DATA INVENTORY, WATERSHED CHARACTERIZATION & SUFFICIENT DATA

Fundamentals of Developing a Water Quality Monitoring Plan

Michael Schramm

2024-03-05

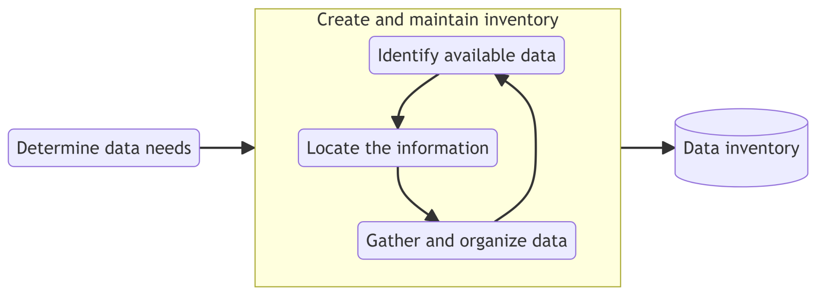
TWRI, Texas A&M AgriLife Research

What are we covering?

- **Data Inventory:** What data do you need, where to obtain it, best practices to maintain it.
- **Watershed Characterization:** What is it and provide an overview of approaches.
- **Data Sufficiency:** Do you have the data you need? Do you have enough data?

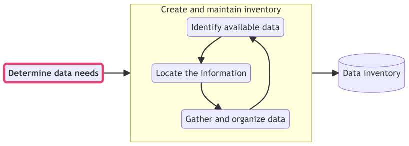
© Texas A&M AgriLife Marketing and Communications; Photo credit: Michael Miller

Major steps



Determine data needs
Create and maintain inventory
Identify available data
Locate the information
Gather and organize data
Data inventory

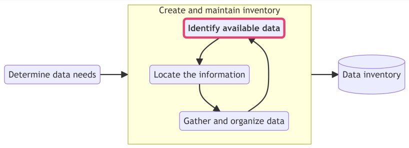
Determine data needs



While you are scoping and planning a project:

- Identify stakeholder concerns
- Identify watershed goals
- Define conceptual models

Identify available data



- Use existing data if/when available!
- Assess the quality/reliability of that data.
- Talk to organizations that have conducted projects in the area.

Locating information

What types of data will you need?

Locating information

What types of data will you need?

- Watershed physical characteristics
- Water quality
- Streamflow
- Climate
- Pollutant sources

Watershed physical characteristics

- Watershed boundaries
- Hydrographic/Topographic
- Soil
- Vegetation, wildlife, etc.

Locating information

Watershed physical characteristics

- USGS NHD, NHD HR: www.usgs.gov/national-hydrography/access-national-hydrography-products
- NLCD, Imperviousness, Tree Canopy, etc.: www.mrlc.gov/
- NRCS Soil Survey:websoilsurvey.nrcs.usda.gov/app/

Locating information

Water quality

© Texas A&M AgriLife Marketing and Communications; Photo credit: Michael Miller

- Routine water quality
- Stormwater/runoff monitoring
- Storm/flow biased water quality
- Benthic and other stream surveys

Locating information

Water quality

© Texas A&M AgriLife Marketing and Communications; Photo credit: Michael Miller

- TCEQ SWQM, Integrated Reports
- CRP Partners
- National Water Quality Portal (successor to STORET): www.waterqualitydata.us/
- Hydroshare: www.hydroshare.org/
- TPWD Bioassessment Studies
- Texas Stream Team Citizen Science Data

Streamflow

© Texas A&M AgriLife Marketing and Communications; Photo credit: Michael Miller

- Daily/instantaneous streamflow
- Inundation/tidal data

Locating information

Streamflow

- National Water Information Service: waterdata.usgs.gov/ nwis
- Water Data for Texas: waterdatafortexas.org
- Hydroshare: www.hydroshare.org/
- NOAA: tidesandcurrents.noaa.gov/

Climate

© Texas A&M AgriLife Marketing and Communications; Photo credit Laura McKenzie

- Precipitation
- Temperature/solar radiation
- Wind direction/speed

Locating information

Climate

- NOAA Climate Data Online: www.ncei.noaa.gov/cdo-web/
- OSU PRISM Climate Group: prism.oregonstate.edu/

Pollution Sources

Photo credit Ryan Gerlich (Texas A&M AgriLife Extension)

- Permitted discharges
- Non-permitted sources

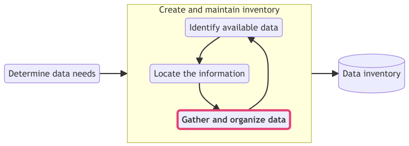
Locating information

Pollution Sources

Photo credit Ryan Gerlich (Texas A&M AgriLife Extension)

- EPA ECHO: echo.epa.gov/
- TCEQ Water Quality General Permits: www2.tceq.texas.gov/wq_dpa/index.cfm
- TCEQ Central Registry: www15.tceq.texas.gov/crpub/
- Non-permitted sources: Often derived from secondary data sources (Census data, 911 addresses, wildlife surveys, etc.)

