

# 1 Data Assimilation Papers

- **Dee et al. (2011):** <http://onlinelibrary.wiley.com/doi/10.1002/qj.828/abstract>

Describes in detail the data assimilation procedure for the ERA-Interim dataset, which begins in January 1989 and proceeds forward in time. In Section 4, they describe all of the input datasets and then the quality control checks that they use. These are mainly described in very broad terms. In Section 4.3.2, I found the following statement: “Figure 13 shows the combined bias corrections applied to reported temperature observations at 50 hPa from radiosondes launched at Bethel, Alaska, at 0000 and at 1200 UTC. The large adjustments in 1989 and 1995 are associated with changes in instrumentation.” So, now we know what happened in 1989 and 1995 that caused the known shifts in that series.

However, this paper should be read more carefully. I haven’t yet found answers to my two big questions, which are (1) what kind of homogenization methods are they using, and (2) do they apply quality control algorithms before homogenization or after? The data assimilation community doesn’t seem to use this word “homogenization” as much. It is only mentioned on page 586 with regard to improving the homogeneity of the ozone assimilation by applying variational bias adjustments, which I believe is described in the next paper (Dee and Uppala 2009).

**Josh Comments:** They don’t seem to discuss the specifics of the quality control algorithms they apply to radiosonde temperatures, likely because they have so many different data inputs and QA processes! They perform “a suite of quality control and data selection steps... (such as) checks for completeness of records, physical feasibility, ...”. Section 4 lays out the general approach to the QC process and then discusses each data input separately. For radiosonde temperature records, they use “the RAOBCORE\_T\_1.3 set of adjustments” from: Haimberger L, Tavolato C, Sperka S. 2008. Toward elimination of the warm bias in historic radiosonde temperature records: Some new results from a comprehensive intercomparison of upper-air data. *J. Climate* 21: 4587-4606. In the Haimberger paper, they detect breakpoints using the SNHT. To me, it sounds like they accept the quality control methods used on the input data without performing further checks. However, during the reanalysis, they mention that they do try to adjust observations that are considered incorrect by comparing them to model forecasts.

- **Dee and Uppala (2009):** <http://onlinelibrary.wiley.com/doi/10.1002/qj.493/abstract>

This paper doesn’t really talk about quality control methods or how outliers may affect their methodology. Instead, they simply present a “variational bias correction” method for satellite data that is designed to detect biases. Biases are when the satellite data is being used to estimate surface temperature, and they systematically either over or underestimate temperature. They also talk about needing overlapping series of concurrent observations in order to calibrate the satellite data appropriately. They say that, “The ERA-Interim data assimilation system is described in more detail by Simmons et al. (2007a, b) and Uppala et al. (2008),” so there may be more there on the interplay between outliers and biases (or shifts in the mean measurements).

In the conclusion, it states, “A key challenge in reanalysis is to properly manage changes in the observing system. This involves difficult technical aspects such as detecting and initializing new

data streams, keeping track of a large variety of instrument types, handling data gaps, monitoring data quality, and correcting data biases. Production of a multi-decadal reanalysis encompassing the modern satellite era requires that most of these tasks are performed automatically.” This supports our argument that making such routines automatic is also important in data assimilation.

- **Tavolato and Isaksen (2014):** <http://onlinelibrary.wiley.com/doi/10.1002/qj.2440/abstract>

This paper mainly deals with QC for outliers and not homogenization. The “innovations” that they talk about are the difference between observational data and some reference dataset. In the conclusion, they say, “In data assimilation, innovations are used extensively for observation QC and provide generally very valuable information (Hollingsworth et al., 1986). But the background forecasts also have errors, sometimes very large ones, so it may be difficult to determine if an observation or the equivalent model value is the outlier.” They describe a robust approach to identifying outliers. This paper was just published in 2014, so we didn’t see it before Ashley’s paper was submitted, but they use the Huber mean, and they investigate in particular the weight that is given to outlying observations in the analysis. They do talk a little bit in Section 4 about stations that have a bias, and by this, I assume that they mean that the mean of the station is higher or lower than it should be.

They also state that, “Any observational study that involves the analysis of historical data records relies on quality control and bias adjustments based on uncertainty assessments of the input data.”

