

Uncertainty due to gap-filling in long-term hydrologic datasets

Craig See¹, Jeremy Hayward¹, Ruth Yanai¹, Mark Green², Doug Moore³
Amey Bailey⁴

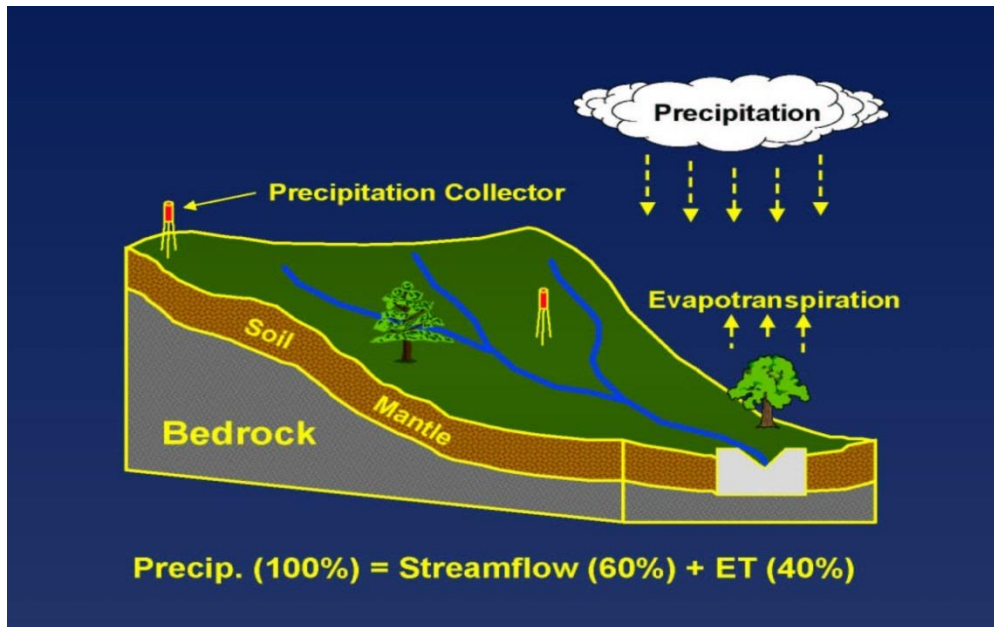
1. SUNY College of Environmental Science and Forestry, Syracuse, NY
2. Plymouth State University, Plymouth, NH
3. University of New Mexico, Albuquerque, NM
4. USDA Forest Service, Northern Research Station

All long term datasets contain gaps



Why Gap Filling?

Sometimes we need a continuous record
(calculating fluxes)



Prepared in cooperation with the U.S. Fish and Wildlife Service

A National Streamflow Network Gap Analysis



(Kiang et. al 2013)

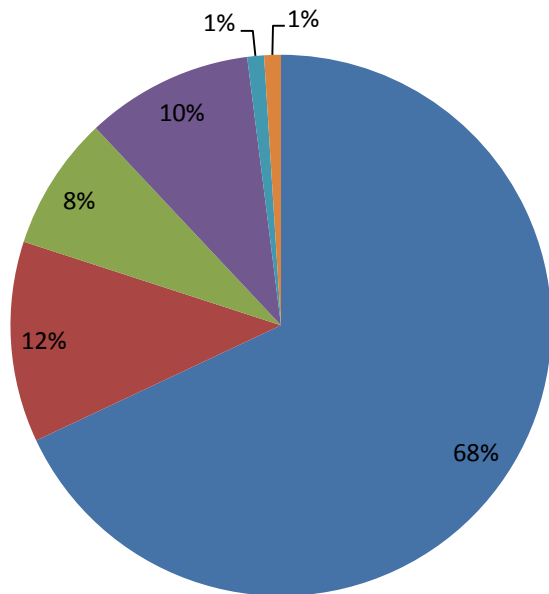
Gap filling methods

- Use of historical averages
- Bayesian Bootstrapping
- Expectation-maximization algorithm
- Use neighboring values
 - Direct substitution
 - Regression

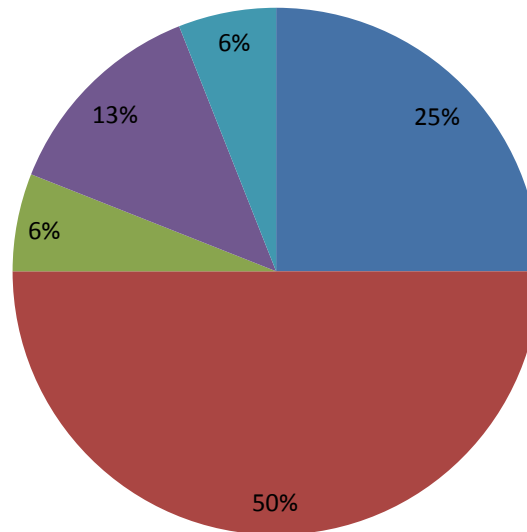
All gap filling methods introduce new error into the final total!

Stream flow Gap Causes

Hubbard Brook

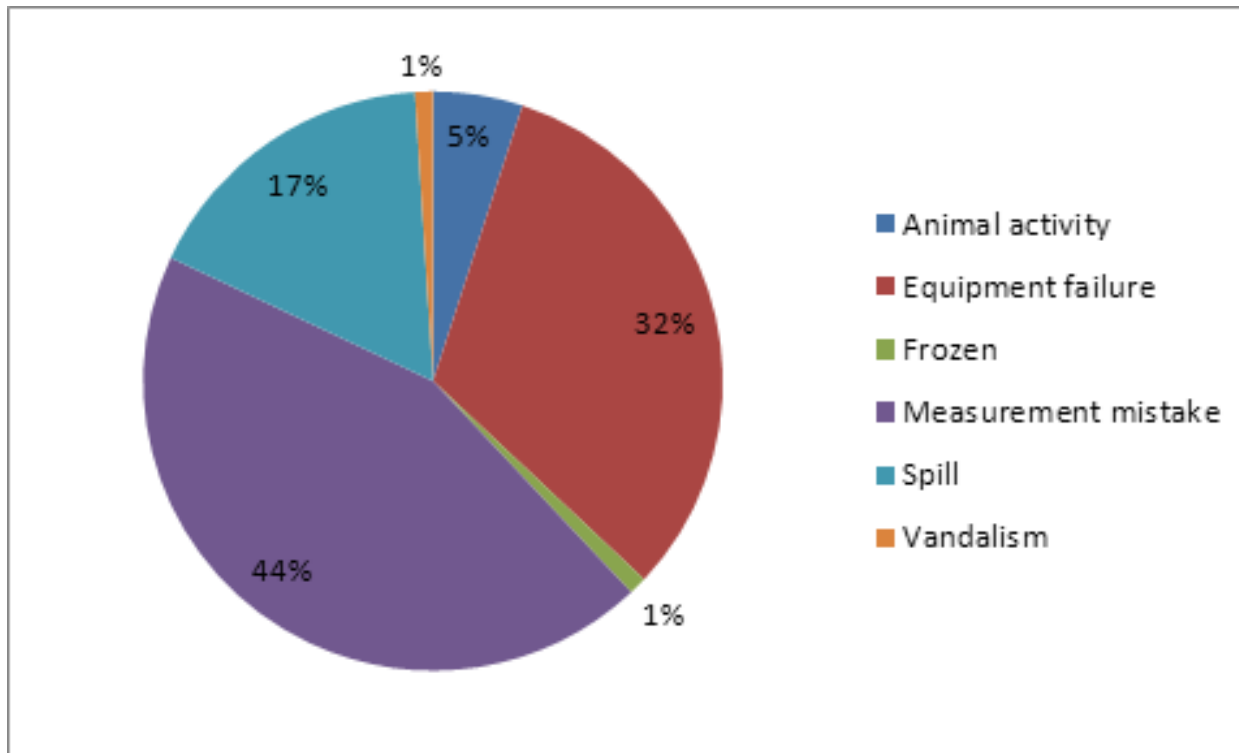


Coweeta



- Equipment failure
- Maintenance
- Operator error
- Ice in v-notch
- Debris in v-notch
- Misc.

