

Ecological Sensor Data Quality Assessed Using Observational Data and Combined Uncertainties

Guy Litt¹, Janae Csavina², Joshua Roberti², and Jesse Vance²

¹Battelle Memorial Institute

²National Ecological Observatory Network

November 22, 2022

Abstract

Delivering long-term, high quality environmental sensor data spanning the continent is a primary goal in the National Ecological Observatory Network's (NEON) Instrumented Systems (IS) group. Some independent observations collected by NEON's Observation System (OS) measure similar data at the same location and time as the in-situ sensors. Coinciding IS and OS measurements facilitate supplementary data quality assessments by vetting IS sensor data (e.g. aquatic pH probe) against corresponding OS data (e.g. water grab sample analyzed in a lab for pH). To assess whether IS data agree with OS measurements, we use uncertainty as a tool to understand data quality. The uncertainty between NEON IS and OS data follow analytical (e.g. summation in quadrature) or numerical (e.g. Monte Carlo) approaches depending on the complexity of the IS-OS comparison algorithms. NEON calculates the IS-OS uncertainties, and applies the expanded uncertainty as control limits for acceptable IS-OS data comparisons. IS-OS comparisons falling outside the uncertainty-based control limits help to (i) explore unaccounted uncertainty in the IS and OS data, and (ii) address issues in the data or sample collection process as ongoing continuous improvement strategies.

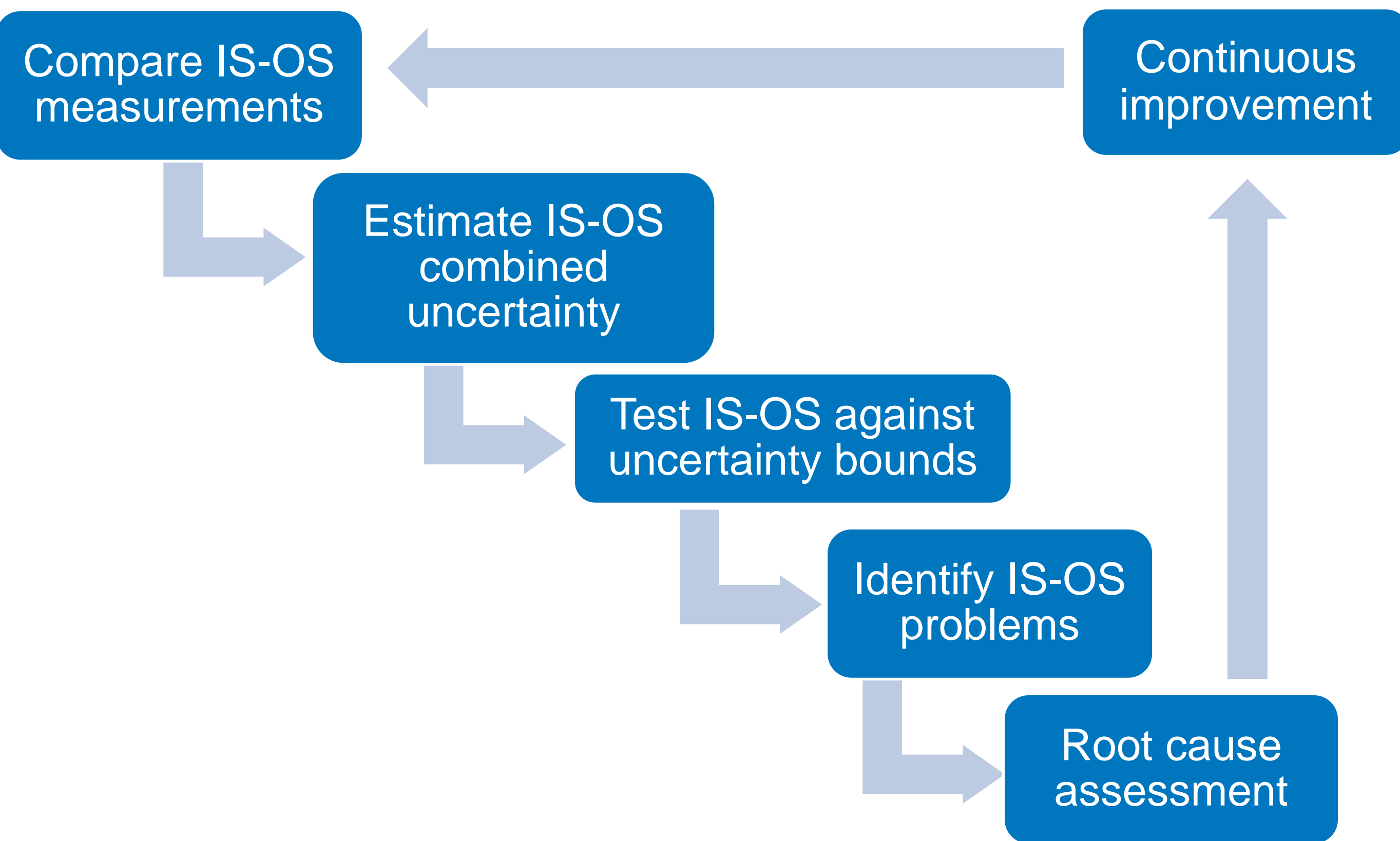
Ecological sensor data quality assessed using observational data and combined uncertainties

G. Litt¹ (litt@battelle.org), J. Csavina¹, J. Roberti¹, J. Vance¹
¹ National Ecological Observatory Network (NEON), 1685 38th St., Suite 100, Boulder, CO 80301, U.S.A.

