Clim. Past Discuss.,
https://doi.org/10.5194/cp-2017-64-AC1, 2017


Climate of the Past Discussions

# *Interactive comment on* "Effects of undetected data quality issues on climatological analyses" *by* Stefan Hunziker et al.

**Stefan Hunziker et al.**

stefan.hunziker@giub.unibe.ch

Received and published: 9 August 2017

We want to thank the anonymous referee #1 for the important and helpful comments and suggestions.

Response to Anonymous Referee #1

Review of Hunziker et al., 'Effects of undetected data quality issues on climatological analyses' The issues raised in this paper are important – in particular, that data quality control methods which focus on daily data may not detect systematic issues that are important at annual or monthly timescales, and that those systematic issues may distort trends and/or homogenisation processes.

The key substantive result of this paper, in my view, was that for TN, the median adjust-



ment in DATAQC-E was very different to that in DATAQC-S (with flow-on consequences for the trends). However, there is no indication in the paper as to why such a difference might have happened. One could reasonably form a null hypothesis that data quality issues might be expected to be randomly distributed in sign and in time; the results found suggest that there is a systematic departure from that null hypothesis, but without any information as to what might be driving the difference, it is very hard to know how to interpret results (or whether they might be applicable to other networks). I think further analysis/discussion of the cause of this difference is important in a revised paper.

Response 1: We agree with the referee that investigating the sources of the differences in the correction factor for TN for DATAQC-S and DATAQC-E would be an important further step. However, finding such explanations would require much further analyses (mostly in-situ), particularly in a sparse network with very limited metadata availability. Reconstructing the station history is nearly impossible in many cases. Therefore, potential sources causing a warm bias in former TN observations cannot be investigated in detail in this paper. Nevertheless, we will include a discussion of this important issue in the article. Furthermore, we will analyze if certain data quality issues occur more/less frequently in recent years and decades and if/how this may impact the adjustments in data homogenization.

Other comments are as follows:

Major comments

P2 line 27 (and elsewhere) – the text at various points suggests that the observed increase in diurnal temperature range may be indicative of a problem. There are cases elsewhere in the world where increases in diurnal range have been associated with drying trends (e.g. Dittus et al 2014 (Aust. Met. Oceanogr. J. 64-261-71) found an association between recent decreases in rainfall and increased frost frequency in parts of southern Australia), and the diurnal temperature range increases found in this study seem to me to be broadly consistent with the observed decreases in precipitation.

Potential rainfall-DTR relationships could also be brought into the discussion in sections 4.3.2 and 4.3.3, as well as being linked to the statement 'stronger increase in TX than TN in the Altiplano' at page 14 line 22.

Response 2: We will discuss this question in more detail in the paper. We will reference the publication of Dittus et al. (2014) and other relevant works. Such rainfall-DTR relationships may potentially be an explanation for the increasing DTR trend signal that still remains after an enhanced QC. The statement in the introduction does not imply that increasing DTR is necessarily the consequence of data quality problems. But the findings in earlier studies draw the attention to regions such as the Central Andean area and motivate to further investigate the causes of the detected trends (climatological or non-climatological). In-depth investigations of potential rainfall-DTR relationships is out of scale for the present publication but would definitely be an interesting research question for a follow-up publication.

P4 line 19-20 – does this mean that TX at these stations is not over a full 24 hours? If so, what is the time window that is used, and how confident are we that this practice has remained unchanged over the period of observations? Would TN also be over less than 24 hours? (if so, this might lead to low minimum temperatures which occur during afternoon storms – a common scenario in the tropics – being missed).

Response 3: We will add some information to these lines to make the issue more clear. At most stations, TX measurements do not cover 24 hours indeed. TX may only cover the afternoon hours (ca. 1200 to 1800 LST), when daily maximum temperatures normally occur. In certain cases however, this may lead to too low daily TX measurements. For TN, however, we are not aware of such practices, and we expect TN to be representative for 24 hours (measurements are normally taken at 0800 LST). To our knowledge, there were no systematic network wide changes in these observation practices. Because of the very limited information about the station histories, different and changing observation practices are possible at individual stations however.



