Earth Syst. Sci. Data Discuss., 6, C260–C262, 2013
www.earth-syst-sci-data-discuss.net/6/C260/2013/

Open Access Earth System Science Data Discussions

Interactive comment on "Description of the ERA-CLIM historical upper-air data" *by* A. Stickler et al.

Anonymous Referee #2

Received and published: 28 December 2013

General Comments

The authors nicely cover the data development process for an ERA-CLM effort, including cataloging what data is available, imaging the hard copy, and finally digitizing the data – primarily focused on historical upper air measurements to supplement, in early times, the existing collections. The project output is well-organized, documented, and accessible from PANGAEA. I am please to read that the over 600 datasets, and more to come, have a consolidation path into the CHUAN archive.

The web-based metadatabase, publicly available, is a great project coordination tool an information source for the future. It appears this was a key element for success.

This is the right sort of paper to include a detailed technical description of the imaging

process (at BERN, and other contributing centers of expertise). It could be helpful to others who might be considering such a task (p. 650). Also the discussion of the digitization process and the successful and not so successful usage of OCR technology was interesting. Truly highlights the depth of this effort.

Using a surface only reanalysis QC check scheme is a new and likely fruitful approach, as should be the feedback from future 3-D reanalysis. The limited application to temperature was a little disappointing, and the outlier criterion was huge. I would have like to see a more significant evaluation of this effort.

The approaches used to QC the data at RIHMI were quite inventive and well described, highlighting the many challenges, e.g. time zone conversions

Specific Comments and Technical corrections.

Line 19, page 642: what is the next version number, so people know what to look for? OK – I see it in Section 5, it will be V2.0.

Line 14,page 643: NETCDF => netCDF

should 'station days' be 'station-days' ? page 645: Strickler et al. 2013a, is not available to see, restructure sentence to not refer to Figures

L 20, p 645: possibly 'proper' => home

L 23, p 645: possibly 'research' => search

L 26, p 645: '9' to nine

L 20, p 647: 'atmospheric transmission data' – this doesn't seem like a common source. Are there meteorological data included? I see some are from mountain top observatories.

L 9, p 648: 'used instruments' => instruments used

L20, p 648: 'and supervise' = 'and to supervise'

Page 646: turn a-i item list into a table – easier and more clearly referenced later, e.g. p. 649.

L 23, p 652: It is worth knowing the conversion constants used. Numerical accuracy, real number truncation, and rounding can have an influence further down when research is done.

3.3 RIHMI – this section is a bit confusing. The first paragraph mentions 'numerous errors in the OCR processing'. Then most of the remaining sections highlight how much it was used.

L 1, p 655: "4" => four: In general all number less than 10 are written in text, e.g. one, two, three,. . .... nine.

Line 12, page 642, line 17, page 644, line 21-22 page 647: mio=million? Probably not an SI unit standard.

Line 14, page 642: focussing on so far less, maybe better => so far focusing on less; similar construct at line 20, page 644

'(http://www.surfacetemperatures.org)' => (ISTI, http://www.surfacetemperatures.org)

Insert 'CHUAN' in line 29, page 644

Sect. => Section on page 645, and throughout the document.

Interactive comment on Earth Syst. Sci. Data Discuss., 6, 641, 2013.