Abstract

Delivering long-term, high quality environmental sensor data spanning the continent is a primary goal in the National Ecological Observatory Network's (NEON) Instrumented Systems (IS) group. Some independent observations collected by NEON's Observation System (OS) measure similar data at the same location and time as the in-situ sensors. Coinciding IS and OS measurements facilitate supplementary data quality assessments by vetting IS sensor data (e.g. aquatic pH probe) against corresponding OS data (e.g. water grab sample analyzed in a lab for pH). To assess whether IS data agree with OS measurements, we use uncertainty as a tool to understand data quality. The uncertainty between NEON IS and OS data follow analytical (e.g. summation in quadrature) or numerical (e.g. Monte Carlo) approaches depending on the complexity of the IS-OS comparison algorithms. NEON calculates the IS-OS uncertainties, and applies the expanded uncertainty as control limits for acceptable IS-OS data comparisons. IS-OS comparisons falling outside the uncertainty-based control limits help to (i) explore unaccounted uncertainty in the IS and OS data, and (ii) address issues in the data or sample collection process as ongoing continuous improvement strategies.

Data Quality Assessment Workflow



Data Quality Assessment Using Uncertainty

Observation (OS) Data



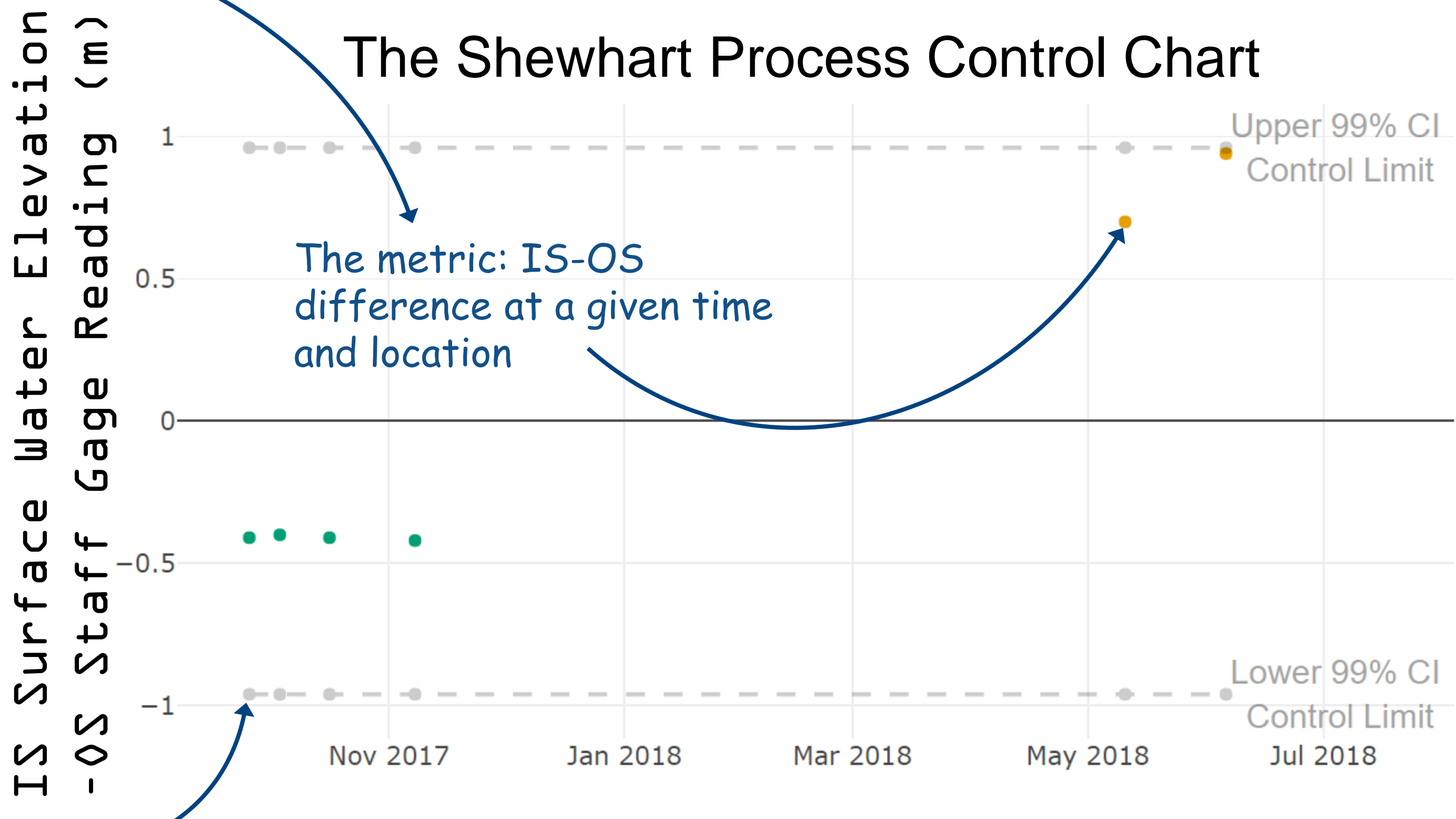
Combined Observation Uncertainty

Instrument (IS) Data



Combined Instrument Uncertainty

Total IS-OS combined uncertainty at 99% confidence interval (CI)

