Journal of Hydrology 534 (2016) 73–86

Contents lists available at ScienceDirect

Journal of Hydrology

journal homepage: www.elsevier.com/locate/jhydrol

Benchmarking laboratory observation uncertainty for in-pipe storm sewer discharge measurements

Marcus F. Aguilar *, Walter M. McDonald, Randel L. Dymond

Department of Civil and Environmental Engineering, Virginia Tech, 200 Patton Hall, Blacksburg, VA 24061, USA

ARTICLE INFO

Article history:
Received 18 August 2015
Received in revised form 7 November 2015
Accepted 26 December 2015
Available online 6 January 2016
This manuscript was handled by Andras Bardossy, Editor-in-Chief, with the assistance of Sheng Yue, Associate Editor

Keywords:
Uncertainty
Hydrologic data
Monitoring
Urban hydrology
Input and calibration data
Laboratory test

SUMMARY

The uncertainty associated with discharge measurement in storm sewer systems is of fundamental importance for hydrologic/hydraulic model calibration and pollutant load estimation, although it is difficult to determine as field benchmarks are generally impractical. This study benchmarks discharge uncertainty in several commonly used sensors by laboratory flume testing with and without a woody debris model. The sensors are then installed in a field location where laboratory benchmarked uncertainty is applied to field measurements. Combined depth and velocity uncertainty from the laboratory ranged from ±0.207–0.710 in., and ±0.176–0.631 fps respectively, and when propagated and applied to discharge estimation in the field, resulted in field discharge uncertainties of between 13% and 256% of the observation. Average daily volume calculation based on these observations had uncertainties of between 58% and 99% of the estimated value, and the uncertainty bounds of storm flow volume and peak flow for nine storm events constituted between 31–84%, and 13–48% of the estimated value respectively. Subsequently, the implications of these observational uncertainties for stormwater best-management practice evaluation, hydrologic modeling, and Total Maximum Daily Load development are considered.

1. Introduction

The value of accurate discharge measurements in urban storm sewer systems was first recognized as rudimentary flood gaging stations appeared in, and upstream of urban areas in the 1970s (Owen, 1979), but has since multiplied with the inclusion of water quality management in the stormwater paradigm (Roy et al., 2008). Discharge measurements paired with constituent concentration data allows for the estimation of pollutant loads, now regulated in many urban areas by the intersection of the National Pollutant Discharge Elimination System's (NPDES), Municipal Separate Storm Sewer System (MS4) program and the Total Maximum Daily Load (TMDL) program (Sections 402 and 303 of the Clean Water Act, respectively). Stormwater managers must now show that their localities are reducing pollutant runoff to achieve limits called Waste Load Allocations (WLAs), and though discharge measurement is necessary for pollutant load estimation, explicit requirements for discharge monitoring are absent from the MS4 and TMDL programs, and the regulation of discharge as a pollutant unto itself was prohibited by the U.S. District Court of Virginia (VDOT v. USEPA, 2013).

As the current regulatory environment does not require discharge monitoring, and may even disincentivize it (Wagner, 2005), only a small proportion of approximately 7000 regulated MS4 entities (USEPA, 2014) monitor discharge. Nevertheless, there are certain localities that have developed monitoring programs either through relationships with the USGS (e.g. Hoogestraat, 2015; Jastram, 2014; Storms et al., 2015), as a department of the local or regional government (e.g. City of Austin, 2009), or as consulting contracts (e.g. Gauron, 2015).

The literature provides thorough guidance on the measurement of discharge in open channels (Turnipseed and Sauer, 2010; USBR, 2001; WMO, 2010), but MS4 permits ascribe the water quality effects of urban stormwater to the underground system's terminal pipe discharging into jurisdictional waters of the U.S. – known as an "outfall." Furthermore, the treatment prescribed for urban stormwater pollution is a combination of programmatic measures and structural controls (Aguilar and Dymond, 2015) whose hydrologic and water quality benefit is yet unknown or uncertain (Barrett, 2008; Taylor and Fletcher, 2007). Detailed guidance addressing the nuances of monitoring storm sewer discharges from MS4 outfalls and stormwater best management practices (BMPs) is needed, and in particular, there is a need for characterization of the

* Corresponding author.
E-mail addresses: mfa1986@vt.edu (M.F. Aguilar), walmcdon@vt.edu (W.M. McDonald), dymond@vt.edu (R.L. Dymond).

http://dx.doi.org/10.1016/j.jhydrol.2015.12.052
0022-1694/

uncertainty associated with in-pipe sensor discharge measurements and its effects on the use of flow data for modeling and pollutant load estimation (Harmel and Smith, 2007). The type of uncertainty associated with sensor measurements is called "measurement" or "observation" uncertainty (McMillan et al., 2012) – the focus of this paper.

The term "uncertainty" should be distinguished from the term "error", which is defined as the difference between the true value and measured value (measurand), which is not operationally helpful since true values are almost never known (Moffat, 1988). Rather, uncertainty is defined as "a parameter associated with the result of a measurement that characterizes the dispersion of values that could reasonably be attributed to the measurement (WMO, 2010)." The two components of observation uncertainty as defined in Coleman and Steele (1995), and applied to hydrologic measurements in Bertrand-Krawjewski and Muste (2008a) are (1) uncertainty due to bias, and (2) precision uncertainty (Fig. 1). While bias and precision are typically thought of as error sources, they are defined as components of uncertainty in this paper, as the true value of the measurand is not known. Bias uncertainty is the systematic difference between the mean of the observations and the benchmark value, while precision uncertainty is the random scatter of observations about the mean, conforming to some probability distribution, and generally described by a simple statistic such as the standard deviation.

As discharge benchmarks are generally not available, there are limited studies that attempt to quantify the components of discharge observation uncertainty with applications to storm sewer field measurements. McMillan et al. (2012) provide a meta-analysis of observation uncertainty for various types of hydrologic measurements, and Lee et al. (2014) apply a standardized uncertainty framework to river flow sensor observations, but neither provide specific information regarding the measurement of discharge in storm sewer pipes. McIntyre and Marshall (2008) and Rehmel (2008) partially fill this gap by comparing acoustic Doppler current profiler observations to the commonly used impeller current meter in nine storm sewer cross sections, and 43 USGS stations respectively, however no attempt was made to perform laboratory benchmarking in these studies. Maheepala et al. (2001) perform flume calibration of flow sensors that are then placed in storm sewer pipes and evaluated, but the procedure and results of laboratory work are not reported. Heiner and Vermeyen (2013) performed laboratory evaluations of nine sensors in a rectangular, circular, and trapezoidal channel, though laboratory constraints allowed comparisons at a limited number of discharge values, and the lab results were not applied to field measurements. The literature on discharge monitoring uncertainty lacks the connection between laboratory benchmarking of sensor uncertainty, application of that uncertainty in the field, and the implications of uncertainty for stormwater monitoring, modeling, analysis, and decision making.

The purpose of this study is to benchmark the uncertainty associated with discharge measurements from several common sensors for their use in storm sewer monitoring and modeling. To do this, uncertainty is determined in the laboratory under controlled conditions, and with the effects of a woody debris model. Laboratory benchmarked uncertainty is then applied to field measurements, and finally the implications of observational uncertainty for urban storm water monitoring and modeling is discussed.

2. Instrumentation

To obtain flow measurements without structural devices (e.g. weirs and flumes), electronic sensors can be used that employ a variety of technologies to measure stage and velocity in open channels and pipes. The sensors used in this study are shown in Table 1, and the technologies employed are discussed in the following sections.

2.1. Depth measurement

Sensors that employ ultrasonic (US) technology are mounted at the top of a pipe, and estimate distance to the water surface by dividing the return time-of-flight of an emitted high frequency sound wave by the velocity of that wave (Angrisani et al., 2009). A shortcoming of ultrasonic sensors is that they require a minimum distance between the sensor and water surface (known as a dead zone or blanking distance) above which the sensor is not able to take measurements (Table 1), constraining the number of potential installation sites. The US instruments tested in the laboratory were the Massa M-300/95 (relabeled as the Telog UT-33u/95) and Global Water WL705, known henceforth as the Massa and GW respectively. These sensors are similar in make, with the primary difference being that the GW includes a data logger that contains the battery power source, while the Massa must be connected to a separate logger for data collection and power.

Sensors that use pressure transducers (PTs) estimate depth using a submerged piezo-resistive chip that is exposed to water pressure and open to the atmosphere through a hose in the communication cable, such that the electrical signal from the chip can be calibrated to water depth. Depending on the sensor design, these electrical signals are processed within the device, or relayed

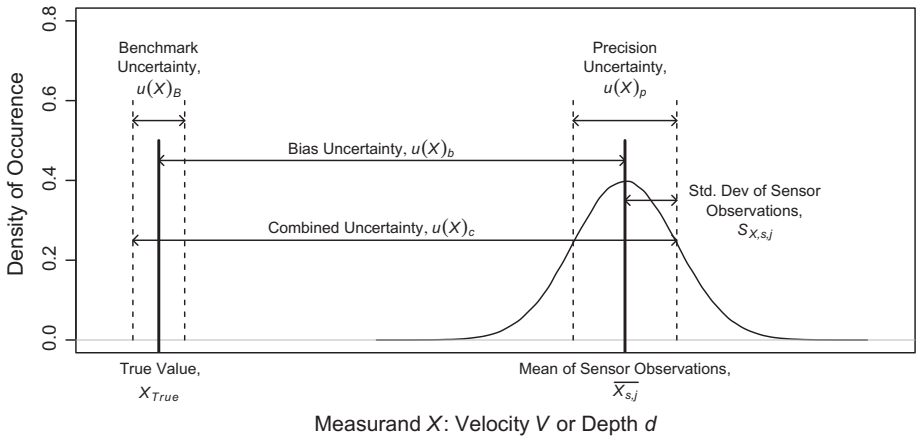

Fig. 1. Uncertainty associated with the observation of a measurand X, adapted from Coleman and Steele (1995).

Table 1
Classification, manufacturer specifications, and cost of sensors.

