Data Quality

Tips for getting it, keeping it, proving it!

Central Coast Ambient Monitoring Program

Why do we care?

- Program goals are key – what do you want to do with the data?
- Data utility increases directly with quality
- High quality data can support 303(d) listing decisions, regulatory decisions, management decisions
- High quality data will be taken seriously
- Useable data is well-documented data

Recent 303(d) listing guidance checks that data used to support a listing be associated with quality assurance program plan

Quick Overview

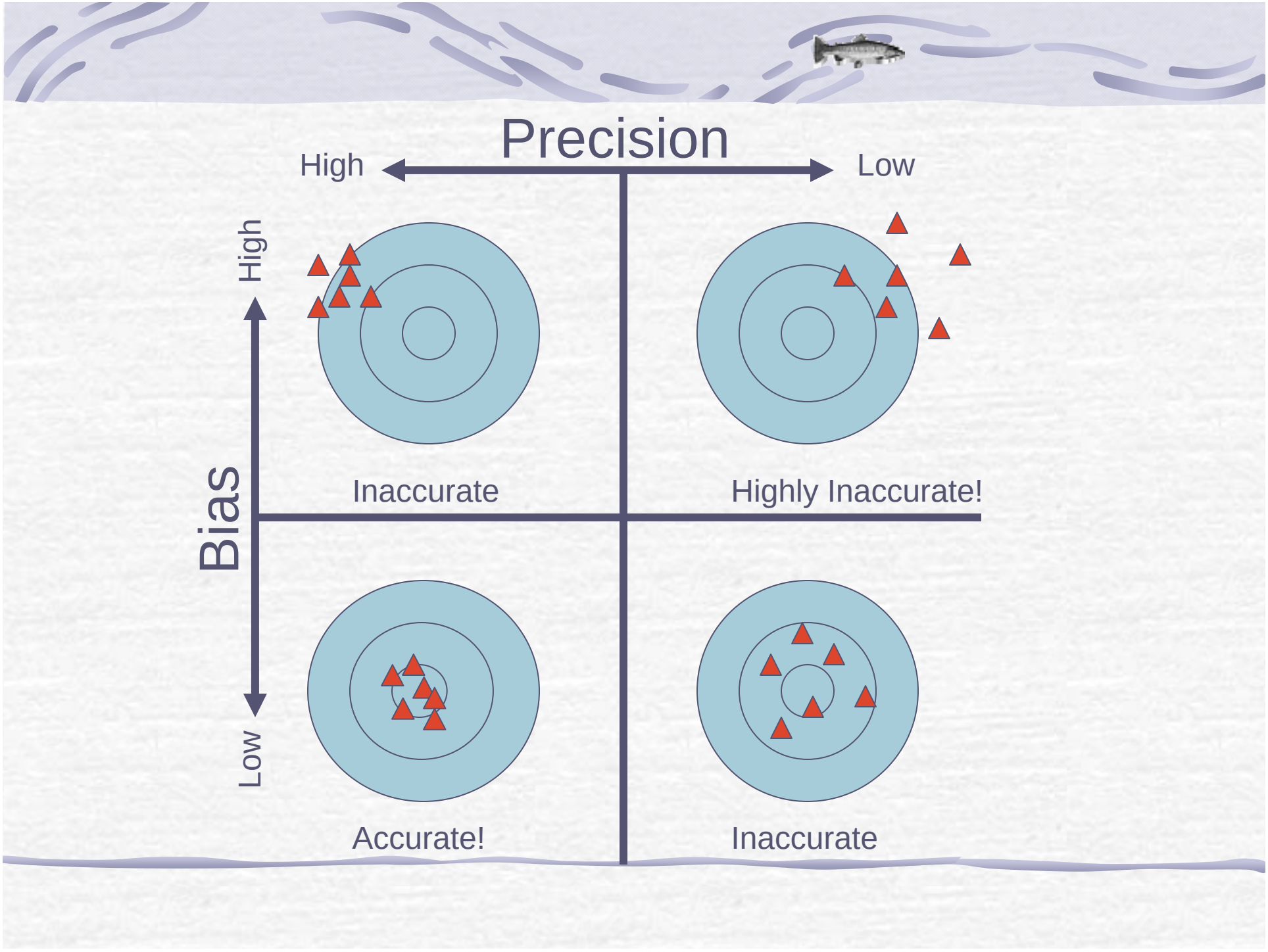
- Data quality objectives
- Data validation and verification
- Quality assurance documents
- Other QA tools