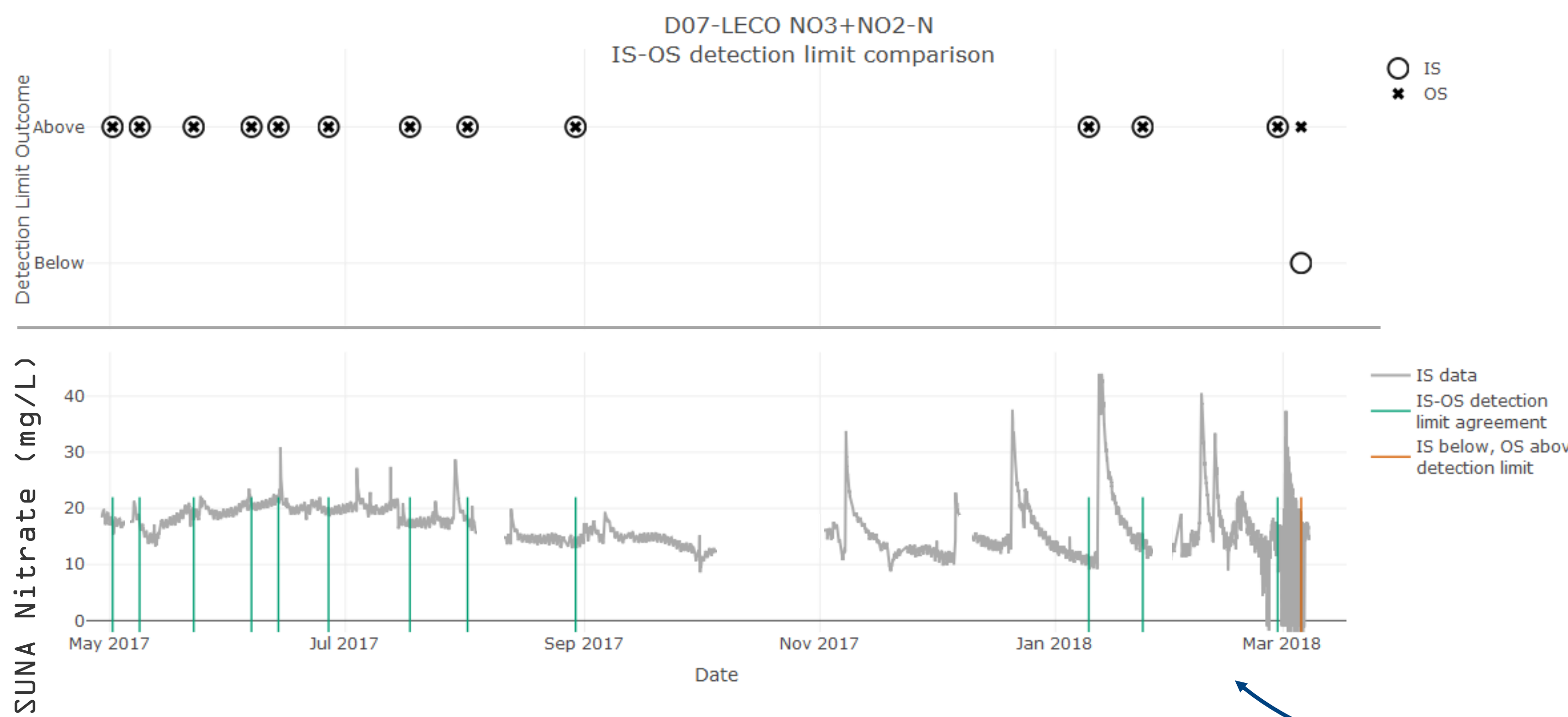


The 'acceptable' bounds to the metric

Uncertainty sources included in assessment:

- Observational sampling method
- Calibration
- Natural variance
- Data acquisition system
- Data transformation
- IS noise during OS sampling time

Detection Limit Assessment



Assessment Challenges:

- Control limits too wide for some data products
- Data publishing lag
- Some IS measurands differ from OS measurands (e.g. IS SUNA Nitrate vs. OS Nitrate+Nitrite)
- Environmental changes to IS-OS relationship
- Measurements below detection limit

NEON's Future QA/QC + Uncertainty Strategies:

- Reduce measurement uncertainty in field (e.g. permanent benchmark installations for elevation data products)
- Overhaul data pipeline
- Assess environmental conditions on IS-OS relationships with more data
- Employ above/below detection limit analyses