Sensor	Depth				Velocity				Source	Cost (USD)
	Type[a]	Range[b]	Precision (in.)	Accuracy	Type[a]	Range (fps)	Precision (fps)	Accuracy		
Massa M-300/95	US	12 in.–13 ft.	0.01	±0.1%	–	–	–	–	Massa (2014)	900
Global Water WL 705	US	4 in.–12 ft.	0.035	<±0.5%	–	–	–	–	Global Water (2014)	800
Teledyne ISCO 2150	PT	0.4 in.–10 ft.	0.012	±0.01 ft	ADV	−5 to 20	0.01	±0.1 fps[c]	ISCO (2011)	3000
FloWav PSA-AV	PT	0.9 in.–15 ft.	N/A	±0.25%	ADV	−5 to 20	N/A	±2%	FloWav (2013)	1800
SonTek Argonaut SW	PT	7.2 in.–16 ft.	N/A	±0.1%	ADCP	−16 to 16	0.003	±1%	Xylem (2009)	5000[d]
Nortek Vectrino II	–	–	–	–	ADCP	0.3–9.8	N/A	±0.5%	Nortek (2013)	10000

[a] Sensor types: ultrasonic (US), pressure transducer (PT), acoustic Doppler velocimeter (ADV), and acoustic Doppler current profiler (ADCP).
[b] Range has a different meaning for ultrasonics and pressure transducers: for ultrasonics, it represents the minimum and maximum distance between the sensor and water surface (blanking distance). For acoustic Dopplers, range represents the minimum and maximum depth of water above the sensor.
[c] The reported accuracy of the ISCO 2150 is ±0.1 fps up to 5 fps, and ±2% between 5 and 20 fps.
[d] The SonTek Argonaut SW has since been discontinued. The price shown is the cost of a SonTek IQ, a comparable unit.

to a recording and telemetry unit (RTU, also known as a data logger) for processing.

PTs in storm sewer applications have a number of shortcomings. Their installation at the bottom of a pipe affects the flow regime immediately around the sensor, subjects the pressure/communications hose to entanglement from debris, and subjects the transducer to burial from sediment. They are also subject to drift; a change in their output over a period of time that is not a function of the measured water level. To address these shortcomings, PTs must be routinely calibrated and maintained. The Teledyne ISCO 2150 ADV, FloWav PSA-AV, and SonTek Argonaut SW (known henceforth as the ISCO, FloWav, and SonTek, respectively) use PTs to sense depth.

2.2. Velocity measurement

Acoustic Doppler current profilers (ADCPs) are a subset of acoustic Doppler velocimeters (ADVs) – instruments that estimate velocity by continuously transmitting sound waves into a control volume, then measuring the Doppler frequency shift caused by reflection off particulates in the water. ADCPs provide multi-dimensional velocity measurements, while ADVs (as defined in this study) integrate samples to provide a one-dimensional water velocity (Larrarte et al., 2008). Generally, these sensors also measure depth in order to estimate a discharge based on a form of the flow continuity equation that incorporates the observed frequency shift, as shown in McIntyre and Marshall (2008).

The ability of these sensors to accurately characterize velocity in a storm sewer pipe is contingent on several factors. First, the position of the sensor on the pipe bottom subjects them to sedimentation at low velocities, and debris accumulation during storm events, potentially leading to data loss or inaccuracy if the velocity transducer is blocked or excessive turbulence is created. Second, the velocity is estimated based on the magnitude of the frequency shift from suspended particles within the conical control volume; measurement error occurs if this volume is not representative of the cross sectional velocity (Bonakdari and Zinatizadeh, 2011), or if there is an uneven vertical distribution of sediment within the volume (McIntyre and Marshall, 2008; Nord et al., 2014). Further sources of error and considerations for field implementation are enumerated in Aguilar and Dymond (2014) and McIntyre and Marshall (2008).

The ADVs/ADCPs used in this study were the ISCO, FloWav, SonTek, and Nortek Vectrino II (known henceforth as the Nortek). The ISCO and FloWav are designed in a similar fashion; they provide a single, one-dimensional velocity measurement for a conical control volume and a single depth at each time interval. They are both bottom mounted, and both require an external data logger and power source. The SonTek has a similar setup, but is larger and measures velocity in two dimensions: parallel to the primary flow direction of the channel and normal to the top face of the sensor. The Nortek is the most dissimilar of the four velocimeters used in this study as it is usually top or side mounted, does not provide a depth measurement, and uses four receiving beams at 60 degrees from the central transmit beam to measure three-dimensional velocity. Due to its sensitivity to damage from external forces commonly found in storm flow situations, it would not be an ideal sensor for autonomous storm flow measurement.

All sensors shown in Table 1 were tested in the laboratory, though several of the sensors were only available for a limited amount of testing, and were not included in the results. Manufacturer names are censored from results, as the objective of this study is to evaluate observation uncertainty, not to recommend one sensor over another.

3. Laboratory procedure

The laboratory setup was a 20 ft. long × 18 in. wide × 18 in. deep experimental channel that recirculates water with a user-controlled pump operated at variable speeds from 0 to 60 Hz at 0.1 Hz increments. To provide a direct benchmark with which to compare sensor observations, the flume provided a venturi-differential manometer configuration with an established rating curve derived from Bernoulli's equation, flow continuity, and hydrostatic pressure. The venturi reduces flow from a diameter of 6.068″ to 3.040″, and has a discharge efficiency coefficient, $C_d = 0.987$. The manometer measures the head difference in the venturi (Δh) to the thousandth of a foot. The equation relating discharge, Q (cfs) with Δh (ft.) is:

$$Q = C_d A_2 \sqrt{\frac{2g\Delta h}{1 - r^2}} \quad (1)$$

where A_2 is the cross sectional area of the contraction in the venturi in ft.^{2}, g is acceleration due to gravity (32.2 ft./s^2), r is the contraction cross section area divided by the entrance cross section area (0.251), and C_d is the discharge efficiency coefficient (0.987). Information regarding the calibration of this venturi meter is from Engineering Laboratory Design (1999). In this paper, the flow as measured by the venturi-manometer configuration is referred to as "manometer discharge".

For each of seven experiment days, the sensors were installed in the flume channel as specified by the manufacturer, and several "tests" were performed where the system was set to an initial channel depth with the pump off (i.e. zero discharge). For each test, the pump was run at increasing frequencies from 0 to 60 Hz, usually at 5 Hz increments, and the system was allowed to reach steady-state at each pump frequency. Steady-state channel depth and manometer discharge were maintained for at least four

minutes – referred to as a "trial", and annotated as j. Depth and manometer discharge were recorded for each trial, and an observation i was taken by each sensor, s, mounted in the channel every minute, referred to as $d_{s,i}$ or $V_{s,i}$ for $i = 1 \rightarrow n$ depth and velocity observations in trial j respectively.

Summary statistics were performed on each trial such that i observations were aggregated into j trials for $j = 1 \rightarrow N$ trials. The mean and standard deviation of sensor depth and velocity observations i in trial j were calculated and annotated as $\overline{X_{s,j}}$ and $S_{X,s,j}$ respectively, where X represents the measurands depth, d and velocity, V, and S represents the standard deviation of i observations of the measurand. The following section describes how $\overline{X_{s,j}}$ and $S_{X,s,j}$ were used to benchmark observation bias and precision uncertainty respectively, and how the uncertainty associated with the benchmark itself was defined. Uncertainty associated with a measurand X will be annotated as $u(X)_k$ where k is one of the three components of uncertainty defined for this study: bias $u(X)_b$, precision $u(X)_p$, and benchmark $u(X)_B$; and where $u(X)_C$ is the combined uncertainty of the three components (Fig. 1).

Measurements taken outside of manufacturer reported sensor limitations (e.g. minimum depth or blanking distance, see Table 1) were removed from analysis, as the sensors were not expected to operate outside of specified conditions. Other erroneous measurements noted during sensor data post-processing were marked and censored from analysis; this will be discussed further in Section 5.

3.1. Benchmark uncertainty

The uncertainty associated with the benchmark measurements ($u(X)_B$) was calculated for the laboratory, as it provided a measure of the extent to which the depth and velocity could be measured given a set of laboratory tools. This is referred to as epistemic uncertainty – the limit to what can be known about a system; its counterpart – aleatory or "natural" uncertainty – is discussed later in this section (definitions from Merz and Thieken (2005)). It was not possible to evaluate benchmark uncertainty in the field, as the depth and velocity were not directly observable, but it was assumed that the instruments used for observation of hydraulic parameters in the field would be at least as uncertain as those used in the laboratory. As a result, laboratory benchmark uncertainty is applied to field measurements as a conservative means of describing the upper limit of what can be directly observed in the field.

The uncertainty associated with the discharge benchmark based on the 0.001 ft. head increments applied to the manometer equation (Eq. (1)) was ±0.013 cfs. The pump frequency-initial channel depth–discharge relationship was created by filling the flume to an initial depth in the channel, running the variable speed pump from 0 to 60 Hz, and estimating discharge based on manometer readings. The initial depth in the flume was then incrementally increased, and the process was repeated. Non-linear least squares regression was performed separately for each initial channel depth to relate manometer discharge to pump frequency based on a power-law relationship. The uncertainty associated with these rating curves was defined as the maximum standard error of the residuals (RSE) of these power-law models (0.01 cfs) where:

$$\mathrm{RSE} = \sqrt{\frac{\sum_{i=1}^{n}\left(X_{o,i} - X_{p,i}\right)^2}{n-2}} \tag{2}$$

and $X_{o,i}$ is the observation of the value by direct measurement, $X_{p,i}$ is the predicted value from the model, and n is the number of observations ($56 \leqslant n \leqslant 92$). The RSE is in the same units as the measurement and provides a metric of the difference between modeled values and observed values at all discharges. The total uncertainty associated with the manometer reading and frequency-discharge rating curve was calculated based on the probable error range defined in Bertrand-Krawjewski and Muste (2008a), Harmel and Smith (2007) and Sauer and Meyer (1992):

$$u^2(y) = \sum_{i=1}^{N} u^2(x)_i \tag{3}$$

where $u(y)$ is the combined uncertainty of N independent components of uncertainty, $u(x)_i$, and for the manometer and rating curve, the combined discharge benchmark uncertainty, $u(Q)_B$ was ±0.016 cfs.

Depth, $d_{o,i}$ was measured in the channel with a depth gauge attached to the plexi-glass channel wall perpendicular to the water surface for a range of depths between zero and twelve inches, with a benchmark uncertainty, $u(d)_B = \pm 0.0625$ in. Benchmark velocity, $V_{o,i}$ was estimated based on the manometer discharge observation divided by cross sectional flow area (as calculated by $d_{o,i}$). The uncertainty in the velocity benchmark is the propagated uncertainty in depth and discharge benchmarks based on the principle from Coleman and Steele (1995):

$$u^2(y) = \sum_{i=1}^{N}\left(\frac{\partial f}{\partial x_i}\right)^2 u^2(x_i) \tag{4}$$

where $u(y)$ is the uncertainty of a value calculated as a function of x_i, and f is that function. The application of this principle to the estimation of velocity uncertainty as a function of depth, discharge, and their respective uncertainties is:

$$u(V)_B = \sqrt{\left(\frac{\partial V}{\partial Q}\right)^2 u(Q)^2 + \left(\frac{\partial V}{\partial d}\right)^2 u(d)^2 + \left(\frac{\partial V}{\partial w}\right)^2 u(w)^2} \tag{5}$$

where V is velocity in fps (estimated as $V = Q/wd$), d is channel depth in ft., w is channel width in ft. (constant and relatively certain) and Q is discharge in cfs. The benchmarked velocity uncertainty, as it is a derived value, varies with manually observed depth and manometer discharge, such that the maximum $u(V)_B$ achieved in the laboratory was ±0.272 fps as Froude number, $Fr \rightarrow 1$. Although it was possible to calculate the velocity benchmark uncertainty in the laboratory, it was not possible to do so in the field, as depth and discharge were not directly observable in the field. As such, the bootstrapped upper 95% confidence interval of the velocity benchmark uncertainty was used, such that $u(V)_B = 0.056$ fps.

