The Comprehensive Historical Upper Air Network (CHUAN)

A. Stickler[1], A. N. Grant[1], T. Ewen[1], T. F. Ross[2], R. S. Vose[2], J. Comeaux[3], P. Bessemoulin[4], K. Jylhä[5], W. K. Adam[6], P. Jeannet[7], A. Nagurny[8], A. M. Sterin[9], R. Allan[10], G. P. Compo[11,12], T. Griesser[1], and S. Brönnimann[1]

Received:

[1]Institute for Atmospheric and Climate Science IACETH, Zurich, Switzerland

[2]NOAA National Climatic Data Center, Asheville, NC, USA

[3]National Center for Atmospheric Research, Boulder, CO, USA

[4]Météo-France, Toulouse, France

[5]Finnish Meteorological Institute, Helsinki, Finland

[6]Deutscher Wetterdienst, Meteorological Observatory Lindenberg, Germany

[7]MeteoSwiss Aerological Station, Payerne, Switzerland

[8]Arctic and Antarctic Research Institute, St. Petersburg, Russia

[9]RIHMI-WDC, Obninsk, Russia

[10]Met Office Hadley Centre, Exeter, United Kingdom

[11]University of Colorado CIRES Climate Diagnostics Center, Boulder, CO, USA

[12]NOAA Earth System Research Laboratory/Physical Sciences Division, Boulder, CO, USA

Corresponding author:
Dr. Alexander Stickler, Institute for Atmospheric and Climate Science, Universitätstrasse 16, ETH Zurich, CHN, CH-8092 Zurich, Switzerland

E-Mail: alexander.stickler@env.ethz.ch

Abstract

In order to better understand variability in weather and climate, it is vital to address past atmospheric circulation. This need requires meteorological information from not just the surface, but at upper levels. Current global upper-level datasets reach back to the 1940s or 1950s only and do not cover some important periods in the first half of the 20th century. Extending the observational record is therefore considered important to analyze climate variability in the past, and verify global climate models used to predict future climate change. Though earlier upper air data from platforms such as radiosonde, aircraft, pilot balloon, registering balloon, and kites are available from various sources, no systematic compilation and quality assessment of upper-level data prior to the International Geophysical Year (1957/58) has ever been performed. Here we present the Comprehensive Historical Upper Air Network (CHUAN). It is a consistent global historical upper air dataset that has been derived from the heterogeneous data available from various sources as well as newly digitized data. The paper describes the CHUAN dataset, the metadata, the quality control procedures, and the relationship to existing datasets. Some examples are given of its usefulness for analyzing weather and climate during the first half of the 20th century. The CHUAN dataset comprises 3987 station records worldwide or about 16.4 million profiles (thereof 12.6 million before 1958 and 5.3 million, primarily from pilot balloons, before 1948). A monthly mean version can be downloaded from the World Wide Web (http://www.historicalupperair.org).

Capsule sentence:

We present a systematic compilation of global upper air data from the first half of the 20th century for weather and climate applications and for possible use in future reanalyses.

To improve our understanding of global weather and climate variability and its change under the influence of global warming, it is vital to extend our knowledge about the atmospheric state and variability in the past. Current reanalysis datasets (NCEP-NCAR 50-Year Reanalysis, Kistler et al. 2001; ERA-40 reanalysis, Uppala et al. 2005) provide detailed information on the atmosphere during the past 60 years. The first half of the 20th century, however, featuring some very prominent climate fluctuations such as the Arctic Warming from the 1920s to the 1940s (Polyakov et al. 2003, Overland et al. 2004, Grant et al. 2009b), the Midwest Dust Bowl droughts in the 1930s (Schubert et al. 2004a, Brönnimann et al. 2009a, Cook et al. 2009), the 1939-42 El Niño (Allan et al. 2003; Brönnimann et al. 2004a,b) and numerous weather extremes, is not covered by these datasets. Also, the roughly 60 years covered by the above-mentioned reanalyses have proven to be too short to sample the whole range of global atmospheric variability, a shortcoming particularly important when it comes to assessing extremes.

During the past few years a number of efforts have been focusing on extending atmospheric datasets back to the early 20th or even 19th century, including a reanalysis using only synoptic surface pressure and monthly sea surface temperature and sea ice data ("The Twentieth Century Reanalysis Project"[1]). The Atmospheric Circulation Reconstructions over the Earth (ACRE) initiative[2] serves as a platform for these activities. In this paper we describe another activity that is complementary to the "Twentieth Century Reanalysis", namely a

[1] See http://www.esrl.noaa.gov/psd/data/20thC_Rean/.

[2] See http://www.met-acre.org/.

compilation of a comprehensive, global historical upper air dataset for use in weather and climate applications (analysis, data and model evaluation) and possible future reanalysis efforts.

