LWSR, whereas the proportion of woody wetlands in the BLT1 drainage may have mitigated changes that occurred at that site. Our combined use of difference and equivalence testing provided an improved understanding of conditions at the sites and helped to identify that more information was needed to assess change for some of the constituents. Additionally, the power analysis suggests that the natural variability of event flow and event concentrations at these sites hinders our ability to detect small or moderate water quality changes in these drainages. Our results indicate that assessing natural variability, the number of sampled events, and statistical power on an ongoing basis will benefit future monitoring efforts. Potential solutions to address these issues include increasing the number of events being sampled, restricting sampling to an index period (e.g., the spring), monitoring a covariate such as soil moisture, or ideally monitoring a comparable control basin.

Acknowledgements

References
APHA (American Public Health Association). 2005. Standard Methods for the Analysis of Water and Wastewater, 21st edition. Washington, DC: American Public Health Association, American Water Works Association, and Water Environment Federation.

Aryal, N., and M.L. Reba. 2017. Transport and transformation of nutrients and sediment in two agricultural watersheds in Northeast Arkansas. Agriculture, Ecosystems & Environment 236:30-42.

Baker, B.H., J.M.P. Czarnecki, A.R. Omer, C.A. Aldridge, R. Kroeger, and J.D. Prevost. 2018. Nutrient and sediment runoff from agricultural landscapes with varying suites of conservation practices in the Mississippi Alluvial Valley. Journal of Soil and Water Conservation 73(1):75-85, doi:10.2489/jswc.73.1.75.

Baker, B.H., R. Kröger, J.D. Prevost, T. Pierce, J.J. Ramirez-Avila, J.M.P. Czarnecki, D. Faust, and C. Flora. 2016. A field-scale investigation of nutrient and sediment reduction efficiencies of a low-technology best management practice: Low-grade weirs. Ecological Engineering 91:240-248.

Bishop, P.L., W.D. Hively, J.R. Stedinger, M.R. Rafferty, J.L. Lojpersberger, and J.A. Bloomfield. 2005. Multivariate analysis of paired watershed data to evaluate agricultural best management practice effects on stream water phosphorus. Journal of Environmental Quality 34:1087-1101.

Buchanan, T.J., and W.P. Somers. 1969. Book 3, Chapter A8: Discharge measurements at gaging stations. In Techniques of Water-Resources Investigations of the United States Geological Survey. Reston, VA: US Geological Survey.

Canty, A., and B. Ripley. 2017. boot: Bootstrap R (S-Plus) Functions. R package version 1.3-19.

Chapman, S.S., B.A. Kleiss, J.M. Omernik, T.L. Foit, and E.O. Murray. 2004. Ecoregions of the Mississippi Alluvial Plain (map scale 1:1,150,000). Reston, VA: US Geological Survey. https://cfpub.epa.gov/si/si_public_record_report.cfm?Lab=NHEERL&dirEntryId=96619.

Clark, M.P., D. Kavetski, and F. Fenicia. 2011. Pursuing the method of multiple working hypotheses for hydrological modeling. Water Resources Research 47(9):W09301.

Corsi, S.R., J.A. Horwatich, T.D. Rutter, and R.T. Bannerman. 2012. Effects of best-management practices in Bower Creek in the East River priority watershed, Wisconsin, 1991–2009. US Geological Survey Scientific Investigations Report 2012-5217. Reston, VA: US Geological Survey.

Corsi, S.R., J.F. Walker, L. Wang, J.A. Horwatich, and R.T. Bannerman. 2005. Effects of best management practices in Otter Creek in the Sheboygan River priority watershed, Wisconsin, 1990–2002: US Geological

Survey Scientific Investigations Report 2005–5009. Reston, VA: US Geological Survey.

Dabney, S.M., R.A. Rebich, and J.W. Pote. 2001. The Mississippi Delta MSEA Program. In Sustaining the Global Farm, paper from the 10th International Soil Conservation Organization Meeting, May 24-29, 1999. West Lafayette, IN: International Soil Conservation Organization in cooperation with the USDA and Purdue University.

