CUMULATIVE UNCERTAINTY IN MEASURED STREAMFLOW AND WATER QUALITY DATA FOR SMALL WATERSHEDS

R. D. Harmel, R. J. Cooper, R. M. Slade, R. L. Haney, J. G. Arnold

ABSTRACT. *The scientific community has not established an adequate understanding of the uncertainty inherent in measured water quality data, which is introduced by four procedural categories: streamflow measurement, sample collection, sample preservation/storage, and laboratory analysis. Although previous research has produced valuable information on relative differences in procedures within these categories, little information is available that compares the procedural categories or presents the cumulative uncertainty in resulting water quality data. As a result, quality control emphasis is often misdirected, and data uncertainty is typically either ignored or accounted for with an arbitrary margin of safety. Faced with the need for scientifically defensible estimates of data uncertainty to support water resource management, the objectives of this research were to: (1) compile selected published information on uncertainty related to measured streamflow and water quality data for small watersheds, (2) use a root mean square error propagation method to compare the uncertainty introduced by each procedural category, and (3) use the error propagation method to determine the cumulative probable uncertainty in measured streamflow, sediment, and nutrient data. Best case, typical, and worst case "data quality" scenarios were examined. Averaged across all constituents, the calculated cumulative probable uncertainty (±%) contributed under typical scenarios ranged from 6% to 19% for streamflow measurement, from 4% to 48% for sample collection, from 2% to 16% for sample preservation/storage, and from 5% to 21% for laboratory analysis. Under typical conditions, errors in storm loads ranged from 8% to 104% for dissolved nutrients, from 8% to 110% for total N and P, and from 7% to 53% for TSS. Results indicated that uncertainty can increase substantially under poor measurement conditions and limited quality control effort. This research provides introductory scientific estimates of uncertainty in measured water quality data. The results and procedures presented should also assist modelers in quantifying the "quality" of calibration and evaluation data sets, determining model accuracy goals, and evaluating model performance.*

Keywords. Error propagation, Nonpoint-source pollution, Nutrient transport, Water quality monitoring.

Although sampling to assess transport of water quality constituents in runoff has been conducted for many years, relatively little information is available on the uncertainty of measured data. The need to understand uncertainty in measured water quality data has recently increased because of the adverse impact of diffuse or nonpoint-source pollution on rivers, lakes, and coastal waters (USEPA, 2000) and the intensified disputes regarding relative contributions of diffuse and point-source pollution (e.g., McFarland and Hauck, 2001). The issue of uncertainty is particularly important in water quality modeling because models are increasingly used to guide decisions regarding water resource policy, management, and regula-

tion (Beck, 1987; Sharpley et al., 2002). It is important that decision makers appreciate the uncertainty in measured water quality data and its effect on model output. The scientific community, however, has not compiled an adequate understanding on the uncertainty of measured runoff water quality data and has not adequately described the effects of uncertainty on water quality management.

Although the issue of uncertainty is germane to many water resource issues such as water quality standards and discharge permitting, the need for scientifically defensible uncertainty estimates is particularly evident in Total Maximum Daily Load (TMDL) projects. In typical TMDLs, measured water quality data are used to calibrate watershed models and evaluate their output, which often are the basis of source load allocation. Although TMDLs are required to include a margin of safety to account for uncertainty in allocations (40 CFR 130.7), scientific guidance on margin of safety determination is limited. As a result, arbitrary margins of safety are often assigned to account for uncertainty, in spite of research that indicates the need for rigorous uncertainty or risk analysis (e.g., Hession et al., 1996). Quantifying the uncertainty associated with measured water quality data (error propagation) is a vital component of scientifically defensible TMDL modeling analyses.

While previous research has produced valuable knowledge of the uncertainty related to various sampling procedures, no complete error propagation analysis, including all error sources, has been conducted on measured water quality

The authors are **R. Daren Harmel, ASABE Member Engineer,** Agricultural Engineer, USDA-ARS, Temple, Texas; **Richard J. Cooper,** Environmental Scientist, The Macaulay Institute, Craigiebuckler, Aberdeen, U.K.; **Raymond M. Slade,** Hydrologist and Surface Water Specialist (retired), USGS, Texas District, Austin, Texas; **Richard L. Haney,** Soil Scientist, and **Jeff G. Arnold,** Agricultural Engineer, USDA-ARS, Temple, Texas. **Corresponding author:** R. Daren Harmel, USDA-ARS, 808 E. Blackland Rd., Temple, TX 76502; phone: 254-770-6521; fax: 254-770-6561; e-mail: dharmel@spa.ars.usda.gov.

data for small watersheds. Uncertainty in measured water quality data is introduced by four procedural categories: streamflow measurement, sample collection, sample preservation/storage, and laboratory analysis. Each of these procedural categories is described in detail in the following section. Data processing and management can also contribute uncertainty to measured data because of missing data, assumptions made to estimate missing values, and mistakes in data management and reporting. This source of uncertainty was not addressed in this research because the uncertainty results from equipment malfunction or personnel mistakes and not from random statistical variation as affected by the appropriate procedures. However, when project-specific quantitative estimates of uncertainty data processing and management are available, that information can be included in the proposed error propagation methodology.

The subsequent discussion and analyses focus on uncertainty related to streamflow measurement, sample collection, sample preservation/storage, and laboratory analysis procedures for measurement of streamflow, nitrogen (N), phosphorus (P), and total suspended sediment (TSS) data from small watersheds. Although the categorization of various watershed scales is difficult because of the variability of hydroclimatic settings and the arbitrary selection of watershed outlet locations, the present research is generally applicable for the field (<50 ha) and small watershed (<10,000 ha) scale. For larger watersheds, streamflow measurement and water quality sampling procedures must be adjusted to adequately quantify discharge and transport processes at that scale. It is important to note that the terms "error" and "uncertainty" are synonymous throughout this article and refer to random statistical variation, as affected by the appropriate and accepted use of each specific instrument or procedure, not equipment malfunction or personnel mistakes.

STREAMFLOW MEASUREMENT

Streamflow (discharge) data are vital in most water quality sampling programs. Constituent concentration data must be coupled with streamflow data to evaluate transport mechanisms and quantify mass loads. Measurement of streamflow typically involves continuous stage (water depth) measurement with a device such as a bubbler, pressure transducer, float sensor (Buchanan and Somers, 1982), or sonic sensor. These stage data are then converted to streamflow rate with an established stage-discharge relationship based on the surveyed cross-sectional channel geometry (Buchanan and Somers, 1976; Brakensiek et al., 1979; Kennedy, 1984; Carter and Davidian, 1989). Although information on the uncertainty associated with individual steps in streamflow measurement exists (e.g., Dickinson, 1967; Pelletier, 1988; Sauer and Meyer, 1992; Schmidt, 2002), such data are seldom used to assess the overall uncertainty in resulting streamflow data.

Stage-discharge relationships are either established by a series of direct streamflow measurements at a range of stages or accompany a pre-calibrated flow control structure. Direct streamflow measurements are conducted by measuring flow depth and velocity at many locations in the stream cross-section. Typically, flow depth and velocity are measured in vertical cross-section segments, each with no more than 5% to 10% of total flow. Pelletier (1988) and Sauer and Meyer (1992) conducted thorough reviews of errors associated with direct streamflow measurement. A brief summary of their findings appears subsequently. Accurate depth measurement with a measuring rod or a sounding weight can be difficult, especially with rapid changes in stage and for channel beds composed of cobbles or covered by soft sediment. In contrast, the error associated with segment width determination is usually negligible. Measurement of flow velocity can also introduce substantial uncertainty. Commercially available current meters typically have a precision rating provided by the manufacturer. The uncertainty introduced by pulsation and turbulence can be largely overcome by taking mean measurements of velocity over time, but this increases the time required for an accurate measurement. The error associated with velocity variability within each vertical section can be decreased by taking the mean of two or more measurements (e.g., at 0.2 and 0.8 of flow depth), as opposed to simply making the common assumption that the mean velocity occurs at 0.6 of flow depth (Carter and Anderson, 1963). Error associated with flow oblique to the cross-section can be reduced with measurement adjustment in the field using established correction factors. Two methods commonly used in the computation of discharge within each vertical segment are the mid-section method and the mean section method (Rantz et al., 1982b), but the mid-section method has been shown to be more accurate (Young, 1950; Hipolito and Leoureiro, 1988).

In natural channels, a stage-discharge relationship is typically established from a series of direct streamflow measurements (Kennedy, 1984; Carter and Davidian, 1989), as described previously. The major source of uncertainty in measured streamflow for natural channels is change in channel dimensions, which can be caused by bed scour/deposition, bank erosion, vegetation changes, and debris deposition. Thus, frequent streamflow measurement and stage-discharge relationship adjustment is required to minimize the uncertainty in measured streamflow data from unstable channels. If this is neglected, the uncertainty can increase substantially. If Manning's equation is used instead of direct streamflow measurements to estimate the stage-discharge relationship, the uncertainty will increase substantially. Thus, the Manning's equation alternative should be avoided if possible.

