Estimating uncertainties on annual nutrient loads using sensor concentrations

François Birgand[1], Chiao-Wen Lin[2]

EGU Meeting – Wien, Österreich, 8-14 April, 2018

@birgandwaterlab

1. Background

- Water quality continuous sensors are revolutionizing our understanding of biogeochemistry because they provide water quality information at the time resolution necessary to track transport and fate of particulate and dissolved materials in the environment
- New sensors often involve miniature water labs, fluorometers and spectrophotometers
- While they certainly provide unprecedented high frequency data, it is important to evaluate how good the 'measured' values are
- All instruments use some sort of calibration method to calculate concentrations from the physically measured. As for all instruments, calibration is key to performance

2. Hypothesis

- We hypothesize that it is possible to calculate the uncertainties associated with new water quality sensors through the variation of their calibration curves

3. Objectives

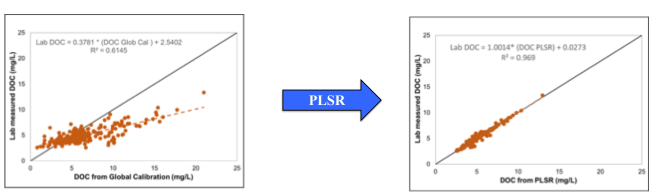
- Show establishment of water quality rating curves from absorbance data
- Calculate the impact of the number, the range and distribution of calibration points on annual loads uncertainties
- Compare these uncertainties to those induced by infrequent sampling
- Propose guidelines to minimize uncertainties

4. Source of data

- Instrument tested: Spectro::lyser from S::CAN®
- Used absorbance measured in a 2nd order stream in the coastal plain of North Carolina over a period of an equivalent of 12 months

5. Establishment of water quality rating curves (WQRC)

- Used Partial Least Square Regression (PLSR) as a central tool to calibrate the probe
- Absorbance spectra or fingerprints are used as ***index data*** and PLSR is used as a **rating method** to obtain WQRC
- We have shown that it is possible to obtain WQRC for a wide variety of parameters, including nitrate, TDN, TN, DOC, TSS, TP, Fe, Si
- Concentrations calculated from 'global calibration' need local calibration otherwise results may be poor



6. Methods used to calculate variability of calibration curves and loads

- **Bootstrap:**
  - Resample original calibration points with replacement
  - Identical sample size with calibration points then PLSR
  - 1000 simulations
- **Random subset:**
  - Subset original calibration points with 12, 24, 52 samples
  - Randomly subset original calibration points then PLSR
  - 1000 simulations
- **Stratified subset:**
  - Subset original calibration points with 12, 24, 52 samples
  - Stratified sampling in several bins then PLSR
  - 1000 simulations
- **M5 algorithm for infrequent samples:**
  - Subset original calibration points with 12, 24, 52 samples
  - Use $L_e = V \times \frac{\sum C_i \times Q_i}{\sum Q_i}$
  - 1000 simulations
- **M6 algorithm for infrequent samples:**
  - Subset original calibration points with 12, 24, 52 samples
  - Use $L_e = \int C_{interp} Q$
  - 1000 simulations

$$\%\ \text{difference (error)} = \frac{L_e - L}{L} \times 100\%$$

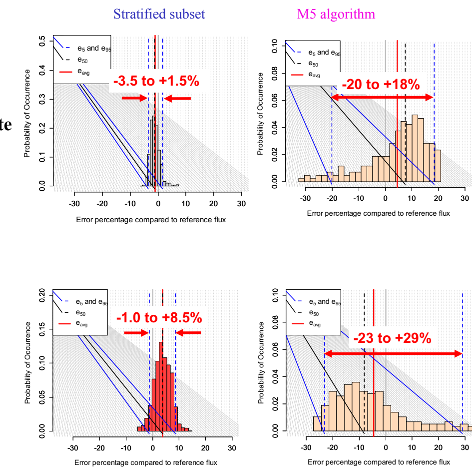
7. Distribution of uncertainties

- 24 samples per year

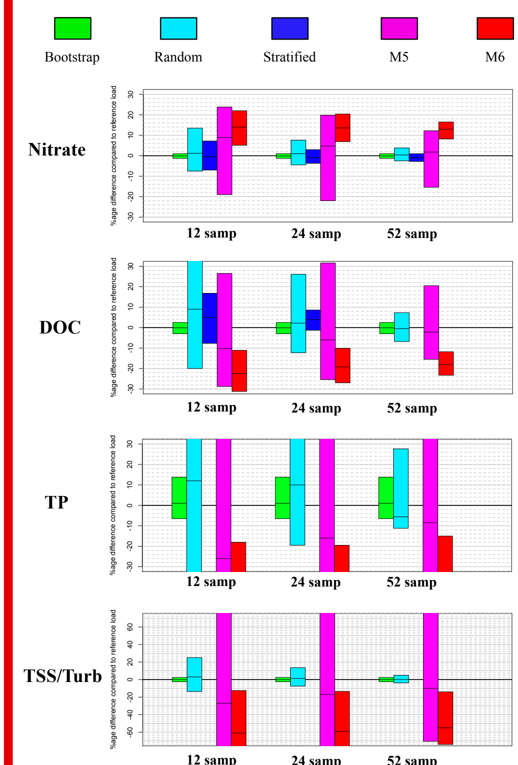
Stratified subset | M5 algorithm

Nitrate

DOC



8. Uncertainty ranges function of number of calibration points and distribution



9. Conclusions

- We are able to obtain very small uncertainties for most parameters, but for TP. Vastly better than with infrequent sampling
- Native distribution of points may induce bias. To reduce bias, one should strive to obtain stratified sampling
- Stratified sampling – the most robust sampling method
- Try to cover as wide a calibration range as possible

1: Bio&Ag Engineering, North Carolina State University, Raleigh, NC, USA 2: Department of Safety, Health, and Environmental Engineering, National Kaohsiung First University of Science and Technology, Kaohsiung City, Taiwan