Data and Measurement Quality Objectives

- Basis of data quality determinations
- Underlies the concept of “SWAMP Compatibility”
- Should be defined in an approved Quality Assurance Program Plan

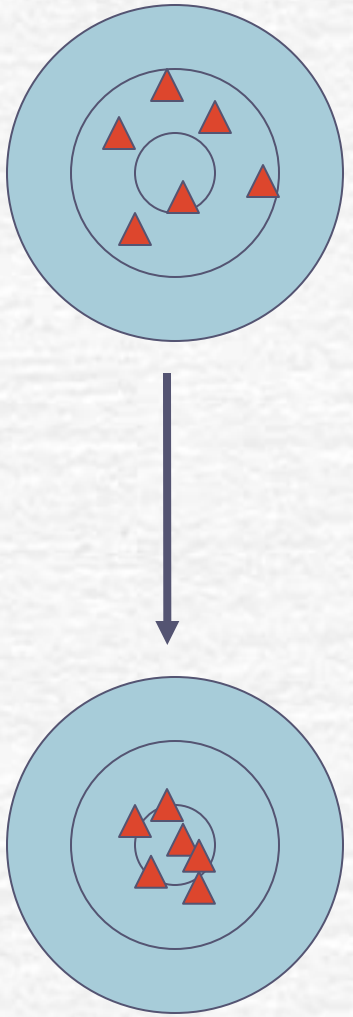
Basic QA terms describing data quality:

- Precision
- Accuracy and bias
- Representativeness
- Completeness
- Comparability
- Sensitivity


Precision
High
Low
Bias
High
Low
Inaccurate
Highly Inaccurate!
Accurate!
Inaccurate

Determining Precision

- Precision is a measure of repeatability and consistency
- Precision is calculated from sample replicates and/or splits
- Precision is dependent on instrument or method, and variability from handling and the environment
- You should be collecting at least 5% duplicates (or one per event) to be “SWAMP compatible”



Determining Precision

- For multiple samples calculate Relative Standard Deviation (RSD):

  $RSD = (s / \bar{x}) * 100 =$ % deviation of measurements

- For pairs of samples calculate Relative Percent Difference (RPD):

  $$RPD = \frac{(X1 - X2) * 100}{(X1 + X2) / 2}$$ where X1 is larger value

  = % deviation of measurements

SWAMP RPD requirement is typically 25%

What's the difference between a field duplicate and a field split?

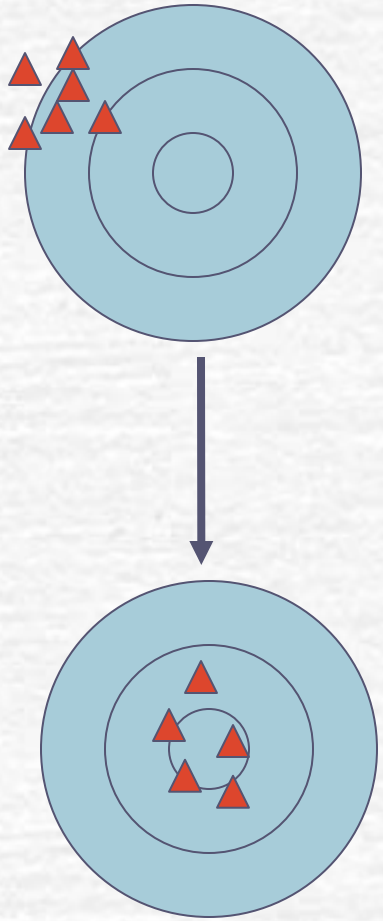
Field duplicates: collected side by side, a combined measurement of environmental variability and sampling and laboratory error

Field splits: collected from a composite sample, factors out environmental variability. A measure of variability associated with sampling and laboratory error. May be sent to different laboratories to measure interlab variability.

