Earth Syst. Sci. Data Discuss.,
https://doi.org/10.5194/essd-2017-101-RC2, 2018

Open Access Earth System Science Data Discussions

Interactive comment on "ReOBS: a new approach to synthetize long-term multi-variable dataset and application to the SIRTA supersite" *by* Marjolaine Chiriaco et al.

Anonymous Referee #2

Received and published: 10 February 2018

This paper introduces a synthetized long-term multi-variable dataset. It provides a set of methods to process ground-based data at an hourly time scale leading to the ReOBS product for the SIRTA supersite. This process put many variables from various of sources into one single netCDF file and homogenizing them into same format and time resolution, which is very easy for community to use. A webpage link is provided for the data download, quicklooks and documentations. I found the data is well-documented and the paper is well-written. I would also be interested to see the future improvement of including radiosonde profiles. This may be a great value added to the current data. Overall, I think it is a valuable contribution to ESSD with only some minor comments.

Line 31: 'type ob'-> 'types of'

Line 106: move the expansion of ReOBS ("Re" stands for . . .) to line 96.

Line 118 and 123: in other part of the text it is about sixty variables, but here is fifty. Please check the consistency.

Line 260: can you elaborate more on how the given weight corresponds to the geometric representativeness? E.g. a geometry map may be overlapped in Figure 3a.

Line 274-283 and Table 3: this part is kind of confusing. Firstly, I don't understand what the authors exactly mean for the first sentence "as the correlation between the adjacent samples increases with the sampling rate". Are you calculating correlation coefficient for different time windows? According to later part, my understanding is that the non-physical jumps are detected by checking the difference (not the correlation) between measurements in two successive time windows (e.g. 5min), is it right? Secondly, In the limits given in Table 3, what do the upper and lower arrows mean? What do the time windows (5min, 15min, 60min) mean? My guess is that the 60min measurements are used to detect the unphysical persistence by calculating the std. dev. of 1min measurements within this 60min window. Does "5min" mean two successive 5-min measurements are used to test the unphysical jumps? How do the two examples in line 281-283 (2 hPa within 1 minute and 0.6C within 1minute) related with the limits given in Table 3?

Line 288-289: "The value is 0 m/s because of frost deposition on the sensor" -> "The value is 0 m/s after 18 UT because of frost deposition on the sensor (shown by low T and high RH in Figure 4c)"

Line 312-323: is there any corrections imposed on the EC-based fluxes (e.g. density correction (Webb et al. 1980), coordinate rotation (Wilczak et al., 2001), etc)?

References:

Webb EK, GI Pearman, and R Leuning. 1980. "Correction of Flux Measurements for

Density Effects Due to Heat and Water Vapour Transfer." Quarterly Journal of the Royal Meteorology Society 106(44):85-100, doi:10.1002/qj.49710644707.

Wilczak JM, SP Oncley, and SA Stage. 2001. "Sonic Anemometer Tilt Correction Algorithms." Boundary-Layer Meteorology 99(1):127-150, doi:10.1023/A:1018966204465.

Line 337: Can you elaborate what kind of profiles are considered noisy? Does <40% are noisy mean >60% profiles are valid?

Line 351: delete "." After "used"

Line 403: the year might be wrong in "Chepfer et al. 201)"

Line 638: timescaless -> timescales

Line 653: "at the time t" this sentence may not be completed.

Table 1: physical bounds and native resolution are not given for "(5) 2-m wind direction". I think there should be some value, right?

Table 2: some variable short names (e.g. tas_SIR, tas_TRP) are not consistent with the variable names in the downloaded data product.

Figure 4b: there are a lot of spikes and it is not clear which are rejected and which are kept.

Figure 7: "norm. T2" is calculated following the equation (1) -> should be equation (2)

Comments related to the data (downloaded from the link provided in the manuscript):

1. The current variables are sorted by alphabet order. This is not convenient to find the variables of site information (lat, lon, time, etc). I would suggest moving those variables to the front or to the end, similar to the ARMBE.

2. Since it is a single product with measurements from many instruments, it is impor-

tant to list the data source (measured from what instrument?) in the attribute of each variable.

3. Some long_names are difficult to understand (e.g., mld, prp, trps, *_l, *_ph). Consider expanding these abbreviations.

4. qc flag: some have flag_meanings in their attributes but some don't

5. std_*: are all these standard deviations within 1-hour time window? Some don't have the "1-hour" in long name.

6. "1-hour std of std_*variable*" should be "1-hour std of *variable*"

7. In some long names of u/v: wing -> wind

8. In global attributes: there are two titles/sources/locations/institutions... and what does the "gps" mean in title/system/source?

Interactive comment on Earth Syst. Sci. Data Discuss., https://doi.org/10.5194/essd-2017-101, 2018.