The reason why current "full" reanalyses (i.e. those including upper air data) do not reach further back than 1957 or 1948 is the general lack of quality controlled, electronically available upper air data prior to the International Geophysical Year (IGY) 1957/58. To overcome this deficiency, a large number of historical upper air datasets from different measurement platforms such as kites, registering balloons, aircraft, pilot balloons and radiosondes have been digitized (e.g. in the framework of the Climate Data Modernization Program[3], at Météo-France, or at ETH Zurich (Brönnimann 2003a,b; Ewen et al. 2008a)) or compiled from existing archives over the past few years. Before 1948, a large fraction of the data stems from pilot balloon observations. The full dataset comprises, amongst other parameters, wind direction and speed, pressure, geopotential height (GPH) and temperature. However, it comes in various formats and qualities, and it is given on partly non-uniform pressure as well as geometrical altitude levels. Furthermore, up to now there has not been a serious attempt to systematically compile all the data and to rigorously assess its quality.

Here we describe a new dataset termed the Comprehensive Historical Upper Air Network (CHUAN), which has been developed as a systematic compilation of the different sources in only three standard formats: one for the data that come on pressure levels from all platforms except radiosondes, one for the data on geometrical altitude levels, and one with 4 additional pressure levels for the radiosonde data. Additionally, we demonstrate the usefulness

[3] See e.g. NOAA/National Climate Data Center (NCDC) Climate Database Modernization Program (CDMP) at http://www.ncdc.noaa.gov/oa/climate/cdmp/cdmp.html and the International Environmental Data Rescue Organization at http://www.iedro.org.

of the new dataset by showing selected examples of 20th century climate events that have been analyzed using the data. The full station record documentation including station name, location/elevation, time coverage, measurement platform, total number of records at the station, data source and quality are provided under www.historicalupperair.org. Detailed information on the evaluation, adjustment and quality assessment of the data can be found in the online documentation and supplement.

THE NEW HISTORICAL UPPER AIR DATASET.

Data Sources.

CHUAN has been compiled from 15 different sources. These, together with information about the measurement platforms, the altitude grid used, the measured variables and whether or not quality checks or adjustments have been applied to the respective dataset, can be found in the online documentation (www.historicalupperair.org). The largest fraction of the data (TD52/53, TD54, TD-6201, CARDS542, see also Eskridge et al. 1995) had already been digitized in earlier years, and was mostly stored on magnetic tape decks. Other parts of the data have been digitized more recently at the University of Arizona, at NOAA/National Climatic Data Center (NCDC) in the framework of the Climate Database Modernization Program (CDMP, Dupigny-Giroux et al. 2007), at the Arctic and Antarctic Research Institute (AARI, St. Petersburg, Russia) and at the Institute for Atmospheric and Climate Science (ETH Zurich, Switzerland). The historical radiosonde data were supplemented up to the present with data from the International Global Radiosonde Archive (IGRA, Durre et al. 2006, see Grant et al. 2009a for details on the merging). Note, however, that we have not systematically compiled data series starting after 1957 but merely extended the radiosonde records with corresponding IGRA data after 1957.

As can be seen from Table 1, the largest portion of the data consists of pilot balloon and radiosonde measurements. In the original data sources the pilot balloon data are given predominantly on geometrical altitude levels above mean sea level, but some observations are given on levels above ground level or on pressure levels. The radiosonde data are given solely on pressure (p) levels while the kite, aircraft and registering balloon data are either on pressure or geometrical altitude levels. The majority of the data including all radiosonde measurements and > 98% of the number of pilot balloon archives have been evaluated, with some pilot balloon, kite, aircraft and registering balloon data on p levels remaining to be checked for quality.

Detailed documentation.

Much more detailed information on each record contained in CHUAN can be found in the detailed station list on www.historicalupperair.org. Besides station number, station name, source dataset, geographical position and elevation, that listing shows the station numbers in the original dataset and, if identifiable, in the WMO and/or Weather-Bureau-Army-Navy (WBAN) system, and whether the record has been merged with other records. The latter information is also given more precisely for the radiosonde data described in Grant et al. (2009a). Records have only been merged whenever identical WMO indices and/or WBAN numbers[4] could be identified and the measurement platform was the same. Nevertheless,

[4] As listed in ftp://ftp.wmo.int/wmo-ddbs/Pub9volA090622.flatfile and ftp://ftp.ncdc.noaa.gov/pub/data/inventories/WBAN.TXT.

potential further merging partners in the surroundings of each station, identified as great circle distances < 25 km calculated from the geographical coordinates, are also listed. Additionally, the covered period, the total number of ascents (number of covered months in the case of the *Monthly Weather Review* datasets, i.e. monthly mean data extracted from the correspondent publication), the measurement platform(s) and information on the evaluation or adjustment processes are included.