Bias and Accuracy…

…how to hit the target!

- Calibrate against standard reference material
- Calibrate against an “expert”
- Spike sample with a known concentration



Calculating accuracy

- SWAMP requires that measured value is between 80% and 120% of the known value

$$PD = \frac{(X1 - X2) * 100}{(X1)}$$

where X1 is known value

= % deviation of measurement from reference

Calibrating instruments

- In order to track equipment accuracy, keep an instrument calibration log
- Instrument drift is determined by calibrating pre- and post- sampling. RPD should be within DQO requirements.
- Be sure to record instrument ID in your data records; in some cases bias can be adjusted out of data if it can be shown to be consistent.

Blanks

- Assessing contamination
- Different types for different purposes
  - Field, equipment, travel, lab
- SWAMP suggests CWQ field blank during periodic field audit. If inadequate performance, increase to 5%
- Blank must be < Minimum Detection Limit

Representativeness

Think through your sampling design

What are you trying to show?

Consider: Statistical adequacy

Seasonality

Time of day

Spatial variability

Completeness

A measure of how many samples you actually take, compared to how many you planned to take.

SWAMP typically requires 90% completeness

$$\%C = \frac{\text{\# valid samples}}{\text{\# total planned}} * 100$$

Comparability

- Requires careful consideration during study design phase
- SWAMP is “performance-based”
  - Specific methods aren’t required
  - Method comparability is determined by ability to meet Data Quality Objectives
- Other factors include site selection, units of measurement, time of sampling, etc.

Measurement Range, Detection Limits and Resolution

- Measurement range should make sense in anticipated sampling environment – do you want to pay for “non-detects”?
- Consider the use of the data in determining project needs
- Meeting SWAMP Target Reporting Limits can increase comparability and utility of data

Taking Action…

...the most important part!!

Section 24 of QAPP is “Reconciliation with Data Quality Objectives”

- Calculate Data Quality Indicators; compare with DQOs
- Take action based on findings
  - Flag or qualify data
  - Disable data, either at the batch or sample level
  - Revisit data quality objectives
  - Reexamine field and/or lab protocols

The closer in time you do this to the actual sampling, the better!

Data Verification and Validation

(a.k.a. Sanity Checking)

Verification: Is the data complete and correct? Does it conform with method requirements?

Validation: Does data meet analyte and sample specific requirements (usually done by a QA officer or external party)?

Step 1: Verification

- Were SOPs followed?
- Were holding times met and QC samples and blanks collected?
- Are records complete? (field data sheets, instrument logs, chain of custody, etc.)
- Is data reported correctly (transcription, conversion, units of measurement, etc.)?
- Is data flagged correctly?