Precipitation Volume Gaps at Hubbard Brook



Precipitation Chemistry Gaps at Sevilletta NWR

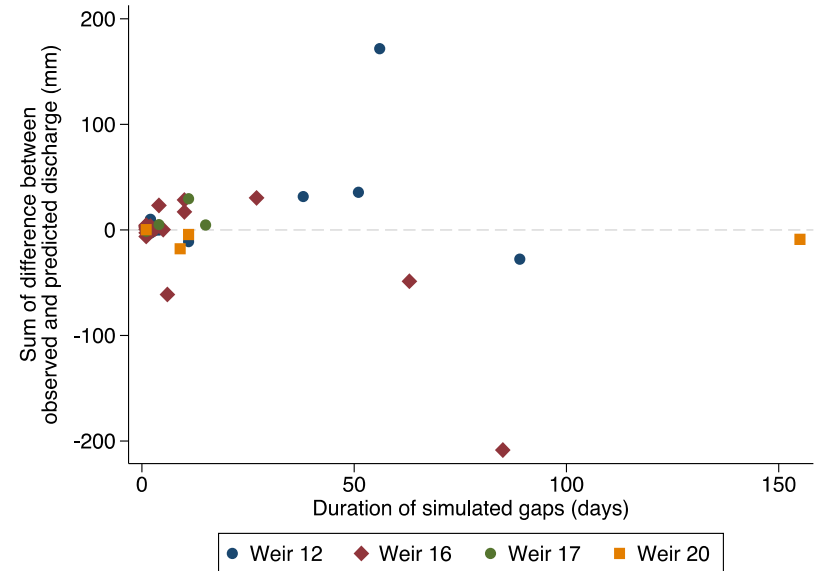
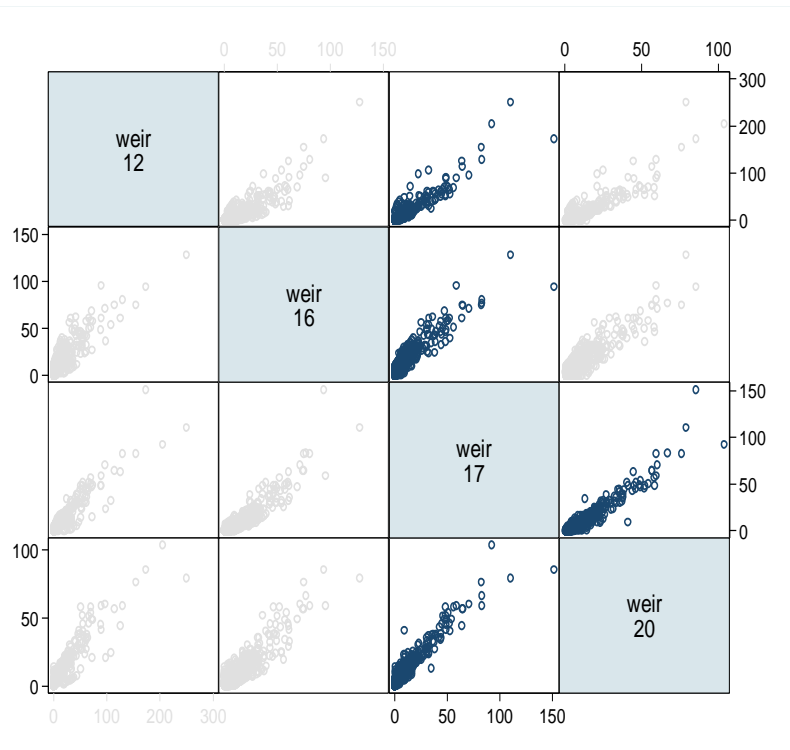


Credit: Odonfiction.wordpress.com



Credit: Doug Moore

Streamflow gaps at Gomadansan Experimental Forest



(Yanai et al. 2013, in review)

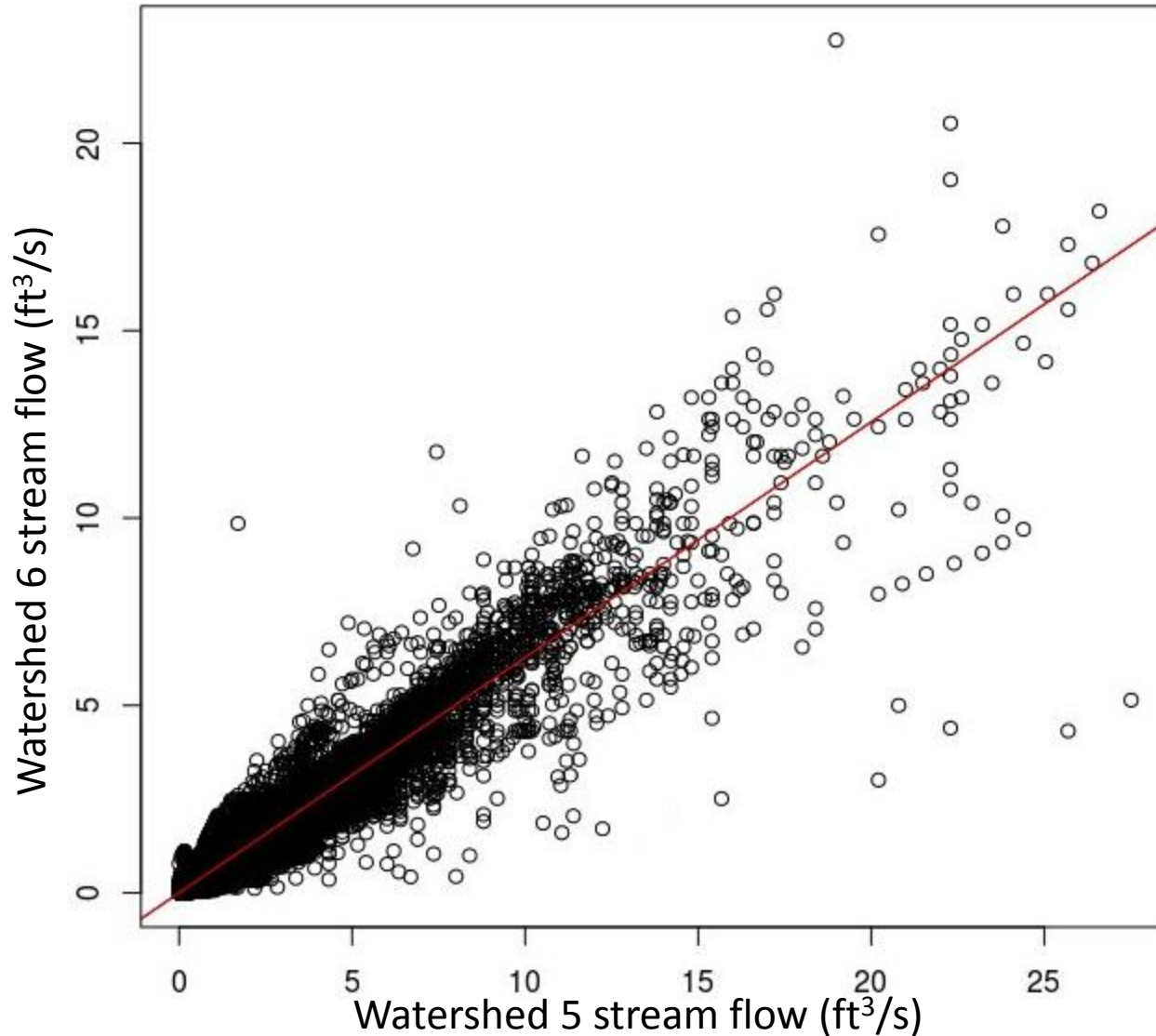
Example: Stream Flow at Hubbard Brook

- Long term record from 2 watersheds
- Model predicting one from the other
- Record of actual gaps occurring
 - Gap length
 - Flow rate at start of gap

