Group 2
Basic QC, level 0 to 1

- Estimate appropriate data ranges from historic local data, e.g., max +- 2SD
- Apply by appropriate time of year.
- Update as the parameter is measured over time.
- Removing vs. flagging “bad” data – some sentiment to remove data collected while sensors are being changed, cleaned, or rebooted.
- Permanent vs. temporary flagging
- Flagging code form? Alphanumeric or just numeric? Cross-site standards?
- Well-established measurements (e.g. air temperature) vs. relatively new types of automated measurements (e.g., soil CO2)
- QC needs to occur immediately as data come from logger, and later as trends become apparent.
- Need descriptive flags – e.g., if data are “bad”, describes why the data are bad (faulty sensor, data transmission)

Variance

- Increased variance can be used as indicator of sensor degredation, fouling.
- Need to look at windows of time
- Standards for frequency of observations?
- Standards for sensitivity/precision of detecting outliers?
- Cross-site standards difficult to establish

QC occurs at >1 processing steps

- Sensor specific
- Level 0 – raw data, no QC, no flagging
- Level ½ - convert data format, units, etc. but no flags
- Level 1 –
  - QC near real time data for immediate problems – several processing steps.
  - At this point, data could be provided to public.
  - May develop web services to query data per user requests (e.g. gap filling or not, coarser timestep summaries, etc.).
- Level 2 – Longer time-series may prompt further exploration of data for problems, e.g., sensor drift, gap filling. Obligation to provide gap filling?

Gap filling

- Should missing time stamps be infilled? (Level ½)
- And if so, would you then flag the blank rows? (Level 1)
- No infilling, but flag bounding rows of a gap?
- Multiple algorithms
  - different researchers will prefer different methods
  - even if the data are provided gap-filled with a preferred method, the researcher may want to download the raw data and check the data processing step (lack trust).

Data qualifiers

- Data qualifiers will be different at each level of data processing.

Timestamp Variable Level 1 flag Level 2 flag…

Past efforts have given subjective code classes, e.g., “good”,”questionable”. For level 1, just mark whether

1) Tests of the data passed or failed (binary).
2) Whether there is ancillary information about the data from a technician
3) The data should be used in analysis or not

Level 2: gap filling, drift detection

Data documentation

- Key – tell the user how data were collected and processed
- Necessary metadata
  - Contact info
  - Geolocation/time range of data collection
  - Data processing workflow
  - Data headers/definitions
  - Flag definitions
  - System requirements
  - Calibration steps/data
  - Information about sensors – basic documentation and log of maintenance over time.
  - Hardware configuration over time
  - Expected file format/contents/size (checksums)
  - Links to other files related to the data

Action Items

- A book or wiki of best practices for data QC and algorithms for data processing (e.g., gap filling)
- Continuation of idea of community network website for choosing, placing, and maintaining sensors, and managing sensor data.