# 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** |

# 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

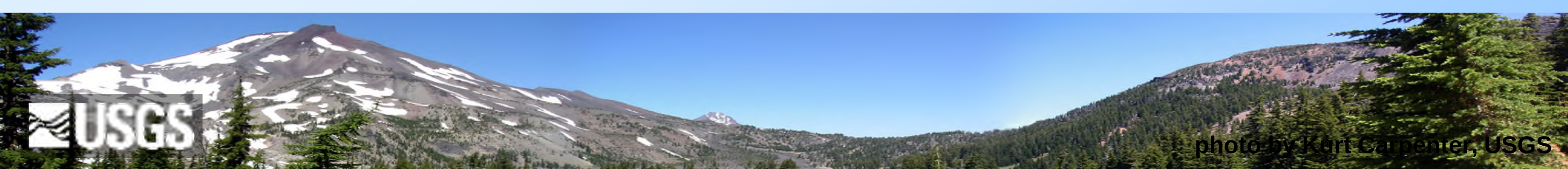
    User error, site-selection errors, site-installation errors
  - Unavoidable Errors

    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

photo by Kurt Carpenter, USGS

# 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.**

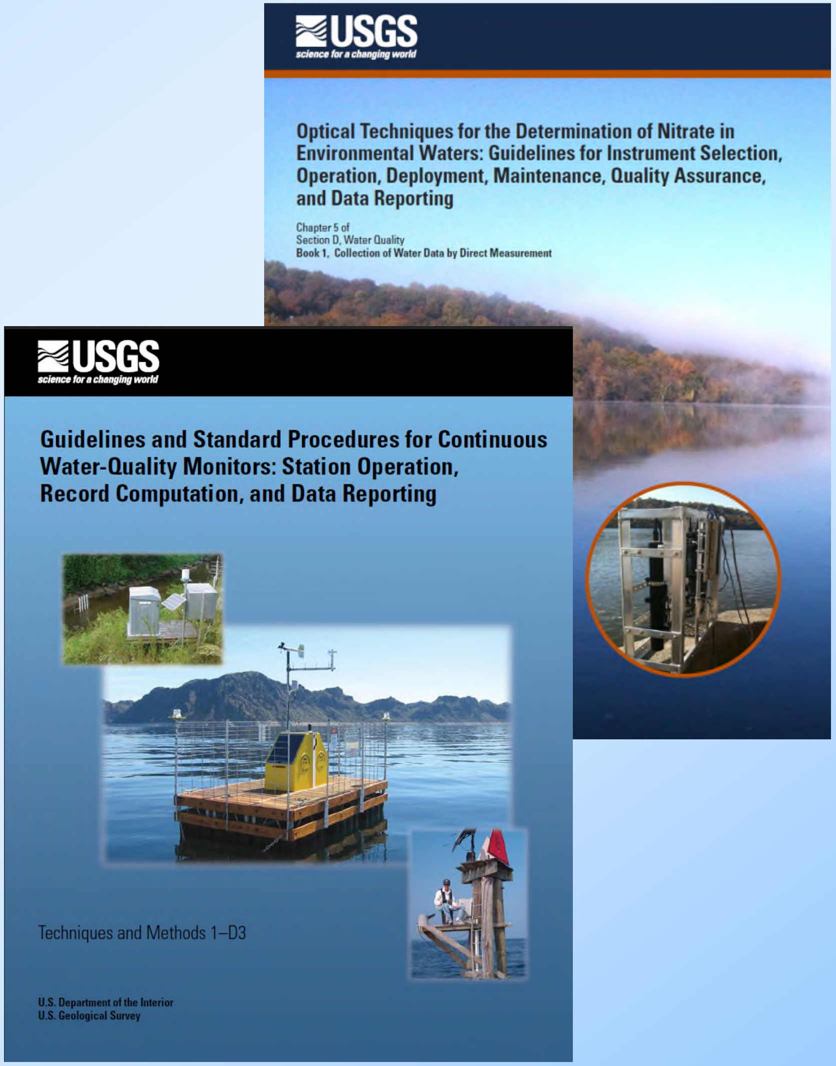


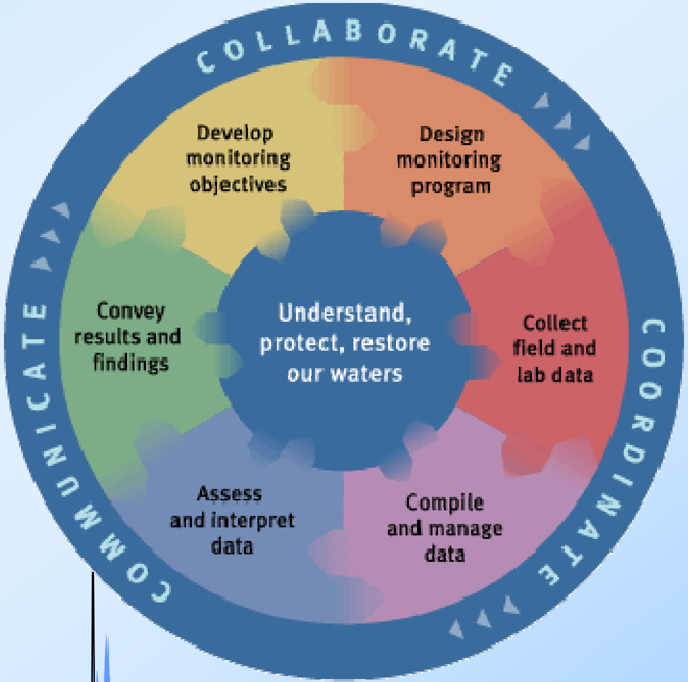


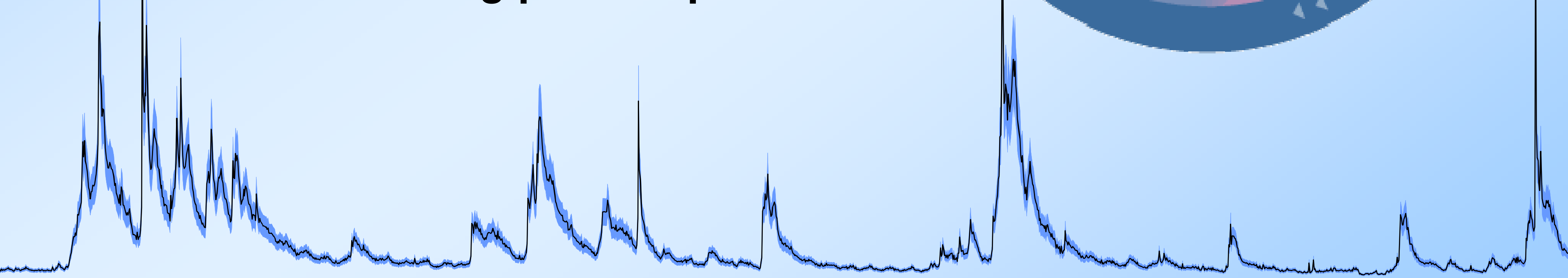
photo by Kurt Carpenter, USGS

# 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



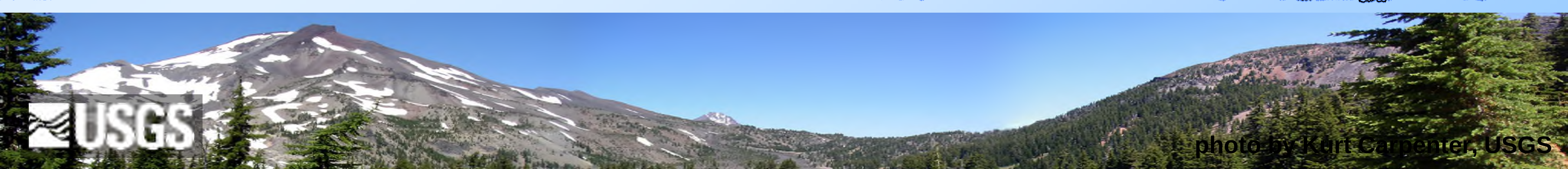



photo by Kurt Carpenter, USGS

# 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.**

```
Analysis performed Monday, 07-Jan-2013 08:44