3.2. Bias uncertainty

The bias uncertainty was defined as the difference between the benchmarked value ($X_{o,j}$) and the mean of i sensor observations for each trial, j ($\overline{X_{s,j}}$), and was calculated as the standard error of the observation residuals using the equation for RSE:

$$u(X)_b = \sqrt{\frac{\sum_{i=1}^{N}\left(X_{o,j} - \overline{X_{s,j}}\right)^2}{N-2}} \tag{6}$$

First, all $\overline{X_{s,j}}$ were plotted against their respective manual observation, $X_{o,j}$ with a line of perfect agreement (1:1 line) for reference. Then, observation residuals ($X_{o,j} - \overline{X_{s,j}}$) were plotted against channel depth, velocity, and Froude number (Fr) to determine the effects of these parameters on the residuals. The 1:1 plots and residual plots were visually inspected, and it was determined how best to model sensor observations as a function of manual observations, and thereby transform sensor observations to better fit manual observations.

Models were built for $X_{o,j}$ as a function of $\overline{X_{s,j}}$ so that sensor observations could be transformed to better fit manual observations, and so this transformation could be applied in the field. These transformation functions were created in three ways depending on the transformation necessary to meet

homoscedasticity and autocorrelation assumptions of linear modeling in the adjusted (i.e. transformed) mean observations ($\overline{X_{s,j,adj}}$): (1) linear least-squares regression, (2) non-linear least-squares regression and (3) addition of observation residual models (as a function of *Fr*) to sensor observations. These transformation functions were applied to the original sensor observations to adjust them toward the line of perfect agreement with direct observations, and to remove heteroscedasticity and autocorrelation. Subsequent linear models were then built for the manual observations as a function of these adjusted sensor observations ($\overline{X_{s,j,adj}}$), and the RSE of this adjusted model was reported as the adjusted bias uncertainty ($u(X)_{b,adj}$).

The date of testing was later included in these models as a categorical variable to determine if sensor depth observations drifted with time, but only for the purpose of commenting on lab conditions, as direct depth observations were not measured in the field. All linear and non-linear least-squares regression modeling was performed in the R software language using the "lm" and "nls" functions respectively (R Core Development Team, 2014).

3.3. Precision uncertainty

Precision uncertainty (Fig. 1) is defined as the random scatter of sensor observations about the mean value of these observations due to electrical limitations of the sensor, and environmental stochasticity [i.e. aleatory uncertainty (Merz and Thieken, 2005)]. Testing in the laboratory was designed to minimize the effects of environmental conditions, although the very presence of the bottom mounted ADVs and ADCPs created non-uniformities in the channel. It was not possible to control for this, and was deemed acceptable as the same effects would be present in the field installation.

This random scatter was evaluated for N trials at steady-state channel depth and velocity as the standard deviation of $n > 3$ sensor observations ($S_{X,s,j}$). These standard deviations were calculated for trials across a range of depths and velocities in the channel, and it was hypothesized that precision uncertainty would increase as $Fr \rightarrow 1$, as the flow regime in the channel was least stable under critical flow conditions. To test this hypothesis, $S_{X,s,j}$ was plotted as a function of the manually observed depth, velocity, and Froude number, and the plots were visually inspected for trends. If no trend was present, the precision uncertainty ($u(X)_p$) was calculated as the median of all $S_{X,s,j}$, as the median is robust against non-normality and outliers.

3.4. The effect of woody debris

Previous experience with storm sewer discharge monitoring, and the results in McCuen et al. (2014) has shown that storm sewer systems are subject to the accumulation of trash and debris, especially in large pipes and at locations where open channels enter daylighted culverts, as there are no preventative measures for removing these detritus before they enter the storm sewer system. Although it was not possible to build an exact scale model of any particular debris jam, the principles from Gippel et al. (1996) were used to approximate a woody debris jam that might exist in a storm sewer pipe.

A modular debris field was constructed from 3/4″ and 1/8″ balsa wood dowels mounted to a ½″ plywood sheet cut to fit the width of the flume channel so that the field could easily be installed and removed from the channel (Fig. 2). The previously described testing procedure was used to benchmark sensor bias and precision uncertainty with debris set in the channel at a fixed distance upstream of each sensor, so that a comparison could be made with the observation uncertainty without debris in the channel. Two days of experimentation were performed where tests were run without debris in the channel, and the same tests were repeated with debris in the channel.

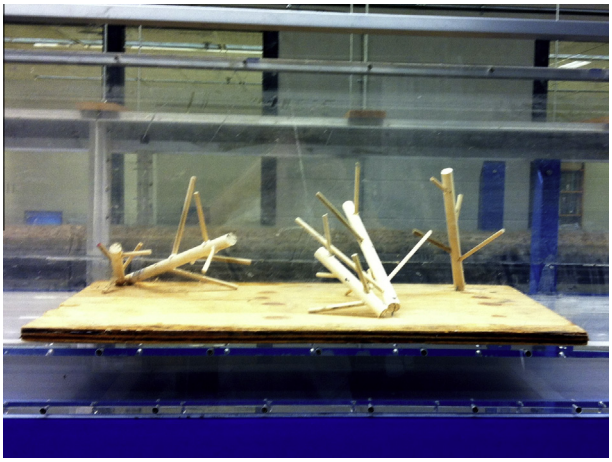

Fig. 2. Model woody debris field used to evaluate sensor uncertainty with obstructions in storm sewer channels.

The presence of the debris model was first used to subset data to determine $u(X)_b$ and $u(X)_p$ of the sensors with and without debris in the channel, and then as a presence/absence categorical variable to determine if the presence of debris had a significant effect ($p < 0.05$) on model specification or coefficients.

3.5. Combined uncertainty

The combined observation uncertainty for a measurand was calculated as the probable error range (Eq. (3)) of adjusted bias, precision, and benchmark uncertainty:

$$u(X)_c = \sqrt{u(X)_{b,adj}^2 + u(X)_p^2 + u(X)_B^2} \tag{7}$$

To evaluate the relative importance of depth and velocity sensor uncertainty on total discharge uncertainty, the principal of error propagation (Eq. (4)) from Bertrand-Krawjewski and Muste (2008a) and Coleman and Steele (1995) was used, such that:

$$u(Q) = \sqrt{u(V)^2(wd)^2 + u(d)^2(Vw)^2} \tag{8}$$

as $Q = Vwd$ for both the flume channel and field installation.

4. Field procedure

4.1. Selection of a field location and installation

As the cost for purchase, operation, and maintenance of instrumentation for water measurements may limit the number of monitoring stations, it is important that installation sites are carefully selected. Certain locations might be considered based on institutional knowledge or regulations, though if a city maintains a storm sewer GIS database, this can be helpful in systematically identifying potential monitoring locations.

The location for the installation of sensors in this study was selected by (1) querying a storm sewer GIS database described in Aguilar and Dymond (2014) and (2) consulting the municipal authority (known as the MS4 operator) responsible for the storm sewer system in which the sensors were to be placed. The GIS was used for a preliminary selection of sites in the MS4, using a series of logical queries. First, all reinforced concrete pipes were selected, as open channels can have unstable cross-sections, and

other materials (e.g. corrugated metal) can produce non-uniformities in the flow. As large fluctuations in depth, velocity, and discharge were desired for this experiment, preference was given by descending pipe size and upstream watershed area. A practical consideration was the accessibility of the sensors for maintenance, as manufacturers recommend weekly maintenance until sufficient understanding of environmental conditions is obtained (ISCO, 2011), and many storm sewer pipes require confined space entry. Locations where the storm sewer system daylights into open streams serve as highly accessible monitoring locations, but are more susceptible to entrance and exit turbulence.

A short-list of potential monitoring locations was created by querying these criteria in the GIS, which was then presented to the MS4 operator. The operator provided insight into locations that might make valuable study areas – for example where future development was proposed, or where flooding complaints had occurred. The inclusion of this institutional knowledge added a layer of validation to the GIS exercise that would have otherwise been missed.

The final site selected was a 70 in. wide × 49 ¾ in. tall rough bottom rectangular reinforced concrete pipe at 0.85% slope, draining a 490 acre watershed near downtown Blacksburg, Virginia. The site was selected as it drained an area that was larger and more urbanized than the other potential sites, and there was future infill development proposed within the contributing watershed. The rectangular pipe was accessed through a manhole in an upstream junction box, which was built as a transition from two parallel 60 in. diameter reinforced concrete pipes upstream of the junction box. The junction box had artificial low-flow channels formed into the bottom to transition baseflow from the two upstream circular pipes, to a single downstream rectangular pipe.

Two ADVs and one US (referred to as Sensor 1, Sensor 2, and Sensor 3 respectively) were installed along the center line of the downstream rectangular pipe at 4 ft., 11.5 ft., and 18.5 ft. downstream of the end of the junction box respectively, as constrained by the lengths of the communication cables. All sensors were mounted to the pipe using concrete screws, and the communication cables were fastened to the pipe bottom, side, or top using plastic clamps to assure that debris would not be caught on the cables. Sensors 2 and 3 were connected to a single RTU, and Sensor 1 used a separate RTU; both of these were hung from a ¾ in. steel all-thread hanger in the manhole above the junction box. Sensors were left in the field location for 2.5 months, recording nine storm events in that time period.

There were several shortcomings of the site. First, the sensors may have been subject to non-uniformities in the flow produced by the transition from the upstream circular pipes and junction box – especially Sensor 3 (the ultrasonic), as it was closest to this transition. As sensors must generally be tethered to an RTU, site access and hydraulic stability are trade-offs unless additional cable lengths can be acquired. The site also did not allow direct observation of the hydraulic conditions during storm events, as this would have been dangerous to the observer. Nevertheless, the site was chosen because it drained the largest area, had the same cross sectional shape as the laboratory channel, and had the best access given the two previous constraints.

4.2. Sensor power considerations

During laboratory and field testing, it was noted that a sensor's power requirements are important considerations. The sensors were designed for field use, and were powered by batteries, which in some cases could be recharged using solar panels (Sensors 2 and 3). However solar recharge was not available in the lab or field setting, and the testing was power intensive as samples were needed at one minute time intervals. Sensor 1 conserved power using a feature that allowed observations at two time intervals given a velocity or depth threshold; it was set to record measurements at 1 minute intervals when depth was above 2 in., and 15 min intervals otherwise. This option was not available for Sensors 2 and 3.