Gather and organize data



Basic Principles:

- Create a project specific data folder or database.
- Maintain living project documentation.
- Use data exploration and summary methods to get a “picture” of the data.

Gather and organize data

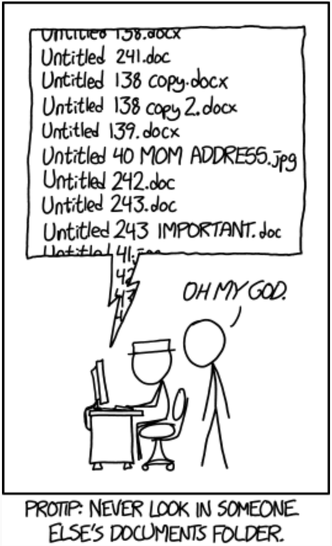
Documentation should identify:

- Types of data (monitoring, geographic)
- Source
- Quality
- Representativeness
- Spatial coverage
- Temporal coverage
- Data gaps
- Units and variable names

Your QAPP is a good starting point.

Best practices:

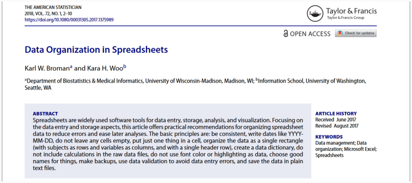
- Consistent terminology and file naming (ex. Nitrate as nitrogen, NO_3-N, or nitrate...).
- Standardized dates (highly suggest YYYY-MM-DD formatting).
- Use non-proprietary files where possible (plain-text, csv, etc.).
- What are your data storage practices?



Gather and organize data

Spreadsheet best practices:

- No empty cells: specify no value, NA, zero, infinity, etc.
- One value per cell.
- Do not use font color, cell color, or comments, as data.



THE AMERICAN STATISTICIAN
2018, VOL. 72, NO. 1, 2–10
https://doi.org/10.1080/00031305.2017.1375989

Taylor & Francis
Taylor & Francis Group

OPEN ACCESS Check for updates

Data Organization in Spreadsheets

Karl W. Broman[a] and Kara H. Woo[b]

[a]Department of Biostatistics & Medical Informatics, University of Wisconsin-Madison, Madison, WI; [b]Information School, University of Washington, Seattle, WA

ABSTRACT
Spreadsheets are widely used software tools for data entry, storage, analysis, and visualization. Focusing on the data entry and storage aspects, this article offers practical recommendations for organizing spreadsheet data to reduce errors and ease later analyses. The basic principles are: be consistent, write dates like YYYY-MM-DD, do not leave any cells empty, put just one thing in a cell, organize the data as a single rectangle (with subjects as rows and variables as columns, and with a single header row), create a data dictionary, do not include calculations in the raw data files, do not use font color or highlighting as data, choose good names for things, make backups, use data validation to avoid data entry errors, and save the data in plain text files.

ARTICLE HISTORY
Received June 2017
Revised August 2017

KEYWORDS
Data management; Data organization; Microsoft Excel; Spreadsheets

Data Organization in Spreadsheets: doi.org/10.1080/00031305.2017.1375989

Watershed Characterization

Photo courtesy of Ed Rhodes, TWRI

What is a Watershed Characterization?

An assessment of past and current conditions within a watershed to support water quality management decisions.

What is a Watershed Characterization?

Can have different objectives:

- Is there a (or where is the) water quality problem?
- Compare and define current and desired conditions.
- Estimate allowable loadings for point source permits or TMDLs
- Evaluate and recommend best management practices
- Assess trends or changes in water quality

(not an exhaustive list…)

How do we do it?

Depends on our objectives!

Objectives determine the approach and data requirements (monitoring requirements).

General approaches

- Reconnaissance
- Plot studies
- Watershed(s) study

Reconnaissance

Objective: Identify if or where there is a water quality problem.

- AKA routine or ambient water quality
- Provides a snapshot of water quality at a given point in time
- Limited statistical application (summary statistics or probability of exceedance)
- TCEQ assessments manual provides baseline methods for comparison.

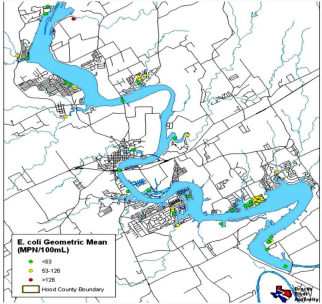


Image: Brazos River Authority, Lake Granbury Watershed Protection Plan

Advantages:

- You might already have enough data (CRP monitoring)
- Large spatial coverage
- Useful to pinpoint problematic areas

Disadvantages:

- Biased towards baseflows
- Possibly low sampling frequency in existing data
- Sampling locations are biased

Plot studies

Objective: Evaluate differences in conditions or treatments

- Experimental trials on multiple plots involving treatment and control replicates. Treatments are small plots where runoff can be routed to a single measurement point per treatment.