Section 3.1.2 – I found this section very difficult to understand as a standalone paper. While it is reasonable to refer people to the IJC paper for full details (I also note that it is an open-access paper), I think more explanation is needed in this section so that readers can have at least a basic understanding of what is happening without having to look up a different paper. A new table would be useful, I think, with a brief explanation of each test – at present, for example, it is not at all obvious to the reader what 'missing temperature interval' or 'PRCP gaps' mean. Some of the checks in this list also aren't fully explained in the IJC paper (moderate and strong irregularities in the data pattern, frequent and large inhomogeneities, strong asymmetric rounding patterns).

Response 4: The referee is right that the tests of this work cannot be found completely in the previous IJC publication of Hunziker et al. (2017). We also agree that the present paper should be understandable without reading another publication. Therefore, we will include a table that summarizes all the tests mentioned in this paper. The explanations will be concise but they will make clear what kind of errors were detected in the dataset. The IJC publication of Hunziker et al. (2017) should just serve to find more details on the data quality issues.

Section 3.1.2 – I would also question how important the 'rounding errors' are. The example given in the IJC paper would affect temperatures by less than 0.1_C (which would not be worth worrying about), but there may be more significant examples.

Response 5: The effects of the rounding errors from degrees Fahrenheit to degrees Celsius are indeed not highly important. However, in these cases there is an accumulation of uncertainties: 1) the observer was not aware from which scale to read the temperature (in most cases, the instruments triggering the error have scales of the two units), and 2), we do not know why and how exactly the rounding error from degrees Fahrenheit to degrees Celsius occurred. Therefore, there are potentially other errors in those measurements, and it seems reasonable to not use such data for climatological analysis.

Section 3.1.2 – flagging of 40% of measurements is a large number. It would be useful to have an indication of which individual tests most commonly led to flagging. With 40% of observations flagged, presumably a substantial number of stations were removed altogether – how many? (Somewhere – maybe section 3.4 – it would be useful to say how many of the original stations were still available for analysis for each of the two methods). It would also be useful to know whether UDQIs were concentrated in one particular era or spread fairly evenly through the time series.

Response 6: We will add additional information on the frequency of the flagging by the tests applied, potentially included in the table described in Response 4. We will also include information on the temporal occurrence of data quality issues. The number of available stations varies between the different analyses. We will provide the most important numbers of available time series for the analyses and the differences resulting from the two data quality control approaches.

Section 3.5 – it needs to be made clear that these trends are for the 1981-2010 period (as it is currently worded, there is the possibility of confusion between the timespan over which trends are calculated and the timespan used as a baseline for normal/index calculation).

Response 7: The wording will be adapted to avoid any confusion between the time period over which trends are calculated and the baseline period for the index calculation. Since we are calculating trends for the 30-year period 1981-2010, the required 30-year baseline period for the index calculation is the same.

P9 line 4-14 – the wording in this paragraph is not as clear as it might be. It is a perfectly reasonable decision to use a monthly method for comparative purposes, since what you are trying to do here is compare one QC method with another. (Also, as far as I know, there are no fully automated daily methods in existence).

Response 8: We will revise the wording of this paragraph.



Section 4.1 – it's probably also worth noting somewhere that TX correlations are much stronger than TN correlations in the valleys (presumably because TN is much more strongly influenced by local topography). Another question which may be worth considering is the extent to which there might be seasonal influences on correlations in a tropical climate – experience from other parts of the world suggest that TN correlation length scales in tropical climates are much shorter in the rainy season than in the dry season (perhaps because in tropical climates in the wet season, low temperatures sometimes occur during rain events).

Response 9: We will include some sentences on the differences between the correlation of TX and TN within the different regions. We agree with the referee that there are likely differences between the correlations detected in the different seasons. However, since the present paper investigates the differences resulting from different QC approaches and not mainly the climatology in the Central Andean area, we consider the seasonal scale of correlations (as well as of trends) an aspect of minor importance. Regarding the length and the large amount of information of the present paper, we prefer to not include analyses on seasonal time scale.

P13 line 19 (and onwards) – another possible effect of UDQIs is that they could also hide real inhomogeneities – either directly, or indirectly through adding noise to a time series (and thus reducing the signal-noise ratio).

Response 10: We will include this important aspect in the discussion.

