Journal of Hydrology 534 (2016) 524–533

Contents lists available at ScienceDirect

Journal of Hydrology

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

Uncertainty in monitoring *E. coli* concentrations in streams and stormwater runoff

R.D. Harmel [a,*], J.M. Hathaway [b], K.L. Wagner [c], J.E. Wolfe [d], R. Karthikeyan [e], W. Francesconi [f], D.T. McCarthy [g]

[a] USDA-ARS, Grassland, Soil and Water Research Laboratory, Temple, TX, United States
[b] Department of Civil and Environmental Engineering, University of Tennessee, Knoxville, TN, United States
[c] Texas Water Resources Institute, Texas A&M AgriLife Research, College Station, TX, United States
[d] Blackland Research and Extension Center, Texas A&M AgriLife Research, Temple, TX, United States
[e] Department of Biological and Agricultural Engineering, Texas A&M University, College Station, TX, United States
[f] International Center for Tropical Agriculture (CIAT), Lima, Peru
[g] Environmental and Public Health Microbiology Laboratory (EPHM Lab), Department of Civil Engineering, Monash University – Clayton Campus, Melbourne, Victoria, Australia

ARTICLE INFO

Article history:
Received 9 April 2015
Received in revised form 4 November 2015
Accepted 19 January 2016
Available online 27 January 2016
This manuscript was handled by Laurent Charlet, Editor-in-Chief, with the assistance of Rizlan Bernier-Latmani, Associate Editor

Keywords:
E. coli
Uncertainty
TMDL
Error propagation
Monitoring

SUMMARY

Microbial contamination of surface waters, a substantial public health concern throughout the world, is typically identified by fecal indicator bacteria such as *Escherichia coli*. Thus, monitoring *E. coli* concentrations is critical to evaluate current conditions, determine restoration effectiveness, and inform model development and calibration. An often overlooked component of these monitoring and modeling activities is understanding the inherent random and systematic uncertainty present in measured data. In this research, a review and subsequent analysis was performed to identify, document, and analyze measurement uncertainty of *E. coli* data collected in stream flow and stormwater runoff as individual discrete samples or throughout a single runoff event. Data on the uncertainty contributed by sample collection, sample preservation/storage, and laboratory analysis in measured *E. coli* concentrations were compiled and analyzed, and differences in sampling method and data quality scenarios were compared. The analysis showed that: (1) manual integrated sampling produced the lowest random and systematic uncertainty in individual samples, but automated sampling typically produced the lowest uncertainty when sampling throughout runoff events; (2) sample collection procedures often contributed the highest amount of uncertainty, although laboratory analysis introduced substantial random uncertainty and preservation/storage introduced substantial systematic uncertainty under some scenarios; and (3) the uncertainty in measured *E. coli* concentrations was greater than that of sediment and nutrients, but the difference was not as great as may be assumed. This comprehensive analysis of uncertainty in *E. coli* concentrations measured in streamflow and runoff should provide valuable insight for designing *E. coli* monitoring projects, reducing uncertainty in quality assurance efforts, regulatory and policy decision making, and fate and transport modeling.

Published by Elsevier B.V.

1. Introduction

The presence of pathogens in surface waters is increasingly a concern in the United States and worldwide, with fecal indicator bacteria (FIB) typically being used to indicate the presence of fecal matter in surface waters and the associated risk of pathogen contamination. Case in point, more stream and river miles were impaired due to pathogens (as inferred by high FIB concentrations) than any other pollutant in the United States Environmental Protection Agency's (USEPA) national summary of data collected from states under sections 305(b) and 303(d) of the Clean Water Act (USEPA, 2014). Since 1995, this has led to more Total Maximum Daily Loads (TMDLs) being developed in the United States for indicator bacteria than any other impairment (USEPA, 2014). Such pollution is not unique to the United States, with similar concerns being present from Australia's Yarra River (Daly et al., 2013) to the Seine River Estuary in France (Garcia-Armisen et al., 2005).

Modeling is a primary component of TMDL development, and similar watershed management plan development worldwide, with models being calibrated and validated using field-collected flow and water quality data. The output from these efforts is used for determining source load allocations (i.e., allowable pollutant

* Corresponding author at: USDA-ARS, 808 E. Blackland Rd., Temple, TX 76502, United States. Tel.: +1 254 541 1875.
E-mail address: daren.harmel@ars.usda.gov (R.D. Harmel).

http://dx.doi.org/10.1016/j.jhydrol.2016.01.040
0022-1694/Published by Elsevier B.V.

loads exported to the impacted surface water by various sources in the watershed). There are inherent errors associated with field monitoring, and TMDLs are required to include some margin of safety in these source load allocations due to the uncertainty present in these data (40 CFR 130.7). Further, optimal water quality monitoring can only be achieved if uncertainty in measurements and alternatives to reduce it are considered in sampling design and implementation (Beven, 2006; Harmel et al., 2006a, 2006b; Rode and Suhr, 2007). This is rarely the case with routine monitoring conducted by regulatory entities, despite the recognition of the importance of measurement uncertainty. In addition, little research has been performed to determine the uncertainty associated with monitoring FIB in streams and stormwater runoff. Due to this lack of information, relatively arbitrary margins of safety are currently employed to account for variability. Studies such as Hession et al. (1996) have indicated that uncertainty and risk analysis are a vital part of TMDL development. Thus, defining the uncertainty associated with FIB monitoring is a critical need that will improve the scientific basis of pathogen regulation, policy, modeling, and watershed plan development and implementation.