## 2 Early Upper Air Observation Overviews

Information on the UADB at NCAR that we are using can be found at <http://rda.ucar.edu/datasets/ds370.1/>. While there is not a peer-reviewed paper, there is a citation on the website (DSS/CISL/NCAR 2014), and the following information:

“The first consolidation step was compositing stations with identical WMO station numbers across all data sources. The UADB-TRH product has nearly 1100 stations that report for 20 years or more, and over 400 stations have time series of 50 years or longer. The UADB-Wind product has about 1600 stations that report for 20 years or more, and over 500 stations have soundings for at least 50 years. Next stations with different WMO station numbers, but that are located at identical latitudes and longitudes, or nearly so (within 40 kilometers) were combined, accounting for WMO station number changes and small position relocations. The result is that even longer station time series are created; extensions occur at over 240 and 250 locations in the UADB-TRH and UADB-Wind products, respectively. The composited and combined products are both available.”

Note that UADB-TRH is the UADB temperature and humidity measurements, and UADB-Wind has the wind measurements.

The temporal range over the entire dataset is July 20, 1922 to February 28, 2015.

- **Stickler et al. (2014):** <http://www.earth-syst-sci-data.net/6/29/2014/essd-6-29-2014.html>

**Ramella Pralungo et al. (2014):** <http://www.earth-syst-sci-data.net/6/185/2014/essd-6-185-2014.html>

In both of these papers, they note that very early upper air measurements have only been made partially available. They define historical measurements made as those pre-1957. Beginning in 1958, the European Re-Analysis (ERA-40) is the primary data source. They say that in 1948, the global upper air coverage reached its current status, but at that time, many of the observations had not been digitized. The CHUAN (Comprehensive Historical Upper-Air Network) dataset contains a large set of pre-1957 upper air data. In Stickler et al. (2014), they catalogue, image, and digitize many more observations that occur in the less well covered regions of the tropics, polar regions, oceans, and in very early upper air data from the US and Europe. In Ramella Pralungo et al. (2014), CHAUN and the IGRA (Integrated Global Radiosonde Archive) are compared and contrasted in terms of when measurements started, and they are merged with the measurements at altitude levels mapped to standard pressure levels.

### 3 Homogenization Method Papers

- **Freitas et al. (2013):** <http://www.met.hu/en/ismeret-tar/kiadvanyok/idojaras/index.php?id=82>

This is a case study that uses both MASH and HOMER. It states that “The former [MASH] was selected for being one of the most widely used by the homogenization community, while the latter [HOMER] was selected because it is one of the most recent homogenization methods ...”

It says that “HOME main objective was to develop a general homogenization method for homogenizing climate and environmental datasets which was accomplished in 2011 with the release of a free software package (HOMER), implemented in R language (HOME, 2011). It should be noted that ACMANT is a modified and automated version of PRODIGE, and that HOMER integrates PRODIGE, ACMANT, and USHCN.”

References for all of the original papers where these methods were introduced are given in this paper.

**PRODIGE** (Caussinus and Mestre 2004),

**MASH** (Szentimerey 2007)

**ACMANT** (Domonkos 2011)

**USHCN** (Menne and Williams 2009)

**Josh Comments** Interesting comment: “with MASH it is obligatory to use available functionalities to fill the missing values and perform automatic correction of outliers. On the other hand, HOMER provides a fast quality control of the data, which includes functions of the CLIMATOL R package (Guijarro, 2011), which allow the user to perform/estimate station density, correlogram, histograms, boxplots, and cluster analysis.” The outlier detection method they used was the “1.5 times IQR” approach as well as “pairwise comparison ... by visual inspection.” Additionally, “MASH has an automatic procedure to detect and correct outliers which is not controlled by the user.”

In the end, they find that there is no clear evidence to choose one over the other. However, the study centers around homogenizing actual time series without simulating data, and so true evaluation of these methods seems lacking.

- **Lindau and Venema (2013):** <http://www.met.hu/en/ismeret-tar/kiadvanyok/idojaras/index.php?id=82>

Not sure how helpful this one will be, but this is one of the only ones that I ran across that specifically addresses the multiple break-point problem. Still, the introduction is good, and there is a section on “dynamic programming,” which was mentioned in one of the referee reports.

**Josh Comments:** I got lost in section 4. In equation 7, they seem to be claiming that averaging standard normal random data over many permutations will provide a particular solution (rather than a distribution). They later use this idea to justify that  $v$  is bounded between 0 and 1, and this argument motivates most of the remainder of the work.

- **Caussinus and Mestre (2004)** <http://onlinelibrary.wiley.com/doi/10.1111/j.1467-9876.2004.05155.x/abstract>

I like this paper because it’s one of the only ones that I’ve seen in a statistics journal with the application of homogenization in long series of temperature data. They say that two of the limitations of the existing homogenization methodologies, which we have also noted, are:

“(a) They rest on the existence of a so-called reference series whose reliability cannot be proved. The different methods for creating such series (Alexandersson, 1986; Førland and Hanssen-Bauer, 1994; Easterling and Peterson, 1993) do not guarantee their quality.”

