Joint NERC Environmental Sensor Network/Sensor NIS Workshop, October 27$^{th}$, 2011

# Working Group 3: QA/QC

# What should we do?

- What is the minimum level of real-time QC we can provide (level 0 to 1)?
  - None.
  - Some (range check, remove logger flags, etc.)
- Foremost, we must identify what QC has been done
  - Provisional data must be labeled clearly
  - Streaming QC methods identified (if any)
    - If none, that should be made clear too

# What can we do?

- Some good practices
  - Range checks
    - Identifies impossible values; not unlikely ones
  - Running Averages
    - Range of calculation is specific to site and sensor type
  - Sensor redundancy
    - Ideal: Triple the sensor, triple the logger!
    - Practical: Cheaper, lower cost, lower resolution sensors, or correlated (proxy) sensors
    - Side Effect: establish user-confidence in data products.

# What tools are available?

- Environments for Streaming QC
  - Matlab GCE toolbox
    - Great GUI, tracks provenance, integrated math environment
  - Kepler
    - Can share agents/workflows with other users, tracks provenance
  - R libraries/statistical programs
    - Code can put in metadata
  - Web service
    - Users submit data for review by server-side application
- Require training & evaluation
  - Workshops, documentation, seminars, webinars, etc.
  - Annotated examples, shared experience, etc.

# Working/Action Item 1

- Creating one-pagers on tool experience
  - IM's background with medium (R, Kepler, Matlab, etc.)
  - Specific methods used
    - Parameter settings/how were these chosen?
  - Dataset information
    - Types of sensors
    - Types of sensor failures addressed (list of 6, other?)
  - Performance
    - How well did the tool work on your problem?
- Compile a survey of reports
- Publish in LTER Newsletter, DataBits

# Gap-Filling

- "Should we?" -- IM, "How do we?" -- Science
- Lots of potential methods
  - Nearest neighbor station (temporal/spatial)
  - Regression-based methods
  - More advanced statistical approaches
- Emphasis on documentation
  - If you do it, describe HOW
- Clearinghouse for tools & approaches.

# Working/Action Item 2

- Creating "Test-Sets" to evaluate gap-filling/QC methods
  - Remove observations from complete records.
  - Inject known modes of sensor failure into data we know is good
- Establish criteria for what a method must provide
  - Uncertainty estimates (conf. intervals, probability values)
  - Documentation
- Establish performance metrics
  - MSE, Type I/II error rates
- Produce a publication that compares several of these methods across these exemplar sets.

# The Reality

- Will we ever get away from provisional data?
  - Not likely the human QC expert will ever be completely removed
- Focus on assisting the QC expert
  - Guide his/her inspection to important components in the data