Reformatting of the data.

CHUAN data have been reformatted into three simple ASCII formats, one each for pressure level and geometrical altitude level data, and one with 4 additional pressure levels for the radiosonde data. The pressure level data have been incorporated without any interpolation and are given on 1000, 950, 925, 900, 850, 800, 750, 700, 650, 600, 550, 500, 450, 400, 350, 300, 250, 200, 175, 150, 125, 100, and 50 hPa surfaces. For the radiosonde data, additional levels are available at 70, 30, 20 and 10 hPa.

The data on geometrical altitude levels have been mapped onto a common, fixed altitude grid relative to the mean sea level (MSL) with a relatively high vertical resolution: 0 m above ground level (AGL), 150, 250, 300, 500, 750, 1000, 1250, …, 15750, 16000, 16250, 16500, 17000, 18000, …, 30000 m above MSL. This ensures that none of the datasets experiences a loss in vertical resolution. However, this also implies that the wind and temperature data originally given on altitude levels AGL (so-called “daily” format of the NOAA CDMP data, some records of TD52/53) had to be linearly interpolated onto the new levels (which is indicated by flags). In general, interpolation was done only above 2000 m AGL to prevent nonlinear effects from the atmospheric boundary layer (ABL) or strong gradients to be distorted. In the case of the CDMP “daily” format data however, interpolation could be safely continued down into the ABL due to the relatively high resolution of about 200 m.

Calculation of monthly means.

Monthly means of all station records have been calculated and (with the exception of the data received from AARI) are being made available to the public on the World Wide Web (http://www.historicalupperair.org). To ensure an adequate number of measurements, monthly averages of temperature, pressure, or GPH have only been calculated if (for a given level and variable) at least 13 days had data or if there were no gaps of more than 7 consecutive days. In one case (early US data from the *Monthly Weather Review*, see detailed online documentation) ascent statistics were used to flag monthly means with likely too little information. For the pilot balloon wind data a minimum of 20 days per month with measurements is required because the visual tracking of the balloons prior to the mid-1940s might introduce a systematic sampling bias (whether or not a balloon is visible depends on the weather). The radiosonde data have been supplemented with IGRA data where available (number 3286-4181) after the end of the primary source record.

Global distribution of measurements.

Figure 1 displays the global distribution of the CHUAN measurement stations labeled according to the measurement platform. CHUAN incorporates a total of 3987 station records (note that this is not the same as the number of stations since different platforms are treated separately and because in some cases parts of the records from identical stations could not be merged due to quality issues, but were kept separate in the dataset). The geographical coverage of the station network changes over time (see Fig. 2). In the pre-1928 period, only the US including overseas military bases, a strip from central Europe to the central Mediterranean, Japan and New Zealand are covered by observations. In the subsequent decade, upper air records from Alaska, Scandinavia, the Eurasian Arctic coasts and Korea are added. Additionally, two records appear in South America. The decade 1938-47 is characterized by a strong increase in the number of records (see also Fig. 3) and a much better global station density. Large parts of the Northern Hemisphere land masses are now covered with the exception of the Sahara, central Greenland and parts of Siberia. The increased density of the measurement network is especially noticeable in the Tropics and the Southern Hemisphere, although vast areas remain uncovered (central South America, Antarctica, central and southern Africa and the southern ocean basins). During 1948-57, the trend to a better coverage continues. More records are now available for Mexico, the Canadian Arctic, Siberia, Africa, western Australia, the southern hemispheric ocean basins and Antarctica. One exception is South America, where coverage decreases along the Brazilian coastline. In general, the highest station density can be found in the Northern Hemisphere, especially in the US and the Caribbean, in Central and Eastern Europe as well as the Mediterranean region, and in India, China and Japan. The lowest coverage can be distinguished over the oceans, especially over the Southern Ocean, over the Amazon basin, parts of Africa, the Arabian Peninsula, Iran, Australia and central Greenland as well as Antarctica.

Evolution of available data over time.

Figure 3 shows the number of available records versus time. The total number of profiles contained in CHUAN is ≈ 16.4 million, of which about 12.6 and 5.3 million profiles are from the time before 1958 and 1948, respectively. The period covered starts on 11 August 1902, when 2 stations in Germany (Frankfurt and Friedrichshafen) launched registering balloons[5] in the framework of the "International Aerological Days", internationally coordinated one-day observational programs (Hoinka 1997). It is only in February 1905, though, that the records resume with regular upper air measurements at the observatory of Lindenberg, about 60 km southeast of Berlin[6]. In the early period until about 1918 kite, aircraft and registering balloon measurements dominate, with very few stations available. Between 1918 and 1928 pilot balloons became the most important contributor to upper air information, while kites still played an important role. During the 20s and especially the 30s, aircraft were a significant, though not dominant contributor to the data. Additionally, during the 30s the use of radiosondes became more widespread. The second half of WWII exhibits a peak in the number of available records, especially for pilot balloons, while the number of radiosonde records shows even a slight decline during the last years of WWII relative to the late 30s/early 40s. However, this decline is followed by a large increase after WWII and a concurrent decrease in the number of pilot balloon records. The total number of upper air records during the second half of WWII is reached again only in the mid 50s.