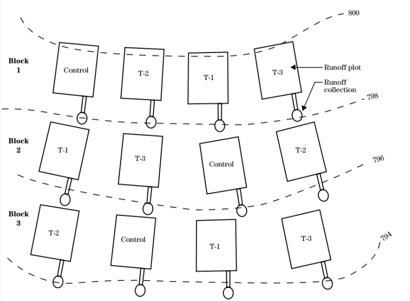


Image USDA NRCS (2003)

Plot studies

Often used to evaluate the impact of BMPs on runoff quality.

Advantages:

- Well designed experiments have strong statistical power (reliability for detecting true effects).
- Ability to control for environmental conditions.

Disadvantages:

- Cost.
- Conditions may not represent real world conditions.
- Results are sometimes (often?) condition specific.

Plot studies

Photo credit: Jason Gerlich, TWRI

Watershed studies

Objective: Evaluate Best Management Practices (BMPs), assess trends.

Multiple potential study designs:

Design	Causal factors
Single watershed - before/after	BMPs, Climate
Single watershed - upstream/downstream	BMP, within watershed factors
Paired watershed - BACI	BMP

Single watershed

Objective: Evaluate conditions before and after implementation, or trend analysis.

- Single monitoring site at the outlet of a watershed.
- Relatively inexpensive.
- Difficult to separate effects due to changes climate or other watershed changes.

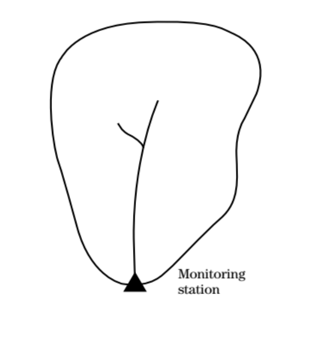


Figure 1: Single watershed (before/after implementation.)

Image USDA NRCS (2003)

Single watershed

Objective: Evaluate conditions before and after implementation, or trend analysis.

- Monitoring before and after implementation.
- Monitor upstream and downstream of implementation.

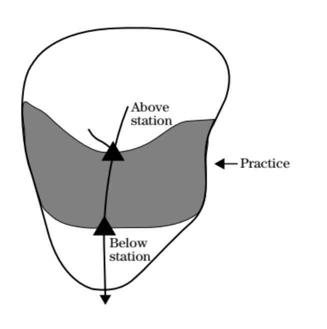


Figure 2: Single watershed (above/below implementation.)

Image USDA NRCS (2003)

Single watershed

Advantages:

- Accounts for some differences within the watershed
- Costs

Disadvantages:

- Downstream measurements are not independent of upstream measurements

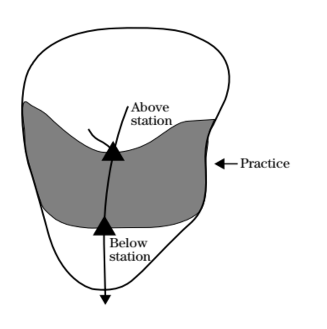


Figure 3: Single watershed (above/below implementation.)

Image USDA NRCS (2003)

Paired watershed

Objective: Evaluate conditions before and after implementation, or trend analysis.

- Monitor before and after in calibration and treatment watersheds
- Two or more watersheds
- One watershed always serves as a control

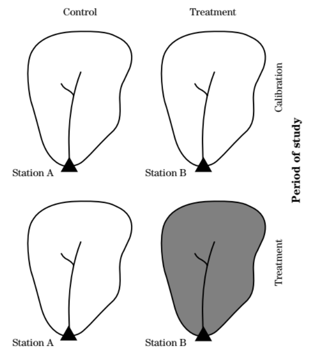


Figure 4: Paired watersheds; before-after control-impact (BACI)

Image USDA NRCS (2003)

Paired watershed

Advantages:

- Can be statistically powerful.
- Can evaluate multiple treatments.
- Helps control for environmentally induced variation not associated with treatment.
- Statistical independence between treatment/pre-treatment measurements.
- Regionally relevant results.

Paired watershed

Disadvantage:

- Requires long study period.
- Lag effects can be gradual masking difference with control (Meals et al., 2010).
- Can be hard to find two watersheds that won't have any other disturbances for the entire study period.
- Logistics.

Trend analysis

Objective: Assess a trend or change in water quality parameters at a single or multiple location.

- Evaluate effect size over time.
- Doesn’t require “reference” watersheds but can be helpful.