In sites with a flow control structure, such as a weir or flume, the structure typically has an associated stage-discharge relationship. If the structure is not pre-calibrated with a known stage-discharge relationship, the direct streamflow measurement procedure discussed previously can be used to establish the relationship. Flow control structures are designed to produce accurate streamflow measurement by providing a unique and consistent relation between stage and flow rate and a constant cross-sectional area. However, improper structure installation resulting in incomplete flow capture (i.e., lateral or hyporheic bypass), unlevel structure alignment, or inappropriate approach channel characteristics can introduce considerable uncertainty (Rantz et al., 1982a). Stage measurement can also introduce uncertainty in measured streamflow. Stage sensor manufacturers provide typical precision estimates, but further error may arise if instrument readings are not adequately calibrated and if the sensor is not placed in a stilling well. Whether structures are pre-calibrated or not, their stage-discharge relationship should be confirmed periodically, at least annually, with direct streamflow measurement.

SAMPLE COLLECTION

In small watersheds, both baseflow and storm sampling may be needed to adequately characterize water quality. To adequately characterize conditions at intermittent or ephemeral sites, manual grab samples of baseflow should be taken periodically at regular time intervals not less than once per month. At that scale, samples can be taken at a single point in the flow, generally in the centroid of flow, because dissolved constituent concentrations typically are assumed to be uniform across the cross-section (Martin et al., 1992; Ging, 1999).

For storm sampling, two techniques are commonly used. The USGS Equal-Width-Increment (EWI) and Equal-Discharge-Increment (EDI) procedures are well-accepted and widely used for storm water sampling (USGS, 1999). These manual procedures take multiple depth-integrated, flow-proportional samples across the stream cross-section and produce accurate measurements of dissolved and particulate concentrations. Less intensive manual sampling, such as point sampling at random times and/or locations during runoff events, can introduce substantial uncertainty because they do not capture the spatial and temporal variability in constituent concentrations. In recent years, automated samplers with a single intake have been increasingly utilized to collect storm water quality samples. Automated samplers are able to use a consistent sampling procedure at multiple sites, to take multiple samples throughout the entire duration of runoff, and to sample within the quick hydrologic response time of small watersheds (Harmel et al., 2003).

Recent field research has been conducted recently on the relative differences between various water quality sampling strategies. Izuno et al. (1998) compared paired time- and flow-based automated sampling strategies on ten sites less than 1865 ha over three years and concluded that no differences existed between flow-based sampling and time-based sampling at intervals from 60 to 180 min. On a 2050 ha site in North Carolina, Stone et al. (2000) conducted a multi-year comparison of four sampling techniques utilizing time- and flow-based automated sampling and manual grab sampling. The sampling techniques created significant differences in the number of samples collected and in calculated N loads. Robertson and Roerish (1999) used intensive sampling to represent true annual loads from eight 1400 to 11,000 ha watersheds in an evaluation of the regression approach, which is commonly used when only limited samples can be collected. For 1-year studies, the authors recommended monthly sampling with storm chasing (traveling to sites to sample during runoff events) because of high precision in spite of 25% to 50% overestimation. For 2- to 3-year studies, the authors concluded that fixed-period semi-monthly sampling alone was adequate to provide precise and accurate load estimation.

A limitation of previous field studies is that they typically compare various estimates of constituent loads and thus evaluate relative differences (precision) in error without regard to possible deviation from the true load (accuracy). It is not possible to compare the accuracy of various sampling strategies without a known (or assumed) true value. As a result, relative comparisons are available, but little is known about the uncertainty of each. The impracticality of collecting all runoff, even from very small areas, for concentration measurement has been realized for some time (Parsons, 1954), and all the alternatives require considerable cost and commitment to make true load measurements. On small watersheds, the practical alternatives involve automated collection of a volumetric portion of the total runoff (e.g., Parsons, 1954; Edwards et al., 1976; Sheridan et al., 1996; Franklin et al., 2001; Bonta, 2002) or frequent subsamples based on small time or flow intervals (King and Harmel, 2004; Harmel and King, 2005).

In a recent study, Harmel and King (2005) addressed uncertainty estimates in measured storm water quality data from small agricultural watersheds with the assumption that the true load was represented by frequent flow-interval sampling, specifically 1.32 mm volumetric depth intervals. Referring to discharge intervals in depth units (such as mm), which represent mean runoff depth over the entire watershed, as opposed to volume units (such as m^3), normalizes discharge over various watershed sizes. This notation allows a consistent transfer of methods and results to watersheds of differing sizes. All 15 of the flow-interval strategies evaluated by Harmel and King (2005), which include sampling intervals of 1.32, 2.64, and 5.28 mm with discrete and composite sampling of 2 to 5 samples per bottle, produced cumulative load errors <10%. The authors concluded that increasing the sampling interval (decreasing the frequency of sample collection) introduced uncertainty but that increasing the number of samples composited into one bottle had no effect. Their analysis also yielded a significant positive correlation between errors in storm load estimates and within-event temporal variability of nutrient and sediment concentrations.

To avoid the expense and time required to conduct field analyses, several researchers have used analytic approaches to explore uncertainty. Miller et al. (2000) and King and Harmel (2003) developed mathematical relations for flow rate and constituent concentrations to compare various time- and flow-interval sampling strategies. The authors concluded that increased sampling intervals result in increased uncertainty and that flow-interval sampling produced less uncertainty than corresponding time-interval strategies. Shih et al. (1994) conducted a similar analysis and concluded that eight time-based samples were generally necessary to achieve load estimates with comparable accuracy of flow-based samples.

SAMPLE PRESERVATION/STORAGE

Sample preservation/storage techniques can also affect the uncertainty of water quality data. Physical, chemical, and biological processes can alter nutrient concentrations during the interval between sample collection and analysis (Lambert et al., 1992; Robards et al., 1994; Jarvie et al., 2002). The increased use of automated samplers has increased the potential for substantial alterations because of the time delay between sample collection and retrieval. Numerous factors, such as the container characteristics, storage environment, chemical preservatives, and filtration methodology, all influence these potential alterations.

Sample collection and storage containers can alter measured nutrient concentrations. This effect is especially is important for dissolved P, which has a high affinity for adsorption onto container walls because of its relatively high charge density (Jarvie et al., 2002). The degree of adsorption depends on the container's pre-treatment, composition, and surface area to volume ratio and on the sample's dissolved P concentration and matrix chemistry. Only phosphate-free detergents should be used to clean collection and storage

containers (Haygarth and Edwards, 2000). Rinsing plastic bottles with dilute acetic or hydrochloric acid (HCl) reduces adsorption by saturating sorption sites, but the opposite is the case for glass bottles (Ryden et al., 1972). Pre-treatment of glass bottles with hydrofluoric acid has been shown to reduce P sorption, but health and safety issues concerning hydrofluoric acid and glass bottles often prevents their use (Jarvie et al., 2002). Acid-washed, low-density polyethylene bottles are generally the most suitable containers for the storage of water samples, although polytetrafluoroethylene bottles further minimize P sorption (Haygarth and Edwards, 2000). The desorption of P from previously used containers can be minimized by thorough pre-washing with dilute HCl and rinsing with deionized water (Latterell et al., 1974). The highest proportional losses of dissolved P occur from low-concentration samples, especially when stored in small bottles with high surface area to volume ratios (Jarvie et al., 2002; Haygarth et al., 1995; Kotlash and Chessman, 1998). For high-concentration P samples, the absolute impact is often insignificant because adsorption is limited by the container's internal surface area (Kotlash and Chessman, 1998). Sample matrix chemistry also affects P sorption, which is related primarily to anion exchange reactions. In high ionic strength samples, the potential is greater for competition between anions in solution for ion exchange sites on the container walls, which leads to reduced P adsorption (Jarvie et al., 2002). In contrast, greater P losses have been observed for low ionic strength solutions (Maher and Woo, 1998).