The manufacturer supplied battery power was insufficient for long term testing of Sensors 2 and 3, so in the laboratory, the data logger for these sensors was connected to a laboratory DC power supply set at a constant 12 V. As the sensors were installed in a manhole covered junction box, there was no practical way to use solar recharge, so instead two 12 V, 21 A-h lead-acid batteries were purchased so that one battery could power the sensors while the other recharged.

To determine the battery rating necessary for Sensors 2 and 3, the total load from the sensors was estimated using the DC power supply in the laboratory. As the sensors required 15 s of approximately 0.20 Amps (A) for measurements every minute, and the floating load was approximately 0.001 A for the remaining 45 s, the total current required for one minute was 0.20 A × 0.25 + 0.001 A × 0.75 = 0.051 A. As it was desired that batteries would last in the field for two weeks without replacement with an additional 10% margin (370 h), it was determined that a battery rated at 0.051 A × 370 h = 19 A h or greater be purchased. Above this rating, the weight of the battery became limiting, as both RTUs would need to be hung from the ¾ in. steel all thread hanger in the junction box manhole.

Similar computations could be performed for other sensor-RTU combinations, though many sensors have proprietary battery mounts, making it difficult to use different power configurations. Sensor 1, for example, used two 6 V lantern batteries mounted inside the RTU in a manner that would have been difficult to use other battery supplies. However, the power demands of Sensor 1, while not directly measured, appeared to be significantly smaller than Sensors 2 and 3, as the lantern batteries were only replaced once in the 2.5 months of field use.

4.3. Application of laboratory uncertainty to field observations

Time series data for the three depth and two velocity sensors installed in the field were downloaded and integrated into a single time series so that analysis could be more easily performed. First, the laboratory defined transformation equations were applied to the field data to determine if these equations reconciled the observations from different sensors to each other. Second, the combined uncertainty of the depth and velocity measurements was applied to the field data using Eq. (8) to determine the total uncertainty of discharge measurements for each sensor. Finally, the level of agreement between the sensors was assessed for each recorded storm event using a variation of the Nash–Sutcliffe Efficiency (NSE) (Nash and Sutcliffe, 1970):

$$E_{a,b} = 1 - \frac{\sum_{t=1}^{T}\left(Q_a^t - Q_b^t\right)^2}{\sum_{t=1}^{T}\left(Q_a^t - \overline{Q}_a\right)^2} \tag{9}$$

where $E_{a,b}$ is the NSE coefficient representing the goodness-of-fit of sensor b's discharge observations (Q_b^t) with respect to sensor a's discharge observations (Q_a^t) and compared to sensor a's mean discharge $\overline{Q}_a$. As the NSE coefficient is commonly used to assess the predictive power of hydrological models by comparing simulated and observed hydrographs, it is used in this paper to compare the hydrographs observed using two different flow sensors in order to comment on how well models should fit measured data. The NSE coefficient can range between $-\infty$ and 1, however values greater than 0.5 are generally deemed acceptable for model simulations (Moriasi et al., 2007).

5. Results and discussion

The results of all laboratory experiments were first plotted as time-series in order to visually inspect the agreement between sensor and manual observations, and several problems were noted in doing so. First, Sensor 1's velocity transducer reported zero values (though not consistently) at manometer derived velocities of 0.3 fps and less, suggesting a minimum detection threshold. Zero velocities and depths were also reported by both Sensor 1 and 2 at random times in the experiments, contrary to channel conditions and serial measurements. These erroneous values were censored from analysis.

It was also noted that at the end of trials when the pump was turned off and velocity in the channel rapidly dropped to zero, Sensor 1's velocity sensor continued to report the previously reported value (i.e. the transducer appeared to be "stuck") until the velocity changed significantly from zero, though it was not possible to directly quantify what allowed the sensor to resume taking observations. These values were also censored from analysis, as this phenomenon was unlikely to occur in the field, since the rate of change of velocity in the field was limited to the ascending and descending limbs of storm hydrographs.

As the sensors tested did not provide any sort of error flagging, data post-processing was necessary to assure that uncertainty benchmarking and field data analyses were not skewed by erroneous data. In general, data validation conformed to recommendations in Bertrand-Krawjewski and Muste (2008b), and it should be noted that as the number of measurements increases, data validation and post-processing can become a limiting factor for large-scale monitoring programs.

This section presents the results of the laboratory benchmarking of bias, precision, and combined uncertainty; the effects of woody debris therein, and the application of laboratory benchmarking to observations in the field. Although benchmark uncertainty is a component of combined uncertainty, benchmark uncertainty is not discussed further in this section, as its estimation and outcomes are included in the laboratory procedure.

5.1. Bias uncertainty

Uncertainty due to bias was defined as the standard error of the residuals (RSE), $X_{o,j} - \overline{X_{s,j}}$ for j trials. This was first calculated based on the RSE of the unadjusted $\overline{X_{s,j}}$, then models were built to transform these values to $\overline{X_{s,j,adj}}$, and the adjusted RSE was calculated. The purpose of this adjustment was to determine the relationship between sensor observations and laboratory benchmarks, so that this relationship could be applied to sensor observations in the field. The transformation functions, unadjusted bias uncertainty, and adjusted bias uncertainty are shown in Table 2.

Fig. 3 shows the unadjusted and adjusted observations, $\overline{X_{s,j}}$ with 60% transparency, and the associated least-squares regression lines demonstrating improved agreement after transformation. Bias uncertainty was reported as the RSE of the adjusted and unadjusted regression lines as shown in Table 2. Observations from the three depth sensors had a linear relationship with the manual observations (Fig. 3A–C), though the velocity sensors did not (Fig. 3D–E); Sensor 2's velocity observations did not appear to have any relationship to manual observations, and as such, no least-squares regression line for the unadjusted observations is shown (Fig. 3E).

Sensor observation residuals were then plotted as a function of manually observed depth, velocity, and Froude number to determine how these parameters affected residuals. Froude number (Fr) plots are shown in Fig. 4, as Fr represents the mixed effect of depth and velocity on residuals. The magnitude of the residuals

Table 2
The transformation functions used to adjust sensor depth (d_s) and velocity (V_s) observations toward manual observations with and without debris, the bias uncertainty of the original observations and the adjusted observations, precision uncertainty, benchmark uncertainty, and combined uncertainty from Eq. (7). Values of uncertainty are in +/− the unit of measurement.

	Sensor	Type	Debris	Number of trials	Transformation function	Unadjusted bias uncertainty	Adjusted bias uncertainty	Precision uncertainty	Benchmark uncertainty	Unadjusted combined uncertainty	Adjusted combined uncertainty
Depth (in.)	1	PT	N	465	$(d_s + 0.106)/1.038$	0.744	0.707	0.031	0.0625	0.747	0.710
			Y[a]	108	$(d_s + 1.27)/1.12$	0.863	0.198	0.040	0.0625	0.866	0.211
	2	PT	N	349	$(d_s + 0.176)/1.063$	0.347	0.279	0.141	0.0625	0.380	0.319
			Y	132	$(d_s + 0.156)/1.053$	0.444	0.393	0.173	0.0625	0.481	0.434
	3	US	N	353	$(d_s - 0.006)/0.993$	0.192	0.192	0.044	0.0625	0.207	0.207
			Y[b]	102	$(d_s - 0.568)/0.864$	0.394	0.404	0.085	0.0625	0.408	0.418
Velocity (fps)	1	ADV	N	287	$0.875 \times V_s^{0.968}$	0.314	0.159	0.050	0.056	0.323	0.176
			Y[c]	92	$0.692 \times V_s^{0.912}$	0.641	0.166	0.059	0.056	0.646	0.185
	2	ADV	N	349	$V_s - 3.42 + 4.65 \times Fr^{1/2}$	1.44	0.588	0.221	0.056	1.458	0.631
			Y[d]	132	$(V_s - 0.390)/0.956$	0.479	0.338	0.179	0.056	0.514	0.387

[a] Debris significantly affects y-intercept, but not slope.
[b] Debris significantly affects y-intercept and slope.
[c] Debris significantly affects nls parameters.
[d] Debris changes the form of the transformation function.

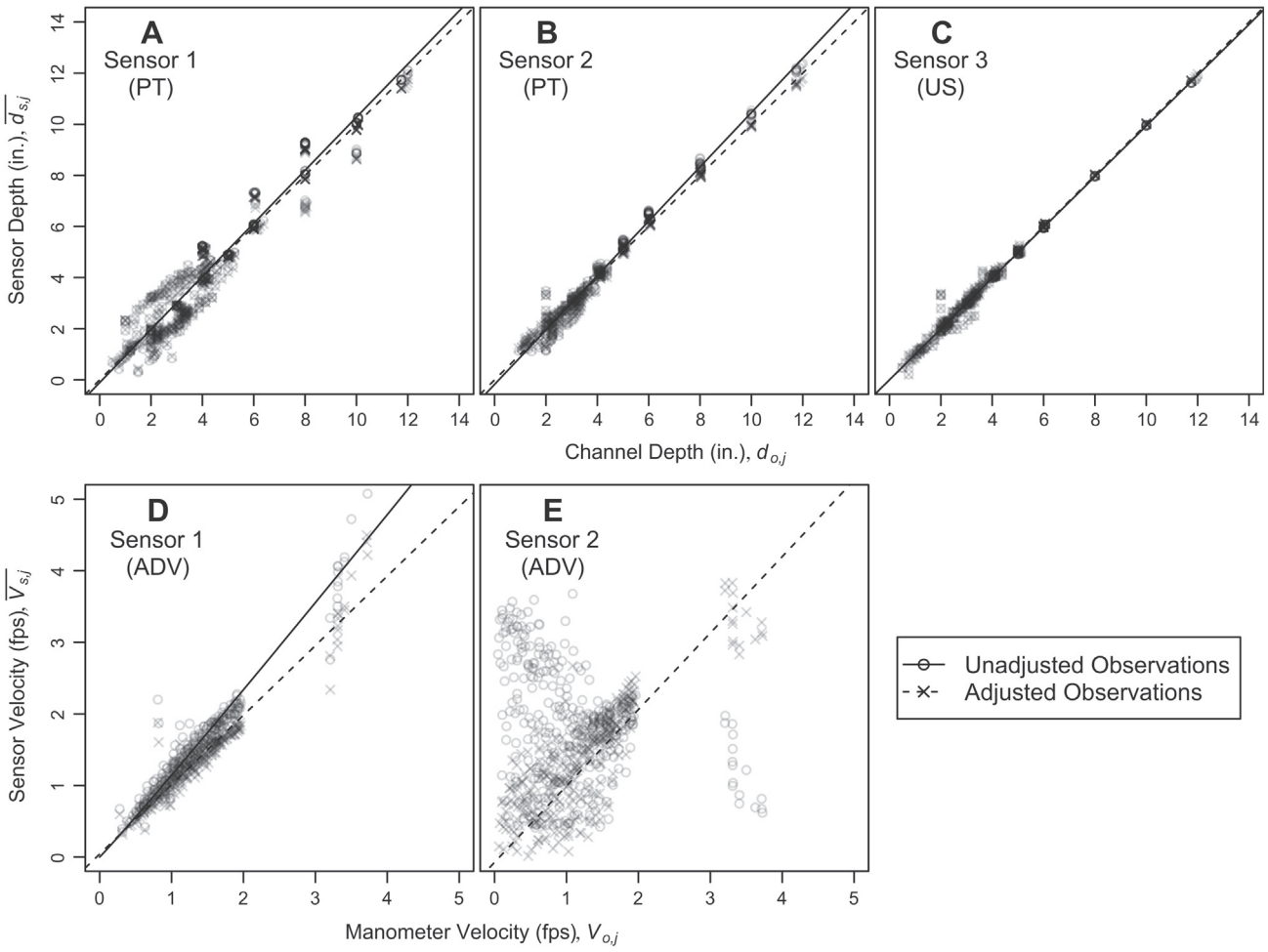

Fig. 3. Sensor depth observations plotted against manual observations (A–C), and sensor velocity observations plotted against manometer derived velocity (D–E) for all trials with least-squares regression lines shown.