“(b) The number of breaks in the series tested is many and unknown. Moreover, an unknown number of outliers may spoil the data.”

Thus, they are also mentioning the fact that having a good “reference series” is not always possible to have and that having outliers in the data can interfere with the homogenization procedure. This will be a good one to read through to get their take on the reference series issue and as support for our message of homogenization in the presence of outliers.

**Josh Comments:** This paper also mentions outlier detection in addition to changepoint detection. Their approach is a bit different from ours: “But, if a detected changepoint remains constant throughout the set of comparisons of a candidate station with its neighbours, it can be attributed to this candidate station. The detection of the outliers follows the same principle.” Detection of changepoints and outliers is done by a penalized maximum likelihood approach, where the penalty increases with the number of changepoints. However, given the computational infeasibility of maximizing such a likelihood, they use reference series and metadata to first determine candidates or possible changepoints/outliers prior to optimizing. It’s also unclear to me how the outliers are determined in this maximization process...

- **Hannart et al. (2014):** <http://onlinelibrary.wiley.com/doi/10.1002/joc.3925/abstract>

These authors are all good, well-known statisticians, and the paper is recent, so it’s worth taking a look at to see how they validate and/or simulate data and results.

**Josh Comments** In the introduction, they provide a good statistical description of homogenization that I found quite interesting.

Detection of break points is often done by differencing neighboring series so as to remove the climate signal. However, detected breaks must then be assigned to "culprit" series and this is often a manual task (although Claude and Menne published a paper on an automated approach). This paper seeks to improve this approach of allocation of the "culprit."

The homogenization method they use is described in another paper: Hannart A, Naveau P. 2009. Bayesian multiple change-points and segmentation: application to homogenization of climatic series. *Water Resour. Res.* 45(10): 1944-1973. My hunch is that it'd be too slow for daily data (given that it's a Bayesian method), but it's worth investigating.

The pre-processing they performed was "Series were quality checked based on the methodology of Vincent et al. (2005)." Given the name "pre-processing", this was certainly done prior to checking for changepoints. The details of the paper are: Observed trends in indices of daily temperature extremes in South America 1960-2000. *J. Clim.* 18(23): 5011-5023

## 4 Daily Homogenization

- **Szentimrey (2013)** : <http://www.met.hu/en/ismeret-tar/kiadvanyok/idojaras/index.php?id=82>

"... homogenization of daily data is still in its infancy and is much more difficult problem than homogenization at monthly or annual scales."

Interesting comment: "There is a common assumption that the correction in mean is sufficient for monthly and annual series, and that the correction of higher order moments is necessary only in the case of daily data series. In general, it is tacitly assumed that the averaging is capable to filter out the inhomogeneities in the higher order moments. However, this assumption is false, for example, if there is an inhomogeneity in the standard deviation of daily data, we may have the same inhomogeneity in monthly data."

**Josh Comments** In their paper, they use a simple example: if two variables  $Y_1, Y_2$  are correlated and standard normal, then  $E(Y_1 - Y_2 | Y_1 \text{ is negative})$  is less than  $E(Y_1 - Y_2)$ . They use this to argue that larger temperature differences (between reference stations) at extreme values are more likely. However, this argument seems problematic to me; I would think that  $Y_1$  and  $Y_2$  are correlated through some latent variable  $Z$ . Additionally, if we see larger values of  $|Y_1 - Y_2|$  at extreme points, then I would think that implies the distribution is changing when  $Z$  is low.

In general, they are very critical of the current homogenization algorithms. They make a valid point regarding the regression approach for adjusting a time-series. It's not clear to me if the method they are analyzing is exactly what is used in HOM, but it does show that the adjustment will not preserve even the second order moment.

- **Toreti et al. (2010)**: <http://journals.ametsoc.org/doi/abs/10.1175/2010JCLI3499.1>

This is probably one of the "Italian" papers on daily homogenization. It extends the HOM method to address the autocorrelation in daily temperatures and to avoid a "subjective choice of regression parameters." They show that in small samples, their HOMAD method applied to daily temperature data improves the correction.

**Josh Comments** The HOM method (and possibly extensions like HOMAD) seem to be very popular in the climate literature. However, I think they are too complex for our study because they require comparison to a reference series. To do this, we’d have to change our simulation approach to simulate multiple time series.

