Impact of Sampling Techniques on Measured Stormwater Quality Data for Small Streams

R. D. Harmel* USDA–ARS
R. M. Slade, Jr. USGS, Austin Community College
R. L. Haney USDA–ARS

Science-based sampling methodologies are needed to enhance water quality characterization for setting appropriate water quality standards, developing Total Maximum Daily Loads, and managing nonpoint source pollution. Storm event sampling, which is vital for adequate assessment of water quality in small (wadeable) streams, is typically conducted by manual grab or integrated sampling or with an automated sampler. Although it is typically assumed that samples from a single point adequately represent mean cross-sectional concentrations, especially for dissolved constituents, this assumption of well-mixed conditions has received limited evaluation. Similarly, the impact of temporal (within-storm) concentration variability is rarely considered. Therefore, this study evaluated differences in stormwater quality measured in small streams with several common sampling techniques, which in essence evaluated within-channel and within-storm concentration variability. Constituent concentrations from manual grab samples and from integrated samples were compared for 31 events, then concentrations were also compared for seven events with automated sample collection. Comparison of sampling techniques indicated varying degrees of concentration variability within channel cross sections for both dissolved and particulate constituents, which is contrary to common assumptions of substantial variability in particulate concentrations and of minimal variability in dissolved concentrations. Results also indicated the potential for substantial within-storm (temporal) concentration variability for both dissolved and particulate constituents. Thus, failing to account for potential cross-sectional and temporal concentration variability in stormwater monitoring projects can introduce additional uncertainty in measured water quality data.

Copyright © 2010 by the American Society of Agronomy, Crop Science Society of America, and Soil Science Society of America. All rights reserved. No part of this periodical may be reproduced or transmitted in any form or by any means, electronic or mechanical, including photocopying, recording, or any information storage and retrieval system, without permission in writing from the publisher.

J. Environ. Qual. 39:1734–1742 (2010)
doi:10.2134/jeq2009.0498
Published online 29 June 2010.
Received 14 Dec. 2009.
*Corresponding author (daren.harmel@ars.usda.gov).
© ASA, CSSA, SSSA
5585 Guilford Rd., Madison, WI 53711 USA

The issue of how best to sample stormwater quality is increasingly relevant because of increased emphasis on improved water quality characterization to support Total Maximum Daily Loads, water quality standards, and nonpoint source pollution control. Characterization of stormwater quality is more difficult than periodic grab sampling, which often focuses on baseflow conditions because runoff events often occur without advance warning, outside conventional work hours, and during adverse weather conditions. As a result, small watershed projects typically utilize automated water quality sampling equipment, so that personnel are not forced to travel to remote sites during relatively short runoff events and manually collect samples under hazardous conditions. The extensive use of automated samplers results from the realization that most projects do not have the resources to maintain an on-call field staff to perform intensive manual storm sampling at multiple sites. Major advantages of automated storm sampling in small streams include its ability to (i) use a consistent sampling procedure at multiple sites, (ii) sample throughout the duration of runoff events, (iii) sample during quick hydrologic response times, and (iv) limit personnel exposure to dangerous conditions (Ging, 1999; Harmel et al., 2006b). Automated samplers are, however, typically limited by their ability to collect samples only at a single fixed intake point, although moveable intakes are occasionally used (e.g., McGuire et al., 1980). Automated sampling equipment is also expensive and requires a considerable financial and personnel resource investment for installation, maintenance, and repair to ensure proper operation.

In contrast, manual storm sampling techniques, whether grab or integrated, require personnel to travel to each sampling site and manually collect samples during runoff events. Wells et al. (1990) and USGS (1999) provide extensive guidance on proper techniques and quality assurance methodology for manual sample collection. Grab sampling at a single collection point at random times during storm events may allow multiple sites to be sampled in storm events, but it does not capture within-channel and temporal concentration variability. Integrated storm sampling typically utilizes the USGS Equal-Width-Increment (EWI) or Equal-Discharge-Increment procedure (Wells et al., 1990; USGS, 1999; Edwards and Glysson, 1999) to produce a single integrated sample obtained throughout the stream cross section. Integrated

R.D. Harmel and R.L. Haney, USDA–ARS, 808 E. Blackland Rd., Temple, TX 76502; R.M. Slade, Jr., USGS (retired), Austin Community College, 11617 River Oaks Tr., Austin, TX 78753. Assigned to Associate Editor Elizabeth Stockdale.

Abbreviations: EWI, Equal-Width-Increment.

sampling provides accurate mean cross-sectional concentration data but requires substantial collection time, which makes it difficult to collect multiple samples from numerous sites. Although bridge access and specialized training and equipment are required, integrated sampling is the preferred sample collection technique on large rivers because of large-magnitude flow depths, channel sizes, and cross-sectional variability.

Martin et al. (1992) explored concentration differences between midchannel surface grab samples and integrated (EWI) samples collected over a 29-mo period from four Kentucky watersheds with drainage areas from 1391 to 13,737 km^2. Their results indicated that concentrations of suspended sediment and certain sediment-associated constituents (total P, Fe, and Mn) were routinely lower in the surface grab samples and that the difference generally increased with increasing flow. In contrast, concentrations of dissolved constituents differed little between the two sampling techniques. Ging (1999) compared concentrations produced by integrated (EWI) and automated sampling on eight streams in Texas with drainage areas from 337 to 4520 km^2. Results showed no directional bias in mean differences between integrated and automated single-intake samples. For 26 constituents analyzed, only dissolved Ca, total P, and dissolved and suspended organic C showed statistically significant differences in median values ($p < 0.10$) from integrated and single-intake automated sample collection.

Aside from these studies, little comparative scientific information is available on water quality (constituent concentration) data resulting from grab, integrated, and automated storm sampling. Therefore, the objective of this study was to evaluate stormwater quality as determined by these common sampling techniques for small streams. The main null hypothesis tested was that stormwater quality does not vary within small-stream cross sections. Specifically, potential differences in dissolved and particulate constituent concentrations were evaluated. The practical implications of these results as they relate to monitoring projects were also presented.