The storage environment can also influence nutrient concentrations. In all cases, the exclusion of light is recommended to inhibit photosynthesis, and thus reduce algal growth and nutrient uptake (Haygarth et al., 1995). The effects of storage temperature and storage time have received considerable attention in recent years. For polluted sites with high nutrient concentrations, little proportional change in nutrients has been observed for up to 6 days of storage without any kind of temperature control (Kotlash and Chessman, 1998). In contrast, up to 90% of ammoniacal N, 50% of oxidized N, 84% of total Kjeldahl N, and 67% of total P may be lost from samples with low nutrient concentrations after 6 h of storage without temperature control (Kotlash and Chessman, 1998). Freezing of samples is impractical in the field and is generally reserved for long-term laboratory storage. It is often practical, however, to refrigerate samples stored in automatic samplers, and this is the preferred method of preservation in many cases. Low temperatures (<4.0°C) reduce microbial activity, thereby reducing microbially mediated nutrient transformations. The effectiveness of refrigeration appears to vary, with some previous studies demonstrating effective preservation for up to 8 days at 4.0°C (Fishman et al., 1986) and others reporting significant changes in nutrient concentrations within 4 to 48 h (Johnson et al., 1975; Lambert et al., 1992; Haygarth et al., 1995). Kotlash and Chessman (1998) reported effective preservation by refrigeration for up to 2 days for a broad range of sites and nutrient concentrations under varying weather conditions.

Chemical preservatives, such as chloroform, inorganic acids, and mercuric chloride, can be used during storage to stabilize nutrient concentrations by stopping metabolic processes and preventing microbial uptake and release (Kotlash and Chessman, 1998; Jarvie et al., 2002). The use of chemical preservatives is, however, limited due to the potential for contamination and colorimetric interference (Haygarth et al., 1995; Maher and Woo, 1998). Preservatives can be deployed in sample collection bottles before sampling, although volatility can limit their longevity. Kotlash and Chessman (1998) reported that acidification produced results comparable to freezing for all N forms; however, the application of chemical preservatives can be problematic. Chloroform can result in the immediate release of dissolved P from particles and algal cells (Fitzgerald and Faust, 1967) and can interfere with colorimetric P determination (Skjemstad and Reeve, 1978). Inorganic acids and mercuric chloride can hydrolyze organic compounds and release dissolved P (Henriksen, 1969). Mercuric chloride can also precipitate bacteria and proteins (Maher and Woo, 1998) and interfere with colorimetric P determination (Skjemstad and Reeve, 1978).

Filtration, which is required for determination of dissolved nutrient species, can also influence measured concentrations (Jarvie et al., 2002). Filtration is commonly conducted once samples have been returned to the laboratory, despite recommendations of immediate field filtration for dissolved P (Lambert et al., 1992; Haygarth et al., 1995; Maher and Woo, 1998). In most cases, samples are passed through 0.45 µm cellulose nitrate acetate filter paper under vacuum pressure, but this can have undesirable effects. First, the choice of a 0.45 µm cutoff is operationally defined; thus, some fine particulate matter and colloidal material will pass through into the filtrate (Lambert et al., 1992). Second, membrane filter papers can yield significant P and N, although this effect can be minimized by pre-washing filter paper with deionized water (Robards et al., 1994). Third, the mechanical effects of filtering under vacuum pressure can alter colloidal material, rupture cells, and release dissolved nutrients (Jarvie et al., 2002). Fourth, reductions in effective pore size by progressive blocking can occur during filtration, which affects passage of fine sediment, colloidal material, and associated nutrients (Jarvie et al., 2002).

LABORATORY ANALYSIS

The main potential sources of uncertainty in the laboratory are associated with sample handling, chemical preparation, analytical method and equipment, personnel expertise, and calibration standards and reference materials (CAEAL, 2003). Although these sources of error have received substantial research attention, the need for quantification and control of analytical error is commonly reported (e.g., CAEAL, 2003; Jarvie et al., 2002; Gordon et al., 2000; Mercurio et al., 2002; Ludtke et al., 2000). Personnel expertise and proper methodology are vital in limiting errors introduced by laboratory analysis. Meyer (2002) lists several well-accepted methods to reduce uncertainty that apply to typical analyses. These include using adequate working techniques and large volumes, minimizing the number of steps, making sample and reference measurements in temporal proximity and with the same instrument, introducing an internal standard, using a certified matrix standard, and performing multiple analyses. Meyer (2002) also refers to the results of Horwitz et al. (1980) to illustrate reduced relative deviation in nutrient analysis compared to other water quality constituents that occur in lower concentrations.

In sediment analysis, sediment is typically separated from collected water samples by drying or filtering. Then, the

concentration is determined from the sediment mass divided by the original sample volume. Whereas only basic mass measurements are necessary to determine sediment concentrations, analysis of dissolved and particulate nutrient concentrations requires more in-depth analytical procedures; however, substantial errors can be introduced if proper methods are not rigorously followed in either case. Total N and P concentrations are determined by analyzing unfiltered water samples, whereas dissolved N and P are measured in the filtrate passing through a 0.45 μm filter. The most widely accepted analytical techniques for determining nutrient concentrations use wet chemistry followed by spectrophotometric (colorimetric) analysis (Robards et al., 1994). These techniques involve blending precise amounts of sample and wet chemicals to cause a colorimetric reaction. The depth of the color, measured with a spectrophotometer, is proportional to the nutrient concentration. Total nutrient concentrations are determined with the same technique except that particulate forms are initially converted into dissolved forms through chemical digestion, which is strongly affected by temperature, pH, and particulate characteristics (Robards et al., 1994). Inductively coupled plasma optical emission spectrometry (ICP-OES) may also be used for the determination of total dissolved P if concentrations are relatively high, but ICP mass spectrometry (ICP-MS) may offer better accuracy for samples with lower P concentrations (Jarvie et al., 2002).

OBJECTIVE

Faced with the need for scientifically defensible estimates of data uncertainty to support water quality management, the present research focused on nutrient and sediment loads measured from small watersheds. The specific objectives were to:

- Compile published information on uncertainty related to measured streamflow and water quality data.
- Utilize a root mean square error propagation method to compare the uncertainty introduced by each procedural category for best and worst case and a range of typical "data quality" scenarios.
- Utilize the error propagation method to calculate the cumulative probable uncertainty in measured streamflow, sediment, and nutrient data for best and worst case and a range of typical "data quality" scenarios.

METHODS

ERROR SOURCE DOCUMENTATION

To accomplish the objectives, results from selected research on uncertainty related to streamflow and water quality data were compiled from available literature and are presented in tables 1 through 4; no new data on uncertainty related to each procedure is presented. The information included was selected based on its relevance to small watersheds but is by no means meant to be comprehensive. Only information on measurement of streamflow, dissolved N and P, total N and P, and TSS is included. Dissolved N and P include dissolved, colloidal, and other forms that typically pass through a 0.45 μm filter. Total N and P include all these forms associated with the water portion of the sample as well as particulate forms.

ERROR PROPAGATION METHODOLOGY

The root mean square error propagation method of Topping (1972), shown in equation 1, was used with reported errors in tables 1 through 4 to estimate the cumulative probable uncertainty for each procedural category (streamflow measurement, sample collection, sample preservation/storage, and laboratory analysis) and for the overall resulting streamflow and water quality data:

$$E_P = \sqrt{\sum_{i=1}^{n}(E_1^2 + E_2^2 + E_3^2 + ... + E_n^2)} \qquad (1)$$

where

E_P	= probable range in error ($\pm\%$)
n	= total number of sources of potential error

E_1, E_2, E_3, E_n = potential sources of error ($\pm\%$).

This error propagation method was selected for two reasons. First, it is a widely accepted method that has been used for similar error calculations related to discharge measurements (Sauer and Meyer, 1992) and water quality constituents (Cuadros-Rodriguez et al., 2002). In addition, it combines all of the potential errors to produce realistic estimates of overall error. This method is valid for measured water quality data because potential errors are typically bi-directional and, therefore, non-additive.

Various "data quality" scenarios were created for each step in measuring water quality loads. Best case, typical, and worst case scenarios were created by arbitrarily selecting procedures from commonly utilized and accepted practices, as identified in tables 1 through 4. The best case scenario represents procedures used with a concentrated effort in quality assurance/quality control (QA/QC) unconstrained by financial and personnel resource limitations and in ideal hydrologic conditions. The worst case represents projects conducted with minimal attention to QA/QC, with limited financial and personnel resources, and in difficult hydrologic conditions. To formulate the best and worst case scenarios, it was assumed either that poor conditions with little QA/QC occurred together to generate large errors or that ideal conditions occurred together with careful QA/QC to minimize errors. Thus, extreme low "best case" or extreme high "worst case" error estimates resulted. The typical scenario represents procedures conducted with a moderate effort at QA/QC and under typical hydrologic conditions. Within the typical scenario, a range of conditions was determined with average values and maximum and minimum uncertainty boundaries to represent a more reasonable "typical" range of uncertainty.

Classification of individual steps into these data quality scenarios (best case, typical, and worst case) was based on professional experience and judgment. The reader is encouraged to apply uncertainty estimates that correspond to specific data sets, if that information is available. In the absence of data set specific uncertainty estimates, the present results can be used to establish reasonable estimates for uncertainty introduced by streamflow measurement, sample collection, sample preservation/storage, and laboratory analysis and for resulting streamflow and water quality data.