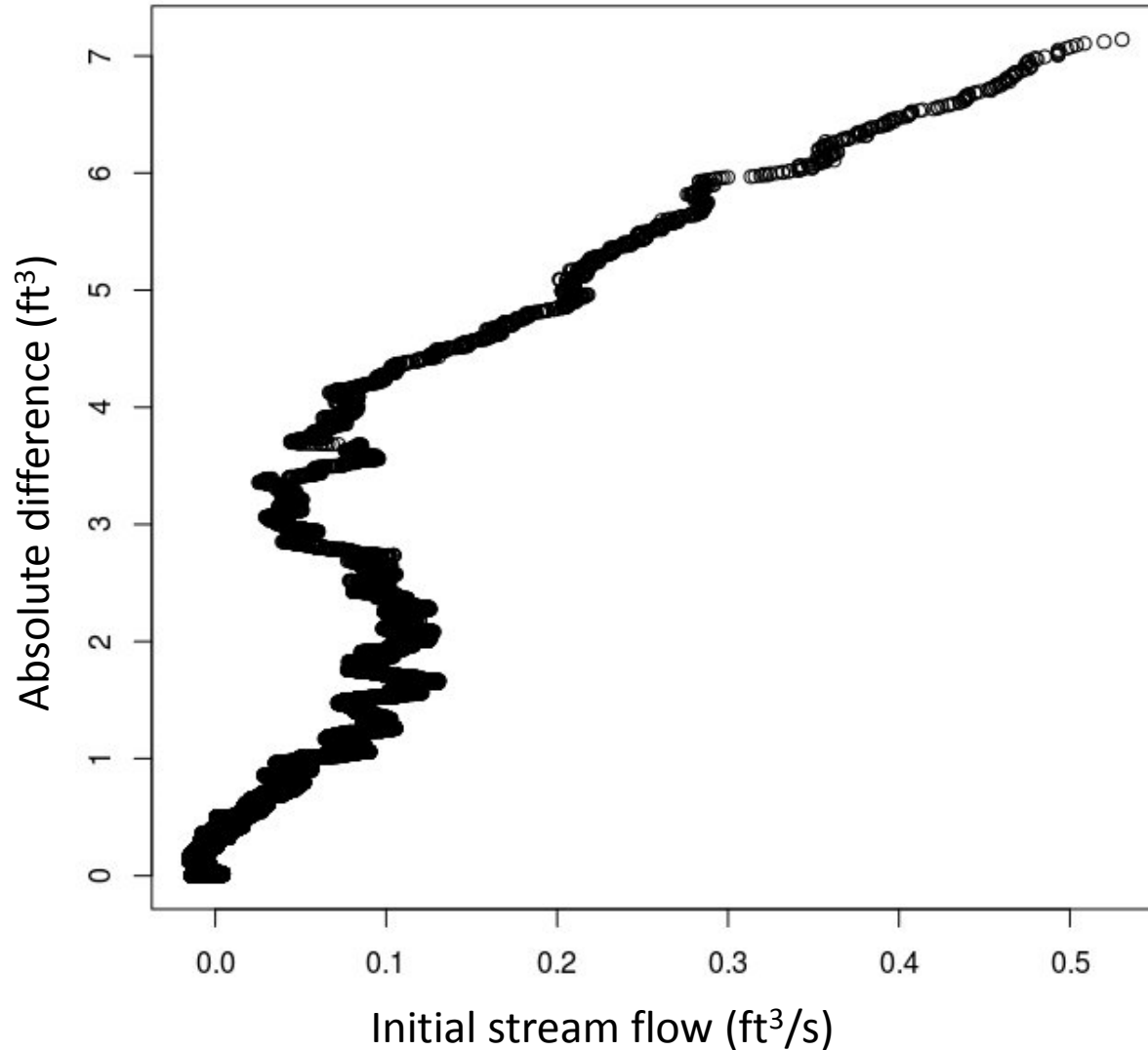


Credit: HBRF

Stream flow at Hubbard Brook 1962-2001

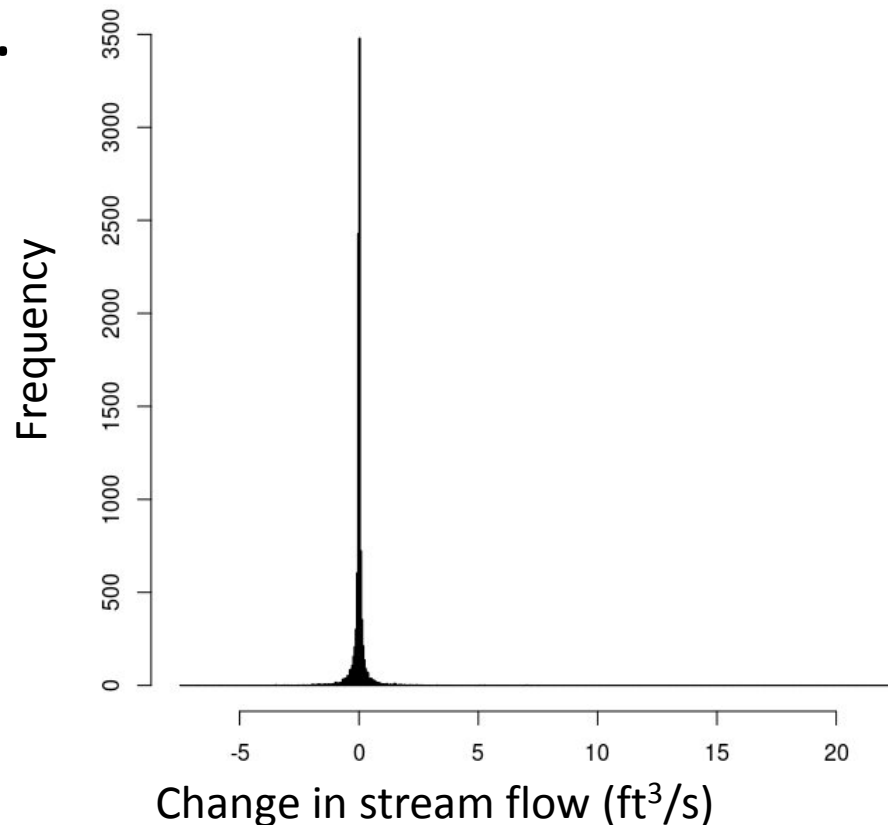


Flow rate during gap impacts uncertainty in total volume (half hour gap)



Methods-flow distribution

Randomly sample 100,000 “fake gaps” from the data to create a distribution of change in flow during a gap.



Creating a sampling pool

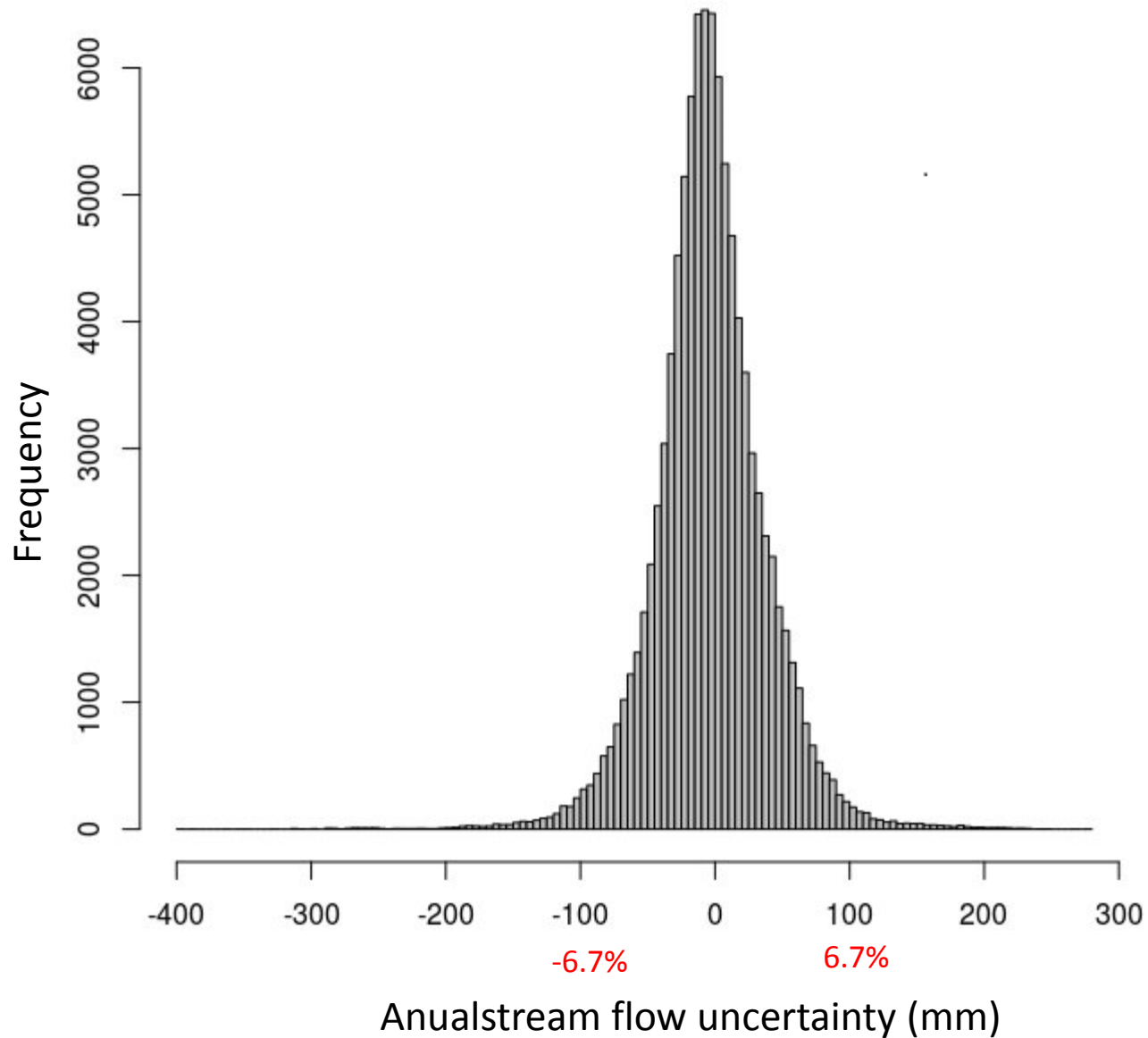
For each “real gap” in the record:

