Climate
of the Past

Open Access

Discussions

Interactive comment on "Expanding HadISD: quality-controlled, sub-daily station data from 1931" *by* R. J. H. Dunn et al.

Anonymous Referee #1

This paper describes an update of an already published dataset. The modifications do not justify a new paper, even though the science behind is good and solid. Perhaps the authors could try to publish in a journal specialized in climate datasets description. Some comments: 4571-6: use a more friendly name for public version control. 4572-1: users might have problems with this as they might rely on a fixed set of stations 4572-15 The selection proess may introduce stations with l5 years of data spread over the 1931-2015 period. Would a station with data only in January between 1931 and 1990 make it to the dataset? Perhaps I am missundertanding the way you dio the selection. It is not clear to me. 4573.10: is there any criteria based on the data itself? For example . . . does the merge extend a station (good) rather than filling little gaps (bad); do you check if correlation over overlapping sections if they exist? Do you check if values of both stations are compatible? 4577-8 & 18: Gaussian vs gaussian. I would recommend not

to capitalize the distribution name. Better write "gaussian distribution" than "gaussian" 4577 : distributional gap and streak test are not well explained. Streaks: does the sentence: "To allow these thresholds to be calulated dynamically, the distribution of repeated values is analysed." Imply that a series with lots of repeated values will have a higher threshold? IF so, this is not convenient. If not, please, explain better. 4580: Neighbour checks: "The closest 20 neighbours within the limits of 500m elevation and 300km distance are obtained for each station. For each of these neighbours, the data overlap with the target is calculated. Also, the correlation between the neighbour and target is obtained after removing the annual and diurnal cycles." : even understanding the need to find neighbours, I wonder if 500 km is not a very large distance, specially at the resolution of this dataset. I am wondering if by removing daiy and anual cycles you are not more likely to correlate stations with very different climates.