The uncertainty, represented by the probable error range (E_p), associated with measured water quality data was compared for each procedural category (streamflow measurement, sample collection, sample preservation/storage,

and laboratory analysis). The error introduced by each step within each procedural category was determined from published information identified in tables 1 through 4. The information included in table 1 is divided into four procedural steps (or sources of uncertainty) involved in streamflow measurement. These include individual streamflow measurements to establish the stage-discharge relationship, application of the stage-discharge relationship, continuous stage measurement, and the effect of streambed condition on stage measurement. Thus, when determining the cumulative uncertainty for the streamflow measurement procedural category, four components, one for each step in table 1, were included in equation 1. The information in table 2 is divided into two methods of sample collection, which are both affected by the minimum flow threshold at which to initiate sample collection. The uncertainty in automated sample collection is related to use of a single sampling intake and the choice of flow- or time-interval sampling. The uncertainty in manual sample collection is related to the choice of integrated or grab sampling. Thus, the uncertainty calculation for sample collection involved applying equation 1 with two uncertainty components (collection method and minimum flow threshold). The overall uncertainty resulting from sample preservation/storage procedures and from laboratory analysis is presented for each constituent in tables 3 and 4. The calculation of uncertainty contributed by sample preservation/storage involved only one component in equation 1. In contrast, the uncertainty contributed by laboratory

analysis included the cumulative effects of each relevant analysis. For example, determination of TSS and dissolved N and P involved only the analysis of constituent concentration (thus one component in eq. 1), but determination of particulate N and P involved analysis of sediment concentration and analysis of sediment nutrient content (thus two components in eq. 1).

The uncertainty estimates for each step (or source of uncertainty) were entered into the root mean square error propagation method (Topping, 1972) to determine the cumulative probable uncertainty within each procedural category. Errors for each category were then propagated with the same method to determine the cumulative probable uncertainty for the resulting streamflow and water quality data.

RESULTS AND DISCUSSION
COMPILATION OF PREVIOUS RESEARCH
Results of selected research pertaining to collection of streamflow and selected water quality constituent data are summarized in tables 1 through 4. The information presented was selected to support the present error propagation analyses and to aid in future analyses. These tables are not meant to be comprehensive but to provide readily accessible estimates of uncertainty related to streamflow measurement, sample collection, sample preservation/storage, and laboratory analysis pertinent to measured water quality data, spe–

Table 1. Uncertainty in streamflow measurement.

	Uncertainty[a]	Reference
Individual Streamflow Measurements		
Velocity-area (direct discharge) method		
Ideal conditions	±2%	Sauer and Meyer (1992)
Ideal conditions (0.2, 0.8d velocity)	±6.1%	Pelletier (1988)
Ideal conditions (0.6d velocity)	±8.5%	Pelletier (1988)
Average conditions	±6%	Sauer and Meyer (1992)
Poor conditions	±20%	Sauer and Meyer (1992)
Ideal conditions	±2%	Boning (1992)
Manning's equation method		
Stable, uniform channel; surveyed reach and cross-section; accurate "n" estimate	±15%	Slade (2004)
Unstable, irregular channel; surveyed reach and cross-section; poor "n" estimate	±35%	Slade (2004)
Stage-Discharge Relationship		
Pre-calibrated flow control structure (properly designed and installed) with periodic current meter checks	±5% to 8%	Slade (2004)
Pre-calibrated flow control structure (properly designed and installed)	±5% to 10%	Slade (2004)
Stable channel with stable control, 8-12 stage-discharge measurements per year	±10%	Slade (2004)
Shifting channel, 8-12 stage-discharge measurements per year	±20%	Slade (2004)
Natural channel, ideal conditions	±6%	Boning (1992)
Continuous Stage Measurement		
Float recorder	±2%	Cooper (2005), unpubl. data
Float recorder	±3 mm (±0.00985 ft)[b]	Herschy (1975)
KPSI series 173 pressure transducer	±0.1%, ±0.022% thermal error	KPSI (2005)
Campbell Scientific SR50-L ultrasonic distance sensor	Larger of ±1 cm or 0.4% of distance to water surface	Campbell Scientific (2003)
ISCO 730 bubbler flow module	±0.035 ft ±0.0003 * (ft) * temp. change from 72°F[c]	Teledyne ISCO (2005)
Effect of Streambed Condition on Stage Measurement		
Stable, firm bed	±0%	Sauer and Meyer (1992)
Mobile, unstable bed	±10%	Sauer and Meyer (1992)

[a] Error estimates are presented as their ±% range for bi-directional error or as their actual % range.
[b] At 10 ft = 0.1%, at 1 ft = 1.0%, at 0.1 ft = 10%, and at 0.01 ft = 100%.
[c] At 72°F and 10 ft = 0.0%, at 1 ft = 0.4%, at 0.1 ft = 3.5%, and at 0.01 ft = 35%.

<div align="center">Table 2. Uncertainty in sample collection.</div>

	Uncertainty[a]	Reference
Sample Collection Method		
Manual sampling		
Integrated sampling (EWI or EDI)		
Frequently during each storm	±5% dissolved; 15% suspended	Slade (2004)
Only once in each storm	±15% dissolved; 30% suspended	Slade (2004)
Grab sampling (single point, random time)	±25% dissolved; >50% suspended	Slade (2004)
Automated sampling		
Effect of single sampling point (sample intake)		
Dissolved N (NH_3, NO_3, NO_2, NO_2+NO_3 forms)	Range of medians 0% to 4% (overall median = 0%)	Martin et al. (1992)
Total N	Range of medians 0% to 0% (overall median = 0%)	Martin et al. (1992)
Dissolved P (PO_4)	Range of medians 0% to 0% (overall median = 0%)	Martin et al. (1992)
Total P	Range of medians 0% to 17% (overall median = 0%)	Martin et al. (1992)
TSS	Range of medians 14% to 33% (overall median = 20%)	Martin et al. (1992)
Flow-interval sampling strategies: f(flow interval)		
2.5-15 mm, with up to 6 composite samples	−6% to +17%	King and Harmel (2003)
1.32-5.28 mm, with up to 5 composite samples	−9% to +3%	Harmel and King (2005)
0.2-1.25 mm	±0% to 22%	Miller et al. (2000)
Time-interval sampling strategies: f(time, number of composite samples)		
5 min, discrete	±0% to 11%	Miller et al. (2000)
5 min, discrete	0%	King and Harmel (2003)
5 min, with up to six composite samples	−5% to 4%	King and Harmel (2003)
30 min, discrete	±3% to 42%	Miller et al. (2000)
30 min, discrete	−2% to 2%	King and Harmel (2003)
30 min, with up to six composite samples	−32% to 25%	King and Harmel (2003)
120 min, discrete	−15% to 13%	King and Harmel (2003)
120 min, with up to six composite samples	−65% to 51%	King and Harmel (2003)
Minimum Flow Threshold		
Disregard flow and concentration outside sampling period, "low" threshold	±1% to 5% (median = 3%)	Professional judgment based on Harmel et al. (2002)
Disregard flow and concentration outside sampling period, "high" threshold	±9% to 81% (median = 35%)	Professional judgment based on Harmel et al. (2002)
Extrapolate flow and concentration outside sampling period, "low" threshold	±2%	Professional judgment based on Harmel et al. (2002)
Extrapolate flow and concentration outside sampling period, "high" threshold	±20%	Professional judgment based on Harmel et al. (2002)

[a] Error estimates are presented as their ±% range for bi-directional error or as their actual % range.