- Randomly create a “fake gap” of same length from the master dataset
- Randomly select a flow-change value from the flow distribution
- Does the max. flow rate inside the fake gap exceed the max. flow rate of the real gap \pm the flow change value?
 - If yes, reject and pick another fake gap
 - If no, include this fake gap in sampling pool

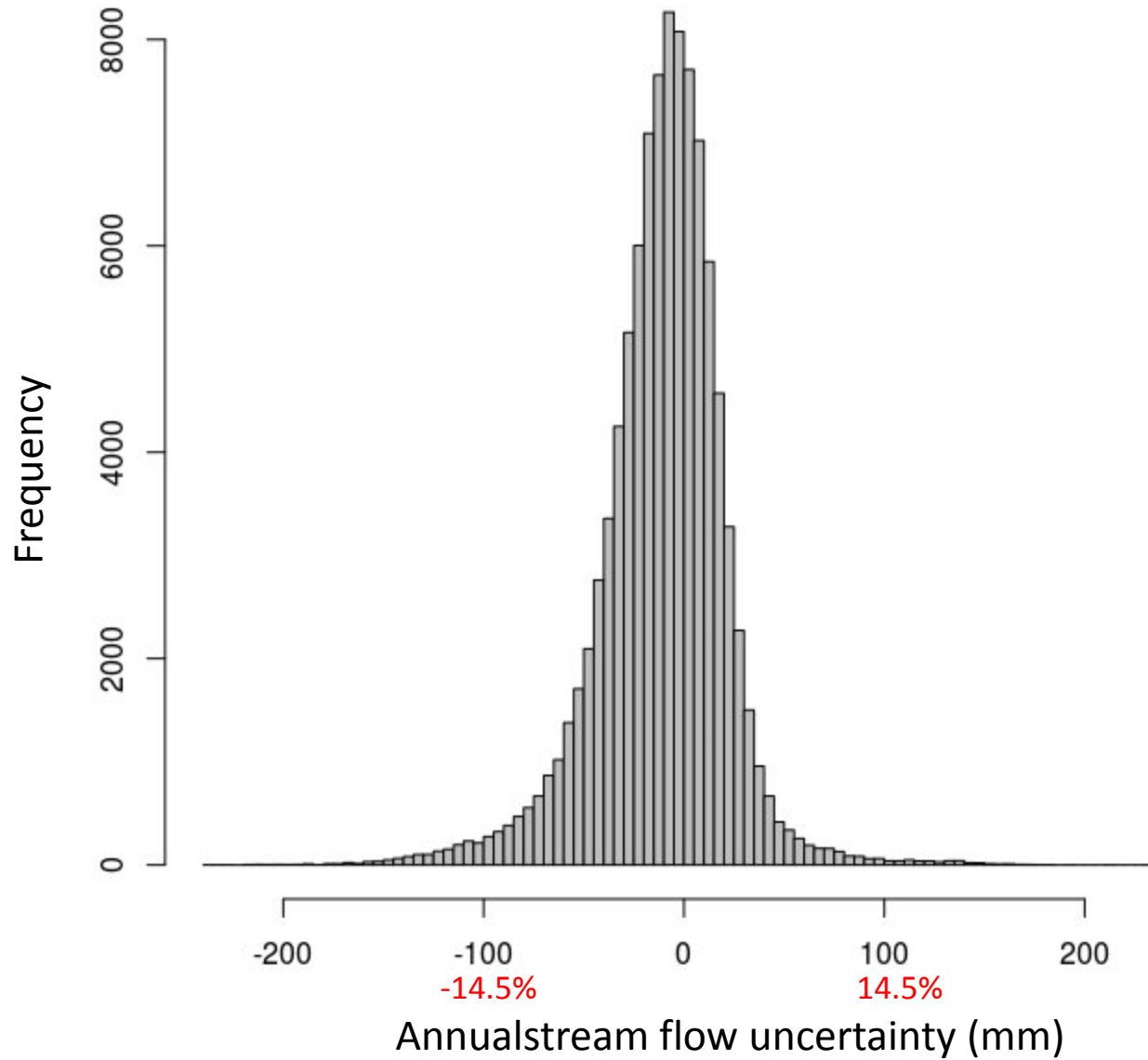
Sampling Pools

- Create a sampling pool of 1000 possible uncertainties for each real gap in the record
- To calculate uncertainty for a year, sum random values pulled from sampling pools for each real gap in that year.
- Do this for 100,0000 iterations.

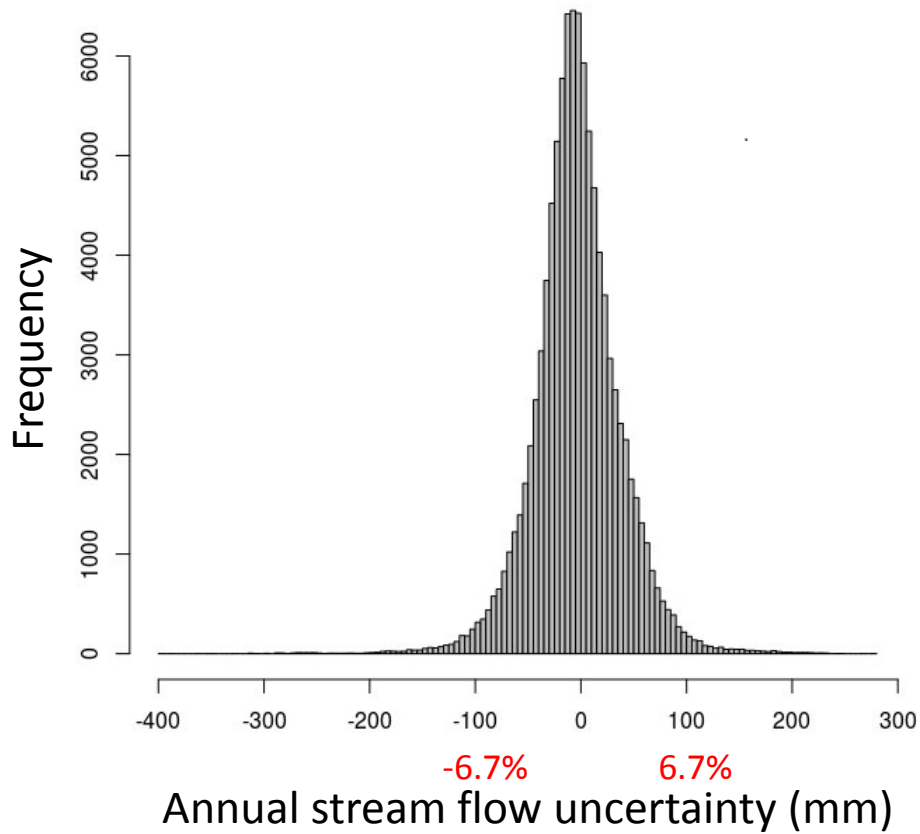
1996 (wet year)