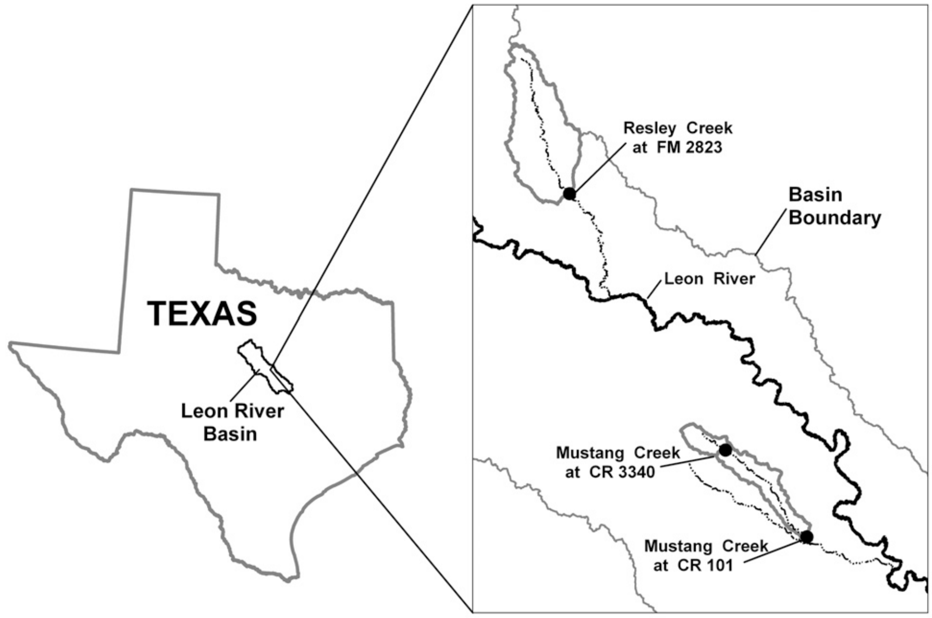

Fig. 1. Location of study sites in central Texas.

Materials and Methods

Site Description

Three small central Texas streams were selected for this study to represent a range of typical sampling site conditions (Fig. 1). As recommended by USEPA (1997) and Harmel et al. (2006b), sites were selected on based on accessibility, channel stability, and presence of a stable flow control point, with preference given to sites with previous monitoring activity. While flow conditions were similar for the sites—intermittent with high flow occurring in spring and fall storm events and with no flow during extended dry periods—the physical site and watershed conditions were quite different. The Resley Creek site at FM2823 (129-km^2 drainage area) was located in a natural channel upstream of a road bridge that exerted little if any hydraulic influence on flow. Resley Creek receives flow from a municipal wastewater treatment plant, but this inflow enters the creek >16 km upstream of the sampling site. The lower Mustang Creek site at CR101 (55-km^2 drainage area) was in a natural channel downstream of a low-head dam and upstream of a low water crossing. The upper Mustang Creek site at CR3340 (15-km^2 drainage area) was located at the upstream edge of a 15.2-m-wide, 2.1-m-tall box culvert. This site had the potential for incomplete mixing as it receives inflow from a spring and from the roadway immediately upstream of the sampling location. The Resley Creek site has limited historical streamflow data but is currently ungauged for flow. The Mustang Creek sites have been gauged since January 2005.

Sample Collection and Analysis

To address typical monitoring conditions, the present study focused on small streams that can be sampled by both manual and automated techniques. Certain agencies such as the USGS with highly trained personnel routinely monitor stream water quality during high flow conditions by utilizing bridge crossings and specialized equipment, but most research and assessment projects do not have the personnel or resources to safely conduct such sampling. Thus, wadeable streams (USEPA, 2006) under wadeable flow conditions were evaluated in a manner similar to that of Ging (1999). A detailed description of the sample collection, storage, and analysis methodology appears subsequently.

Manual Sample Collection

Between March 2006 and May 2008, stormwater quality was assessed with both manual grab sampling and integrated sampling techniques. There was no portion of the storm hydrograph that was targeted for sampling. Similar to the USGS approach, staff were mobilized and simply attempted to arrive at the site during the event, which can even be difficult in short-lived events at remote sites. Upon arrival at each site, stormwater was sampled if the flow rate was sufficient for sampling and if the stream was safely wadeable. Under these conditions, the following procedural steps were utilized in order:

- The water depth (stage) read from a staff plate was recorded.
- The flow rate was determined with a Flow Tracker Handheld Acoustic Doppler Velocimeter (SonTek, San Diego, CA) by the established USGS methodology (Buchanan and Somers, 1976).
- The stage was again recorded.
- Water quality samples were collected with two manual techniques, grab and integrated sampling. Sample collection was performed as rapidly as possible (within 10 min or less) to minimize temporal influences.
- Grab samples (1 L) were collected at each of three locations in the channel (channel thalweg, 50 cm in from the left edge of flow, and 50 cm in from the right edge of flow) and at two depths for each location (water surface and 15 cm from the stream bottom). If flow depth was <30 cm, only samples 15 cm from the stream bottom were collected at each location.
- Then, the USGS EWI technique (USGS, 1999) with a DH-81 sampler (USGS Hydrologic Instrumentation Facility, Stennis Space Center, MS) was used to collect cross-sectionally integrated samples within 8 to 10 equal-width vertical stream sections. With this technique, water is "continuously" collected as the sampler is moved vertically in these cross-section intervals. The resulting composite sample was churned in the field and subsampled to produce a 1-L sample. Between each site, the sampler and churn were rinsed three times with deionized water and three times with stream water.
- A handheld YSI 650MDS logger with a YSI 600XL multiparameter water quality probe (YSI Inc., Yellow Springs, OH) was used to determine dissolved oxygen (DO), pH, specific conductance, and temperature at the vertical midpoint of depth at the left and right banks and the channel thalweg.
- The stage was recorded a last time. Stage, and thus flow, was measured three times to confirm that flow variability was negligible during sample collection. If substantial change in flow would have occurred, then temporal changes in concentrations as well as spatial (cross-sectional) changes would have confounded the analyses.

Automated Sample Collection

Seven of the 31 storm events at the Mustang Creek CR3340 and CR101 sites were also sampled with automated equipment. At these sites, an ISCO 6712 automated sampler with an ISCO 730 bubbler water level meter (Teledyne-ISCO, Lincoln, NE) collected frequent discrete stormwater quality samples (sampling interval 2.54-mm volumetric depth) based on protocols by Harmel et al. (2003, 2006b). The sampler intake position was fixed approximately 1 to 10 cm above the streambed to avoid burial in deposited sediment.

Sample Storage and Analysis

Immediately after collection, samples were chilled with ice and transported to the laboratory. Samples were stored at 4°C before analysis. Once in the laboratory, the samples were shaken and a 25-mL aliquot was removed for total P analysis. Total P was determined with a Varian 700-ES inductively coupled plasma optical emission spectrophotometer (Varian, Inc., Palo Alto, CA) at a 215.407-μm wavelength.