<div align="center">Table 3. Uncertainty in sample preservation/storage.</div>

Preservation/Storage Technique		Uncertainty[a]	Reference
NH_3-N	Iced, analyzed within 6 h	−50% to 1% (median = −18%)	Kotlash and Chessman (1998)
	Acidified to ph <2, analyzed within 6 h	−58% to 4% (median = −8%)	Kotlash and Chessman (1998)
	Refrigerated, analyzed within 54 h	−79% to 83% (median = −16%)	Kotlash and Chessman (1998)
	Unpreserved, analyzed within 192 h	−90% to 67% (median = −38%)	Kotlash and Chessman (1998)
	Unpreserved, analyzed within 96 h	±5% to 77% (median = 3%)	Cooper (2005), unpublished data
NO_3-N	Iced, analyzed within 6 h	±0% (median = 0%)	Kotlash and Chessman (1998)
	Acidified to ph <2, analyzed within 6 h	−6% to 20% (median = −1%)	Kotlash and Chessman (1998)
	Refrigerated, analyzed within 54 h	−47% to 14% (median = −2%)	Kotlash and Chessman (1998)
	Unpreserved, analyzed within 192 h	−65% to 71% (median = −2%)	Kotlash and Chessman (1998)
	Unpreserved, analyzed within 96 h	±7% to 30% (median = 1%)	Cooper (2005), unpublished data
TKN	Iced, analyzed within 6 h	−14% to 22% (median = 3%)	Kotlash and Chessman (1998)
	Acidified to ph <2, analyzed within 6 h	−16% to 49% (median = −1%)	Kotlash and Chessman (1998)
	Refrigerated, analyzed within 54 h	−28% to 32% (median = −9%)	Kotlash and Chessman (1998)
	Unpreserved, analyzed within 192 h	−84% to 20% (median = −26%)	Kotlash and Chessman (1998)
Total P	Iced, analyzed within 6 h	−13% to 69% (median = −7%)	Kotlash and Chessman (1998)
	Refrigerated, analyzed within 54 h	−7% to 92% (median = 7%)	Kotlash and Chessman (1998)
	Unpreserved, analyzed within 192 h	−64% to 9% (median = −11%)	Kotlash and Chessman (1998)
Filterable P[b]	Iced, analyzed within 6 h	−47% to 267% (median = −7%)	Kotlash and Chessman (1998)
	Refrigerated, analyzed within 54 h	−52% to 600% (median = 8%)	Kotlash and Chessman (1998)
	Unpreserved, analyzed within 192 h	−39% to 20% (median = −17%)	Kotlash and Chessman (1998)

[a] Error estimates are presented as their ±% range for bi-directional error or as their actual % range.
[b] Dissolved P, passed through a 0.45 µm filter membrane.

Table 4. Uncertainty in laboratory analysis.

Sample Analysis	Uncertainty[a]	Reference
TSS		
Sandy sediment	95% CI range = −9.8% to 5.1%, range of medians = −4.9% to −2.5%	Gordon et al. (2000)
Fine sediment	95% CI range = −5.3% to 4.4%, range of medians = −1.3% to −0.4%	Gordon et al. (2000)
Dissolved (Solution) N, P		
Total P (ICP)	Range of medians = ±4% to 210%	Miller and Kotuby-Amacher (2005), unpublished data
Total P (alkaline persulfate digestion)	Range = −24% to 22% (avg. = 2%)	Mercurio et al. (2002)
Total P (micro-Kjeldahl)	Range of medians = −1% to 3%	Ludtke et al. (2000)
PO_4-P spec.	Range of medians = ±5% to 400%	Miller and Kotuby-Amacher (2005), unpublished data
PO_4-P (colorimetric)	Range of medians = −5% to 9%	Ludtke et al. (2000)
PO_4-P (colorimetric)	Range = −14% to 22% (avg. = 8%)	Mercurio et al. (2002)
TKN	Range of medians = ±11% to 82%	Miller and Kotuby-Amacher (2005), unpublished data
Total N (alkaline persulfate digestion)	Range = −20% to 20% (avg. = −3%)	Mercurio et al. (2002)
NO_3-N	Range of medians = ±7% to 400%	Miller and Kotuby-Amacher (2005), unpublished data
NO_3-N (colorimetric)	Range −4% to 9% (avg. = 7%)	Mercurio et al. (2002)
NO_3+NO_2-N (colorimetric)	Range of medians = −6% to 3%	Ludtke et al. (2000)
NH_4-N	Range of medians = ±15% to 200%	Miller and Kotuby-Amacher (2005), unpublished data
NH_4-N (colorimetric)	Range −22% to 26% (avg. = −8%)	Mercurio et al. (2002)
NH_4-N (colorimetric)	Range of medians = −7% to 12%	Ludtke et al. (2000)
Particulate N, P		
TKN	Range of medians = ±6% to 15%	Miller and Kotuby-Amacher (2005), unpublished data
Total N (combustion)	Range of medians = ±4% to 30%	Miller and Kotuby-Amacher (2005), unpublished data
Total N (combustion)	Range −1% to 12% (avg. = 2%)	Mercurio et al. (2002)
Total P (HCl digestion)	Range −2% to 16% (avg. = 7%)	Mercurio et al. (2002)
Whole Water Sample		
TKN (micro-Kjeldahl)	Range of medians = −24% to 0%	Ludtke et al. (2000)
TKP (micro-Kjeldahl)	Range of medians = 0% to 4%	Ludtke et al. (2000)

[a] Error estimates are presented as their ±% range for bi-directional error or as their actual % range.

cifically streamflow and N, P, and TSS loads. Each entry contains a specific procedure or potential source of error, the associated uncertainty, and an appropriate reference. A majority of the error estimates are presented as a ±% range, which reflects their bi-directional nature, but others are presented as the % range of error values. Regardless of the notation, all potential error sources were assumed to be bi-directional in error propagation calculations.

ERRORS FOR EACH PROCEDURAL CATEGORY

Error propagation (cumulative probable uncertainty) results contributed by each procedural category are presented in figure 1 and discussed for worst case, typical, and best case scenarios. These results are pertinent to storm load (mass) determination and are based on the assumptions made to create the various data quality scenarios. Storm load determination requires all four procedural categories; thus, uncertainty contributed by streamflow measurement, sample collection, sample preservation/storage, and laboratory analysis are analyzed.

Streamflow Measurement

The cumulative probable uncertainty results presented for streamflow measurement were determined by propagation of uncertainty introduced by individual streamflow measurements to establish the stage-discharge relationship, application of the stage-discharge relationship, continuous stage measurements, and the effect of streambed condition on stage measurement (table 1). The probable error range (E_P, ±%) for measured streamflow was estimated to be 42% for the worst case scenario. The worst case scenario involved

streamflow estimation with Manning's equation with a stage-discharge relationship for an unstable, mobile bed and a shifting channel. For the typical scenario, the E_P ranged from 6% to 19% for a range of individual streamflow measurement techniques, channel types, and channel conditions (fig. 1). The estimated E_P was 3% for the best case scenario, which included flow measurement under ideal hydrologic conditions, specifically a pre-calibrated flow control structure (stable bed and channel) and a stilling well for stage measurement.

Sample Collection

The cumulative probable uncertainty results presented for sample collection were determined by propagation of uncertainty introduced by individual steps within two sample collection methods and by the minimum flow threshold at which to initiate sample collection (table 2). The uncertainty in automated sample collection is related to use of a single sampling intake and the choice of flow- or time-interval sampling. The uncertainty in manual sample collection is related to the choice of integrated or grab sampling. In terms of uncertainty introduced by sample collection, the estimated E_P was 104% for dissolved constituents and 109% for TSS and total N and P in a worst case scenario. This scenario involved liberal estimates of error associated with sample collection at a single point, infrequent time-interval sampling at a high minimum flow threshold, and disregard of conditions outside the sampling period. Under a range of typical scenarios, which included moderate errors associated with frequent flow- or time-interval sample collection at a single point and estimation of conditions outside a high flow

threshold, the E_P ranged from 4% to 47% for dissolved constituents and from 4% to 50% for TSS and total N and P (fig. 1). Under the best case scenario, which included conservative error estimates associated with frequent flow- or time-interval sample collection at a single point and estimation of conditions outside a low flow threshold, E_P was 1% for dissolved and sediment-associated constituents.

Sample Preservation/Storage and Laboratory Analysis

Error propagation was unnecessary for sample preservation/storage because combined errors (not errors for individual steps) in this procedural category were reported for individual constituents (table 3). In terms of uncertainty introduced by sample storage/preservation, estimated E_P values ranged from 20% to 90% for dissolved constituents and from 9% to 84% for total nutrients in a worst case scenario. The worst case scenario involved unpreserved, unrefrigerated sample storage for 144 h and then refrigerated storage for 48 h prior to analysis. Under a typical scenario, which involved refrigerated sample storage for 54 h prior to analysis, E_P estimates ranged from 2% to 16% for dissolved nutrients and from 7% to 9% for total nutrients (fig. 1). Under a best case scenario, which included iced sample storage for 6 h prior to analysis, E_P ranged from 0% to 2% and from 1% to 3% for dissolved and total nutrients, respectively. The concentration of TSS was not affected by preservation and storage procedures.

Although the overall uncertainty related to laboratory analysis is presented for each constituent in table 4, the uncertainty in sediment, dissolved N and P, and particulate N and P is included, where appropriate, in the determination of constituent concentrations. In terms of uncertainty introduced by laboratory analysis, estimated E_P values were 200% to 400% for dissolved constituents, 88% to 211% for total nutrients, and 10% for TSS in a worst case scenario. The worst case scenario involved liberal estimates of error for constituents present in very low concentrations. For a range of typical scenarios with moderate error estimates for low constituent concentrations, estimated E_P values ranged from 4% to 26% for dissolved nutrients, from 3% to 32% for total

nutrients, and from 1% to 5% for TSS (fig. 1). The best case scenario with conservative error estimates for constituents present in moderate concentrations resulted in E_P estimates of less than 2%, 4%, 1%, respectively, for dissolved nutrients, total nutrients, and TSS.