Minor comments

P2 line 14 – should be 'lose significance'.

Response 11: Will be corrected.

P6 line 30 – 'Spearman' should be capitalised.

Response 12: Will be corrected.

P7 line 9 – ‘usually becomes slightly negative’ – do you mean ‘weaker’? I wouldn’t expect a day shift to reduce a strong positive correlation to below 0.

Response 13: Shifting two highly correlated time series by one day indeed often results in a slightly negative correlation of the first differences in this tropical area.

P8 line 1 – should be ‘performances’

Response 14: Will be corrected.

P10 line 30 – 10.2_C seems very large – is this actually a data quality problem large enough to trigger homogeneity checks?

Response 15: Data quality issues often trigger the statistical detection of inhomogeneities. This very large adjustment did occur in the time series shown below (Fig. 1) between the segments 1990-1994 and 1997-2007. This time series is (at least temporarily) affected by several systematic data quality issues: strong asymmetric rounding patterns (e.g. around 2000), moderate and strong irregularities in the data patterns (not clearly classifiable but highly suspicious patterns, such as strong changes in the variance, truncations, or missing temperature ranges), and frequent and large inhomogeneities. The low quality of the data suggests that the inhomogeneity rather originates from an observer error (e.g. erroneous reading of the instrument) than from other impacts (e.g. station relocation).

P11 line 23 – ‘inflation of the trend spread’ – this doesn’t make sense as the trend spread is decreasing, not increasing. I guess you could use ‘deflation’, but I think just ‘decrease’ is fine.

Response 16: Will be corrected.

P13 line 13 – should read ‘Peruvian Andes and Switzerland’.

Response 17: Will be corrected.

P15 line 5 – should read ‘trends of a few climate change indices’



Response 18: Will be corrected.

P15 line 7 – should read ‘drawing of clear conclusions’

Response 19: Will be corrected.

P15 line 16-18 – while missed observations of a few millimetres may have a negligible effect on monthly sums, they still affect indices which use number of rain days > 1 mm as a basis (e.g. SDII).

Response 20: We will reformulate the sentence. Here, we want to stress the fact that UDQI may not have the same effect in different regions. Missing observations of small PRCP events (e.g. up to 2 mm) may not affect the monthly precipitation sums in a wet region, but in a dry area these missed observations may bias monthly PRCP sums.

References – I note that the Gubler et al paper is now published.

---

Interactive comment on Clim. Past Discuss., https://doi.org/10.5194/cp-2017-64, 2017.

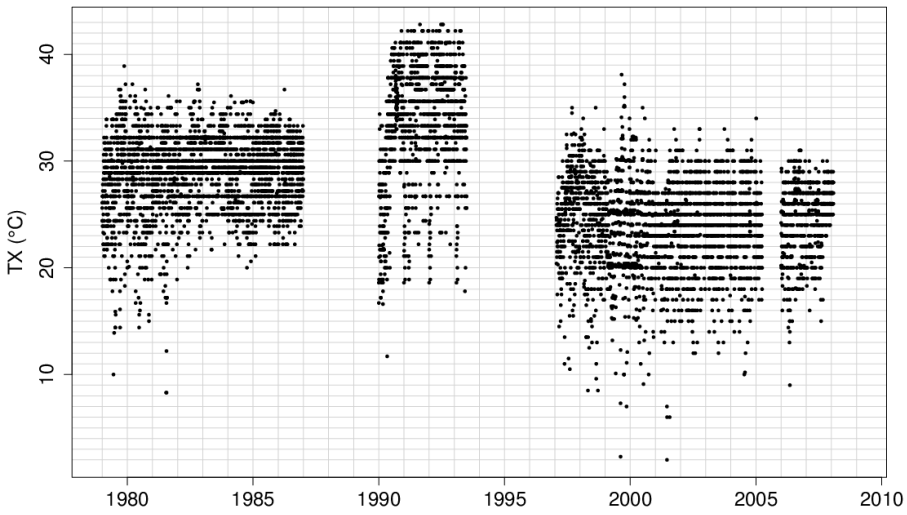


**Fig. 1.** Time series for which the largest temperature adjustment values were calculated in the data homogenization process.