Station Name:   FANNO CREEK AT DURHAM, OR
Station ID:     14206950
Parameter:      pH (std units)
PCODE:          00400
ADAPS DD:       6
Begin Date:     01-Oct-2009
End Date:       30-Sep-2010
Rating Method: Daily rating based on minimum unit-value rating for the day
DV Filter:      Rating provided regardless of existence of daily values

Verified 100.0% of 8731 computed data points, with a maximum
verification error of 0.02 std units.

SAC = Sum of the Absolute values of the Correct[illegible]
SAP = Sum of the Absolute values of [illegible] correcti[illegible]

The following data-qualit[illegible] were determ[illegible] on
the basis of the [illegible] values:

EXCE[illegible] 79.8% of the [illegible], 79.2% of [illegible] ratings
 Criteria:  SAC <= [illegible] units
 Daily range[illegible]
   20[illegible] to 2010/03/26
   201[illegible]/04/21 to 20[illegible]8
   2010/05/13 to 201[illegible]/08/24

GOOD -- 18.6% of the unit values, 18.6% of daily ratings
 Criteria:  0.2 std units < SAC <= 0.5 std units
 Daily ranges:
   2010/03/27 to 2010/04/20
   2010/04/29 to 2010/05/10
   2010/08/25 to 2010/09/24

FAIR -- 1.6% of the unit values, 2.2% of daily ratings
 Criteria:  0.5 std units < SAC <= 0.8 std units
 Daily ranges:
   2010/05/11 to 2010/05/12
   2010/09/25 to 2010/09/30

POOR -- 0.0% of the unit values, 0.0% of daily ratings
 Criteria:  0.8 std units < SAC <= 2 std units
 Daily ranges:
   none
```

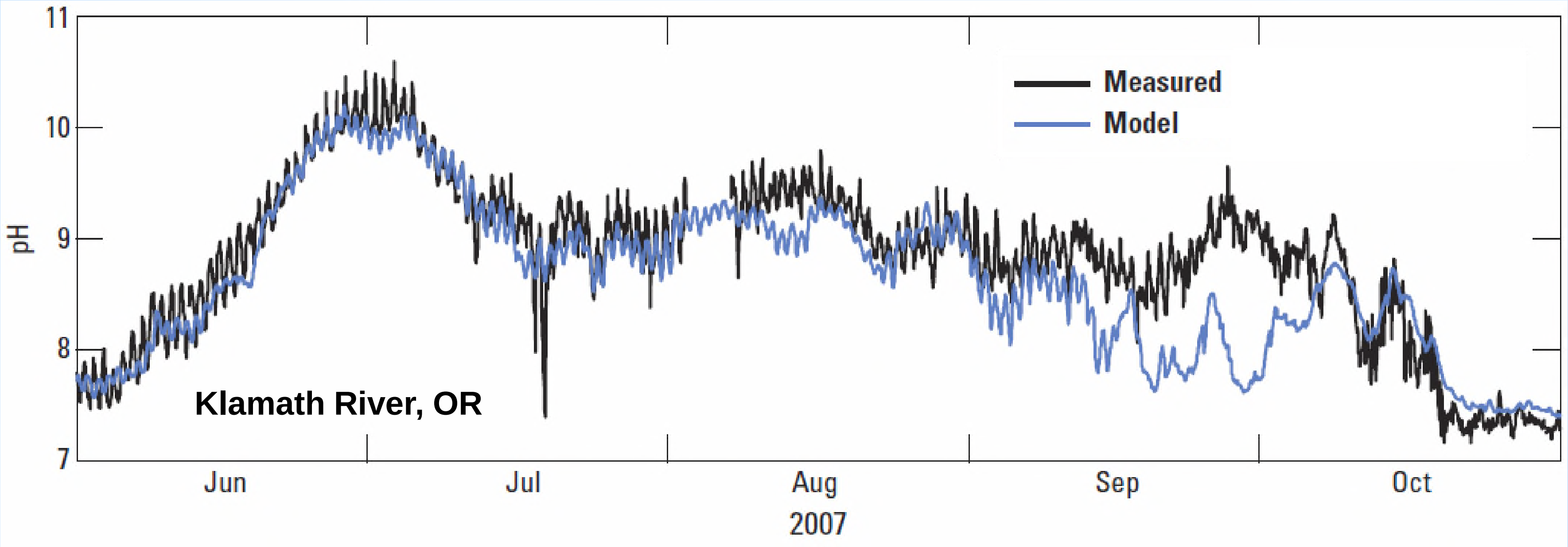
**The size of data corrections does not denote uncertainty. We can do better!**