COMPARISON OF PROCEDURAL CATEGORIES

Although sample preservation/storage and laboratory analysis are typically emphasized in QA/QC efforts, the scientific validity for this emphasis over other sources of error has not been established. Results of the present research indicate that substantial error can be introduced by each of the procedural categories (fig. 1). In the worst case scenario, sample collection and laboratory analysis can contribute more uncertainty than streamflow measurement and sample preservation/storage. The potentially large uncertainty introduced by sample analysis is influenced by inflated relative errors for low-magnitude known values, e.g., very low constituent concentrations. Kotlash and Chessman (1998) noted this influence on nutrient concentrations in unpolluted streams, which are very sensitive to small absolute errors and result in high relative errors. In the range of typical scenarios, sample collection had the greatest potential to contribute large amounts of uncertainty because of the cumulative effects of substantial potential error associated with a single sampler intake, automated sampler operation, and the minimum flow threshold (table 1). In the best case scenario, the uncertainty was less than 5% for each procedural category.

These findings indicate that changes in current QA/QC methodology are needed. Although sample preservation/storage and laboratory analysis are typically emphasized, the uncertainty introduced by streamflow measurement and sample collection should not be ignored in QA/QC plans aimed at reducing uncertainty in measured water quality data. In fact, sample collection has the potential to introduce the most uncertainty under typical scenarios with moderate QA/QC effort and average hydrologic conditions (fig. 1). Based on field experience, Martin et al. (1992) made a similar
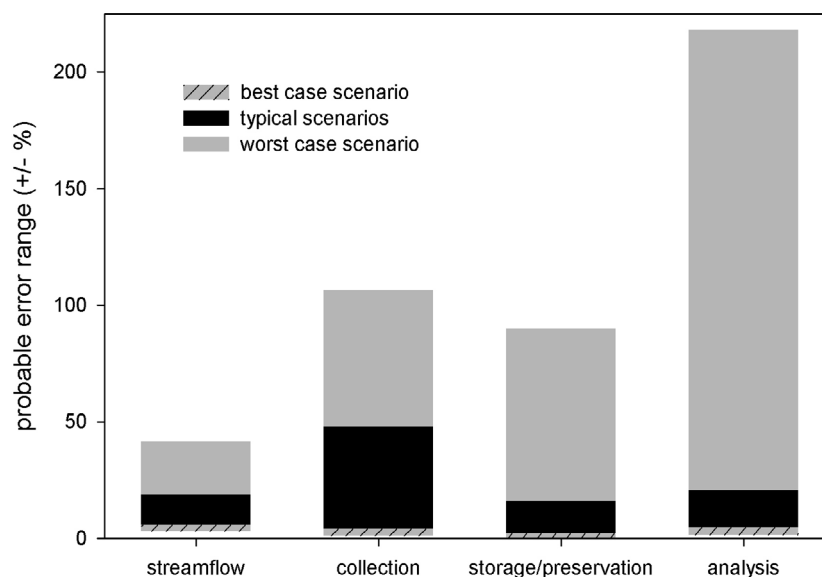

Figure 1. Comparison of probable uncertainty contributed by each procedural category for best case, typical, and worst case "data quality" scenarios; data presented were averaged across constituent type.
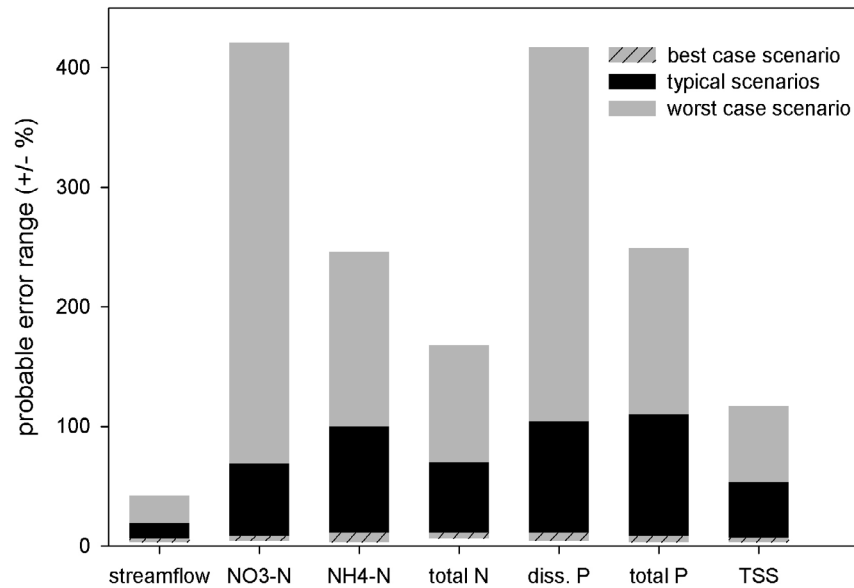
Figure 2. Cumulative probable uncertainty, represented by the probable error range (E_P), for streamflow and nutrient and sediment storm loads for worst and best case scenarios and a range of typical scenarios.

observation that sample collection can be the largest source of uncertainty in measured water quality. The authors also recognized the relative lack of discussion on sample collection methodology in publications that address standard methods for evaluating water quality.

UNCERTAINTY IN RESULTING STREAMFLOW AND WATER QUALITY DATA

The uncertainty in each step of the procedural categories was propagated with the Topping (1972) method in equation 1 to estimate the cumulative probable uncertainty in the resulting water quality data (table 5, fig. 2). These results indicate that streamflow data are much more reliable than nutrient and sediment load data. Estimated E_P values ranged from 8% to 104% for dissolved nutrients, from 8% to 110% for total nutrients, and from 7% to 53% for TSS under typical conditions and ranged from 117% to 421% in the worst case scenario. Estimated E_P values for streamflow were substantially lower, ranging from 6% to 19% under typical conditions and reaching only 42% in the worst case scenario. The reduced error for streamflow was expected because determination of constituent loads involves additional procedural steps and potential sources of error. Errors in TSS loads were typically less than errors in nutrient loads because sediment mass is not subject to post-collection transformation. In typical scenarios, uncertainty in NO_3-N data was less than that in NH_4-N and dissolved P data because NO_3-N often occurs in higher concentrations; however, all dissolved constituent data can be quite uncertain in the worst case

scenario (fig. 2) due to the difficulty of accurately measuring very low concentrations.

The cumulative probable uncertainty results presented thus far apply to measured storm loads, which require streamflow measurement, sample collection, sample preservation/storage, and laboratory analysis. However, streamflow measurement is unnecessary when storm water quality concentration data, not load data, are required. Thus, the cumulative probable uncertainty in measured constituent concentrations in stormflow is reduced due to the elimination of streamflow uncertainty. When uncertainty contributed by streamflow measurement was eliminated, reductions in E_P ranged from 2% to 8% in the worst case scenario, from 2% to 3% in typical scenarios, and from 1% to 2% in the best case scenario, compared to storm load uncertainty. These small reductions in uncertainty for measured constituent concentrations in stormflow are attributed to the finding that streamflow measurement typically contributes less uncertainty than the other procedural categories (fig. 1).

Determination of constituent concentrations at baseflow involves even less uncertainty because errors associated with storm sampling are eliminated. In contrast to the relatively low reduction in errors for stormflow concentrations, the error reductions associated with baseflow sampling were substantial because storm sample collection can introduce substantial uncertainty in measured data. For baseflow constituent concentrations, reductions in E_P ranged from 15% to 83% in the worst case scenario, from 3% to 35% in typical scenarios, and from 1% to 3% in the best case scenario, compared to storm load uncertainty.

**Table 5. Cumulative probable uncertainty, represented by the probable error range (E_P), for streamflow
and nutrient and sediment storm loads for worst case, best case, and typical scenarios.**

	Streamflow (%)	NO_3-N (%)	NH_4-N (%)	Total N (%)	Diss. P (%)	Total P (%)	TSS (%)
Worst case scenario	42	421	246	168	417	249	117
Typical scenario maximum	19	69	100	70	104	110	53
Typical scenario average	10	17	31	29	23	30	18
Typical scenario minimum	6	8	11	11	11	8	7
Best case scenario	3	4	3	6	4	3	3

IMPLICATIONS FOR WATERSHED AND WATER QUALITY MODELING

The uncertainty inherent in data used to calibrate and evaluate watershed and water quality model output is commonly acknowledged but seldom considered in modeling analyses because such estimates of uncertainty were previously unavailable. With the methodology presented, modelers can now quantify the "quality" of their calibration and evaluation data sets and realistically judge model performance. Models should be expected to produce output within the uncertainty limits inherent in measured data, not to produce outputs with low deviation from measured data. If models are judged solely on their ability to produce values similar to measured data, then the model in a sense may be precise but may not be accurately reproducing actual hydrological and water quality conditions.