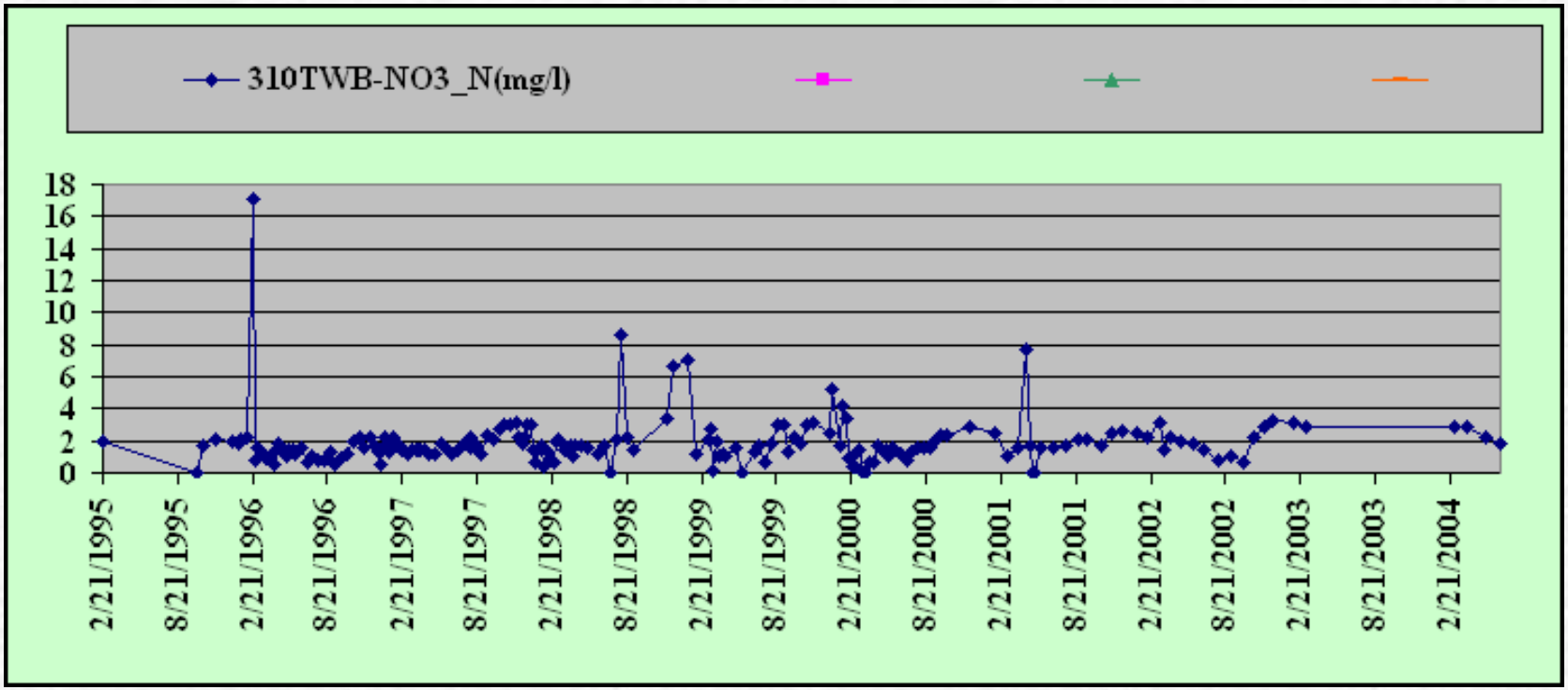
Look at your data; talk to your samplers!

Step 2: Validation

- Were project needs met?
- Have Data Quality Objectives been met? If not, why not?
- Summarize deviations in field procedures and impacts on data quality
- Summarize QC deficiencies and impacts on data quality
- Qualify data as necessary
- Prepare Validation Report