of the three sensor's depth observations were invariant with manually observed depth (Fig. 4A–C), so a linear regression was used to transform $\overline{d_{s,j}}$ to vary about the line of perfect agreement. Sensor 1's velocity observations expressed a non-linear relationship with manometer derived velocity, and as such a non-linear least-squares regression model was created to transform Sensor 1's velocity observations (Fig. 3D). Sensor 2's velocity observations did not appear to have any correlation with the manometer derived velocity (Fig. 3E), although observation residuals appeared to vary systematically with *Fr* (Fig. 4E). To adjust Sensor 2's velocity observations, a linear least-squares regression model was developed for the observation residuals as a function of *Fr* (Fig. 4E), and the adjusted velocity observations were calculated as the sum of the unadjusted observations and the output of this linear model built for observation residuals as a function of Froude number.

After initial experiments, it was noted that Sensor 1's pressure transducer (PT) systematically overestimated depth, as the sensor vs. manual depth observation line had a slope of approximately 1, but the *y*-intercept was 1.07 in. (i.e. the sensor systematically overestimated depth by 1.07 in.). This was thought to be caused either by sensor drift, or a mounting configuration in a stainless steel bracket with silicone caulk that prevented the sensor's pressure transducer from being properly exposed to water pressure. The sensor was re-mounted in an improved bracket without caulk, and calibrated by placing the sensor in a known depth of water, and setting the sensor's depth observation to the known depth. This shifted the intercept of the depth observations model by 0.74 in. toward the actual values. This was the only manual calibration or modification to any sensor that was performed, though Sensor 1's depth observations continued to shift between experiments, with a range of *y*-intercepts from −1.12 to +1.07 in. When the date of the experiment was included as a categorical variable to shift the *y*-intercept for different dates, the unadjusted bias uncertainty was reduced to ±0.276 in.

Sensor 2 and 3's depth observations also showed significant shifts between experiments (Sensor 3's to a lesser degree), though there was no exogenous factor that could be found to cause this shift. The range of *y*-intercepts (i.e. measurement drift) for different experiment dates for Sensor 2's PT ranged from −0.20 to 0.71 and Sensor 3's ultrasonic (US) transducer from −0.02 to 0.11. As it was not possible to apply this information to the field, the overall *y*-intercept was used to adjust field measurements, and the sensor drift noted is included in bias uncertainty.

5.2. Precision uncertainty

Sensor precision uncertainty was evaluated as the standard deviation of $n > 3$ sensor observations ($S_{X,s,j}$) when depth and velocity in the channel were at steady-state. It was hypothesized that precision uncertainty would increase as $Fr \rightarrow 1$ from either direction, as sensors aggregate many measurements into a single reported value of depth and velocity under these non-uniform conditions. Plots of $S_{X,s,j}$ as a function of $Fr_{o,j}$ at these steady-state conditions are shown in Fig. 5, but were inconclusive to this end. It was

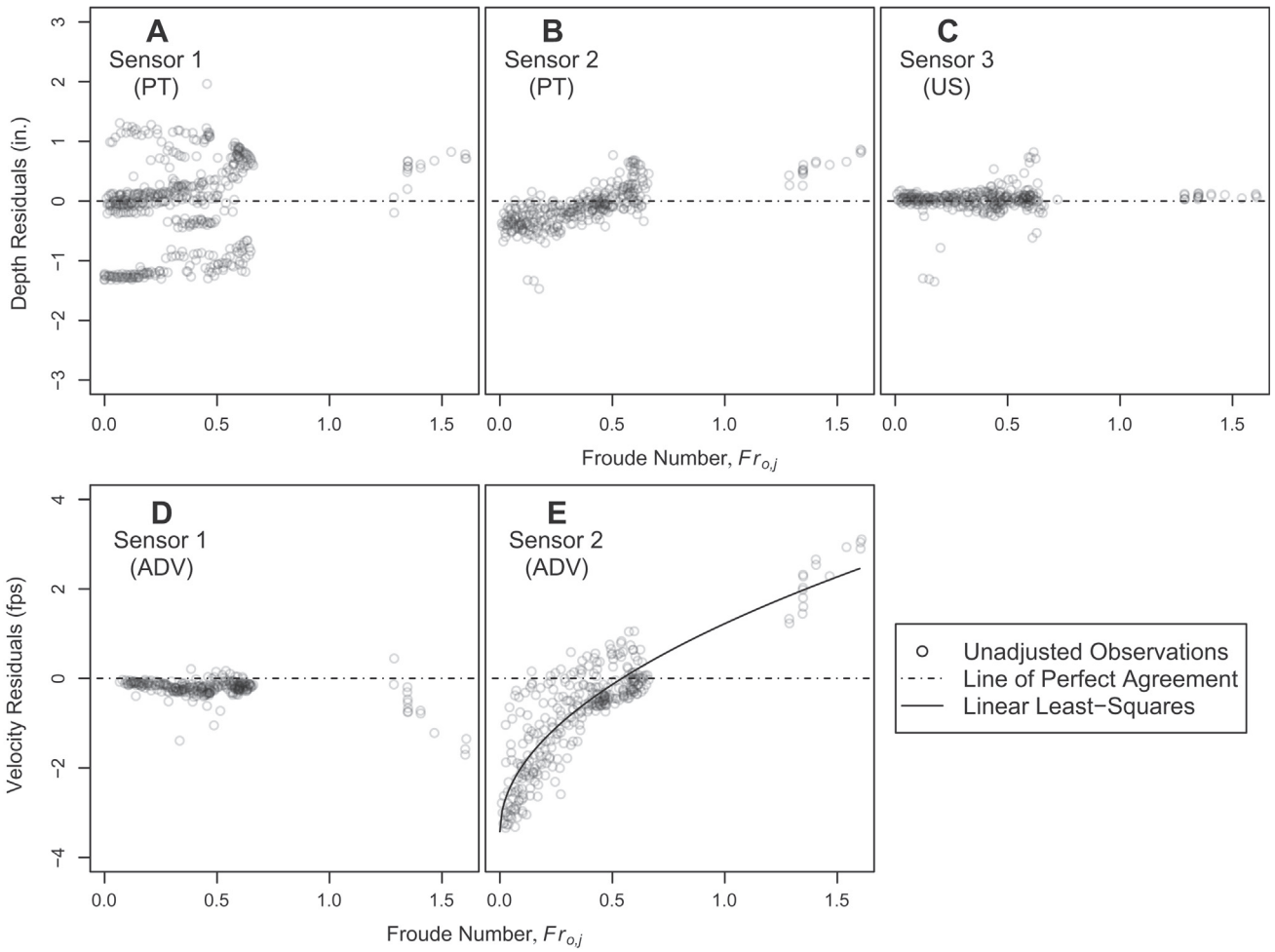

Fig. 4. Sensor observation residuals, $X_{o,j} - \overline{X_{s,j}}$ for depth (A–C) and velocity (D–E) with line of perfect agreement. (E) Sensor 2's velocity residuals varied with Froude number; the least-squares regression line is shown.

not possible to evaluate $S_{X,s,j}$ at $0.72 < Fr < 1.28$, as the channel entered gradually varied flow conditions in this range, and the varying depth and velocity along the length of the channel prevented comparison to manual observations. As a result of this gap in the data, precision uncertainty is reported as the median of all $S_{X,s,j}$ for each sensor, with the caveat that near critical flow, sensor precision uncertainty is likely to increase based on the principles of the measurement devices, but not substantiated by experimental results.

5.3. The effect of woody debris

The effect of the presence of a woody debris field in a storm sewer on the uncertainty of sensor observations was tested by placing a balsa wood model designed to approximately simulate a debris jam in the flume at a set distance upstream from the sensors. The effects of the debris jam were evaluated by assessing how the presence of debris affected observation residuals, and if the presence of debris had a significant impact on sensor uncertainty. The significance of the debris field, transformation functions, and associated uncertainties are shown in Table 2.

The bias uncertainty associated with Sensor 1's depth sensor significantly increased with upstream debris, though the transformation function adjusted these observations so that $u(d)_{b,adj}$ was significantly smaller than the "no debris" model, since there were only two days of debris testing, and there was minimal sensor drift between these two days. The presence of debris jam in the channel resulted in insignificantly higher bias and precision uncertainties for Sensor 2's depth transducer, although the debris had a significant effect on Sensor 3's depth observations, as Sensor 3 overestimated at channel depths below three inches. These overestimations at low values had high leverage for the transformation model, and as such, the bias uncertainty of the adjusted model is actually worse than the original model.

Sensor 1's velocity transducer was significantly less certain with debris in the channel, with unadjusted bias uncertainty increasing by 104% over the unadjusted no-debris bias uncertainty due to systematic overestimation, but when adjusted, bias uncertainty was nearly that of the sensor with no debris in the channel. The presence of debris in the channel had the opposite effect on Sensor 2's velocity observation uncertainty – it significantly decreased both bias and precision uncertainty. There was a more visible agreement between Sensor 2's $\overline{V_{s,j}}$ and $V_{o,j}$ with debris in the channel, though it was not possible to determine why this was the case.

The precision uncertainty for all other transducers increased with debris in the channel, an effect that would be expected given the decrease in flow uniformity. However the decrease in precision was small relative to the increase in bias due to debris, suggesting that while non-uniformities in the flow do not significantly reduce sensor observation repeatability, they make it more difficult for the sensor to report average measures of depth and velocity that are characteristic of the hydraulic conditions.