Figures 4 and 5 give an impression of the vertical distribution of the data over time, broken down into daytime and nighttime observations. Time series of the number of observations on different altitude levels above the mean sea level (ASL) are given in the right panels of the figures. For the daytime observations (Fig. 4), the general picture looks similar

[5] Registering balloons are weather balloons carrying registering instruments without being equipped with a radio transmitter.

[6] First kite and balloon ascents in Berlin were made by Aßmann from 1900 on. Starting in 1902, regular upper air measurements were performed at the observatory in Tegel. 1905 the observatory was transferred away from the city to Lindenberg due to the better meteorological as well as administrative conditions found there. The data obtained during the ascents in Tegel have not yet been digitized and processed.

than for the total number of available records over time, even though the peak in the mid 40s stands out even more. For the nighttime observations (Fig. 5) there is a clear dip around 1925 which is not seen in the daytime observations. Also, the relative number of nighttime observations is very small before about 1913. In general, the number of nighttime observations is significantly smaller than the number of daytime observations (note the different scales in both figures). However, the difference is less than about 30% from 1938 on, while it temporarily reaches over 50% before that time. Integrated over time (1902-56), the difference between the number of nighttime and daytime observations is not so much apparent: The upper parts of the vertical profiles (above 10 km ASL, see left panels) are dominated by the radiosondes which show practically no predominance of the number of daytime ascents relative to the number of nighttime ascents. In contrast, the lower parts of the profiles, dominated by pilot balloon measurements, exhibit a clear predominance of daytime ascents.

Generally speaking, the altitude reached by the different observation platforms increases with time (right panels). This is probably mainly connected to the growing importance of radiosondes but also to the continuing development of the measurement technique. It is interesting to note that during the period 1936-48 there was a decline in the mean altitude level reached, particularly for the nighttime observations. More than 85% of the observations before 1947 do not reach levels above 10 km ASL and about 50% do not reach 3.5 km ASL before that time (even more between 1936 and 1948). The majority of all ascents before 1918 reaches altitudes less than 3.5 km ASL. The respective data between 1905 and 1917 comes exclusively from the observatory of Lindenberg, Germany.

Furthermore, both figures reveal the annual cycle of the number of observations with maxima in boreal summer and minima in boreal winter, especially at higher levels. This might be explained by a combination of the dominance of the Northern Hemisphere in CHUAN and generally more clouds during the winter season which hamper the tracking of pilot balloons, the dominant measurement platform in the dataset.

EXAMPLES OF USE.

There is a wide range of possible applications for CHUAN. In fact the data have already been used for various applications, e.g. they have been used to derive and evaluate data products such as statistical reconstructions (Brönnimann and Luterbacher, 2004; Ewen et al. 2008b; Brönnimann et al., 2009a,b; Griesser et al. 2009), to evaluate and assess reanalysis datasets (Grant et al. 2009a; Gil Compo 2009, personal communication), to analyze important climate events such as the early 20th century Arctic warming (Grant et al. 2009b), the "Dust Bowl" droughts (Brönnimann et al. 2009a) and the 1939-1942 El Niño (Brönnimann et al. 2004a,b), and to examine the variability and trends of major indices of northern hemispheric circulation during the 20th century (Brönnimann et al. 2009b). In the following we briefly summarize this work and give additional examples of applications such as a case study of a severe storm.

Evaluation of reanalyses.

Grant et al. (2009a), who homogenized the radiosonde data included here and applied the physics and station history based adjustments described in the online documentation and supplement, have shown that these adjustments lead to large changes and improved internal consistency of GPH as well as T fields over large parts of Eurasia compared to the unadjusted data. They also demonstrated that the assimilation of these erroneous data into the NCEP-NCAR 50-Year Reanalysis has lead to a formerly unnoticed, widespread warm bias in the 1950s (their Fig. 9).