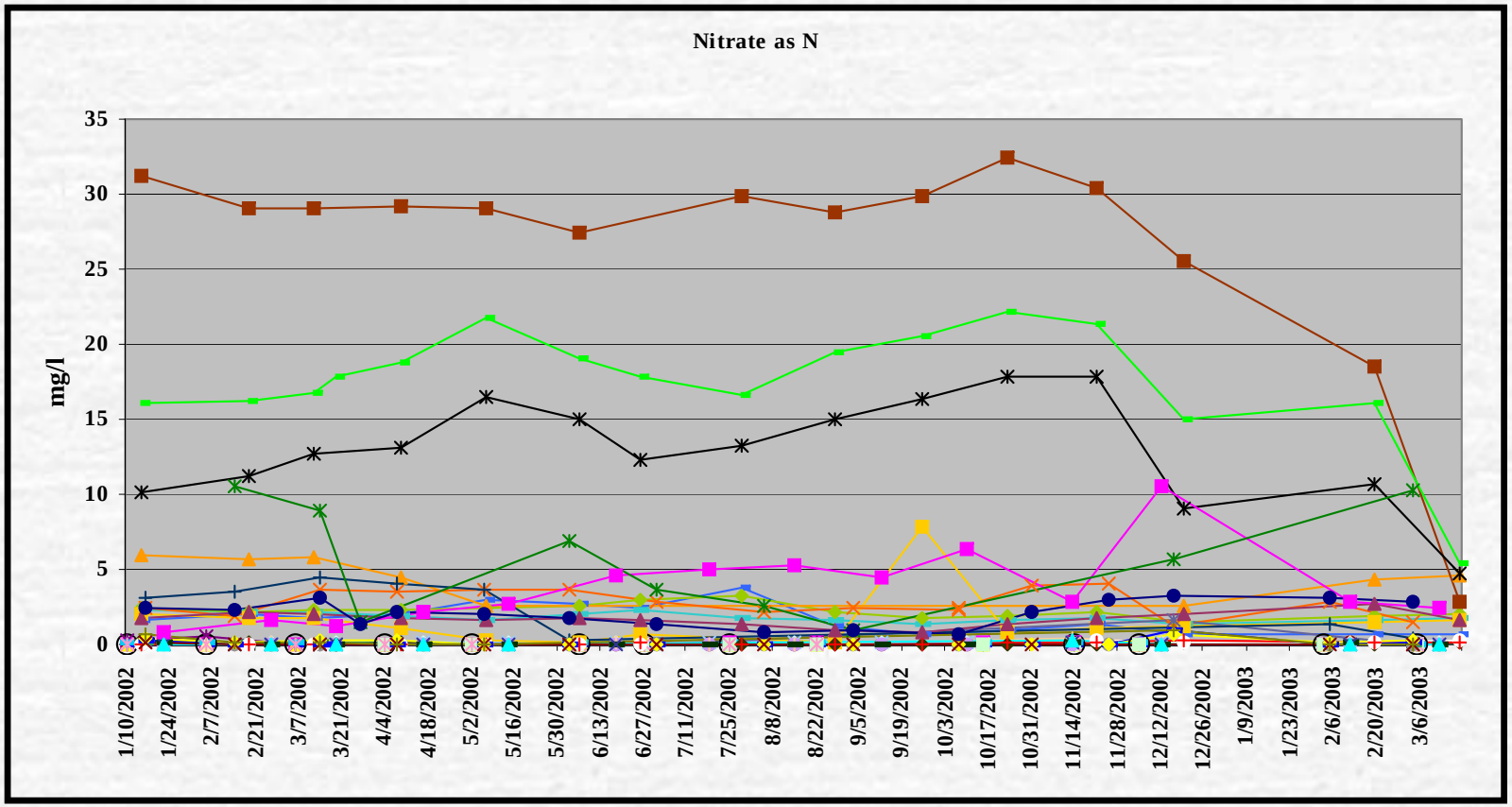
Having said all this…

- Look at your data early and often!
- Sort for zeros and outliers
- Plot the data – get some context!



Use the Smell Test!

Use your knowledge of the sampling area to determine if data is reasonable; don't take information at face value.



Use common sense

- If goal is to look for exceedance of water quality standards, disqualifying data for DQO irregularities at or near the detection limit may not make sense.

  e.g. How important is the 25% difference between 0.01 and 0.0125 mg/L Nitrate as N?

- If goal is to calculate loading, error at or near the detection limit may be more important.

Developing a Quality Assurance Project Plan

- SWAMP QA compatibility now required for grant funded projects
- Templates available on SWAMP website
- Following the template as closely as possible will reduce pain and speed the review process
- The Clean Water Team has extensive material online related to QAPP development and Data Quality Management

Resources

- EPA Volunteer Monitor's Guide to QAPPs and other supporting documents
  http://www.epa.gov/OWOW/monitoring/volunteer/
- CWT Guidance Compendium
  http://www.waterboards.ca.gov/nps/cwtguidance.html
- SWAMP Quality Assurance URL
  http://www.swrcb.ca.gov/swamp/qapp.html