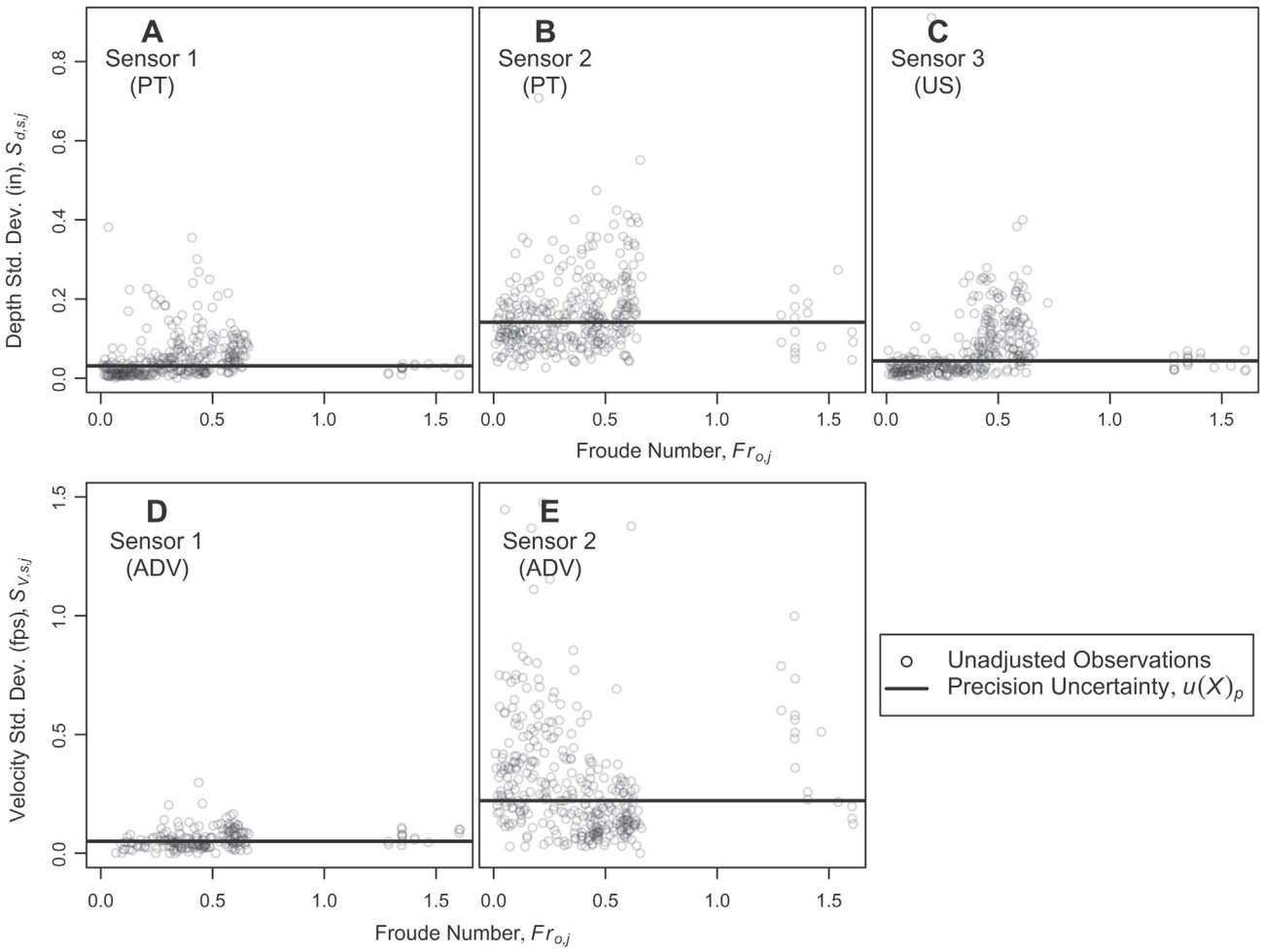

Fig. 5. Standard deviation of i sensor observations for each trial, $S_{s,j}$ plotted against manometer derived Froude number, $Fr_{o,j}$ with the median value shown representing precision uncertainty.

5.4. Combined uncertainty

The precision, bias, and benchmark uncertainty was combined using the probable error range (Eq. (7)) to arrive at a total laboratory uncertainty for each of the sensors for application to the field (Table 2). Sensor 1 had the highest $u(d)_c$ of 0.710 in. of all the depth sensors when there was no debris in the channel, however it was apparent that this was due to sensor drift between experiment dates. When the data was controlled for sensor drift, the unadjusted bias uncertainty was reduced to 0.276 in., but this factor could not be included in the transformation function because it was not possible to determine how to shift Sensor 1's depth measurements in the field.

Sensor 3 (an ultrasonic) had better bias uncertainty than both the pressure transducers, but the precision was approximately the same as Sensor 1's. As a result, Sensor 3 may provide a more reliable depth measurement than the other two, as the bias was low and minimal transformation was required, however when the water surface in the pipe is <12 in. from Sensor 3's transducer, observations from one of the pressure transducers will need to be used. This is also true of low depth conditions when there is debris in the channel, though it is generally not possible to identify these conditions without direct observation.

Sensor 1's velocity bias and precision was lower than Sensor 2's, and as a result had lower combined uncertainty with and without debris in the channel. It did not appear that there was any relationship between Sensor 2's observations and manually observed velocity (Fig. 3E) until the residuals were plotted against depth, velocity, and Froude number (Fig. 4E), but it was not clear why the residuals of Sensor 2's observations would vary with these parameters. With debris in the channel, there was at least a visible relationship between Sensor 2 observations and manual velocity observations, but with marginally better precision.

5.5. Application to field observations

The sensors were deployed in the field for 2.5 months (March 12 to May 29, 2015) during which time they recorded 9 separate storm events. The transformation equations were applied to the sensor data to determine the adjusted discharges, and the combined uncertainties from Table 2 were used with Eq. (8) to estimate the propagated uncertainty of discharge measurements. The transformation equation for Sensor 2's velocity requires a manually observed Froude number, but since there was no way to directly observe the depth or velocity in the field, the Froude number calculated using Sensor 1's adjusted observations was used as a surrogate.

While the transformation equations reduced the uncertainty in the field measurements, the NSE of the observed hydrographs showed only minor improvements for 4 of the 9 storms and an overall decrease in the average discharge NSE across all storms. This lack of improvement in the agreement by adjusting sensor

Table 3
Minimum, maximum and mean Nash–Sutcliffe efficiency (NSE) values for nine storm events using Sensor 1 as the "observed" data.

Storm	Depth				Velocity		Discharge	
	PT Unadj	PT Adj	US Unadj	US Adj	ADV Unadj	ADV Adj	Original	Adjusted
Min	−21.88	−21.26	−8.73	−9.69	−1.65	−2.82	−2.70	−15.71
Mean	−3.05	−2.95	−3.14	−3.51	−0.37	0.07	−0.49	−2.04
Max	0.55	0.55	−0.60	−0.66	0.42	0.77	0.55	0.44

discharge estimates was largely attributed to the influence that adjusted depth measurements have on the discharge, which showed only a marginal improvement in the average NSE (+0.1) for the PT and were worse for the US (Table 3). Although the adjusted velocity readings had an improved fit in 7 out of 9 storms, the velocity transformation function is inherently biased through the input of the adjusted Sensor 1 Froude number and would not be useful in most cases where there is not a second velocity measurement device.

These results do not support the hypothesis that adjusting field depth and velocity data using laboratory benchmarked values will result in better agreement between the sensors. One possible reason for this is that the physical conditions in the laboratory were restricted to the capabilities of the system and the power of the pump. Therefore, the depth and velocities simulated in the experimental flume may not be completely representative of field depth and velocity combinations. Another reason could be that the flow conditions between the two sensor locations vary during storm events, where limitations due to the site selection, such as the upstream pipe transition and junction box, result in non-uniformities in the flow. The following results should be viewed within the context of these field limitations.

The level of agreement between the sensors during storm events was first evaluated using the NSE for the unadjusted and adjusted time series (Table 3). The NSE of the depth, velocity and flow was computed for each storm event using Eq. (9), with Sensor 1 as the observed data and Sensors 2 and 3 as the simulated data. The average NSE was less than zero for all permutations of adjusted and unadjusted depth, velocity and discharge, except for the adjusted velocity, though this agreement is a result of the use of Sensor 1's depth and velocity observations for transformation.

The overall level of disagreement between the sensors is substantial, as a hydrologic model is considered to be acceptably calibrated when modeled hydrographs match observed hydrographs with NSE $\geqslant 0.5$. Only the unadjusted discharge data from the 3/14/2015 storm met this criteria (NSE = 0.55). These results suggest that when calibrating a stormflow model to observed discharges measured by an ADV, a modeler should consider whether the uncertainty in the flow measurement data merits a lowering of the acceptable NSE value.

The storm peak, volume, and their uncertainties were also evaluated using the unadjusted and adjusted observations. Fig. 6 shows the peak flow, volume, and the range of cumulative uncertainty estimated using each of the sensor's observations for the nine

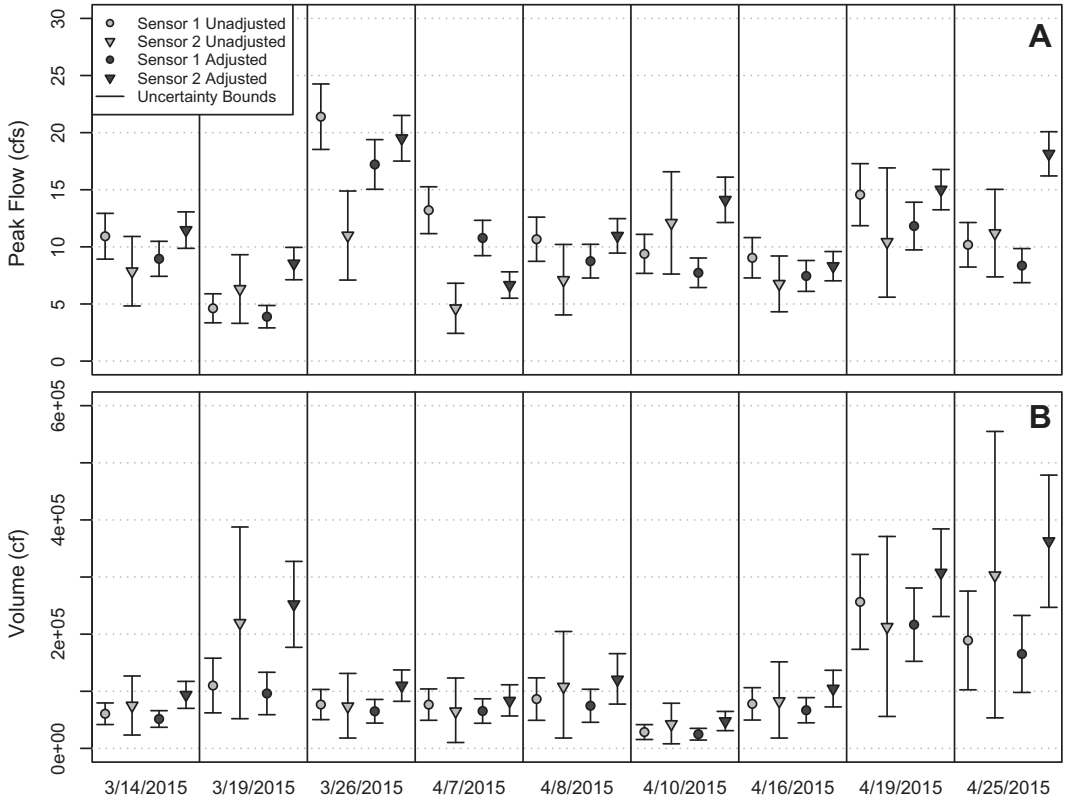

Fig. 6. (A) Peak flow and (B) total runoff volume calculated for nine storm events based on original and adjusted discharge observations from Sensor's 1 and 2 with the upper and lower bounds of laboratory benchmarked uncertainty shown.

storm events, and illustrates the effect of combined uncertainty on hydrograph characteristics that are commonly used in hydrologic modeling. For example, for the storm event that occurred on 3/26/2015, the unadjusted peak flow uncertainty varied between 7.1 and 24.3 cfs for the two sensors, and the unadjusted volume uncertainty for the storm event on 4/25/2015 ranged between 53,000 cf and 554,000 cf. Fig. 6 also illustrates the effect that the transformation functions had on the range of uncertainty, as adjustments reduced the magnitude of peak and volume uncertainty for individual sensors for all storms. Overall, the uncertainty bounds constituted 13–48% of unadjusted peak flow, 10–25% of adjusted peak flow, 31–84% of unadjusted volume, and 25–41% of adjusted volume.