Then, the sediment concentration, represented by the total settleable solids concentration, in each sample was determined by allowing the sample to settle for 3 to 5 d at 4°C and decanting off a majority of the solution. The sediment slurry was dried at 116°C for 18 to 24 h. The sediment concentration was calculated as the mass of dried sediment divided by the measured volume of the collected sample. The liquid portion of each sample was analyzed for dissolved nitrate plus nitrite N (NO_3 + NO_2–N, noted hereafter as NO_3–N), ammonium N (NH_4–N), and ortho-phosphate P (PO_4–P). The samples were analyzed for NO_3–N, NH_4–N, and PO_4–P with an O.I. Analytical Flow IV colorimetric rapid flow analyzer (O.I. Analytical, College Station, TX).

This method of decanting off the liquid portion for dissolved constituent analysis and allowing the particulate constituents to settle out is a viable alternative to filtration. The settling method can increase the amount of particulates captured and makes it unnecessary to remove particulates from the filter, both of which can be difficult with filtration.

Data Analysis

Potential differences in measured dissolved and particulate constituent concentrations were evaluated for two manual sampling methods, grab and integrated (EWI), for every measured event. Concentrations produced by the two sampling techniques were compared, using the common assumption that integrated sampling produced the "true" value. Thus, the ability of grab sampling to produce the actual cross-sectional mean concentration was evaluated by determining the percent error relative to the integrated sample concentration. Relative errors for each constituent (sediment, total P, NO_3–N, NH_4–N, and PO_4–P) were first analyzed with results grouped across all sampling locations at all sites (Fig. 2). Then, results were compared for surface vs. near-bottom grab samples and for edge vs. thalweg grab samples to detect potential concentration gradients within the channel (Fig. 3 and 4). In these comparisons, the inherent measurement error was assumed to be the same for all samples analyzed for a given constituent and collected with a given storm, which is reasonable because all samples were collected with the same protocol. The USGS "rule of thumb" for determining whether a stream is well mixed (Wilde and Radtke, 2005) was also applied as an additional indicator of concentration variability. This "rule of thumb" suggests that if four parameter probe values (pH, temperature, specific conductance, and DO) taken throughout a stream cross section differ by <5%, then a single sampling point at the centroid of flow adequately represents the mean cross-sectional concentration for dissolved constituents. In addition to these graphical and mathematical comparisons, which are well suited to illustrate differences and similarities between the various sampling

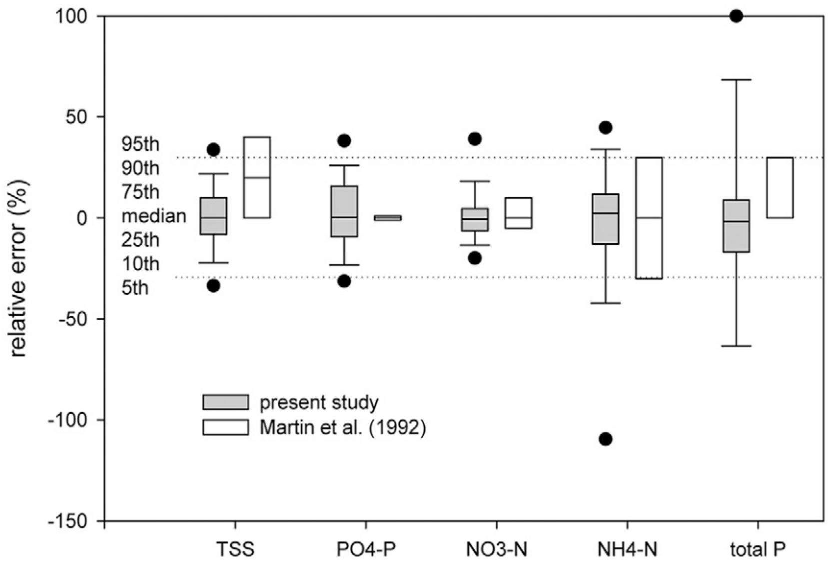

Fig. 2. Relative (%) errors for grab sample concentrations compared to integrated sample concentrations at all three study sites in central Texas. The fifth percentile value for total P does not appear because it is less than −150%. The ±30% relative error lines and results from Martin et al. (1992) are presented for comparison. TSS, total settleable solids.

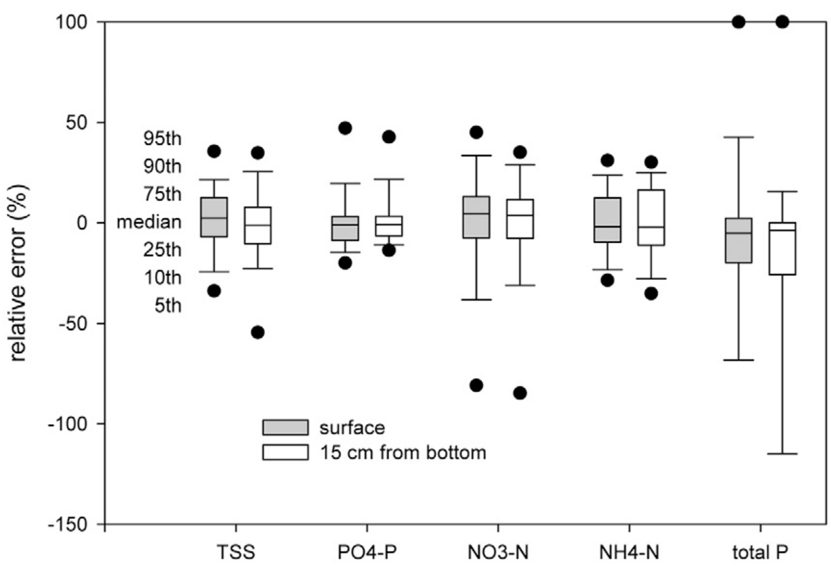

Fig. 3. Relative errors for surface and near-bottom grab sample concentrations in central Texas. The fifth percentile values for total P do not appear because they are less than −150%. TSS, total settleable solids.

techniques, statistical analyses were also applied. Specifically, a one-sample *t* test (α = 0.05, significance level) was used to determine whether the grab sample mean concentration for each constituent in each sampling event was significantly different from the "true" mean from the integrated sample.