Davison, A.C., and D.V. Hinkley. 1997. Bootstrap Methods and Their Applications. Cambridge: Cambridge University Press.

Dixon, P.M., and J.H.K. Pechmann. 2005. A statistical test to show negligible trend. Ecology 86(7):1751-1756.

Faulkner, S., W. Barrow Jr., B. Keeland, S. Walls, and D. Telesco. 2011. Effects of conservation practices on wetland ecosystem services in the Mississippi Alluvial Valley. Ecological Applications 21(3)Supplement:S31-S48.

Fishman, M.J., and L.C. Friedman. 1989. Book 5, Chapter A1: Methods for determination of inorganic substances in water and fluvial sediments. In US Geological Survey Techniques of Water-Resources Investigations. Reston, VA: US Geological Survey.

Harrigan, S., C. Murphy, J. Hall, R.L. Wilby, and J. Sweeny. 2014. Attribution of detected changes in streamflow using multiple working hypotheses. Hydrology and Earth System Sciences 18:1935-1952.

Hypoxia Task Force. 2017. Mississippi River/Gulf of Mexico Watershed Nutrient Task Force: 2017 Report to Congress. Washington, DC: US Environmental Protection Agency. https://www.epa.gov/ms-htf/hypoxia-task-force-reports-congress.

Kennedy, E.J. 1983. Book 3, Chapter A13: Computation of continuous records of streamflow. In Technique of Water-Resources Investigations of the United States Geological Survey. Reston, VA: US Geological Survey.

Kibler, K.M., D.D. Tullos, and G.M. Kondolf. 2011. Learning from dam removal monitoring: Challenges to selecting experimental design and establishing significance of outcomes. River Research and Applications 27(8):967-975.

King, S.L., and B.D. Keeland. 1999. Evaluation of reforestation in the Lower Mississippi River Alluvial Valley. Restoration Ecology 7(4):348-359.

King, S.L., D.J. Twedt, and R.R. Wilson. 2006. The role of the Wetland Reserve Program in conservation efforts in the Mississippi River Alluvial Valley. Wildlife Society Bulletin 34(4):914-920.

Krzywinski, M., and N. Altman. 2013. Power and sample size. Nature 10:1139-1140.

Levine, C.R., R.D. Yanai, G.G. Lampman, D.A. Burns, C.T. Driscoll, G.B. Lawrence, J.A. Lynch, and N. Schoch. 2014. Evaluating the efficiency of environmental monitoring programs. Ecological Indicators 39:94-101.

Littlejohn, K.A., B.H. Poganski, R. Kroger, and J.J. Ramirez-Avila. 2014. Effectiveness of low-grade weirs for nutrient removal in an agricultural landscape in the

Lower Mississippi Alluvial Valley. Agricultural Water Management 131:79-86.

Lizotte, R.E., S.S. Knight, M.A. Locke, and R.L. Bingner. 2014. Influence of integrated watershed-scale agricultural conservation practices on lake water quality. Journal of Soil and Water Conservation 69(2):160-170, doi:10.2489/jswc.69.2.160.

Lizotte, R.E., L.M.W.Yasarer, M.A. Locke, R.L. Bingner, and S.S. Knight. 2017. Lake nutrient responses to integrated conservation practices in an agricultural watershed. Journal of Environmental Quality 46:330-338.

Loftis, J.C., L.H. MacDonald, S. Streett, H.K. Iyer, and K. Bunte. 2001. Detecting cumulative watershed effects: The statistical power of pairing. Journal of Hydrology 251:49-64.

McBride, G.B., J.C. Loftis, and N.C. Adkins. 1993. What do significance tests really tell us about the environment? Environmental Management 17(4):423-432.

MDEQ (Mississippi Department of Environmental Quality). 2006. Bee Lake Implementation Plan, 1-67. Prepared for Yazoo Basin Team. Stoneville, MS: Delta F.A.R.M. and Delta Wildlife, Inc.