It should be noted that the uncertainties applied to the field data are the combined uncertainties developed in the laboratory without the woody debris field. While it is not known if and to what extent debris existed or affected the sensors during these events, it can be assumed that the field flow conditions are more turbulent than in the laboratory; therefore, the application of laboratory benchmarked uncertainties are conservative. This may be the reason that for some storms, the peak and volume uncertainty bounds for the two sensors do not overlap.

The effect of uncertainty between the sensors over the duration of a storm is further illustrated in Fig. 7. This figure represents the adjusted depth, velocity, and discharge, and the maximum upper and lower uncertainty bounds of the storm with the best agreement between sensors (4/16/2015). Although the sensors had the highest level of statistical agreement for this storm event, and adjustments to the storm data reduced the peak and volume uncertainty of the sensors by an average of 11% and 28% respectively, the stage, velocity and discharge still exhibit a large range of uncertainty over the duration of the storm.

The demonstrated extent of discharge uncertainty has implications for watershed modeling and calibration. For example, when calibrating a watershed model, the behavior and performance of the model is evaluated through the comparison of simulated and observed storm peaks and volumes. These uncertainties suggest that as with the NSE, the peak and volume results from a simulation should be viewed within the context of the uncertainty of the data. One way to apply these results could be to use the range of uncertainty illustrated in Fig. 7 to calibrate a watershed model, where the simulated hydrograph would need to fall within the upper and lower uncertainty bounds of the discharge time series (Harmel and Smith, 2007).

Data from the field observations were also evaluated to determine the magnitude of uncertainty associated with average daily flow volumes over the entire study period as a percentage of the adjusted and unadjusted observations (Table 4). This shows a considerable amount of uncertainty, with the unadjusted uncertainty bounds constituting between 63% and 99% of average daily volume. However, the magnitude of these values may be affected by the baseflow stage and velocity conditions at the field site which, during dry periods could be less than the combined uncertainty. To test this hypothesis, the average volumes and uncertainties using only discharge data during storm events were computed, and it was found that this resulted in an average decrease of 28% of the uncertainty as a percentage of the average volume.

The average daily volumes of the unadjusted and adjusted equations vary widely between the three sensors. For example, Sensor 2's unadjusted average daily volume is 2.1 times greater than Sensor 1's, and its adjusted volume is 2.2 times greater. As with the stormflow measurements, these results suggest that although the uncertainty is reduced from the unadjusted to the adjusted data, the adjustments to the stage and velocity measurements do not produce more agreeable average volumes over the study period. These results illustrate the impact that uncertainty can have on calculations and subsequent decisions that are dependent upon flow volumes, such as the estimation of pollutant loads, or the volume reduction capabilities of a stormwater BMP.

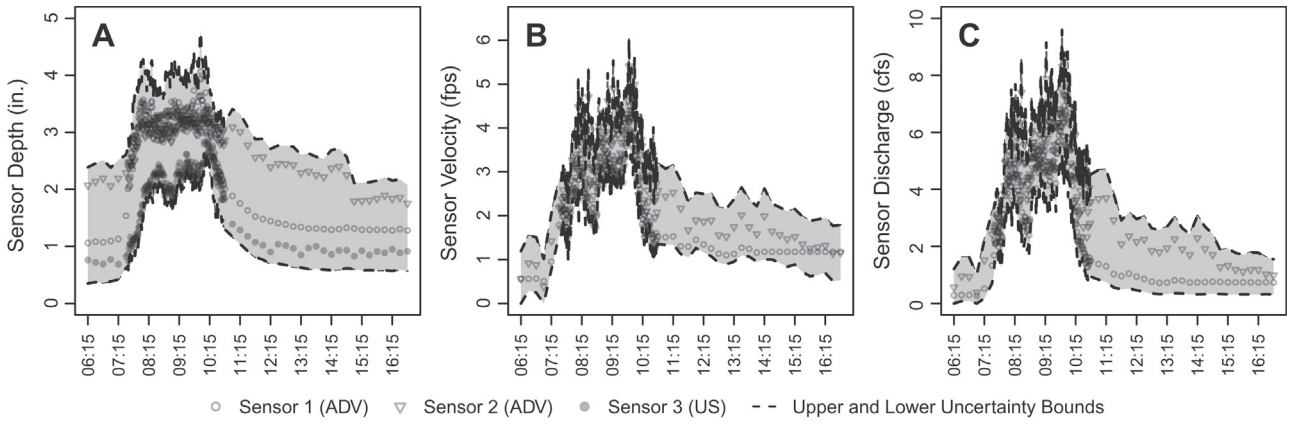

Fig. 7. Adjusted (A) depth, (B) velocity, and (C) discharge time series for storm event on 4/16/2015 with observations from three depth sensors, two velocity sensors, and the upper and lower bounds of combined uncertainty from the laboratory. The adjusted discharge observations for this storm event had the highest Nash–Sutcliffe Efficiency of any event recorded.

Table 4
Average daily volume summary.

	Unadjusted		Adjusted		Uncertainty as % of Average	
	Average Volume (cf)	Uncertainty (±cf)	Average Volume (cf)	Uncertainty (±cf)	Unadjusted (%)	Adjusted (%)
Sensor 1	70,364	44,646	62,681	36,122	63	58
Sensor 2	147,823	145,772	138,768	97,336	99	70
Sensor 3	65,660	53,440	74,783	53,108	81	71

6. Conclusions

The propagation of the laboratory benchmarked depth and velocity uncertainty to unadjusted field discharge measurements resulted in discharge uncertainties of between 13% and 107% of the observation during storm events and baseflow conditions respectively for Sensor 1, and between 34% and 256% for Sensor 2. McIntyre and Marshall (2008) report ADV uncertainty as ±20% of the observation, and Heiner and Vermeyen (2013) report deviations of −54.6% to +62.4% of venturi meter discharges for all ADVs and ADCPs tested, with some of the sensors producing considerably better results. These values contrast with the manufacturer reported discharge uncertainties of ±2.4–26% for Sensor 1 based on Eq. (8), and ±2% for Sensor 2 based on the probable error range.

This study was constrained to the use of three depth sensors and two velocity sensors in the field, but discharge measurements with lower uncertainties than those reported in this study appear to be a possibility; further experimentation is needed to benchmark the relationship between laboratory and in-pipe uncertainty using additional sensors. Furthermore, ultrasonics and pressure transducers are used widely as depth ranging devices for structural discharge measurement techniques (e.g. weirs and flumes); further research is needed to determine the effects of depth observation uncertainty on discharge estimation using these devices. It should not necessarily be inferred that the technology underlying the sensors used in this study is not suitable for measuring storm sewer discharge, but rather that conditions in storm sewers provide challenges to accurate water flow measurement that laboratory results may not be able to account for.

This was evident in the lack of agreement between depth observations in the field, even after the transformation equations from the laboratory were applied. It was thought that these adjustments would reconcile sensor observations to each other in the field, but the transformation had little effect, and in some cases caused further disagreement. This could be an artifact of the hydraulic conditions in the rectangular culvert caused by the upstream transition from dual circular culverts, though this was not possible to determine. Storm sewer operators who choose to use autonomous flow sensors in difficult (or dangerous) to access locations will have the same problems verifying sensor measurements with direct observations. As such, the balance between ease-of-access and hydraulic stability should be considered during planning so additional equipment can be purchased if necessary.

The implications of the uncertainties reported in this research for stormwater science vary based on how a sensor is used. An inflow vs. outflow study of a stormwater BMP would need to consider the uncertainty of the instruments used while determining if there was a statistical difference between inflow and outflow volume, peak flow, and pollutant load (if measured). If a sensor is used as the baseline for model calibration and validation, there are two important implications. First, uncertainty of the sensor should be considered when deciding how "good," a goodness-of-fit metric needs to be (Dotto et al., 2014; Harmel et al., 2010). Second, if the model is being used to test the effect of a treatment in a pre- vs. post-BMP implementation (i.e. longitudinal) or a treatment vs. control (i.e. paired watershed) study, the uncertainty of the sensor may subsume any effects that the treatment is hypothesized to have.

For urban stormwater management, the importance of observation uncertainty in BMP studies and hydrologic modeling is compounded in the TMDL program, where numeric effluent limits called waste load allocations (WLAs) are developed for MS4 jurisdictions based on calibrated hydrologic models, and stormwater BMPs with a regulator-defined reduction efficiency are prescribed as the means by which these WLAs are met. Margins of safety as percentages added to TMDLs (Dilks and Freedman, 2004), and volume reduction percentages ascribed to BMPs (Geosyntec Consultants and Wright Water Engineers, 2011) are relatively inconsequential if the instruments used for TMDL model calibration and BMP testing have discharge observation uncertainties as reported in this study and others. As such, both the precedent for stormwater regulatory compliance in urban areas, and the compliance methods therein are contingent on discharge measurements that are attributed a level of certainty that the results of this study suggest may not be appropriate.

References

Aguilar, M.F., Dymond, R.L., 2014. Innovative technologies for stormwater management programs in small urbanized areas. J. Water Resour. Plan. Manage. 140 (11), 04014029.

Aguilar, M.F., Dymond, R.L., 2015. Evaluation of variability in response to the NPDES Phase II stormwater program in Virginia. J. Sustain. Water Built Environ., 1–13 http://dx.doi.org/10.1061/JSWBAY.0000802, 04015006.