Based on the present results regarding various "data quality" scenarios, model results within 10% to 31% of measured values are within the average uncertainty range of water quality data measured with a typical QA/QC effort. However, where high quality calibration and evaluation data (collected in good hydrologic conditions with dedicated QA/QC) are available, model performance can be held to a higher standard, (such as ±10%). In the opposite case, when model calibration and evaluation data are collected in a worst case scenario with difficult hydrologic conditions and little effort at minimizing uncertainty, the uncertainty values can exceed 40% for streamflow, 100% for TSS, and 150% to 400% for nutrients. Such data sets may offer little advantage over absence of available data for model evaluation, and it can be argued that uncalibrated model application is appropriate.

The results of the present research clearly demonstrate the benefits of uncertainty estimates associated with model calibration and evaluation data; however, another important issue remains. How should uncertainty estimates for measured data be combined with model uncertainty to determine overall probable error values for projects, such as TMDLs, in which calibrated watershed models are used to allocate loads? Only when this question is satisfactorily answered can arbitrary margins of safety be replaced with rigorous, scientifically defensible, potentially more cost-effective uncertainty estimation and risk assessment.

A similar approach to that utilized in the present study, i.e., cumulative probable combination of overall errors inherent in measured data with uncertainty in modeling procedures, is one option to achieve this enhancement. Model uncertainty associated with factors such as model parameterization, process representation, and equifinality can be quantified with sensitivity analyses using techniques such as generalized likelihood uncertainty analysis (Beven and Binley, 1992). A discussion of how such techniques could be coupled with quantitative error estimates for measured data is beyond the scope of this article. It is clear, however, that consideration of both measured data and model uncertainty remains a major challenge in watershed modeling projects with regulatory, legal, or policy implications.

SUMMARY AND CONCLUSIONS

Several noteworthy findings and recommendations resulted from the present analysis of uncertainty in measured streamflow and water quality data. These results are applicable at the field and small watershed scale but must be adjusted for application at the large watershed and river basin scale.

- Selected information from previous research related to uncertainty in measured streamflow and water quality data was presented in tables 1 through 4 to support the present analysis and subsequent analyses. However, in subsequent error propagation analyses, this information should be supplemented with available project-specific uncertainty data.

- Although sample preservation/storage and laboratory analysis are typically emphasized in current QA/QC efforts, the present research indicated that substantial error can be introduced by each procedural category. Thus, changes in QA/QC methodology are needed to appropriately consider potential uncertainty in streamflow measurement and sample collection, especially since storm sample collection has the potential to introduce the most uncertainty in typical scenarios.

- Based on the assumptions made to create a range of typical "data quality" scenarios, the cumulative probable uncertainty for measured storm loads ranged from 8% to 69% for NO_3-N, from 11% to 100% for NH_4-N, from 11% to 70% for total N, from 11% to 104% for dissolved P, from 8% to 110% for total P, and from 7% to 53% for TSS. However, the uncertainty can increase substantially for poor measurement conditions and limited quality control effort.

- The cumulative probable uncertainty for measured storm water quality concentrations, which does not include uncertainty in streamflow measurement, was reduced only slightly compared to storm load data because streamflow measurement typically contributes less uncertainty than the other procedures. However, for measured baseflow constituent concentrations, which do not require streamflow measurement or storm water sample collection, reductions in E_P ranged from 15% to 83% in the worst case scenario, from 3% to 35% in typical scenarios, and from 1% to 3% in the best case scenario.

- In the absence of site- or project-specific uncertainty data, the estimates presented can be used to establish reasonable initial uncertainty values for small watershed streamflow and water quality data. However, if site- or project-specific data are available for the procedural categories, that data should be incorporated into the proposed error propagation methodology to better estimate actual uncertainty.

With the results and methodology presented, the water resource community can better assess the uncertainty or "quality" of available data sets for use in water quality management. The information should be especially useful to water quality modelers, thus allowing them to make realistic, science-based evaluations of model performance based on the uncertainty present in calibration and evaluation data sets. Policy, regulatory, research, and legal interests will, as a result, have a quantified confidence in model output; therefore, they will be better informed to make appropriate decisions based on model results.

preservation and storage effects. We also are thankful to Robert Miller and Janice Kotuby-Amacher for sharing their very useful data on the uncertainty of various laboratory analyses. Under their direction, the North America Proficiency Testing program (NAPT) has contributed valuable data and insight into the uncertainty associated with water, soil, and plant analysis.

REFERENCES

Beck, M. B. 1987. Water quality modeling: A review of the analysis of uncertainty. *Water Resources Res.* 23(8): 1393-1442.

Beven, K., and A. Binley. 1992. The future of distributed models: Model calibration and uncertainty prediction. *Hydrological Processes* 6: 279-298.

Boning, C. W. 1992. Policy statement on stage accuracy. Technical Memorandum No. 93-07. Washington, D.C.: USGS, Office of Water.

Bonta, J. V. 2002. Modification and performance of the Coshocton wheel with the modified drop-box weir. *J. Soil and Water Cons.* 57(6): 364-373.

Brakensiek, D. L., H. B. Osborn, and W. J. Rawls, coordinators. 1979. *Field Manual for Research in Agricultural Hydrology.* USDA Agriculture Handbook No. 224. Washington, D.C.: USDA.

Buchanan, T. J., and W. P. Somers. 1976. Book 3, Chapter A8: Discharge measurements at gaging stations. *Techniques of Water-Resources Investigations of the United States Geological Survey.* Washington, D.C.: USGS.

Buchanan, T. J., and W. P. Somers. 1982. Book 3, Chapter A7: Stage measurement at gaging stations. *Techniques of Water-Resources Investigations of the United States Geological Survey.* Washington, D.C.: USGS.

CAEAL. 2003. Policy on the estimation of uncertainty of measurement in environmental testing. Ottawa, Ontario: Canadian Association For Environmental Analytical Laboratories Available at: www.caeal.ca/P19_CAEAL_Unce_ Pol.pdf.

Campbell Scientific. 2003. SR50-L ultrasonic distance sensor specifications. Logan, Utah: Campbell Scientific, Inc. Available at: www.campbellsci.com.

Carter, R. W., and I. E. Anderson. 1963. Accuracy of current meter measurements. *J. Hydraulics Division, Proc. ASCE* 4(1): 105-115.

Carter, R. W., and J. Davidian. 1989. Book 3, Chapter A6: General procedure for gaging streams. *Techniques of Water-Resources Investigations of the United States Geological Survey.* Washington, D.C.: USGS.

Cooper, R. J. 2005. Unpublished data. Craigiebuckler, Aberdeen, U.K.: The Macaulay Institute.

Cuadros-Rodriquez, L., M. E. Hernandez Torres, E. Almansa Lopez, F. J. Egea Gonzalez, F. J. Arrebola Liebanas, and J. L. Martinez Vidal. 2002. Assessment of uncertainty in pesticide multiresidue analytical methods: Main sources and estimation. *Anal. Chim. Acta* 454(2): 297-314.

Dickinson, W. T. 1967. Accuracy of discharge determinations. Hydrology Papers, No. 20. Fort Collins, Colo.: Colorado State University.

Edwards, W. M., H. E. Frank, T. E. King, and D. R. Gallwitz. 1976. Runoff sampling: Coshocton vane proportional sampler. ARS-NC-50. Washington, D.C.: USDA-ARS.

Fishman, M. J., L. J. Schroder, and M. W. Shockley. 1986. Evaluation of methods for preservation of water samples for nutrient analysis. *Intl. J. Environ. Studies* 26(3): 231-238.

Fitzgerald, G. P., and S. L. Faust. 1967. Effect of water sample preservation methods on the release of phosphorus from algae. *Limnol. Oceanogr.* 12(2): 332-334.

Franklin, D. H., M. L. Cabrera, J. L. Steiner, D. M. Endale, and W. P. Miller. 2001. Evaluation of percent flow captured by a small in-field runoff collector. *Trans. ASAE* 44(3): 551-554.

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. Washington, D.C.: USGS.

Gordon, J. D., C. A. Newland, and S. T. Gagliardi. 2000. Laboratory performance in the sediment laboratory quality-assurance project, 1996-98. USGS Water Resources Investigations Report 99-4184. Washington, D.C.: USGS.

Harmel, R. D., and K. W. King. 2005. Uncertainty in measured sediment and nutrient flux in runoff from small agricultural watersheds. *Trans. ASAE* 48(5): 1713-1721.