MDEQ. 2007. Lake Washington Watershed Implementation Plan, 1-98. Prepared for Lake Washington Watershed Implementation Team. Little Rock, AR: FTN Associates, Ltd.

MDEQ. 2013. MDEQ master SOP compendium of analytical services for Field Services Division. Jackson, MS: Mississippi Department of Environmental Quality.

Merz, S., S. Vorogushyn, S. Uhlemann, J. Delgado, and Y. Hundecha. 2012. HESS Opinions: More efforts and scientific rigour are needed to attribute trends in flood time series. Hydrology and Earth System Sciences 16:1379-1387.

Mulla, D.J., A.S. Birr, N.R. Kitchen, and M.B. David. 2008. Limitations of evaluating the effectiveness of agricultural management practices at reducing nutrient losses to surface waters. *In* Final Report: Gulf Hypoxia and Local Water Quality Concerns Workshop, 189-212. St. Joseph, MI: American Society of Agricultural and Biological Engineer.

Murphy, J., M. Hicks, and S. Stocks. 2019. Hydrologic event-based water-quality and hydrology data for 3 oxbow tributaries in northwestern Mississippi, 2007-2016. US Geological Survey data release. Reston, VA: US Geological Survey. https://doi.org/10.5066/F75H7FJJ.

R Core Team. 2017. R: A language and environment for statistical computing. Vienna, Austria: R Foundation for Statistical Computing. https://www.R-project.org/.

Rantz, S.E. et al. 1982. Measurement and Computation of Streamflow: Volume 1. Measurement of Stage and Discharge. US Geological Survey Water-Supply Paper 2175. Reston, VA: US Geological Survey.

Ryberg, K.R. 2017. Structural equation model of total phosphorus loads in the Red River of the North Basin, USA and Canada. Journal of Environmental Quality 46:1072-1080.

Sharpley, A.N., P.J.A. Kleinman, A.L. Heathwaite, W.J. Gburek, G.J. Folmar, and J.P. Schmidt. 2008. Phosphorus loss from an agricultural watershed as a function of storm size. Journal of Environmental Quality 37:362-368.

Sharpley, A.N., M.B. Daniels, L. Berry, and C. Hallmark. 2015. Discovery Farms: Documenting water quality benefits of on-farm conservation management and empowering farmers. Acta Agriculturae Scandinavica, Section B – Soil Plant Science 65(Suppl. 2):186-198.

Tomer, M.D., and M.A. Locke. 2011. The challenge of documenting water quality benefits of conservation practices: A review of USDA-ARS's conservation effects assessment project watershed studies. Water Science and Technology 64(1):300-310.

US Congress. 2008. Section 1619 of the Food, Conservation, and Energy Act of 2008 (2008 Farm Bill). H.R. 6124. https://www.congress.gov/bill/110th-congress/house-bill/6124.

USDA NRCS (Natural Resources Conservation Service). 2017. Conservation Practice Adoption on Cultivated Cropland Acres: Effects on Instream Nutrient and Sediment Dynamics and Delivery in Western Lake Erie Basin, 2003-06 and 2012, 77. Special Study Report. Washington, DC: USDA Natural Resources Conservation Service.

USEPA (US Environmental Protection Agency). 1983. Methods for chemical analysis of water and wastes. EPA-600/4-79-020. Washington, DC: US Government Printing Office.

USEPA. 1986. Test Methods for Evaluating Solid Waste Physical/Chemical Methods, 3rd edition. SW-846. Washington, DC: US Government Printing Office.

USDA NASS (National Agricultural Statistics Service). 2007-2016. CropScape - Cropland Data Layer. https://nassgeodata.gmu.edu/CropScape/. Washington, DC: USDA National Agricultural Statistics Service.

USGS (US Geological Survey). 2018. National Water Information System—Web interface. USGS Water Data for the Nation. Reston, VA: US Geological Survey. http://dx.doi.org/10.5066/F7P55KJN.