Fecal indicator bacteria are generally used instead of specific pathogens because of the large number of potential waterborne pathogens, substantial time required and expense of pathogen analyses, analytical expertise required to perform such analyses, difficulty determining which pathogens to target, and longer survivability of indicators (EPA, 2003). Various FIB, including fecal coliform, *Escherichia coli*, and enterococci, are utilized to assess compliance with water quality standards related to fecal contamination with the FIB of choice varying regionally and by water body type. In 1986, the USEPA published a report recommending *E. coli* or enterococci as a preferred FIB for fresh waters (USEPA, 1986). Subsequently, *E. coli* has been more frequently utilized and researched in fresh waters and is the focal point of this study.

Previous efforts to elucidate the uncertainty associated with water quality sampling and analysis have focused on nutrients and sediment (Harmel et al., 2006b, 2009). Harmel et al. (2006b) compiled error sources associated with flow measurement, sample collection, sample preservation/storage, and laboratory analysis for total suspended solids and various nutrient species. The total error accompanying these elements was compiled using the root mean square error propagation methodology (Topping, 1972). Harmel et al. (2006b) estimated the uncertainty of storm concentrations to be ±15% for total suspended sediment, ±14% for NO_3-N, ±20% for PO_4-P, ±27% for total N, and ±29% for total P. Using a similar methodology, McCarthy et al. (2008) conducted the only known comprehensive uncertainty analysis of field-collected *E. coli* data. Their results showed an average uncertainty of ±33% and a range of ±15–67%. However, because uncertainty varies based on the method of data collection, storage, and analysis, further research is needed to understand the uncertainty of additional monitoring regimes not analyzed by McCarthy et al. (2008). The Harmel et al. (2006b) and McCarthy et al. (2008) studies noted that "random" effects or sources of uncertainty are typically bi-directional and appropriately represented by the normal distribution.

The objective of this study was to expand on previous urban stormwater work by McCarthy et al. (2008) by compiling a more comprehensive collection of uncertainty data related to *E. coli* concentrations measured in streamflow and runoff. Specifically, uncertainty contributed by sample collection, sample preservation and storage, and laboratory analysis in measured *E. coli* data were compiled and presented using the theoretical framework established by Harmel et al. (2006b) and McCarthy et al. (2008). Similarly, the differences in sampling method and sample type (individual discrete and runoff event) were compared.

Similar to Harmel et al. (2006b), the analysis applies principally to edge-of-field runoff (<50 ha) and streamflow in small watersheds (<10,000 ha). On larger streams and rivers with perennial flow, additional considerations such as diurnal fluctuations, groundwater contribution, freshwater and saltwater interaction, and point sources such as waste water treatment plant outfalls would need to be considered. Lastly, the terms "error" and uncertainty are used synonymously herein to represent random and systematic statistical variation. Human error and equipment malfunction are not considered.

2. Materials and methods

2.1. Compilation of uncertainty data

An exhaustive literature search was performed to collect and compile data pertaining to measurement uncertainty for determination of *E. coli* concentrations in runoff and streamflow from small watersheds (inclusion of sources of spatial and temporal variability that contribute to uncertainty in data sets from long-term and/or multi-location monitoring projects was outside the scope of the present analysis). Then uncertainty estimates were determined as described in Table 1. These data/results were used to populate Tables 2–4, which present uncertainty estimates for steps/procedures within the major procedural categories (i.e., sample collection, sample preservation/storage, laboratory analysis) established by Harmel et al. (2006b). The distributional parameters presented in Tables 2–4 (usually the average and standard deviation) were used in the subsequent estimation of uncertainty contributed by each of the procedural categories and in the overall measured *E. coli* concentrations.

Harmel et al. (2006b, 2009) assumed that measurement uncertainty in water quality data collection was random, bi-directional (equally likely to be positive or negative), and normally distributed. These assumptions are valid for sources of "random" uncertainty in the present analysis of the uncertainty associated with individual *E. coli* concentrations, whether individual discrete samples or throughout a single storm runoff event. *It is important to note that this assumption does not apply to populations or sets of E. coli data, which are often asymmetric.* In contrast to Harmel et al. (2006b), the present analysis also assessed several sources of "systematic" uncertainty that introduced directional bias and are not appropriately represented by the normal distribution. To accommodate both types of uncertainty, uncertainty sources were separated based on whether they introduce random or systematic uncertainty (Tables 2–4).

2.2. Estimation of uncertainty in each procedural category and in measured E. coli concentrations

With the uncertainty estimates for individual steps or procedural categories, the random uncertainty in each procedural category and in measured *E. coli* concentrations was estimated with the method of Topping (1972) adapted as shown in Eq. (4). These results represent the cumulative random uncertainty such that over-estimation and under-estimation are equally likely; therefore, the resulting uncertainty is presented as ±%.

$$\pm\%\ unc. = \frac{\Delta E.\ coli}{E.\ coli} = \sqrt{\left(\frac{\Delta x_1}{x_1}\right)^2 + \left(\frac{\Delta x_2}{x_2}\right)^2 + \left(\frac{\Delta x_3}{x_3}\right)^2 + \cdots + \left(\frac{\Delta x_n}{x_n}\right)^2} \tag{4}$$

Then, the influence of systematic uncertainty was included as the sum of uncertainty in individual steps or processes that contributed to over- or under-estimation. The systematic uncertainty thus shifted the random uncertainty by the appropriate direction to achieve an overall uncertainty estimate.