In seven of these storm sampling events, samples were taken with three sampling techniques (automated, manual integrated [EWI], and manual grab), although direct comparisons of the concentration results were difficult because of several confounding influences introduced by these sampling techniques. As stated previously, grab samples and integrated samples were taken within a few minutes of each other, so the major difference between the two techniques was the location of sample collection within the cross section. In contrast, automated sampling produced samples from a single location within each cross section but produced numerous samples for each event. Another difficulty of comparing constituent concentrations measured in storm events with automated, manual integrated, and manual grab sampling is the lack of an accepted procedure that produces a "true" value. Obviously, multiple grab samples (within the cross section) and integrated samples better represent within-channel variability, but these techniques do not capture temporal variability unless they are repeated during each storm event. The opposite is true for automated sampling, which better captures temporal concentration variability, but may not capture cross-sectional variability. Because of these confounding factors, mainly graphical techniques were relied on to compare the three techniques, although statistical methods were also utilized to enhance the comparison. Two-sample *t* tests were used to evaluate potential differences in mean concentrations from automated and grab sampling for each constituent in each sampling event (α = 0.05, significance level), and coefficient of variation (CV) values were calculated for grab sample and automated sample results to assess potential differences in the magnitude of cross-sectional (within-channel) and temporal (within-storm) variability.

Results and Discussion

Characteristics of Storm Sampling Events

A total of 31 storm events were sampled during the 27-mo study period. Summary flow and water quality data for these events appear in Table 1. The Resley Creek and lower Mustang Creek sites produced adequate sampling depths more often than did the upper Mustang Creek site, which as described above, was located in a very wide (15.2 m) box culvert that often produced shallow flow conditions even in storm events. As per study design, samples were not collected during extreme high flow conditions (>2-yr return period peak flow rates based on regional relationships documented by Raines [1998]) due to safety concerns.

Comparison of Integrated and Grab Sample Concentrations

A total 146 grab samples were collected in these 31 storm events, and their concentrations were compared to those of corresponding integrated samples. No distinct differences in relative error were evident between the three sites in spite of substantial variation in channel geometry, flow conditions, and constituent concentrations; therefore, results were grouped across all sites (Fig. 2). More than ±5% error occurred for 52% of the NO_3–N samples, for 74% of the NH_4–N samples,

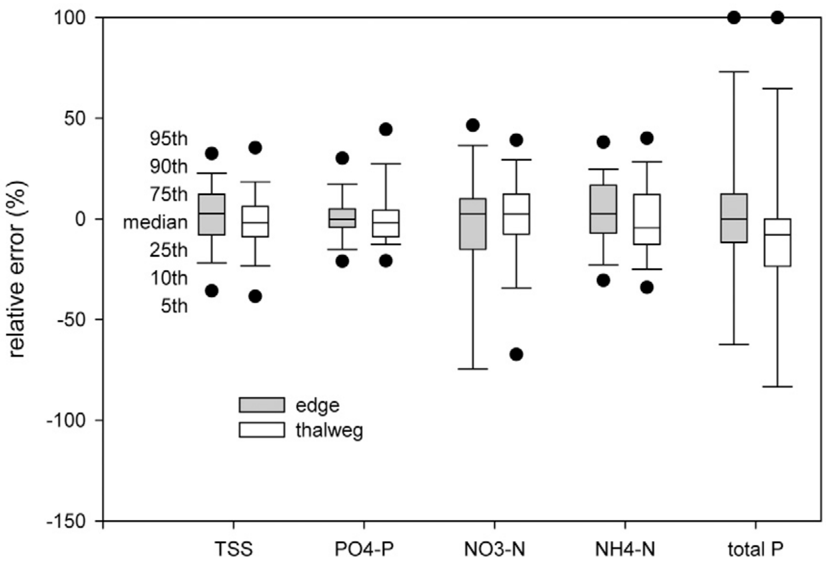

Fig. 4. Relative errors for edge of stream and thalweg grab sample concentrations in central Texas. The fifth percentile values for NO_3–N and total P do not appear because they are less than −150%. TSS, total settleable solids.

and for 81% of the PO_4–P samples. In fact, 7–24% of the samples had more than ±30% error for dissolved constituents. For the particulate constituents, more than ±5% error occurred for 68% of the total P samples and for 71% of the sediment samples. In addition, 32 and 12% of the total P and sediment samples had more than ±30% error. These results indicate the potential for substantial concentration variability within small-stream cross sections for both dissolved and particulate constituents. While the magnitudes of these errors for particulate constituents were not surprising because of presumed vertical and horizontal concentration gradients, the magnitudes of errors for dissolved constituents were larger than expected.

The distribution of relative errors for NO_3–N and NH_4–N were similar to those reported for larger watersheds (1400–14,000 km^2) by Martin et al. (1992), but the distributions for PO_4–P, total P, and sediment were noticeably different (Fig. 2). Percent errors for total P and sediment were positively biased in Martin et al. (1992) on streams with mean depths of 1.1 to 27.4 m but were symmetrical in the present study with mean stream depths <1 m. The relative errors for PO_4–P were smaller in Martin et al. (1992), but both studies produced median errors near 0%.

Since individual comparisons yielded at times large relative errors, potential differences due to sampling location (surface vs. near-bottom and stream edge vs. thalweg) were examined graphically (Fig. 3 and 4). As expected for dissolved constituents, this examination indicated little bias and little vertical or horizontal stratification. Surprisingly, sediment concentrations showed little bias and little difference between surface and near-bottom grab samples and between edge of stream and thalweg grab samples. Higher total P concentrations relative to those of integrated samples did, however, occur in the stream thalweg and near the stream bottom. The reason for differing results for total P and sediment is unknown.

When one-sample *t* tests were applied to each constituent in each storm event, 39 of 155 overall comparisons (25%) showed significant differences between mean grab sample concentrations and integrated sample concentrations (Table 2). It was no surprise that mean concentrations from grab samples at multiple points within the cross section were closer to the "true" mean (as represented by the integrated sample concentration) than were individual grab samples, but even three to six grab samples often produced concentrations that were significantly different than the "true" mean.

One possible contributor to differences in measured concentrations may be uncertainty contributed by sample collection, storage/preservation, and analysis, which commonly ranges from ±14 to 28% and even higher (Harmel et al., 2006a, 2009). Although uncertainty does contribute to these differences, it is probably not the sole cause of differences that might indeed be real. As noted by Martin et al. (1992) and Harmel et al. (2006a), sample collection rarely receives the same attention as sample preservation, storage, and analysis procedures in standard methods and quality assurance programs, even

Table 1. Summary data for storm sampling events presented as mean (standard deviation) of all values for each site in central Texas.

