

## ***Interactive comment on “HadISDH: an updated land surface specific humidity product for climate monitoring” by K. M. Willett et al.***

**Anonymous Referee #2**

Received and published: 2 January 2013

In this manuscript the authors described an updated monthly data set of surface specific humidity over global land from 1973 onwards, with a focus on estimates of the uncertainties associated with sampling, homogenization and measurements. It has a very detailed description of the uncertainty estimates and made some comparison with other similar products in terms of recent humidity trends. The analysis was done carefully and the uncertainty estimates are helpful. I think this will become a very useful dataset for the climate community, and this paper provides a useful document for this data set. I recommend publication of this manuscript with some minor revisions.

Specific comments:

1. Fig. 10a shows small differences between this and other similar products, including some without homogenization (e.g., Dai 2006). Does this mean the surface  $q$  data  
C3014

are relatively homogeneous and the adjustments made in this analysis have only small impacts? It would be nice if you can compare global and regional time series of  $q$  from the cases with and without the homogenization, and make some comments regarding the impact of the adjustments on regional and global  $q$  trends.

2. Fig. 10b-e show that the uncertainty associated with the incomplete area-coverage overwhelmingly dominates over the other uncertainties that are a major focus of this study (sections 4.1 and 4.2). I understand that these uncertainty estimates are useful at grid box levels, but they seems to be less important for regional and global trend analysis? If that's true, you may want to point out that in the Abstract and Summary.

3. Also, Fig.10b-e seems to suggest that the area-coverage are fairly constant since 1973. Is that true? A few maps showing the number of stations with each grid box with monthly  $q$  data for select years (e.g., 1973, 1990, 2011) would be helpful.

4. It appears that the station data won't be part of the released data set (p. 14, lines 24-28). Thus, the community will have to rely on the gridded data. The gridding used in this analysis appears to be very simple: a simple average of the station data within a 5deg x 5deg box without searching for nearby stations when few or no stations exist inside the box. This may not be the best approach as monthly surface  $q$  (like monthly  $T$ ) has a spatial correlation distance of  $\sim 1000$ km. It would be nice if the station data are included in the released data set, so people can grid the data using other approaches. For many applications, station data (not the gridded data) will be needed.

5. p. 6, lines 24-30: adjustments did not account for seasonal differences in inhomogeneities, and thus this homogenized monthly data set may be used best for annual trend analysis, may not be suitable for seasonal trend analysis despite of its monthly resolution. People will likely use it for seasonal change analysis. Should the authors warn such applications in advance? Of course, if the authors do not find major impacts of the adjustments from homogenization, then maybe the data are ok for seasonal trend analysis too?

6. p. 3, lines 37-44: Why can't you define  $e$  always with respect to water? Isn't  $e$  in eq. (3) with respect to water?

7. p. 4, top paragr.: Why don't you surface pressure ( $P_s$ ), which is available at most stations?

8. p. 12, lines 21-23: Could you provide more details on how you derive the certainties for the regionally averaged  $q$  values from the uncertainties at each grid box?

9. It is unclear whether gridboxes without observations for individual months will have values (of climatological data), like in CRU TS data sets. It is best to assign a missing data code to these boxes because many users will take the climatological values as real observations.

---

Interactive comment on Clim. Past Discuss., 8, 5133, 2012.