Angrisani, L., Baccigalupi, A., Schiano, R., Moriello, L., 2009. Ultrasonic-based distance measurement through discrete extended kalman filter. In: Moreno, V. M., Pigazo, A. (Eds.), Kalman Filter Recent Advances and Applications. InTech, pp. 269–296. http://dx.doi.org/10.5772/56411.

Barrett, M.E., 2008. Comparison of BMP performance using the International BMP Database. J. Irrigation Drain. Eng. 134 (5), 556–561.

Bertrand-Krawjewski, J.L., Muste, M., 2008a. Understanding and managing uncertainty. In: Fletcher, T.D., Deletic, A. (Eds.), Data Requirements for Integrated Urban Water Management. Taylor & Francis, The Netherlands, pp. 65–90.

Bertrand-Krawjewski, J.L., Muste, M., 2008b. Data validation: principles and implementation. In: Fletcher, T.D., Deletic, A. (Eds.), Data Requirements for Integrated Urban Water Management. Taylor & Francis, The Netherlands, pp. 103–124.

Bonakdari, H., Zinatizadeh, A.A., 2011. Influence of position and type of Doppler flow meters on flow-rate measurement in sewers using computational fluid dynamic. Flow Meas. Instrum. 22 (3), 225–234. http://dx.doi.org/10.1016/j.flowmeasinst.2011.03.001.

City of Austin, 2009. Stormwater Runoff Quality and Quantity from Small Watersheds in Austin, TX: Updated through 2008. Report Number CM-09-03. Watershed Protection Department, City of Austin, Austin, TX.

Coleman, H.W., Steele, W.G., 1995. Engineering application of experimental uncertainty analysis. AIAA J. 33 (10), 1888–1896. http://dx.doi.org/10.2514/3.12742.

Dilks, D.W., Freedman, P.L., 2004. Improved consideration of the margin of safety in total maximum daily load development. J. Environ. Eng. 130 (6), 690–694. http://dx.doi.org/10.1061/(ASCE)0733-9372(2004) 130:6(690).

Dotto, C.B.S., Kleidorfer, M., Deletic, A., Rauch, W., McCarthy, D.T., 2014. Impacts of measured data uncertainty on urban stormwater models. J. Hydrol. 508, 28–42. http://dx.doi.org/10.1016/j.jhydrol.2013.10.025.

Engineering Laboratory Design, 1999. 20'0" Sediment Demonstration Channel – Installation, Operation, and Maintenance Instructions. Lake City, MN.

FloWav, 2013. Pipeline Model PSA-AV. Product Datasheet. Retrieved from <http://thinkhd.com/dev/flowav/wp-content/uploads/2013/06/Pipeline-Model-PSA-AV-v19.pdf>.

Gauron, T., 2015. State promulgates TMDLs based on sporadic grab sampling. World Water: Stormwater Manage. 3 (3), 19–21.

Geosyntec Consultants and Wright Water Engineers, 2011. International Stormwater Best Management Practices (BMP) Database. Technical Summary: Volume Reductions. WERF, FHWA, ASCE-EWRI. Retrieved from <http://www.bmpdatabase.org/Docs/Volume Reduction Technical Summary Jan 2011.pdf>.

Gippel, C.J., O'Neill, I.C., Finlayson, B.L., Schnatz, I., 1996. Hydraulic guidelines for the re-introduction and management of large woody debris in lowland rivers. Regulated Rivers: Res. Manage. 12, 223–236.

Global Water, 2014. WL705 Ultrasonic Water Level Sensor. Product Datasheet. Retrieved from <http://www.globalw.com/downloads/WL700/WL705B.pdf>.

Harmel, R.D., Smith, P.K., 2007. Consideration of measurement uncertainty in the evaluation of goodness-of-fit in hydrologic and water quality modeling. J. Hydrol. 337 (3–4), 326–336. http://dx.doi.org/10.1016/j.jhydrol.2007.01.043.

Harmel, R.D., Smith, P.K., Migliaccio, K.W., 2010. Modifying goodness-of-fit indicators to incorporate both measurement and model uncertainty in model calibration and validation. Trans. ASABE 53 (1), 55–63.

Heiner, B.J., Vermeyen, T.B., 2013. Laboratory Evaluation of Open Channel Area-Velocity Flow Meters. In: Einhellig, R.F., Frizell, K.W. (Eds.), Hydraulic Laboratory Report HL-2012-03. U.S. Bureau of Reclamation, Denver, CO.

Hoogestraat, G.K., 2015. Water-Quality Characteristics of Stormwater Runoff in Rapid City, South Dakota, 2008-14: Scientific Investigations Report 2015-5069. U.S. Geological Survey, Reston, VA.

ISCO, 2011. 2150 Area Velocity Flow Module and Sensor – Installation and Operation Guide. Teledyne ISCO.

Jastram, J.D., 2014. Streamflow, Water Quality, and Aquatic Macroinvertebrates of Selected Streams in Fairfax County, Virginia, 2007-12: Scientific Investigations Report 2014-5073. U.S. Geological Survey, Reston, VA.

Larrarte, F., Bernard Bardiaux, J., Battaglia, P., Joannis, C., 2008. Acoustic Doppler flow-meters: a proposal to characterize their technical parameters. Flow Meas. Instrum. 19 (5), 261–267. http://dx.doi.org/10.1016/j.flowmeasinst.2008.01.001.

Lee, K., Ho, H.-C., Marian, M., Wu, C.-H., 2014. Uncertainty in open channel discharge measurements acquired with StreamPro ADCP. J. Hydrol. 509, 101–114. http://dx.doi.org/10.1016/j.jhydrol.2013.11.031.

Maheepala, U.K., Takyi, A.K., Perera, B.J., 2001. Hydrological data monitoring for urban stormwater drainage systems. J. Hydrol. 245 (1–4), 32–47. http://dx.doi.org/10.1016/S0022-1694(01)00342-0.

Massa, 2014. Model M-300/95 Low Cost Sensor. Product Datasheet. Retrieved from <http://www.massa.com/wp-content/uploads/2014/11/M-300-95-Datasheet-141107.pdf>.

McCuen, R.H., Sharkey, M., Walker, A., 2014. Trash deposition in rivers and streams: guidelines for prediction of counts, volumes, and masses. J. Hydrol. Eng. 04014001, 1–10. http://dx.doi.org/10.1061/(ASCE)HE.1943-5584.0000911.

McIntyre, N., Marshall, M., 2008. Field verification of bed-mounted ADV meters. Proc. ICE – Water Manage. 161 (4), 199–206.

McMillan, H., Krueger, T., Freer, J., 2012. Benchmarking observational uncertainties for hydrology: rainfall, river discharge and water quality. Hydrol. Process. 26 (26), 4078–4111. http://dx.doi.org/10.1002/hyp.9384.

Merz, B., Thieken, A.H., 2005. Separating natural and epistemic uncertainty in flood frequency analysis. J. Hydrol. 309 (1–4), 114–132. http://dx.doi.org/10.1016/j.jhydrol.2004.11.015.

Moffat, R.J., 1988. Describing the uncertainties in experimental results. Exp. Thermal Fluid Sci. 1 (1), 3–17. http://dx.doi.org/10.1016/0894-1777(88)90043-X.

Moriasi, D.N., Arnold, J.G., Van Liew, M.W., Bingner, R.L., Harmel, R.D., Veith, T.L., 2007. Model evaluation guidelines for systematic quantification of accuracy in watershed simulations. Trans. ASABE 50 (3), 885–900.

Nash, J., Sutcliffe, J., 1970. River flow forecasting through conceptual models Part I—a discussion of principles. J. Hydrol. 10 (3), 282–290.

Nord, G., Gallart, F., Gratiot, N., Soler, M., Reid, I., Vachtman, D., Latron, J., Martín-Vide, J.P., Laronne, J.B., 2014. Applicability of acoustic Doppler devices for flow velocity measurements and discharge estimation in flows with sediment transport. J. Hydrol. 509, 504–518. http://dx.doi.org/10.1016/j.jhydrol.2013.11.020.

Nortek, 2013. Vectrino II. Product Datasheet. Retrieved from <http://www.nortek-as.com/lib/brochures/vectrino-ii> (December 11, 2013).

Owen, H.J., 1979. Information for Local Officials on Flood Warning Systems. National Weather Service, Palo Alto, California. Retrieved from <http://hdl.handle.net/2027/uc1.31822017076480>.

R Core Development Team, 2014. R Language and Environment for Statistical Computing Version 3.0.2. Vienna Austria: R Foundation for Statistical Computing. Retrieved from <http://www.r-project.org>.

Rehmel, M., 2008. Application of acoustic Doppler velocimeters for streamflow measurements. J. Hydraulic Eng. 133 (12), 1433–1438.

Roy, A.H., Wenger, S.J., Fletcher, T.D., Walsh, C.J., Ladson, A.R., Shuster, W.D., Thurston, H.W., Brown, R.R., 2008. Impediments and solutions to sustainable, watershed-scale urban stormwater management: lessons from Australia and the United States. Environ. Manage. 42 (2), 344–359.

Sauer, V.B., Meyer, R.W., 1992. Determination of Error in Individual Discharge Measurements. Open-File Report 92-144. U.S. Geological Survey, Norcross, GA.

Storms, E.F., Oelsner, G.P., Locke, E.A., Stevens, M.R., Romero, O.C., 2015. Summary of Urban Stormwater Quality in Albuquerque, New Mexico, 2003-12 Scientific Investigations Report 2015-5006. U.S. Geological Survey, Reston, VA.

Taylor, A.C., Fletcher, T.D., 2007. Nonstructural urban stormwater quality measures: building a knowledge base to improve their use. Environ. Manage. 39 (5), 663–677.

Turnipseed, D.P., Sauer, V.B., 2010. Discharge Measurements at Gaging Stations. U.S. Geological Survey Techniques and Methods Book 3. Reston, VA. p. 87.

USBR, 2001. Water Measurement Manual. U.S. Bureau of Reclamation.

USEPA, 2014. Municipal Separate Storm Sewer System Permits – Post-Construction Performance Standards & Water Quality-Based Requirements – A Compendium of Permitting Approaches. Office of Water – Water Permits Division, Washington D.C. Retrieved from <http://water.epa.gov/polwaste/npdes/stormwater/upload/sw_ms4_compendium.pdf>.

VDOT v. USEPA, 2013. Virginia Department of Transportation, et al., versus USEPA, et al. United States District Court for the Eastern District of Virginia – 1:12-CV-775.

Wagner, W.E., 2005. Stormy regulation: the problems that result when stormwater (and other) regulatory programs neglect to account for limitations in scientific and technical information. Chapman Law Rev. 9 (191).

WMO, 2010. Manual on Stream Gauging – Vol I – Fieldwork. WMO-No. 1044 (Vol. I). World Meteorological Organization, Geneva, Switzerland.

Xylem, 2009. SonTek Argonaut-SW Shallow Water. Product Datasheet. Retrieved from <http://www.water-monitor.com/documents/ArgSWbrochurerev1.pdf>.