Data type	Upper Mustang Creek site (CR3340)	Lower Mustang Creek site (CR101)	Resley Creek site (FM2823)
Sampling events (*n*)	5	13	13
Flow rate (m^3 s^{-1})	1.02 (0.40)	1.13 (1.20)	2.01 (2.95)
Depth (m)	0.15 (0.03)	0.26 (0.11)	0.23 (0.10)
Width (m)	15.2 (0.0)	9.3 (5.7)	8.2 (3.5)
Temperature (°C)	22.7 (3.0)	20.2 (3.8)	17.6 (5.3)
Specific conductance (μS cm^{-1})	301 (82)	358 (111)	567 (230)
DO† (mg L^{-1})	9.5 (4.1)	9.4 (1.8)	8.9 (1.4)
pH	7.6 (0.3)	7.6 (0.3)	7.6 (0.5)
Sediment (mg L^{-1})	223 (199)	246 (175)	351 (256)
NO_3–N (mg L^{-1})	0.43 (0.24)	0.24 (0.23)	0.84 (0.29)
NH_4–N (mg L^{-1})	0.09 (0.08)	0.06 (0.02)	0.26 (0.23)
PO_4–P (mg L^{-1})	0.08 (0.05)	0.07 (0.04)	0.31 (0.16)
Total P (mg L^{-1})	0.21 (0.46)	0.09 (0.12)	0.52 (0.38)

† DO, dissolved oxygen.

Table 2. Number of significant differences out of the total number of comparisons between mean grab sample concentrations and integrated sample concentrations in central Texas, as determined with one-sample *t* tests ($\alpha = 0.05$).

Data type	Upper Mustang Creek site (CR3340)	Lower Mustang Creek site (CR101)	Resley Creek site (FM2823)	Constituent totals
Sediment (mg L^{-1})	0/5	2/13	2/13	4/31
NO_3–N (mg L^{-1})	2/5	3/13	1/13	6/31
NH_4–N (mg L^{-1})	1/5	3/13	3/13	7/31
PO_4–P (mg L^{-1})	0/5	7/13	5/13	12/31
Total P (mg L^{-1})	2/5	3/13	5/13	10/31
Site totals	5/25	18/65	16/65	Overall total 39/155

though sample collection can be the largest source of uncertainty in measured water quality data (Harmel et al., 2009).

Practical Implications

When comparing corresponding grab and integrated storm samples (collected at approximately the same time with very little if any difference in flow rate), potential differences are driven by the degree of concentration variability within the cross section; therefore, the location of sample collection is the important consideration. In the present study, single grab samples at six different locations within each cross section produced more than ±5% error in 69% of the comparisons and more than ±30% error in 17% of the comparisons, which indicates the potential for considerable concentration variability. In contrast, little difference was observed between surface and near-bottom grab samples and between edge of stream and thalweg grab samples (except for total P), which indicates well-mixed conditions with little vertical or horizontal concentration variability. The USGS "rule of thumb" (Wilde and Radtke, 2005) also indicated well-mixed conditions, as only 12 of the 319 "rule of thumb" comparisons yielded differences >5%.

Although these comparisons yielded differing results, they do indicate the potential for considerable variability in both dissolved and particulate constituent concentrations within small-stream cross sections. Thus, this variability should be captured to the greatest extent possible or at least accounted for in reporting uncertainty associated with stormwater quality data. Integrated sampling, which best captures this cross-sectional variability, is the preferred manual sampling method on small streams that can be safely waded or sampled from a bridge (USGS, 1999; Ging, 1999); nothing in this study contradicts that recommendation. However, even the USGS protocol allows for grab sampling at the centroid of flow in small well-mixed streams, especially at sites that cannot be safely sampled with integrated techniques (USGS, 1999). When grab sampling is employed, Martin et al. (1992) advised that a thorough evaluation of cross-sectional variability be conducted and that vertical gradients in particulate constituents be considered. In addition, the present study showed that the mean of multiple grab samples better represents the "true" cross-sectional mean concentration than a single grab sample, although some significant differences did occur. All of these considerations should be kept in mind when using single grab samples to characterize stormwater quality.

Comparison of Automated, Integrated, and Grab Sample Concentrations

For seven of these storm events, samples were also taken with an automated sampler. Similar to the results discussed previously, graphical comparison of concentrations produced by these sampling techniques on a storm-by-storm basis showed that cross-sectional variability was quite large for some sampling events but quite small for others (notice the spread of PO_4–P and sediment concentrations produced by grab samples, for example, in Fig. 5 and 6). As expected, graphical comparison also illustrated considerable temporal variability within storm events (notice the spread of concentrations produced by automated sampling in Fig. 5 and 6). As a result of the inconsistent nature of concentration variability within channel cross sections and within storm events, the three sampling techniques at times produced similar PO_4–P, NO_3–N, NH_4–N, and sediment concentrations (e.g., Storms #1 and #5 in Fig. 5); however, they produced obviously different concentrations for other storms (e.g., Storm #6 in Fig. 6). Two-sample *t* tests of mean concentrations produced by automated sampling and manual grab sampling also confirmed the inconsistent nature of concentration variability by indicating significant differences

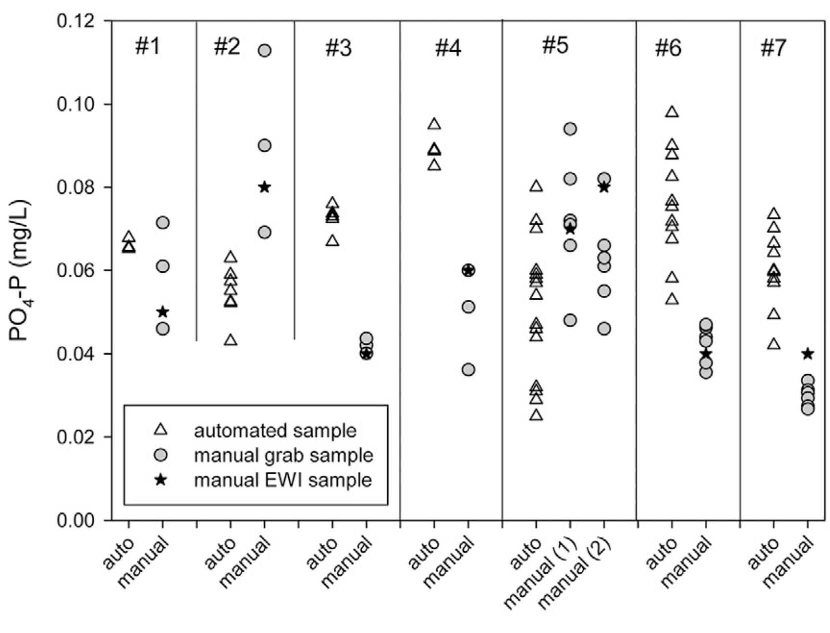

Fig. 5. PO_4–P concentrations produced by each sampling technique (automated, integrated, and grab) for seven storm sampling events in central Texas. Two sets of manual samples were collected for Storm #5. EWI, USGS Equal-Width-Increment procedure.

for some constituents in individual storms but not for others.