photo by Kurt Carpenter, USGS

# Why Do We Care About Data Uncertainty?

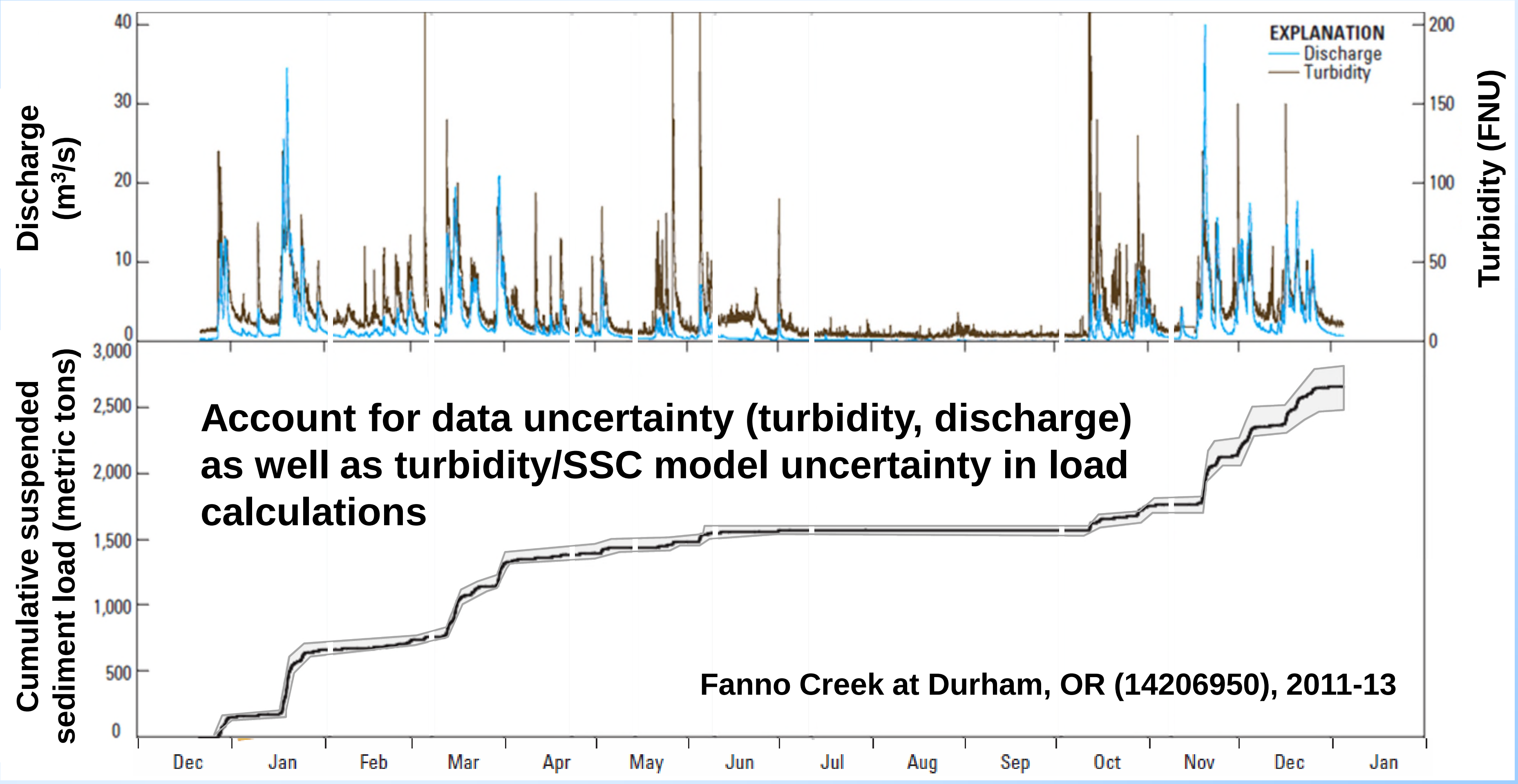
## 3. Improve & extend the value & applicability of data

