Review ESSD-2017-58

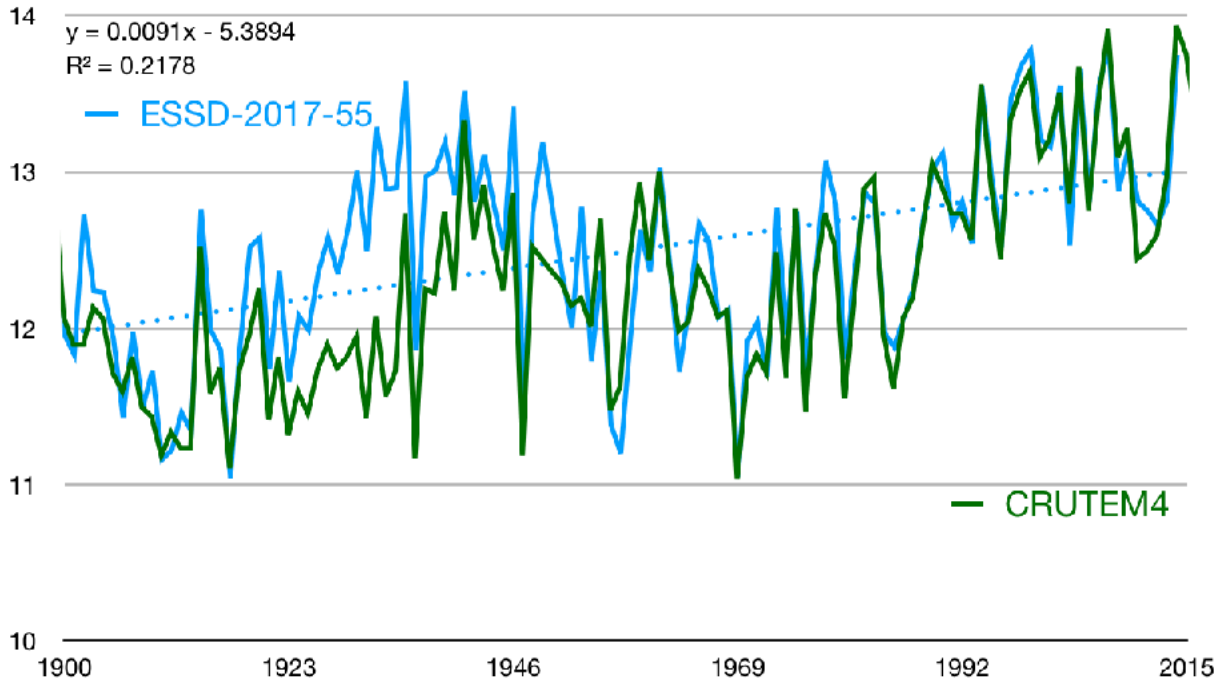
Surface Air Temperature Qingdao

Important and positive to see these types of data sets emerge, both for the location and the century-long time scales.

Easy access to a very clean .txt file. Easy to reproduce Figure 6. A .csv file might represent a more familiar format to many users?

For comparison and reference purposes we should have the relevant WMO station number? If not Qingdao then nearest reliable station? Or perhaps CMA has its own station numbers and perhaps reference stations? If so, helpful to know how this data fits within the CMA system?

The descriptive manuscript seems to lack information on validation and uncertainties? For validation I looked at CRUTEM4 (for access, start from https://doi.org/10.5194/essd-6-61-2014). I plotted the CRUTEM4 gridbox anomaly values for Qingdao (derived from multiple stations within that grid box) with these ESSD-2017-55 data (see below).



One sees, not surprisingly, remarkably good correspondence, no doubt because for most time periods CRUTEM4 and this data set access identical source data. But the time period 1920 to 1940 suggests some discrepancies. Either CRUTEM4 seems low or Willmott and Matsuura seem high? Do the authors have an explanation? Do the authors have sufficient confidence in the quality of this data to suggest a revision to CRUTEM4? Could this particular station differ substantially from the grid-box average over those two decades?

In this data set we find, monthly and annually, precise temperature values with no hints of uncertainty. But uncertainties must have arisen in, at minimum: a) the original measurements at all time periods; b) the digitisation process (presumably involving optical character recognition) of the German charts; c) the Willmott and Matsuura interpolation and gridding processes; d) the more recent (and, one presumes, more accurate) CMA data processing; and e) the quality control and homogenisation processes described here. The authors should at least compile that information from the cited references as a guide and perhaps a caution to users of this time series? One would like to see error bars or coloured uncertainty ranges (presumably decreasing with time) on, for example, Figure 6. In my re-plot of data for Figure 6 I included a (not very high) correlation coefficient. The authors could do likewise and explain for readers and users the

various reasons (natural variability plus measurement uncertainty) for the values they expect and achieve.

One could exploit this data for additional information? Perhaps a full analysis belongs in a separate science paper but, because of thrice daily data from 1899 to 1905 and hourly data from 1905 to 1914, the authors could at least hint or promote the possibilities of comparing early 20th century with present day daily temperature ranges. They could also look at differential warming, nocturnal vs diurnal. The authors mention rapid industrialisation in and around Qingdao. Do the sub-daily data then and now reflect those changes?