Quantifying Uncertainty in Time-Series Water-Quality Data

Stewart Rounds
USGS
Oregon Water Science Center
Portland, OR

with help & insights from many others:

USGS: Stacey Archfield, Janice Fulford, David Holtschlag, Brian Pellerin, Pat Rasmussen, Susan Wherry
YSI / Xylem: Rob Ellison
Aquatic Informatics: Stu Hamilton, Brian Gouge

U.S. Department of the Interior
U.S. Geological Survey
Thinking About and Quantifying Errors

Error

- The difference between a measured value and a known reference or true value
- Composed of random and predictable components
  - Predictable components are repeatable & correctable
- Many sources of error
  - Avoidable Errors
    - User error, site-selection errors, site-installation errors
  - Unavoidable Errors
    - Signal noise/variability, site variability, interferences
    - Probe fouling, probe failure, calibration drift
    - Deviation from lab verification samples or other references
    - Surrogate model error
Error vs Uncertainty

Error

- The difference between a measured value and a known reference or true value
- Composed of random and predictable components
  - Predictable components are repeatable & correctable
- Many sources of error

Uncertainty

- A description of the degree of accuracy of the final corrected data
  - Can be expressed in many ways, including statistical representations (std. dev.) or a simple half-width of probable intervals
What are the Objectives?

• Set forth general & usable methods for computing data uncertainty for water-quality time-series data, within a context that can be applied to other datasets.

• Produce a peer-reviewed report describing the methods (with examples) and recommendations for implementation and future work.
Why Do We Care About Data Uncertainty?

1. Quantify uncertainty in order to manage it
   - Identify & assess measurement errors in order to avoid, minimize, and correct for them (we already do this)
   - Improve datasets and/or decrease costs by focusing resources where uncertainty is greatest
   - Optimize data-collection and model building plans & protocols
Why Do We Care About Data Uncertainty?

2. Communicate the accuracy of datasets → Improves acceptance and appropriateness of use

Currently, USGS rates data:

- Qualitatively
  (Excellent / Good / Fair / Poor)

- Based on:
  1. Sum of absolute values of all data corrections, and
  2. Assessment of hydrographer

- Ratings are not provided to the public with all data downloads
  Instead, data rounding is used as a poor substitute for a quantitative uncertainty.
Why Do We Care About Data Uncertainty?

3. Improve & extend the value & applicability of data

- Model predictions can be compared to time-series data in a way that takes both model error and data uncertainty into account
  - Would help to determine the significance of model-data deviations

Photo by Kurt Carpenter, USGS

Klamath River, OR
Why Do We Care About Data Uncertainty?

3. Improve & extend the value & applicability of data

Account for data uncertainty (turbidity, discharge) as well as turbidity/SSC model uncertainty in load calculations

Fanno Creek at Durham, OR (14206950), 2011-13
4. Create a better framework for comparisons to benchmarks & standards

• If uncertainties are known, we can compute the probability that a criterion is exceeded.
• Provides the basis for a new and better way to assess impairment and apply regulations

\[
\log(\text{As}) = f(\text{WT, Q})
\]

Computed Dissolved Arsenic Concentration (µg/L)

Percent Chance of Exceedance

Discharge (ft^3/s)

Data from nrtwq.usgs.gov

EXPLANATION
- Discharge
- Measured or computed water-quality constituent
- 90-percent prediction interval for computed value
- Value obtained from discrete sampling and analysis
- Load calculated using laboratory analysis and discharge
- Water-quality criteria

photo by Kurt Carpenter, USGS
The Vision: Include Uncertainties with Data

Now:

• Time-series graphs from USGS show no uncertainty bands
• Data downloads are:
  \[ \text{Date, Time, Time Zone, Value, Flag/Remark} \]

Future:

• Time-series graphs from USGS show uncertainty bands and comparison to standards w/ probability of exceedance
• Data downloads are:
  \[ \text{Date, Time, Time Zone, Value, Uncertainty, Flag/Remark} \]

This vision may sound simple, but I believe it is:

A game-changer. A paradigm shift. An opportunity.

Let’s make it happen!
Contact, and Thanks

Stewart Rounds
USGS Oregon Water Science Center
sarounds@usgs.gov
503-251-3280

Data Uncertainty Team
USGS: Stacey Archfield, Janice Fulford, David Holtschlag, Brian Pellerin, Pat Rasmussen, Susan Wherry
YSI / Xylem: Rob Ellison
Aquatic Informatics: Stu Hamilton, Brian Gouge