Using single ascent pilot balloon data from CHUAN, we find strong indications that there exist significant and spatially systematic differences between the NCEP-NCAR 50-Year Reanalysis and wind observations over the West African Monsoon region. The spatial organization of these differences, calculated with respect to observations taken inside time windows of max. 2 h around the reanalysis times, suggests a too strong and too shallow monsoon flow in boreal summer (JJA), and might point to a deficiency of the reanalysis as opposed to one of the pilot balloon measurements. As an example, Fig. 6 shows fields of the difference on the 600, 700 and 850 hPa pressure levels for this season. For the central part of sub-Saharan Africa in general, a northerly to northeasterly bias of the reanalysis relative to observations is found for MAM (not shown) and JJA with a maximum, depending on the season, between 500 and 700 hPa, and a westerly to southwesterly bias in all seasons between 850 and 925 hPa (maximum JJA). The seasonal mean difference reaches values of about 5 m s^{-1} in some regions. More details can be found in Stickler et al. 2010 (in preparation).

Analysis of the "Dust Bowl" droughts.

Brönnimann et al. (2009a) analyzed the 3-dimensional regional as well as large-scale atmospheric circulation during the "Dust Bowl" droughts in the 1930s in the US Midwest based on pilot balloon as well as aircraft data from the newly developed dataset. The study found that the Great Plains Low Level Jet, responsible for the largest part of moisture advection into the region during summer (Rasmusson 1967), was significantly weakened on its eastern flank, shallower and penetrated less far to the north during the 1930s compared to the much wetter first half of the 1940s.

It is interesting to note that model simulations up to now have failed to reproduce the exact regional structure of the precipitation anomalies, e.g. they produce a drought also in northern Mexico which was not observed (Schubert et al. 2004b; Cook et al. 2008; Seager et al. 2005, 2008). The forcings responsible for the ‘‘Dust Bowl’’ droughts, as found in model simulations, are expected to proceed through changes in atmospheric circulation (Brönnimann et al. 2009a). The wind data from the CHUAN dataset in the 1930s therefore act as a test case for atmospheric models and might contribute to a further understanding of the exact mechanism of “Dust Bowl”-like droughts.

The Armistice Day Storm of 11 November 1940.

The new dataset can also be used to analyze extreme events like strong midlatitude winter storms, and has a more general potential for providing insight into significant past events on the synoptic time scale. As an example, one of the most severe winter storms ever recorded in the central US was the so-called Armistice Day Storm of 11 November 1940. A meteorological description of the storm was given quite early (Knarr 1941), but the respective upper air data now included in CHUAN were not available in digital form until recently. In Fig. 7 we display wind fields at 1000 m MSL during the development and culmination phase of the storm. A comparison of the wind fields depicted in Fig. 7 with synoptic weather charts (as provided e.g.

by Knarr 1941) shows generally an excellent agreement: The wind observations at the stations match very well their location with respect to the frontal systems.

CONCLUSIONS AND OUTLOOK.

We have presented the Comprehensive Historical Upper Air Network (CHUAN), a new upper air dataset that spans the period 1902 to 2007, encompassing the period prior to upper air based reanalyses available to date. A monthly mean version can be downloaded from the World Wide Web (http://www.historicalupperair.org). CHUAN has been adjusted and evaluated. Further details can be found in Brönnimann (2003b), Ewen et al. (2008a) and Grant et al. (2009a). These newly available data should be useful for analyzing the climate of the first half of the 20th century by providing quantitative knowledge about a period known to cover some of the most extreme climate events in the 20th century such as the “Dust Bowl” droughts in the 1930s (Schubert et al. 2004a,b; Brönnimann et al. 2009a; Cook et al. 2009), the November 1940 Armistice Day Storm or the 1939-42 El Niño (Allan et al. 2003; Brönnimann et al. 2004a,b). We have demonstrated that the new upper air dataset is a valuable tool not only for the analysis of climatological and interannual variability aspects, but also of the synoptic development on sub-daily timescales. It is hoped that these data may help to eventually deepen our understanding of global climate variability. They might also become important for intended full (i.e. including upper air information) reanalysis projects going further back in time. Furthermore, the data have already proven highly valuable for the evaluation of surface based reanalyses like the “Twentieth Century Reanalysis project” (Gil Compo 2009, personal communication). A similar positive impact can be expected from the evaluation of statistical climate reconstructions and of climate model simulations. Considerable digitization and homogenization work dealing with historical upper air data is still ongoing (e.g. in the framework of the NOAA/NCDC Climate Data Modernization Program) or planned. Besides the data themselves, it would be very beneficial if the metadata, helpful to apply corrections in the first place, were more readily available for future quality assessment. All datasets used for the compilation of CHUAN and discussed in this paper will be made available at NCDC, including the single ascent data.

Acknowledgements.