- 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




# Why Do We Care About Data Uncertainty?

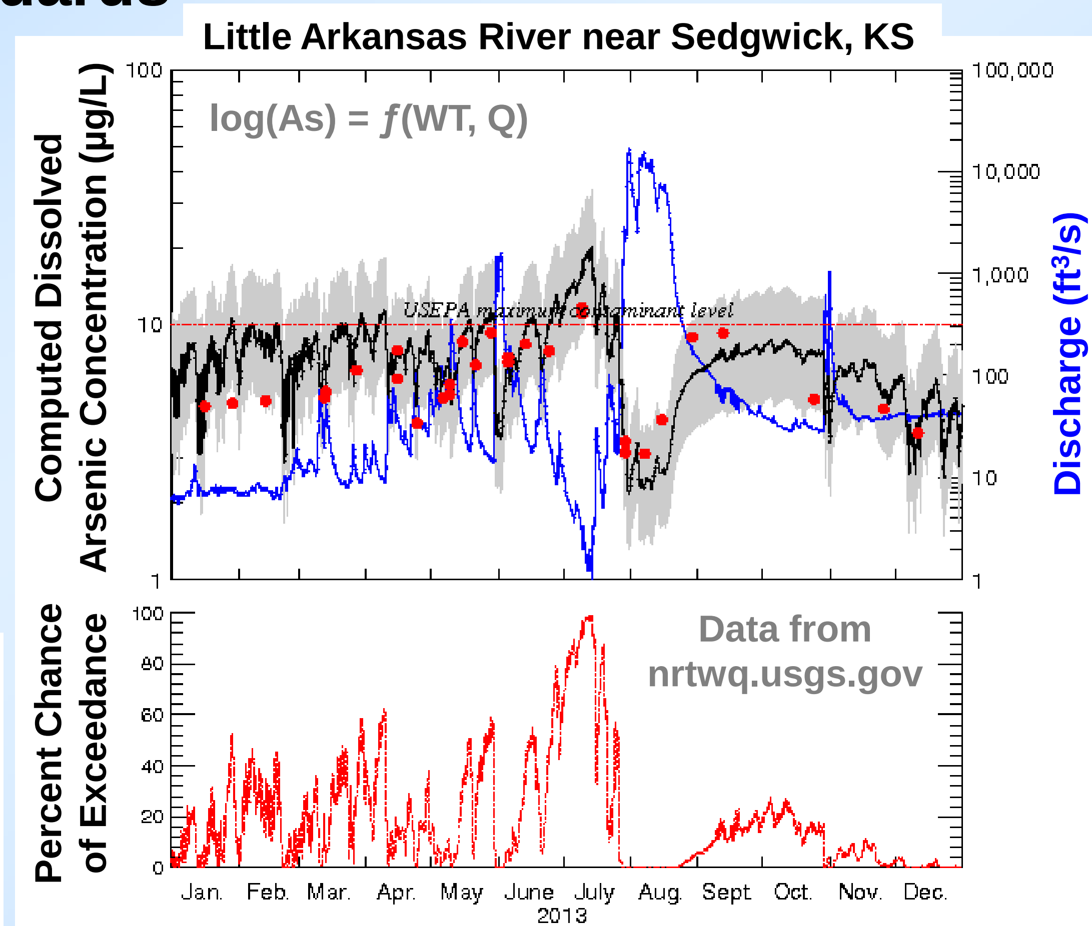
## 3. Improve & extend the value & applicability of data

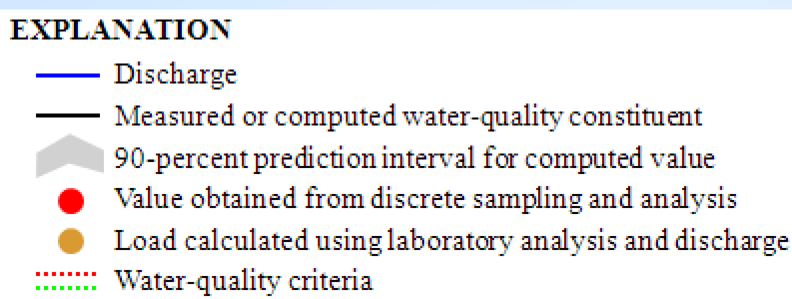


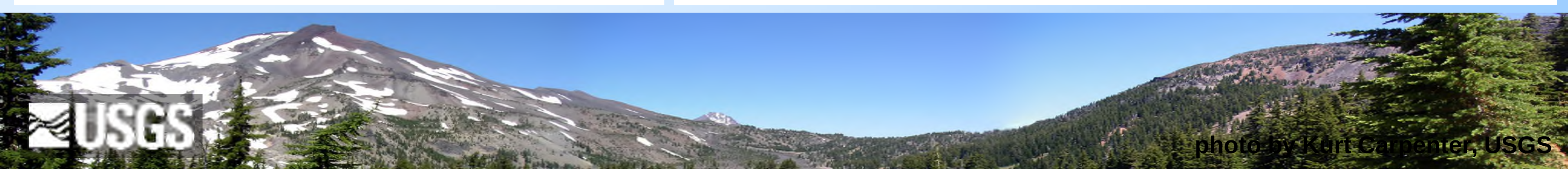

# Why Do We Care About Data Uncertainty?

## 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



# The Vision: Include Uncertainties with Data

## Now:

- **Time-series graphs from USGS show no uncertainty bands**
- **Data downloads are:**

  ***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:**

  ***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!**

photo by Kurt Carpenter, USGS

# 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 |