I took a look at the code posted online for HOMER and HOM/SPLIDHOM. It is freely available, and there is a decent amount of documentation, but it seems like it also assumes a very specific file type to read in data. I’m not entirely sure how much work it would take to figure out how to use it, but it may be tricky to figure out what assumptions are made about data structure/output/... I also emailed Toreti to see if he can provide us with the code he used for HOMAD (it’s mentioned in the paper that the code is available in an R package from the first author).

- **Mestre et al. (2011):** <http://journals.ametsoc.org/doi/abs/10.1175/2011JAMC2641.1>

These authors say that daily temperatures are needed to study features such as frequency and intensity of extreme events. They introduce a new method called SPLIDHOM for homogenization of daily data that adjusts not only the mean but also the higher order moments of the time series.

**Josh Comments** I thought that HOM also corrects higher order moments (in Szentimrey, 2013 it states “higher order moments (HOM) method by ...”). My understanding of this algorithm is that it is an alternative algorithm to HOM. In particular, the authors state “This is a significant difference from HOM, in which the LOESS smoothing parameter is fitted empirically, as stated by the authors themselves.” It’s also based on comparison to a reference series...

## References

- Caussinus, H. and Mestre, O. (2004) “Detection and correction of artificial shifts in climate series,” *Journal of the Royal Statistical Society: Series C (Applied Statistics)*, 53: 405–425.
- Data Support Section/Computational and Information Systems Laboratory/National Center for Atmospheric Research/University Corporation for Atmospheric Research. (2014, updated monthly) “NCAR Upper Air Database, 1920-ongoing,” *Research Data Archive at the National Center for Atmospheric Research, Computational and Information Systems Laboratory*. <http://rda.ucar.edu/datasets/ds370.1/>. Accessed Jul 03, 2015.
- Dee, D. P., et al. (2011) “The ERA-Interim reanalysis: Configuration and performance of the data assimilation system,” *Quarterly Journal of the Royal Meteorological Society*, 137: 553–587.
- Dee, D. P. and Uppala, S. (2009) “Variational bias correction of satellite radiance data in the ERA-Interim reanalysis,” *Quarterly Journal of the Royal Meteorological Society*, 135: 1830–1841.
- Domonkos, P. (2011) “Adapted Caussinus-Mestre algorithm for networks of temperature series (AC-MANT),” *International Journal of Geosciences* 2: 293–309.
- Freitas, L., Gonzalez Pereira, M., Caramelo, L., Mendes, M. and Filipe Nunes, L. (2013) “Homogeneity of monthly air temperature in Portugal with HOMER and MASH,” *Quarterly Journal of the Hungarian Meteorological Service*, 117: 69–90.
- Hannart, A., Mestre, O., and Naveau, P. (2014) “An automatized homogenization procedure via pairwise comparisons with application to Argentinean temperature series,” *International Journal of Climatology*, 34: 3528–3545.

- Lindau, R. and Venema, V. (2013) “On the multiple breakpoint problem and the number of significant breaks in homogenization of climate records,” *Quarterly Journal of the Hungarian Meteorological Service*, 117: 1–34.
- Menne, M. J. and Williams Jr., C. N. (2009) “Homogenization of temperature series via pairwise comparisons,” *Journal of Climate*, 22: 1700–1717.
- Mestre, O., Gruber, C., Prieur, C., Caussinus, H., and Jourdain, S. (2011) “SPLIDHOM: A method for homogenization of daily temperature observations,” *Journal of Applied Meteorology and Climatology*, 50: 2343–2358.
- Ramella Pralungo, L., Haimberger, L., Stickler, A., and Brönnimann, S. (2014) “A global radiosonde and tracked balloon archive on 16 pressure levels (GRASP) back to 1905 Part 1: Merging and interpolation to 00:00 and 12:00 GMT,” *Earth System Science Data*, 6: 185–200.
- Stickler, A., et al. (2014) “Description of the ERA-CLIM historical upper-air data,” *Earth System Science Data*, 6: 29–48.
- Szentimrey, T. (2007) “Manual of homogenization software MASHv3.02,” Hungarian Meteorological Service.
- Szentimrey, T. (2013) “Theoretical questions of daily data homogenization,” *Quarterly Journal of the Hungarian Meteorological Service*, 117: 113–122.
- Tavolato, C. and Isaksen, L. (2014) “On the use of a Huber norm for observation quality control in the ECMWF 4D-Var,” *Quarterly Journal of the Royal Meteorological Society, Early View Online*, DOI: 10.1002/qj.2440.
- Toreti, A., Kuglitsch, F. G., Xoplaki, E., Luterbacher, J., and Wanner, H. (2010) “A novel method for the homogenization of daily temperature series and its relevance for climate change analysis,” *Journal of Climate*, 23: 5325–5331.