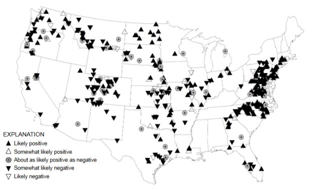


Image Helsel et al. (2020)

Trend analysis

Advantages:

- Can use a single station.
- Useful for practices with long-lag times.
- Useful for assessing larger watersheds with many implementation activities.

Disadvantages:

- Typically requires long-term data (10-20 years).
- High chance of changes in watershed and lab methodologies.
- Trends are sensitive to the chosen time period.

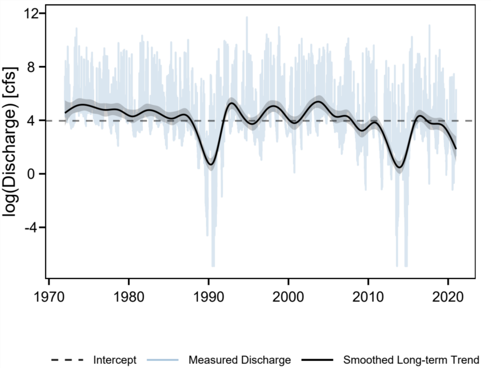


Image M. Schramm (2023)

Data Sufficiency

Sufficient data

Depends on:

- Statistical design
- Variability of measurement parameters
- Relevant effect size

Sufficient data

Things to consider:

- Adequate number of sites
- Adequate number of samples
- Frequency of samples
- Monitoring duration

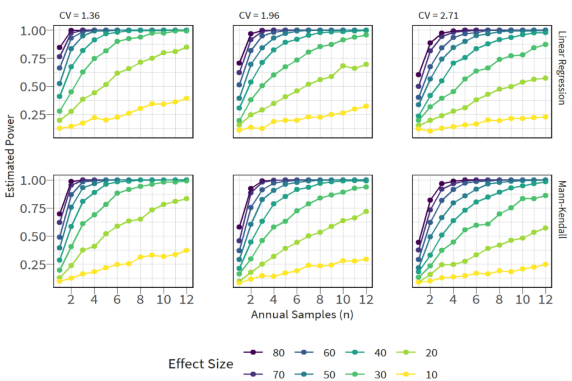
Sufficient data

Available tools:

- TCEQ assessment criteria (for ambient water quality)
- Statistical power analysis: Helsel et al. (2020) Chapter 13

\note<1>{We could have a whole data on how much data you need, but I'm going to point you to two different resources. First, TCEQ's assessment procedures are very good for general characterization or reconasiance type approaches. They provide clear guidance for duration and sample size needed to make statisically based decisions according to the state's numeric criteria. If you are using a watershed or plot trial, chapter 13 of the USGS stastical methods in water reources handbook, is excellent. This discusses how to do something called power analysis to estimate the statstical power of your approach based on the number of samples and expected variance. Stastical power is essentially how likely your stasstical test is to correctly idenfity an effect when there actually is an effect. We

Sufficient data



Example power analysis for trend tests on *E. coli* concentrations over 10-years (M. P. Schramm, 2021).

\note<1>{Here is an example of a power analysis I conducted for detecting 10-year trends in E. coli concentrations. If you pick one of these plots, the x-axis shows the number of annual samples needed to achieve the desired stastical power on the y-axis. Each line represents a different effect size, so the light yellow lines represent if we wanted to detect a 10 percent reduction in bacteria, even at 12 samples a year, our power is only around

Contact us

We’d love to talk about all things water

- 979.314.2356
- michael.schramm@ag.tamu.edu
- twri.tamu.edu

Helsel, D. R., Hirsch, R. M., Ryberg, K. R., Archfield, S. A., & Gilroy, E. J. (2020). ***Statistical methods in water resources: U.S. Geological Survey techniques and methods, book 4, chapter A3.*** USGS. https://doi.org/10.3133/tm4a3

Meals, D. W., Dressing, S. A., & Davenport, T. E. (2010).**Lag time in water quality response to best management practices: A review.** *Journal of Environment Quality*, *39*(1), 85. https://doi.org/10.2134/jeq2009.0108

Schramm, M. (2023).**Assessing linkages between watershed nutrient loading and water quality in a subtropical estuary with generalized additive models.** *PeerJ*. https://doi.org/10.7717/peerj.16073

Schramm, M. P. (2021).**Estimating statistical power for detecting long term trends in surface water *Escherichia coli* concentrations.** *Texas Water Journal*, *12*(1), 140–150. https://doi.org/10.21423/txj.v12i1.7126

USDA NRCS. (2003). *National Water Quality Handbook Part 614* (tech. rep.).
USDA NRCS.
https://archive.epa.gov/water/archive/web/pdf/stelprdb1044775.pdf