As shown in Fig. 7, the flow rate at the time of sampling for manual techniques tended to be less than for automated sampling because of the difficulty of mobilizing staff and reaching remote sites and due to safety concerns associated with manual sampling at high flows. The differences in flow rate between the manual and automated sample collection did at times appear to contribute to differences in sediment concentrations (notice Storms #5 and #6 in Fig. 6 and 7), but flow rate differences did not translate into corresponding differences in measured concentrations of dissolved constituents. This is attributed to poor correlations between flow rate and dissolved concentrations (average p = 0.24, average adjusted R^2 = 3%) and to better correlations between flow rate and particulate concentrations (average p = 0.44, average adjusted R^2 = 40%).

Because of the potential for considerable cross-sectional (within-channel) and temporal (within-storm) variability, their magnitudes were compared to determine which was more important to capture in sample collection. For NO_3–N concentrations, the within-storm variability (average CV = 0.52) was greater than the cross-sectional variability (average CV = 0.08) in every storm. Similarly, for sediment concentrations, the within-storm variability (average CV = 0.47) was greater than the cross-sectional variability (average CV = 0.19) in all but one event. In contrast, CV values for NH_4–N and PO_4–P, did not clearly indicate whether within-storm or within-channel variability was larger. For NH_4–N, the average CV values were 0.24 and 0.22, and for PO_4–P, the average CV values were 0.12 and 0.17.

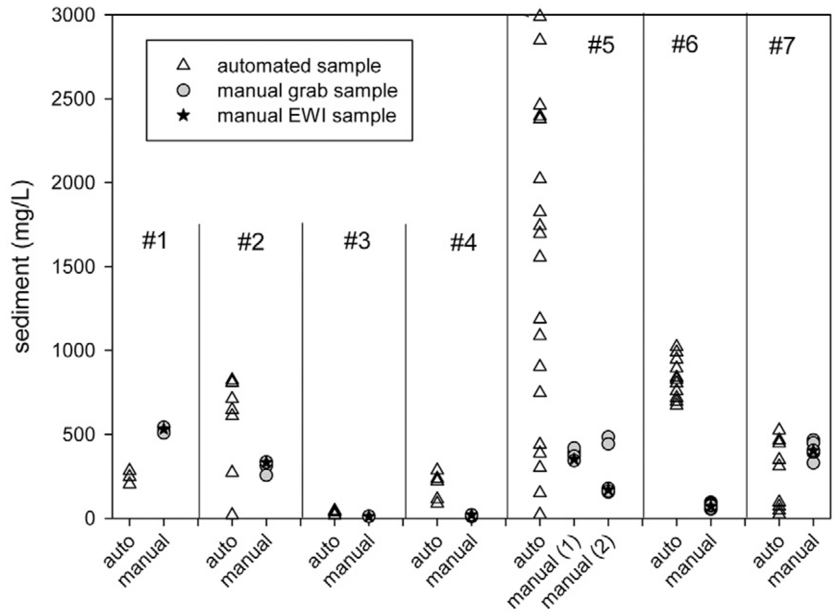

Fig. 6. Sediment concentrations produced by each sampling technique (automated, integrated, and grab) for seven storm sampling events in central Texas. Two sets of manual samples were collected for Storm #5. EWI, USGS Equal-Width-Increment procedure.

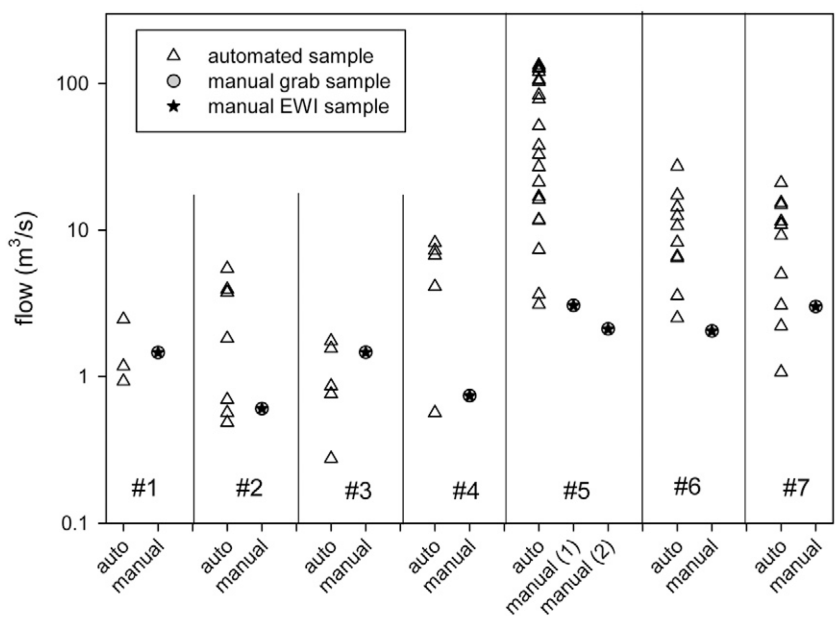

Fig. 7. Flow rate at the time of sample collection for each sampling technique (automated, integrated, and grab) for seven storm sampling events in central Texas. Notice that the flow rate is the same for integrated and grab sampling, since they were taken at approximately the same time. Two sets of manual samples were collected for Storm #5. EWI, USGS Equal-Width-Increment procedure.

Practical Implications

When comparing grab, integrated, and automated storm sampling techniques, both the time and location of sample collection are important considerations. In the present study, concentrations from manual sampling (grab and integrated) were at times similar and at times very different from concentrations produced by automated sampling, presumably due to the interaction between within-channel and within-storm variability and the influence of flow rate on particulate concentrations.

