

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

D. Papale (Referee)

darpap@unitus.it

Received and published: 25 January 2013

Taylor and Loescher in this paper present a number of tools to perform data quality evaluation and flagging in measurements acquired by environmental sensors. The topic is of absolute importance in the actual situation where observatory networks are under construction in US and Europe and the huge amount of data that will be collected require automated and objective tools to ensure the highest possible quality. Although the efforts in the direction of a standardized and automatic QAQC procedure in micrometeorological measurements already started years ago within the regional networks (CarboEurope, Ameriflux, Fluxnet-Canada) and then in FLUXNET, the defi-

C7667

inition of new fully automatic and objective methods is still needed.

The paper is of interest for Biogeosciences but in my opinion there are some changes needed to make it more clear and interesting; papers that propose techniques (like this one) need to be precise in the method description in order to avoid possible misunderstanding in their application. Here below the points that I suggest to change and/or improve:

1) It would be better to define at the beginning (in section 2.1, or in table 2) that when you talk about mean and standard deviation you don't refer to the average and sigma of the data but, for example for the sigma_test, the “standard deviation of the standard deviations” (page 18185, lines 5-6). This would improve the clearness.

2) I found the differences between null_test and gap_test not very clear; where is the threshold in number of gaps between the null and the gap tests? In addition it is not clear which action is suggested if a period is flagged for the null_test (that means that there are a number of single measurement points missing higher than expected): the whole period is flagged?

3) The previous point leads to a more general question: also the others test are based on analysis of “designed periods of time”. This is quite arbitrary and could lead to different conclusions if period length is different also because the “flagging” is related to the whole period. I suggest that more information and guidelines are given in the period definition; for example one option to explore could be that if a period doesn't pass one of the tests, it could be then analyzed using shorter periods in order to better identify the measurements points that create the problem.

4) The method proposed to calculate mean and standard deviation based on temporally and spatially adjacent observations is interesting (Eq. 4 and 5) but it presents two problems that should be better discussed: the arbitrary definition of the parameters z and t and the fact that in this way errors in the others sensors used are added on. I'm not sure that it is worthwhile to take this approach.

C7668

5) The Result section should be called something like “Test example”

6) One of the main point that I think should be changed in the paper is the example presented. With more than 10 years of micrometeorological measurements and a quite dense network, I would like to see the example based on a long term dataset (to test the sigma and delta tests, not applied now), with adjacent sites (to test the equations 3 and 4) and possibly with a second sensor that could be taken as reference “true” value to be used as validation. I don’t think that it will be difficult to find such dataset (I can help if needed to find a site), for example one could do the analysis using sonic temperature and dataset to be quality checked and thermometer measurements as reference in a site cluster. The use of data coming from exclusively from a NEON tower is not relevant for the scope of the paper that wants to present a methodology.

7) I found not very clear the sentence at page 18190, lines6-8

8) My final comment is more a personal feeling I had reading the paper that however I want to share with the authors: in the paper you refer many times to NEON (cited 24 times in the text). It is important to describe what is NEON to give a background about why you are proposing these tools. However I think that these references and the sentences related to “what NEON will do” are too many, and I had to come back on the title of the manuscript while reading it to check if you were presenting a general method (as in the title) or NEON and his approach. In my opinion a paper less focused on what NEON planned to do (e.g. section 4.2) and more on the method proposed would be better and increase the clearness because more focused.

Dario Papale

PS: thanks for the acknowledgment, but there is an error in the name.

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