Harmel, R. D., K. W. King, J. E. Wolfe, and H. A. Torbert. 2002. Minimum flow considerations for automated storm sampling on small watersheds. *Texas J. Science* 54(2): 177-188.

Harmel, R. D., K. W. King, and R. M. Slade. 2003. Automated storm water sampling on small watersheds. *Applied Eng. in Agric.* 19(6): 667-674.

Haygarth, P. M., and A. C. Edwards. 2000. Sample collection, handling, preparation, and storage. In *Methods of Phosphorus Analysis for Soils, Sediments, Residuals, and Waters.* Southern Cooperative Series Bulletin No. 396. Available at: www.soil.ncsu.edu/sera17/publications/sera17-2/pm_cover.htm.

Haygarth, P. M., C. D. Ashby, and S. C. Jarvis. 1995. Short-term changes in the molybdate reactive phosphorus of stored waters. *J. Environ. Qual.* 24(6): 1133-1140.

Henriksen, A. 1969. Preservation of water samples for phosphorus and nitrogen determination. *Vatten* 25: 247-254.

Herschy, R. W. 1975. The accuracy of existing and new methods of river gauging. PhD diss. Reading, U.K.: University of Reading, Department of Geography.

Hession, W. C., D. E. Storm, C. T. Haan, K. H. Reckhow, and M. D. Smolen. 1996. Risk analysis of total maximum daily loads in an uncertain environment using EUTROMOD. *J. Lake and Reservoir Mgmt.* 12(3): 331-347.

Hipolito, J. N., and J. M. Leoureiro. 1988. Analysis of some velocity-area methods for calculating open-channel flow. *Hydrological Sci. J.* 33: 311-320.

Horwitz, W., L. R. Kamps, and K. W. Boyer. 1980. Quality assurance in the analysis of foods and trace constituents. *J. AOAC* 63(6): 1344-1354.

Izuno, F. T., R. W. Rice, R. M. Garcia, L. T. Capone, and D. Downey. 1998. Time versus flow composite water sampling for regulatory purposes in the Everglades Agricultural Area. *Applied Eng. in Agric.* 14(3): 257-266.

Jarvie, H. P., P. J. A. Withers, and C. Neal. 2002. Review of robust measurement of phosphorus in river water: Sampling, storage, fractionation, and sensitivity. *Hydrology and Earth System Sci.* 6(1): 113-132.

Johnson, A. H., D. R. Bouldin, and G. W. Hergert. 1975. Some observations concerning preparation and storage of stream samples for dissolved inorganic phosphate analysis. *Water Resources Res.* 11: 559-562.

Kennedy, E. J. 1984. Book 3, Chapter A10: Discharge ratings at gaging stations. *Techniques of Water-Resources Investigations of the United States Geological Survey.* Washington, D.C.: USGS.

King, K. W., and R. D. Harmel. 2003. Considerations in selecting a water quality sampling strategy. *Trans. ASAE* 46(1): 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(5): 1457-1463.

Kotlash, A. R., and B. C. Chessman. 1998. Effects of water sample preservation and storage on nitrogen and phosphorus determinations: Implications for the use of automated sampling equipment. *Water Res.* 32(12): 3731-3737.

KPSI. 2005. Series 169 and 172 submersible level transducer. Hampton, Va.: Pressure Systems, Inc. Available at: www.pressuresystems.com.

Lambert, D., W. A. Maher, and I. Hogg. 1992. Changes in phosphorus fractions during storage of lake water. *Water Res.* 26(5): 645-648.

Latterell, J. J., D. R. Timmons, R. F. Holt, and E. M. Sherstad. 1974. Sorption of orthophosphate on the surface of water sample containers. *Water Resources Res.* 10: 865-869.

Ludtke, A. S., M. T. Woodworth, and P. S. Marsh. 2000. Quality assurance results for routine water analysis in U.S. geological survey laboratories, water year 1998. USGS Water Resources Investigations Report 00-4176. Washington, D.C.: USGS.

Maher, W., and L. Woo. 1998. Procedures for the storage and digestion of natural waters for the determination of filterable reactive phosphorus, total filterable phosphorus, and total phosphorus. *Anal. Chim. Acta* 375(1-2): 5-47.

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(7): 866-876.

McFarland, A. M. S., and L. M. Hauck. 2001. Determining nutrient export coefficients and source loading uncertainty using in-stream monitoring data. *J. American Water Resources Assoc.* 37(1): 223-236.

Mercurio, G., J. Perot, N. Roth, and M. Southerland. 2002. Maryland biological stream survey 2000: Quality assessment report. Springfield, Va.: Versar, Inc., and Baltimore, Md.: Maryland Department of Natural Resources.

Meyer, V. R. 2002. Minimizing the effect of sample preparation on measurement uncertainty. *LC-GC Europe* (July 2002): 2-5.

Miller, P. S., B. A. Engel, and R. H. Mohtar. 2000. Sampling theory and mass load estimation from watershed water quality data. ASAE Paper No. 003050. St. Joseph, Mich.: ASAE.

Miller, R. O., and J. Kotuby-Amacher. 2005. Unpublished data. NAPT. Fort Collins, Colo.: Colorado State University and Logan, Utah: Utah State University.

Parsons, D. A. 1954. Coshocton-type runoff samplers: Laboratory investigations. SCS-TP-124. Washington, D.C.: USDA-SCS.

Pelletier, P. M. 1988. Uncertainties in the single determination of river discharge: A literature review. *Canadian J. Civil Eng.* 15(5): 834-850.

Rantz, S. E., et al. 1982a. Volume 1: Measurement of stage and discharge. In *Measurement and Computation of Streamflow*, 1-284. USGS Water-Supply Paper 2175. Washington, D.C.: USGS.

Rantz, S. E., et al. 1982b. Volume 2: Computation of discharge. In *Measurement and Computation of Streamflow*, 284-601. USGS Water-Supply Paper 2175. Washington, D.C.: USGS.

Robards K., I. D. Kelvie, R. L. Benson, P. J. Worsfold, N. J. Blundell, and H. Casey. 1994. Determination of carbon, phosphorus, nitrogen, and silicon species in waters. *Anal. Chim. Acta* 287(3): 147-190.

Robertson, D. M., and E. D. Roerish. 1999. Influence of various water quality sampling strategies on load estimates for small streams. *Water Resources Res.* 35(12): 3747-3759.

Ryden, J. C., J. K. Sayers, and R. F. Harris. 1972. Sorption of inorganic phosphate by laboratory ware: Implications in environmental phosphorus techniques. *Analyst* 97: 903-908.

Sauer, V. B., and R. W. Meyer. 1992. Determination of error in individual discharge measurements. USGS Open File Report 92-144. Washington, D.C.: USGS.

Schmidt, A. R. 2002. Analysis of stage-discharge relations for open-channel flows and their associated uncertainties. PhD. diss. Urbana-Champaign, Ill.: University of Illinois at Urbana-Champaign, Department of Civil and Environmental Engineering.

Sharpley, A. N, P. J. A. Kleinman, R. W. McDowell, M. Gitau, and R. B. Bryant. 2002. Modeling phosphorus transport in agricultural watersheds: Processes and possibilities. *J. Soil and Water Cons.* 57(6): 425-439.

Sheridan, J. M., R. R. Lowrance, and H. H. Henry. 1996. Surface flow sampler for riparian studies. *Applied Eng. in Agric.* 12(2): 183-188.

Shih, G., W. Abtew, and J. Obeysekera. 1994. Accuracy of nutrient runoff load calculations using time-composite sampling. *Trans. ASAE* 37(2): 419-429.

Skjemstad, J. O., and R. Reeve. 1978. The automatic determination of ppb levels of ammonia, nitrate plus nitrite, and phosphate in water in the presence of added mercury II chloride. *J. Environ. Qual.* 7(1): 137-141.

Slade, R. M. 2004. General Methods, Information, and Sources for Collecting and Analyzing Water-Resources Data. CD-ROM. Copyright 2004 Raymond M. Slade, Jr.

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(2): 281-290.

Teledyne ISCO. 2005. 730 Bubbler Flow Module. Lincoln, Neb.: Teledyne ISCO, Inc. Available at: www.isco.com.

Topping, J. 1972. *Errors of Observation and Their Treatment.* 4th ed. London, U.K.: Chapman and Hall.

USEPA. 2000. National water quality inventory 1998 report to Congress. USEPA 841-R-00-001. Washington, D.C.: USEPA, Office of Water.

USGS. 1999. Book 9, Section A: National field manual for the collection of water-quality data. *Techniques of Water-Resources Investigations of the United States Geological Survey.* Washington, D.C.: USGS.

Young, K. B. 1950. A comparative study of mean-section and mid-section methods for computation of discharge measurements. USGS Open File Report. Washington, D.C.: USGS.