AS, ANG, TE, TG and SB were funded by the Swiss National Science Foundation (PP002-102731). GPC was supported by the NOAA Climate Program Office. ERA-40 data were downloaded from the ECMWF server. NCEP-NCAR 50-Year Reanalysis was downloaded from the NCAR website. Thanks go to SMHI and NMI for providing large parts of the Scandinavian radiosonde data. We would like to acknowledge COST action ES0601. Rob Allan is supported with funds from the Queensland Climate Change Centre of Excellence (QCCCE) in Australia, together with the joint DECC, Defra and MoD Integrated Climate Programme (ICP) in the UK – GA01101, CBC/2B/0417_Annex C5.

References

Alexandersson, H., 1986: A homogeneity test applied to Swedish precipitation data. *J. Climatol.*, **6**, 661-675.

Allan, R. J., C. J. C. Reason, J. A. Lindesay, and T. J. Ansell, 2003: 'Protracted' ENSO episodes and their impacts in the Indian Ocean region. *Deep-Sea Research, Part II*, **50**, 2331-2347.

Brönnimann, S., 2003a: Description of the 1939-1944 upper air data set (UA39_44) Version 1.0. *Tech. Rep.*, Lunar and Planetary Laboratory, Univ. of Arizona, Tucson, USA.

Brönnimann, S., 2003b: A historical upper air data set for the 1939-1944 period. *Int. J. Climatol.*, **23**, 769-791.

Brönnimann, S., and J. Luterbacher, 2004: Reconstructing Northern Hemisphere upper-level fields during World War II. *Climate Dyn.,* **22**, 499-510.

Brönnimann, S., J. Luterbacher, J. Staehelin, T. M. Svendby, G. Hansen, and T. Svenøe, 2004a: Extreme climate of the global troposphere and stratosphere in 1940-42 related to El Niño. *Nature*, **431**, 971-974.

Brönnimann, S., J. Luterbacher, J. Staehelin, and T. Svendby, 2004b: An extreme anomaly in stratospheric ozone over Europe in 1940-42. *Geophys. Res. Lett.*, **31**, L08101, doi: 10.1029/2004GL019611.

Brönnimann, S., A. Stickler, T. Griesser, T. Ewen, A. N. Grant, A. M. Fischer, M. Schraner, T. Peter, E. Rozanov, and T. Ross, 2009a: Exceptional atmospheric circulation during the "Dust Bowl". *Geophys. Res. Lett.*, **36**, L08802, doi:10.1029/2009GL037612.

Brönnimann, S., A. Stickler, T. Griesser, A. M. Fischer, A. Grant, T. Ewen, T. Zhou, M. Schraner, E. Rozanov, and T. Peter, 2009b: Variability of Large-Scale Atmospheric Circulation Indices for the Northern Hemisphere During the Past 100 Years. *Met. Z.*, in press.

Compo, G. P., J. S. Whitaker, and P. D. Sardeshmukh, 2006: Feasibility of a 100-Year Reanalysis Using Only Surface Pressure Data. *Bull. Amer. Meteor. Soc.*, **87**, 175–190.

Cook, B. I., R. L. Miller, and R. Seager, 2008: Dust and sea surface temperature forcing of the 1930s ”Dust Bowl” drought. *Geophys. Res. Lett.,* **35***,* L08710, doi:10.1029/2008GL033486.

Cook, B. I., R. L. Miller, and R. Seager, 2009: Amplification of the North American “Dust Bowl” drought through human-induced land degradation. *Proc. Natl. Acad. Sci.,* **106**, 4997-5001.

Dupigny-Giroux, L. A., T. F. Ross, J. D. Elms, R. Truesdell, and S. R. Doty, 2007: RESOURCES - NOAA's Climate Database Modernization Program: Rescuing, Archiving, and Digitizing History. *Bull. Amer. Meteor. Soc.*, **88**, 1015–1017.

Durre, I., T. C. Peterson, and R. S. Vose, 2002: Evaluation of the Effect of the Luers-Eskridge Radiation Adjustments on Radiosonde Temperature Homogeneity. *J. Climate*, **15**, 1335-1347.

Durre, I., R. S. Vose, and D. B. Wuertz, 2006: Overview of the Integrated Global Radiosonde Archive. *J. Climate*, **19**, 53-68.

Eskridge, R. E., O. A. Alduchov, I. V. Chemykh, Z. Panmao, A. C. Polansky, and S. R. Doty, 1995: A Comprehensive Aerological Reference Data Set (CARDS): Rough and systematic errors. *Bull. Amer. Meteor. Soc.*, **76**, 1759-1775.

Ewen, T., A. Grant, and S. Brönnimann, 2008a: A Monthly Upper-Air Dataset for North America Back to 1922 from the *Monthly Weather Review*. *Mon. Wea. Rev.*, **136**, 1792-1805.

Ewen, T., S. Brönnimann, and J. Annis, 2008b: An Extended Pacific-North American Index from Upper-Air Historical Data Back to 1922. *J. Climate*, **21**, 1295-1308.