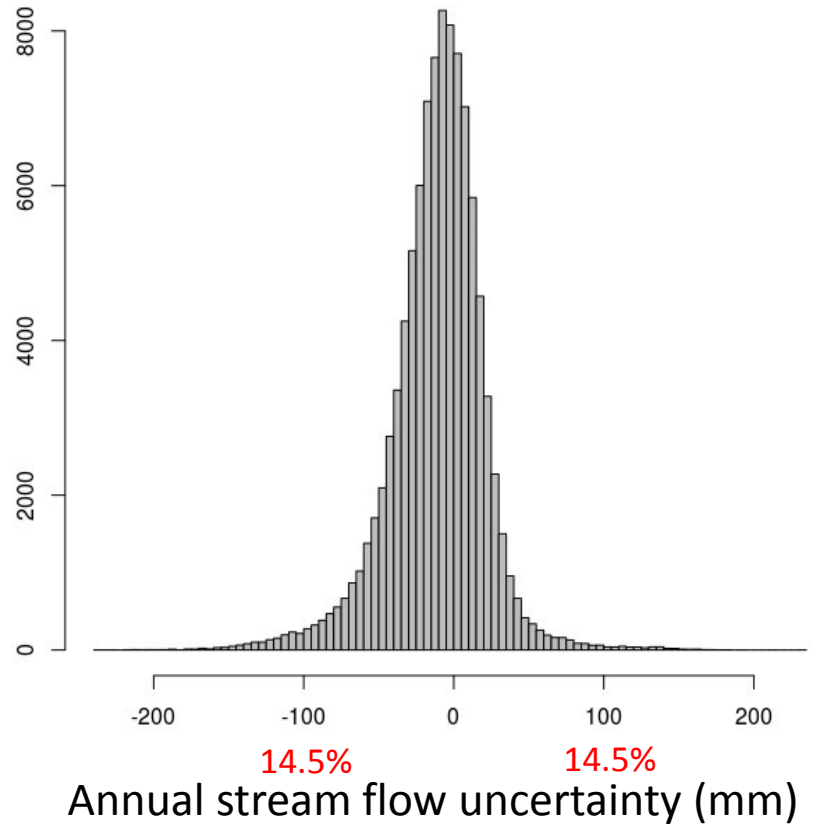
2001 (dry year)



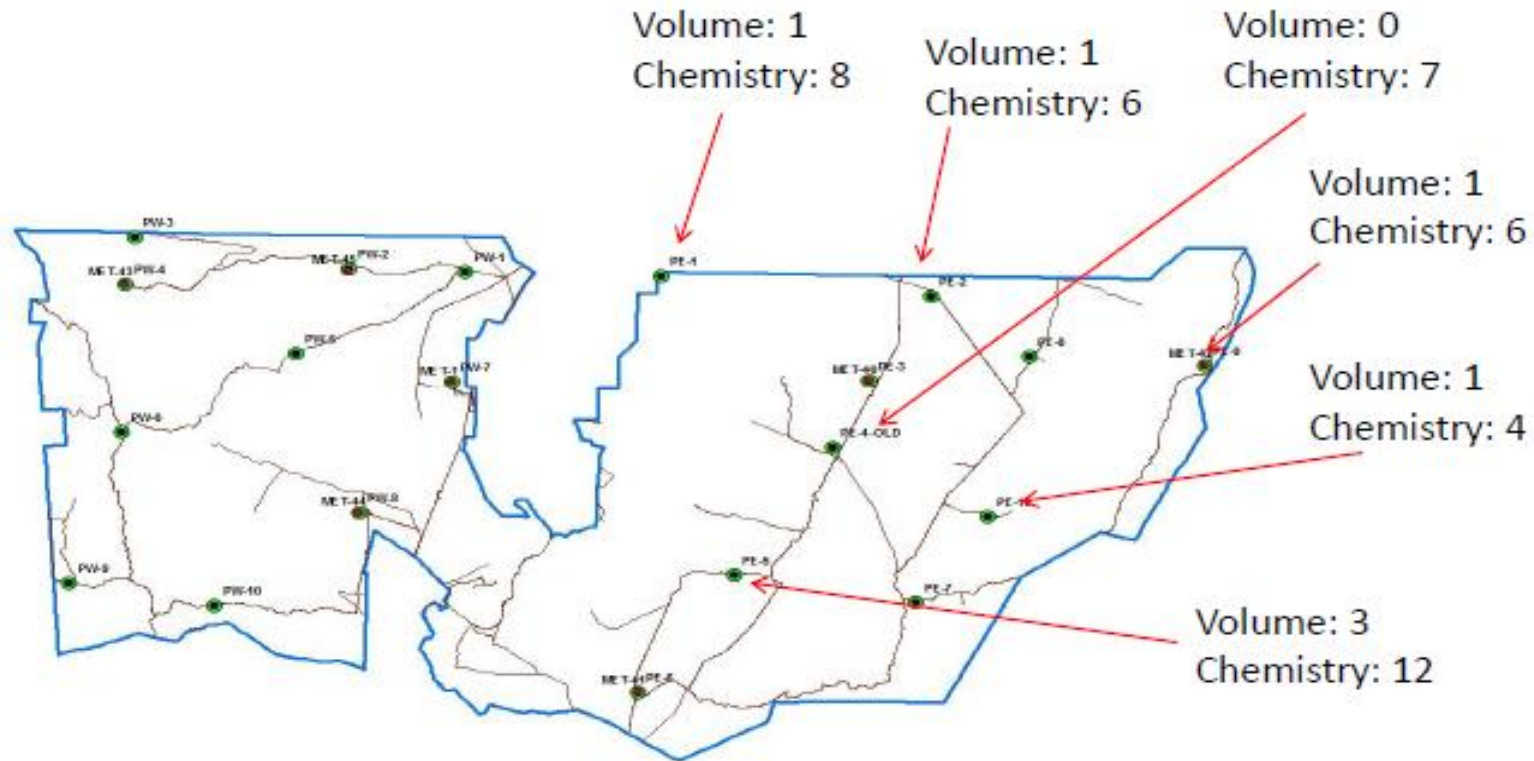
Wet Year (1996)



Dry Year (2001)

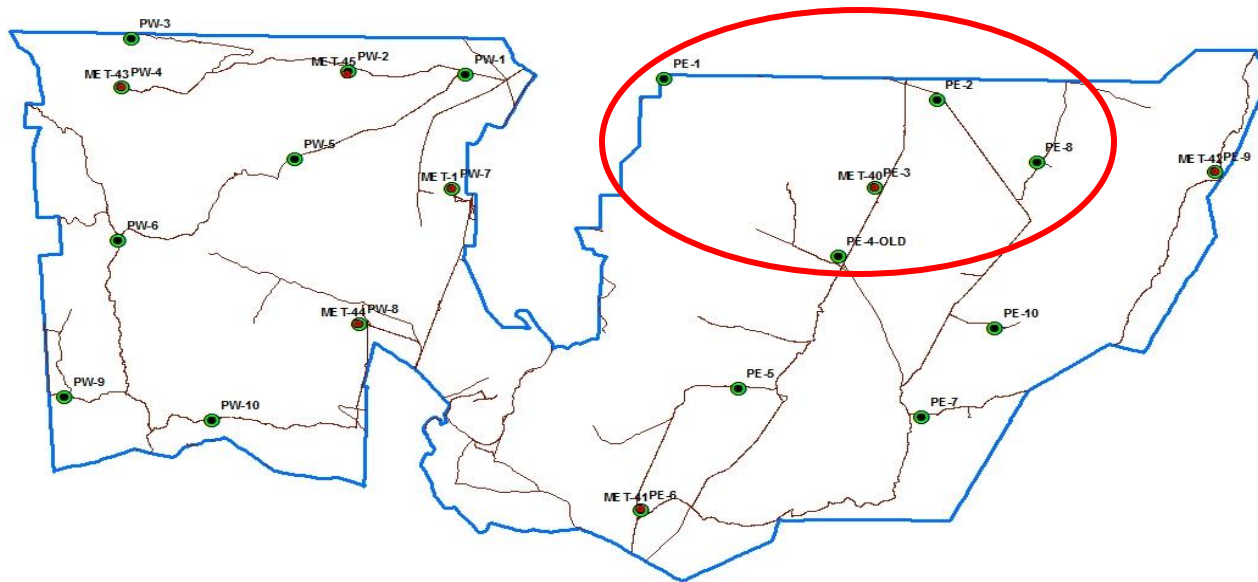


Precipitation gaps at Sevilleta (1989-1996)



How do we incorporate “gap uncertainty” into annual nitrate deposition estimates?