These results coupled with those of Ging (1999) have important practical implications related to sampling stormwater quality in small streams. As stated previously, integrated sampling is often the preferred USGS sampling technique because it provides very accurate determinations of concentrations at a point in time. However, a single integrated sample does not necessarily accurately estimate the "true" storm mean concentration because it does not capture within-storm (temporal) concentration variability, which in the present study tended to be greater than the cross-sectional variability. Therefore, if integrated sampling is employed, collection of multiple integrated samples throughout the storm duration is recommended to reduce the uncertainty of measured concentrations, although time constraints and logistical difficulties for such sampling at multiple sites can be restrictive. If integrated or automated sampling cannot be conducted because of resource limitations or site conditions, then collection of multiple grab samples from various locations in the cross section throughout the storm event is recommended. Single grab samples are discouraged because they do not capture either the cross-sectional or temporal variability and likely will not produce an accurate estimate of the actual concentration.

Automated samplers are increasingly used because typical projects do not have the resources to maintain an adequate on-call field staff to perform intensive manual storm sampling at multiple sites. As presented in Harmel et al. (2006b), automated samplers do require vigilant and proactive maintenance. Weekly or biweekly checks of the power supply, stage recorder, dessicant strength, and sampler lines can minimize malfunctions during storm events (during prolonged dry periods, fuses and the sample pump should also be regularly checked). Sampling efforts that do not use an intensive proactive maintenance procedure often suffer from frequent malfunctions and thus missed data and missed samples. While frequent sampling to capture temporal concentration variability is certainly straightforward with automated samplers, the single-intake set up of typical automated samplers prevents them from capturing cross-sectional concentration variability and thus introduces uncertainty into the measured concentration. Although other factors that affect uncertainty in data produced by automated samplers (sampling threshold, sampling interval, sample type—discrete or composite) have been evaluated and discussed (e.g., Shih et al., 1994; Miller et al., 2000, 2007; Stone et al., 2000; Harmel et al., 2002, 2003, 2006b; King and Harmel, 2003, 2004; Harmel and King, 2005), the only known evaluations of a single-intake (sample collection) point are the present study, McGuire et al. (1980), and Ging (1999). While previous publications (Martin et al., 1992; Ging, 1999; Harmel et al., 2003, 2006b) urged caution in the use of single-intake autosamplers for collection of suspended sediment and sediment-associated constituents, only McGuire et al. (1980) expressed similar caution for dissolved constituents. The present study showed that considerable differences can occur between automated (single-intake) and integrated samples for both dissolved and particulate constituents. Development of vertical intakes that extend throughout the water column and capture vertical concentration gradients would decrease these differences; however, no such intakes are commercially available (although floating intakes are occasionally used, see McGuire et al., 1980). Ahyerre et al. (2001) did develop a mobile sampling system to sample on a fine vertical resolution, but the system requires manual operation.

Another alternative is to vertically orient sampler intakes, which betters capture vertical gradients at sites with adequate flow depth to completely submerge the intake; however, many ephemeral sites often have much too shallow flow. Based on Harmel et al. (2002), sampling should be initiated at a "low" flow threshold. With a horizontally oriented intake this threshold can be as low as 1.2 to 2.5 cm, but with a vertically oriented intake ~15 cm of depth is required to submerge the intake. Thus, an alternative solution involves development of a relationship between concentrations at the sampler intake and mean concentrations as determined by integrated sampling at a range of discharges (e.g., Ging, 1999). With such a relationship, concentrations at the intake can be adjusted to represent mean concentrations for the total cross section.

Whichever sampling technique is used, single or infrequent sample collection during storm events is discouraged. This is an important consideration because the idea of low-frequency sampling coupled with regression methods is occasionally proposed for small streams to reduce costs and technical difficulties associated with intensive storm sampling. Regression methods, which utilize relationships between measured concentrations and flow rates (e.g., Cohn et al., 1989), have been successfully applied to large rivers (e.g., Robertson, 2003; Haggard et al., 2003). Toor et al. (2008), however, demonstrated poor results for small watersheds because of poor correlations between mean daily flow and measured concentrations.

Conclusions

The objective of this study was to evaluate the impact of common sampling techniques on measured stormwater quality data for small streams. When applying these results, it is important to consider that they apply most directly to "wadeable" streams. Grab sampling results demonstrated the possibility of more than ±5% and even more than ±30% error compared with the often preferred method of integrated sampling for both dissolved and particulate constituents due to cross-sectional concentration variability. In contrast, comparisons of constituent concentrations between stream edge and thalweg samples and between near-bottom and surface samples indicated little vertical or horizontal stratification except for total P. Similarly, the USGS "rule of thumb" indicated well-mixed conditions (little cross-sectional variability). Comparisons of grab, integrated, and automated storm sampling for seven sampling events indicated that measured concentrations can be similar or quite different between the sampling techniques, depending on the time and location of sample collection, presumably due to the complex interaction between within-channel and within-storm variability and the influence of flow rate on constituent transport.

These potential temporal (within-storm) and cross-sectional (within-channel) differences in constituent concentrations should not be ignored in measuring stormwater quality. If integrated sampling is used, then the resulting concentration is likely an accurate estimate of the "true" cross-sectional mean concentration at that moment but is not necessarily an accurate estimate of the "true" storm mean concentration due to temporal variability. Similarly, a single grab sample cannot be assumed to accurately represent the cross-sectional mean concentration at a moment in time or the "true" within-storm mean concentration. Thus, if manual sampling is employed, then collection of multiple grab samples within the cross section or multiple integrated samples throughout the storm duration is recommended to reduce the uncertainty of measured concentrations; however, the time constraints and logistical difficulties of employing either of these manual sampling techniques at multiple sites should be kept in mind. If automated sampling is selected, then frequent sampling to capture temporal concentration variability is not difficult, but accounting for cross-sectional concentration variability can be because of the single-intake setup typical in automated samplers. To overcome this deficiency and increase the accuracy of concentration data produced by automated samplers, capturing potential concentration gradients is required. This improvement can be accomplished with a vertically integrated intake or with a relationship between intake concentrations and mean concentrations.

Whichever sampling technique is chosen, storm sampling on small streams is a difficult endeavor. Runoff events are often short-lived; sampling conditions are at times hazardous; travel time can be substantial; and equipment and personnel are expensive. In addition, the resulting data are only estimates with inherent uncertainty (Harmel et al., 2006a, 2009). Thus, the value of decreased uncertainty vs. the increased cost of additional samples or improved techniques should be carefully weighed. In spite of these difficulties, accurate quantification of stormwater quality is increasingly important for research, management, decision-making, and regulation. It is hoped that the present research contributes to increased awareness of the implications and uncertainties of various sampling techniques for small streams and helps improve the quality of measured data collected in such studies.