Grant, A. N., S. Brönnimann, T. Ewen, and A. Nagurny, 2009a: A New Look at Radiosonde Data Prior to 1958. *J. Climate*, doi: 10.1175/2008JCLI2539.1, in press.

Grant, A. N., S. Brönnimann, T. Ewen, T. Griesser, and A. Stickler, 2009b: The early twentieth century warm period in the European Arctic. *Met. Z.,* in press.

Griesser, T., S. Brönnimann, A. Grant, T. Ewen, A. Stickler, and J. Comeaux, 2009: Reconstruction of global monthly upper-level temperature and geopotential height fields back to 1880. Submitted to *J. Climate.*

Hoinka, K. P., 1997: The tropopause: discovery, definition and demarcation. *Met. Z.,* **6**, 281-303.

Kistler, R., E. Kalnay, W. Collins, S. Saha, G. White, J. Woollen, M. Chelliah, W. Ebisuzaki, M. Kanamitsu, V. Kousky, H. van den Dool, R. Jenne, and M. Fiorino, 2001: The NCEP–NCAR 50-Year Reanalysis: Monthly Means CD-ROM and Documentation. *Bull. Amer. Met. Soc.*, **82**, 247-268.

Knarr, A. J., 1941: The Midwest storm of November 11, 1940. *Mon. Wea. Rev.,* **69**, 169-178.

Lanzante, J., 1996: Resistant, robust and non-parametric techniques for the analysis of climate data: theory and examples, including applications to historical radiosonde station data. *Int. J. Climatol.*, **16**, 1197-1226.

NCAR, 2002: The TD54 World Rawinsondes, Most 1946-72 - Parts 1 and 2, *Tech. Rep.*, http://dss.ucar.edu/docs/papers-scanned/papers.html documents RJ0187,RJ0188.

NCAR/UCAR, 2002: Early wind aloft data, TD52 and TD53. *Tech. Rep.*, http://dss.ucar.edu/docs/papers-scanned/papers.html, Documents RJ0167, RJ0168.

OMI (Organisation Météorologique Internationale), 1951: *Comparaison Mondiale des Radiosondes. Acte Final*. Vol. I. Station Centrale Suisse de Météorologie.

OMM (Organisation Météorologique Mondiale), 1955: *Comparaison Mondiale des Radiosondes. World Comparison of Radiosondes. Acte Final*. Vol. III. Station Centrale Suisse de Météorologie.

Overland, J. E., M. C. Spillane, D. B. Percival, M. Wang, and H. O. Mofjeld, 2004: Seasonal and Regional Variation of Pan-Arctic Surface Air Temperature over the Instrumental Record. *J. Climate*, **17**, 3263-3282.

Polyakov, I. V., R. V. Bekryaev, G. V. Alekseev, U. S. Bhatt, R. L. Colony, M. A. Johnson, A. P. Maskshtas, and D. Walsh, 2003: Variability and Trends of Air Temperature and Pressure in the Maritime Arctic 1875-2000. *J. Climate,* **16**, 2067-2077.

Rasmusson, E. M., 1967: Atmospheric water vapor transport and the water balance of North America: Part 1. Characteristics of the water vapor flux field, *Mon. Wea. Rev.*, **95**, 403–426.

Schubert, S. D., M. J. Suarez, P. J. Pegion, R. D. Koster, and J. T. Bacmeister, 2004a: On the cause of the 1930s Dust Bowl. *Science*, **303**, 1855-1859.

Schubert, S. D., M. J. Suarez, P. J. Pegion, R. D. Koster, and R. T. Bacmeister, 2004b: Causes of long-term drought in the U.S. Great Plains. *J. Climate,* **17***,* 485-503.

Seager, R., Y. Kushnir, C. Herweijer, N. Naik, and J. Velez, 2005: Modeling of tropical forcing of persistent droughts and pluvials over western North America: 1856–2000. *J. Climate,* **18**, 4065-4088.

Seager, R., Y. Kushnir, M. Ting, M. Cane, N. Naik, and J. Miller, 2008: Would advance knowledge of 1930s SSTs have allowed prediction of the Dust Bowl drought?. *J. Climate,* **21***,* 3261–3281.

Uppala, S. M., P. W. Kållberg, A. J. Simmons, U. Andrae, V. da Costa Bechtold, M. Fiorino, J. K. Gibson, J. Haseler, A. Hernandez, G. A. Kelly, X. Li, K. Onogi, S. Saarinen, N. Sokka, R. P. Allan, E. Andersson, K. Arpe, M. A. Balmaseda, A. C. M. Beljaars, L. van de BERG, J. Bidlot, N. Bormann, S. Caires, F. Chevallier, A. Dethof, M. Dragosavac, M. Fisher, M. Fuentes, S. Hagemann, E. Hólm, B. J. Hoskins, L. Isaksen, P. A. E. M. Janssen, R. Jenne, A. P. McNally, J.-F. Mahtouf, J.-J. Morcrette, N. A. Rayner, R. W. Saunders, P. Simon, A. Sterl, K. E. Trenberth, A. Untch, D. Vasiljevic, P. Viterbo, and J. Woollen, 2005: The ERA-40 re-analysis. *Q. J. R. Meteorol. Soc.,* **131**, 2961-3012.