Precipitation gaps at Sevilleta

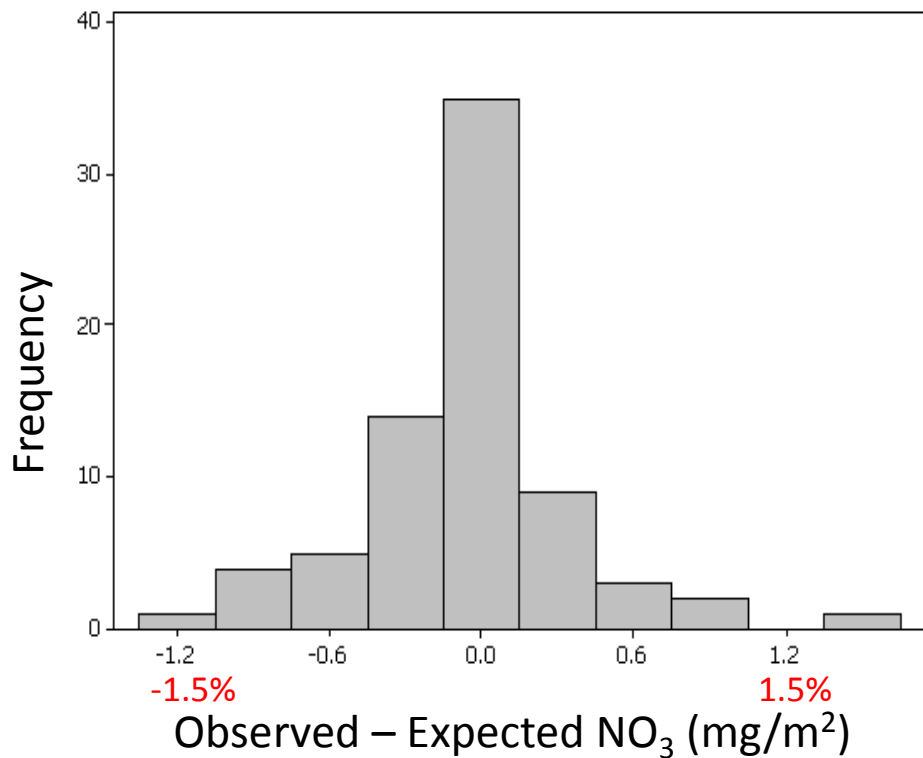


How do we incorporate “gap uncertainty” into annual nitrate deposition estimates?

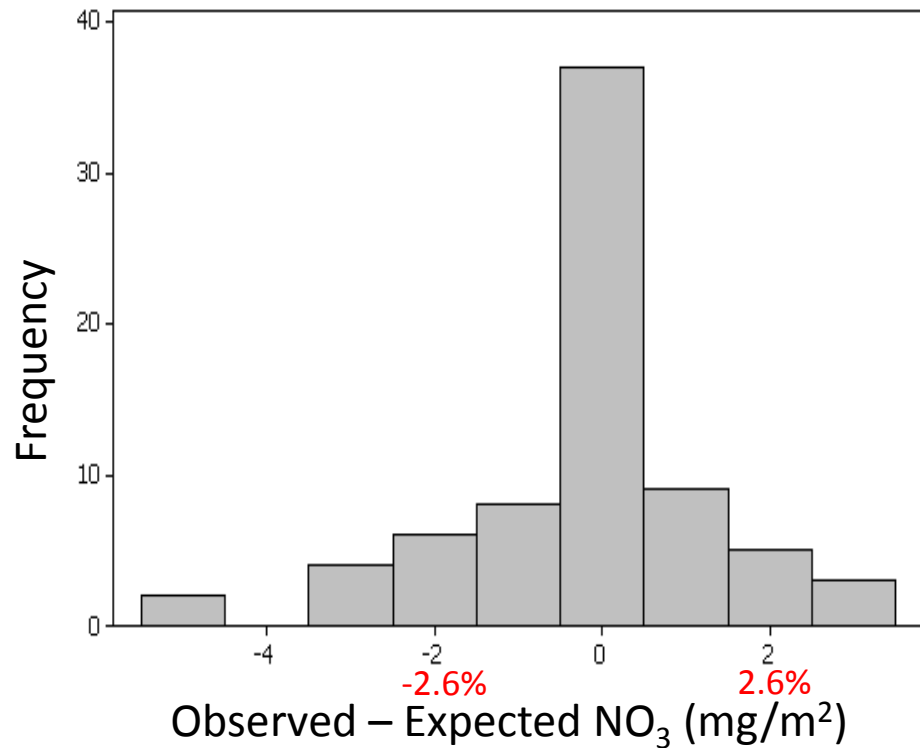
Methods-Wet Deposition Gaps

- Construct a variogram based on observations that include all gauges (84 observations)
- Use basic kriging interpolation to estimate deposition for all observations
- Re-run analysis separately for all combinations of funnels.
- Create distribution of observed-expected for each combination

