

Interactive comment on “Automated quality control methods for sensor data: a novel observatory approach” by J. R. Taylor and H. L. Loescher

Anonymous Referee #1

Received and published: 4 January 2013

GENERAL COMMENTS

I find this paper generally well written. The topic covered is an important building stone for the NEON project and the success of the outlined QA approach is crucial to the delivery of a high quality data product. In this paper, validity of some of the underlying concepts needs to be established and applicability of the QC approach needs to be proven.

DETAILED COMMENTS

Use of central limit theorem: Use of the central limit theorem really only works well at representing the distribution of a statistic if the nominal distribution is nearly normal -

C7037

which temperature statistics (max, min, variability) generally are. For time series which have statistics that exhibit bimodal or more complex distributions the resulting randomly sampled distributions will not accurately represent the limits of the true data. (An example can be seen at the web site (<http://www.cs.uic.edu/~wilkinson/Applets/clt.html>) if one selects the arcsine or binomial choices.)

For many variables this will not be an issue, but for variable which can exhibit distinctly bimodal characteristics - for example precipitation and radiation and other closely associated factors - using the central limit theorem will result in distinctly un-representative parameter values.

Page 18184 lines 21-24: Cases such as the examples given are not 'indefinable'. It is more likely that their definition is trivial, such as a maximum wind direction of 360 degrees. A measured wind direction greater than 360 degrees would clearly indicate an error in either the instrumentation or algorithms used to determine wind direction, and as such would benefit from a plausibility maximum of 360 degrees.

Page 18185 lines 11-13: The concept of multi-stage, or cross-referenced QC is quite important. Statistical tests are not sufficient on their own and other comparative tests are required. For the reader, knowing what comparisons would be beneficial is paramount as it dictates what instrumentation needs to be deployed, and where it needs to be placed and how frequently it needs to be sampled in order to be useful for QC testing. It would be beneficial if this paper could present the physical concepts required for such cross-referenced QC and give a concrete example. Essentially, this means that sensors used for cross-reference QC must measure or be influenced by a common environmental property. For example a precipitation sensor could be used in QC of wetness sensors, humidity sensors, many radiation sensors, and soil moisture sensors to name a few. What readers need to know is how such QC relationships should be established.

Page 18186 lines 5-6: It is not clear how values are being flagged in the step test.

C7038

Because the step test uses the differential of the time series a large step value will actually represent two values from the time series, one which is nominally good and an adjacent value which is nominal bad and separated from the good value by a relatively large change in value. Does the step test flag both the good and bad values from the original time series – which would mean that 2.5

Page 18186 lines 10-24: The null/gap tests seem to be of a different nature than the preceding statistical tests. Although not clear from the text, it appears as if you are measuring the number of missing values in a statistical interval and assigning a single QC flag to that interval based on the result. In the preceding tests, I assume that the QC flags are applied directly to the measured time series values and not to any associated interval of derived statistics.

There also seems to be a logical fault to the null test. By enforcing a $\mu + 2\sigma$ upper limit on the number of missing values would imply that a time series which generally has very few missing values could result in statistical intervals consisting of primarily good data being flagged as bad because of a small number of missing values.

Page 18188: The temporally and spatial weighted statistics defined by equations 4-7 would likely be relevant to derived statistics such as hourly means, daily maximum or such. It is not clear how such weighting would be applied to time series collected at higher frequency (eg 10Hz or 0.1 Hz) as the weighting functions may only be relevant or much smaller spatial and temporal scales. A concrete set of examples are needed in order to show how such a weighted statistical test value could be employed with the measured NEON time series data.

Page 18190 line 23: Should ‘...5.2 x 10⁶ data points.’ Really be ‘... 5.2 x 10⁶ data points.’ ?

Data driven approach: The authors promote a ‘data driven approach’ to quality control which minimizes the subjective nature of QC inherent in human driven approaches (page 18192 line 21). This is a laudable goal but the example and caveats given in this

C7039

paper suggest to this reviewer that we may still be some way from achieving such a goal. Specifically: the lower limit variance test was not applied, the Delta test was not applied, and an arbitrary gap test size was employed.

Page 18193 line 3: Can the authors specify references for the ‘other QC approaches’ which employ ‘numerous assumptions’

Page 18194: The wording of this section is very in-definite. The use of the words: ‘ideally’, ‘should’, and ‘potentially’ give the impression that NEON doesn’t really have a plan worked out for quality control of the data.

Page 18194 line 22: Will NEON make available non-quality controlled data to allow the community to test/apply their own quality measures?

Page 18195 line 22-23: The sentence ‘By adopting a uniform-basis for data QC, future network interoperability can be assured.’ is somewhat of an overstatement. Simple uniformity to QC does not guarantee interoperability; certainly it is possible to have uniformly bad or inappropriate QC.

Page 18202: Should the sigma test include tests for both “ $\mu - 2\sigma$ ” and “ $\mu + 2\sigma$ ” as opposed to the calculation specified in the table?

Page 18209: The clipping of the large (but obviously good) PRT values in this figure emphasize that the blind application of ‘data driven’ QC tests has the potential be very inappropriate. It may be that the tests outlined in this paper do not scale well to the short time series used in this example. If so, then the authors need to use a longer time series example or better, to devise test methods that do scale to short or long time series, so that readers can benefit from this work.

Interactive comment on Biogeosciences Discuss., 9, 18175, 2012.

C7040