Acknowledgments

This research was a component of the USDA Conservation Effects Assessment Project. Trixie Delisle, a former USGS hydrological technician, performed all field sample collection. Her dedication to the project and her expertise on sampling methodology are greatly acknowledged. Mention of trade names or commercial products is solely for the purpose of providing specific information and does not imply recommendation or endorsement by the United States Department of Agriculture.

References

Ahyerre, M., G. Chebbo, and M. Saad. 2001. Nature and dynamics of water sediment interface in combined sewers. J. Environ. Eng. 127:233–239.

Buchanan, T.J., and W.P. Somers. 1976. Chapter A8: Discharge measurements at gaging stations. *In* Book 3: Techniques of water-resources investigations of the United States Geological Survey. Available at http://pubs.usgs.gov/twri/twri3a8/ (verified 11 May 2010). USGS, Arlington, VA.

Cohn, T.A., L.L. DeLong, E.J. Gilroy, R.M. Hirsch, and D.K. Wells. 1989. Estimating constituent loads. Water Resour. Res. 25:937–942.

Edwards, T.K., and D. Glysson. 1999. Chapter C2: Field methods for measurement of fluvial sediment. *In* Book 3: Techniques of water-resources investigations of the United States Geological Survey. Available at http://pubs.usgs.gov/twri/twri3-c2/ (verified 11 May 2010). USGS, Arlington, VA.

Ging, P. 1999. Water-quality assessment of south-central Texas: Comparison of water quality in surface-water samples collected manually and by automated samplers. USGS Fact Sheet FS-172-99. USGS, Austin, TX.

Haggard, B.E., T.S. Soerens, W.R. Green, and R.P. Richards. 2003. Using regression methods to estimate stream phosphorus loads at the Illinois River, Arkansas. Appl. Eng. Agric. 19:187–194.

Harmel, R.D., R.J. Cooper, R.M. Slade, R.L. Haney, and J.G. Arnold. 2006a. Cumulative uncertainty in measured streamflow and water quality data for small watersheds. Trans. ASABE 49:689–701.

Harmel, R.D., and K.W. King. 2005. Uncertainty in measured sediment and nutrient flux in runoff from small agricultural watersheds. Trans. ASAE 48:1713–1721.

Harmel, R.D., K.W. King, B.E. Haggard, D.G. Wren, and J.M. Sheridan. 2006b. Practical guidance for discharge and water quality data collection on small watersheds. Trans. ASABE 49:937–948.

Harmel, R.D., K.W. King, and R.M. Slade. 2003. Automated storm water sampling on small watersheds. Appl. Eng. Agric. 19:667–674.

Harmel, R.D., K.W. King, J.E. Wolfe, and H.A. Torbert. 2002. Minimum flow considerations for automated storm sampling on small watersheds. Tex. J. Sci. 54:177–188.

Harmel, R.D., D.R. Smith, K.W. King, and R.M. Slade. 2009. Estimating storm discharge and water quality data uncertainty: A software tool for monitoring and modeling applications. Environ. Model. Softw. 24:832–842.

King, K.W., and R.D. Harmel. 2003. Considerations in selecting a water quality sampling strategy. Trans. ASAE 46:63–73.

King, K.W., and R.D. Harmel. 2004. Comparison of time-based sampling strategies to determine nitrogen loading in plot-scale runoff. Trans. ASAE 47:1457–1463.

Martin, G.R., J.L. Smoot, and K.D. White. 1992. A comparison of surface-grab and cross sectionally integrated stream-water-quality sampling methods. Water Environ. Res. 64:866–876.

McGuire, P.E., T.C. Daniel, D. Stoffel, and B. Andraski. 1980. Sample intake position and loading rates from nonpoint source pollution. Environ. Manage. 4:73–77.

Miller, P.S., B.A. Engel, and R.H. Mohtar. 2000. Sampling theory and mass load estimation from watershed water quality data. ASAE Pap. 003050. ASAE, St. Joseph, MI.

Miller, P.S., R.H. Mohtar, and B.A. Engel. 2007. Water quality monitoring strategies and their effects on mass load calculation. Trans. ASABE 50:817–829.

Raines, T.H. 1998. Peak-discharge frequency and potential extreme peak discharge for natural streams in the Brazos River basin, Texas. USGS Water-Resources Invest. Rep. 98-4178. USGS, Austin, TX.

Robertson, D.M. 2003. Influence of different temporal sampling strategies on estimating total phosphorus and suspended sediment concentration and transport in small streams. J. Am. Water Resour. Assoc. 39:1281–1308.

Shih, G., W. Abtew, and J. Obeysekera. 1994. Accuracy of nutrient runoff load calculations using time-composite sampling. Trans. ASAE 37:419–429.

Stone, K.C., P.G. Hunt, J.M. Novak, M.H. Johnson, and D.W. Watts. 2000. Flow-proportional, time-composited, and grab sample estimation of nitrogen export from an Eastern Coastal Plain watershed. Trans. ASAE 43:281–290.

Toor, G., R.D. Harmel, B.E. Haggard, and G. Schmidt. 2008. Evaluation of regression methodology with low-frequency water quality sampling to estimate constituent loads for ephemeral watersheds in Texas. J. Environ. Qual. 37:1847–1854.

USEPA. 1997. Monitoring guidance for determining the effectiveness of nonpoint-source controls. EPA 841-B-96-004. USEPA, Washington, DC.

USEPA. 2006. Wadeable streams assessment: A collaborative survey of the nation's streams. EPA Rep. 841-B-06-002. USEPA, Washington, DC.

USGS. 1999. Section A: National field manual for collection of water-quality data. *In* Book 9: Handbooks for water-resources investigations. Techniques of water-resources investigations of the United States Geological Survey. Available at http://water.usgs.gov/owq/FieldManual/ (verified 11 May 2010). USGS, Arlington, VA.

Wells, F.C., W.J. Gibbons, and M.E. Dorsey. 1990. Guidelines for collection and field analysis of water-quality samples from streams in Texas. USGS Open-File Rep. 90-127. USGS, Austin, TX.

Wilde, F.D., and D.B. Radtke. 2005. Chapter A6: Field measurements. *In* Book 9: Handbooks for water-resources investigations. National field manual for the collection of water-quality data. Techniques of water-resources investigations of the United States Geological Survey. Available at http://water.usgs.gov/owq/FieldManual/Chapter6/Ch6_contents.html (verified 11 May 2010). USGS, Washington, DC.