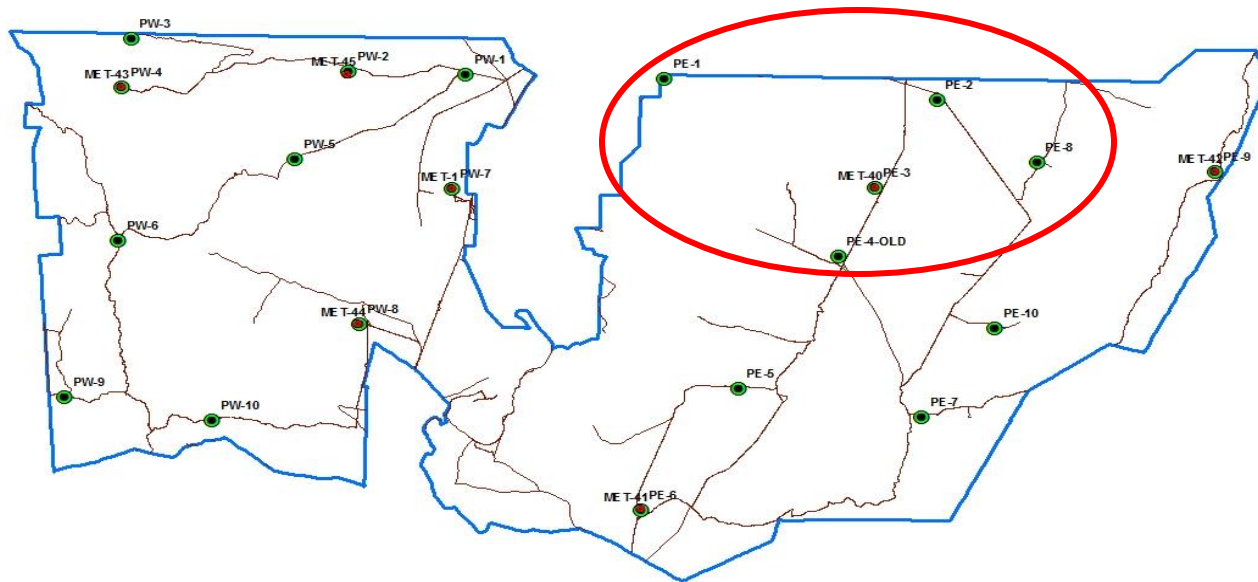
Gauge 3 removed



Gauge 8 removed



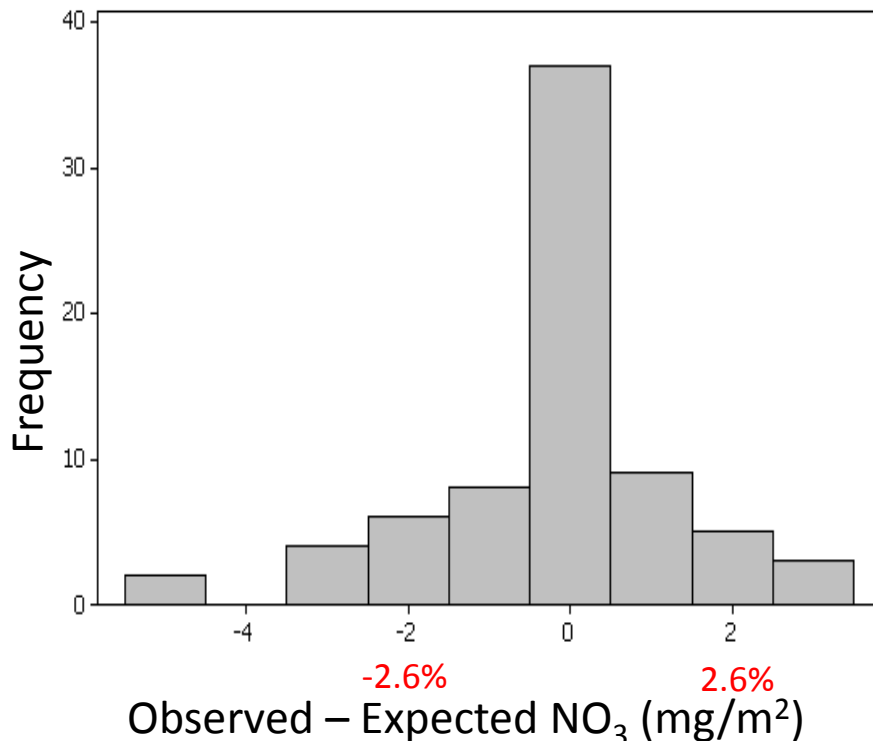
Precipitation gaps at Sevilleta



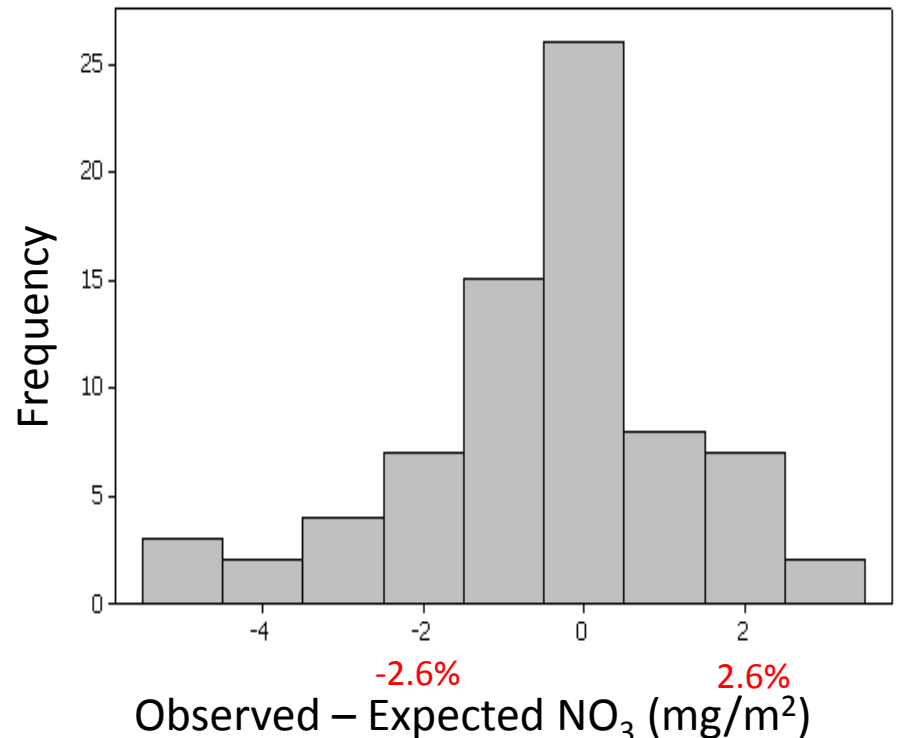
How do we incorporate “gap uncertainty” into annual nitrate deposition estimates?

Removing multiple gauges rarely changes the range of possible errors, but increases the chance of higher error within that range for a given event.

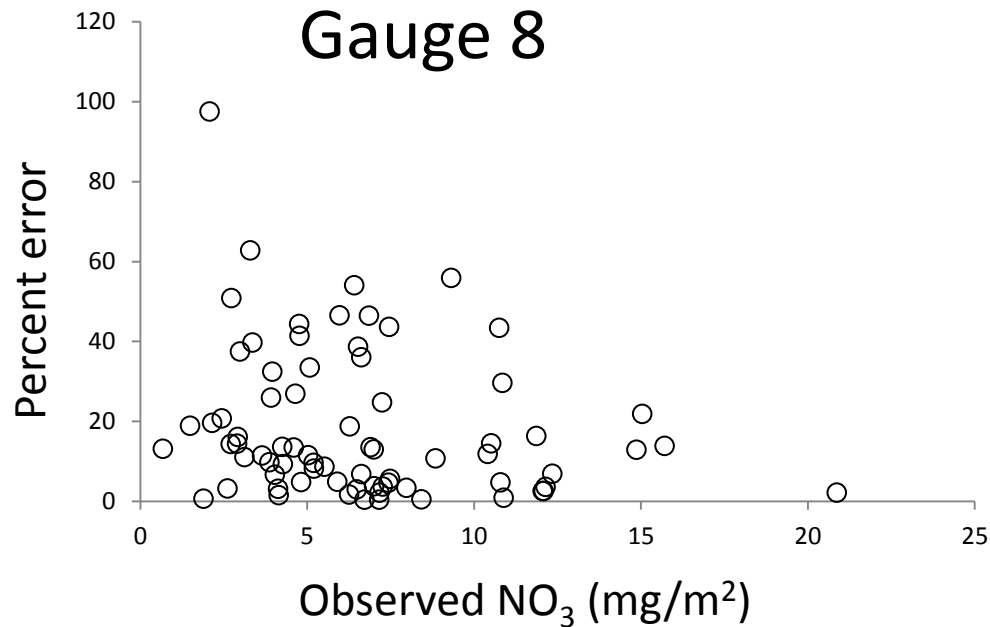
Gauge 8 removed



Gauge 3&8 removed



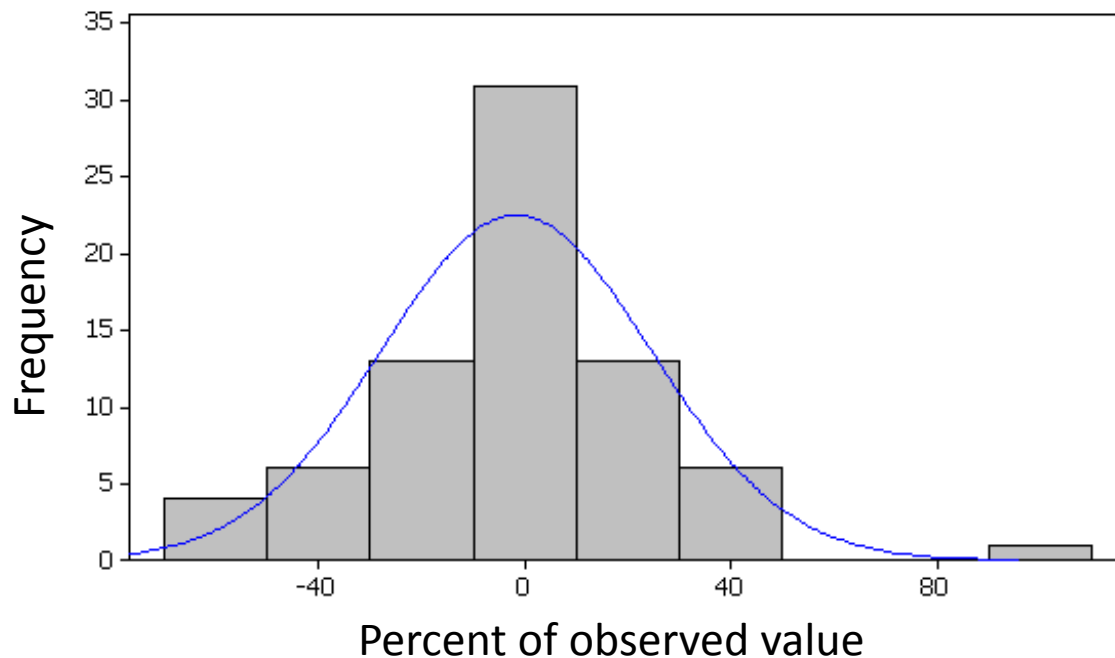
There is no correlation between amount of deposition from a collection, and the percent error in gap filling with kriging.



Its OK to express your sampling pool as a percentage.

To account for differences in collection, express the distribution as a percent of N deposition during a given event

Gauge 8



Use these percentages and the deposition estimate for a real gap to create a sampling pool!

Summary

- Data gaps can contribute significant uncertainty to estimates of ecosystem nutrient inputs and outputs.
- With long term datasets, this uncertainty can be estimated empirically

Advantages:

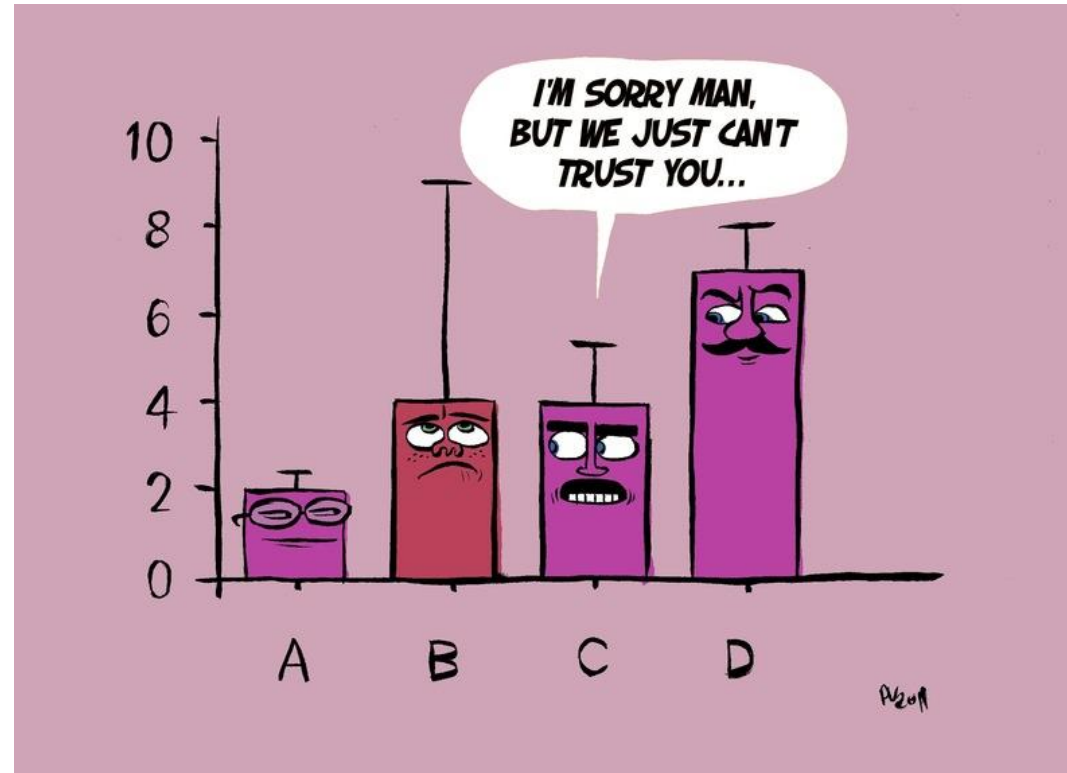
- Doesn't rely on parametric estimates
- Can use any model
- Easy to understand

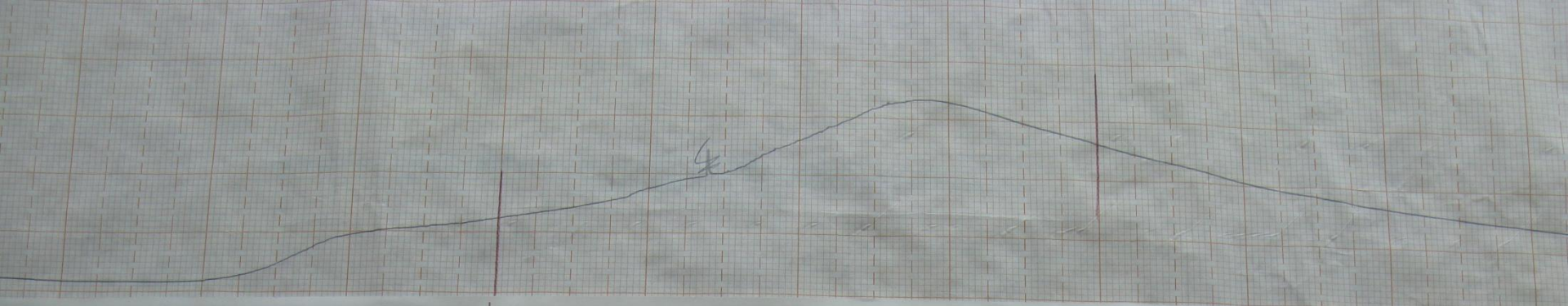
Disadvantages:

- Requires data
- Computationally intensive
- The past doesn't always predict the future!

Thank you!

- Co-authors
- John Campbell
- Franklin Diggs
- Becka Walling

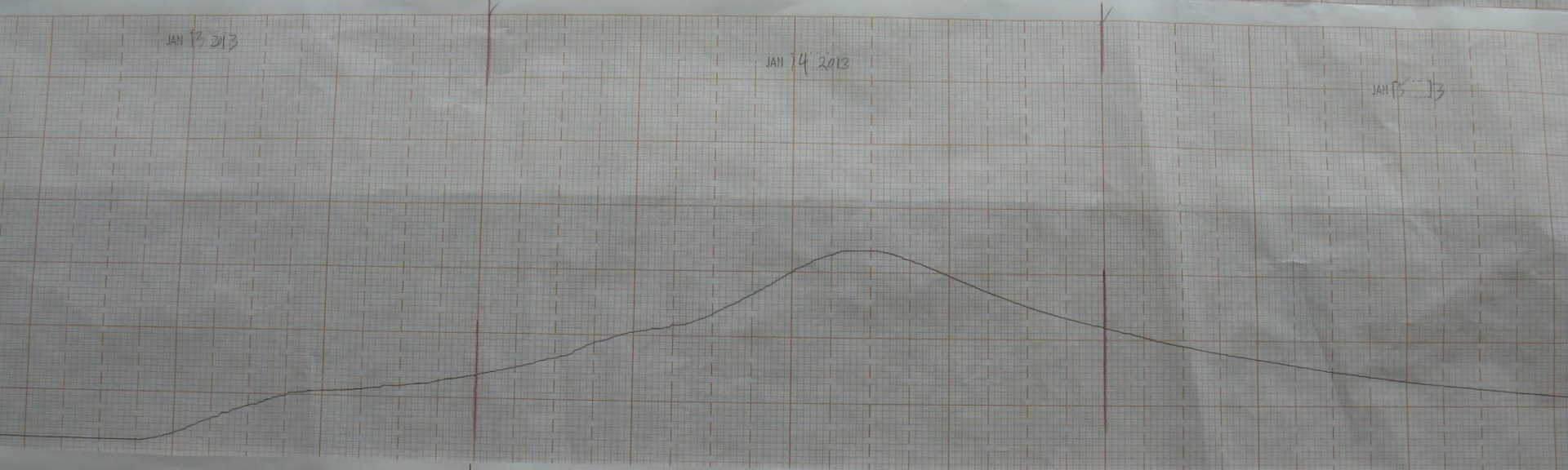




JAN 13 2013

JAN 14 2013

JAN 15 13



JAN 13 13

JAN 14 13

JAN 15 13