Wang, C., S.-K. Lee, and D. B. Enfield, 2008: Climate response to anomalously large and small Atlantic warm pools during the summer. *J. Climate,* **21***,* 2437-2450.

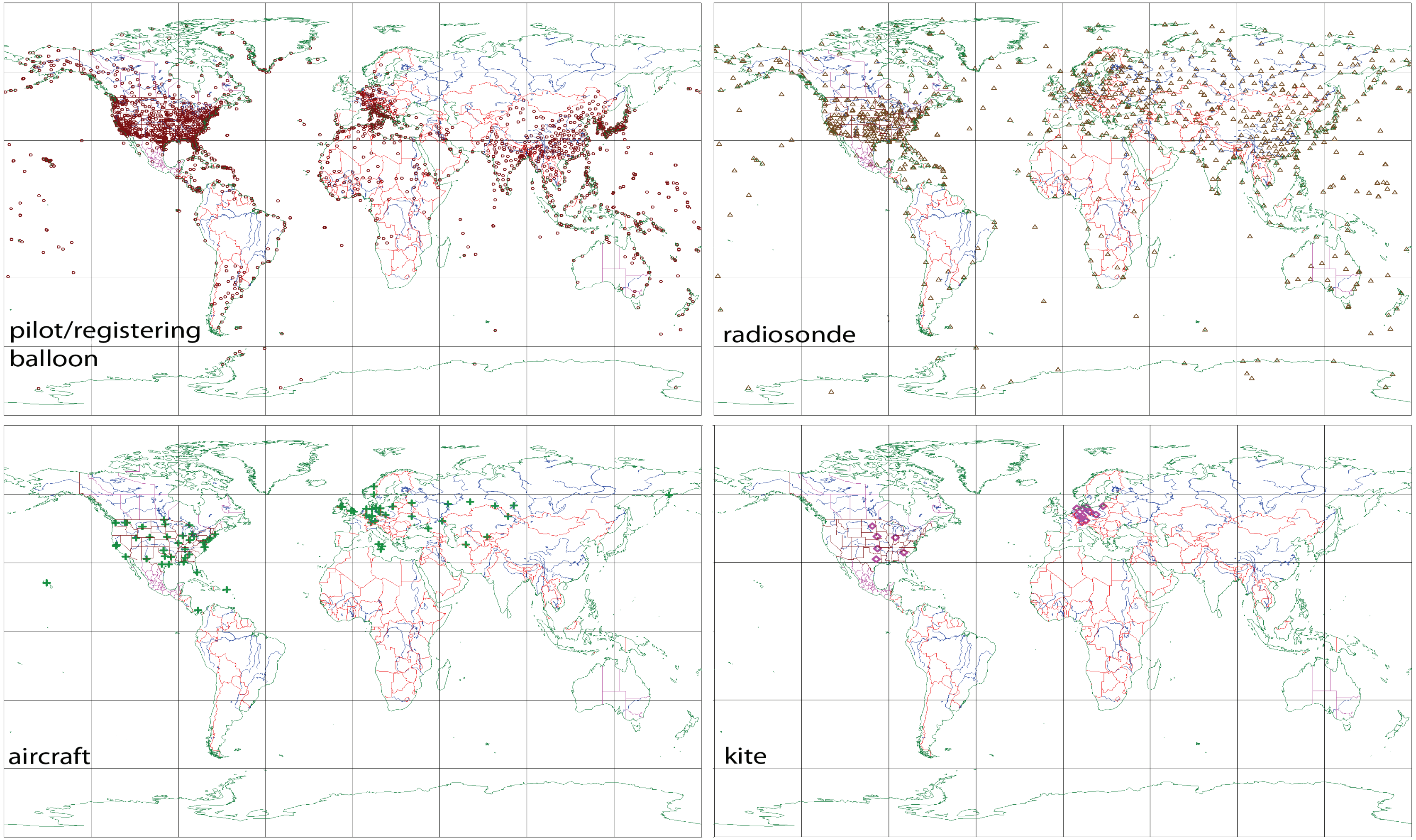

Figure 1: Maps showing the global distribution of all CHUAN upper air stations (1902-57). For clarity, measurement platforms are presented separately in the four panels.

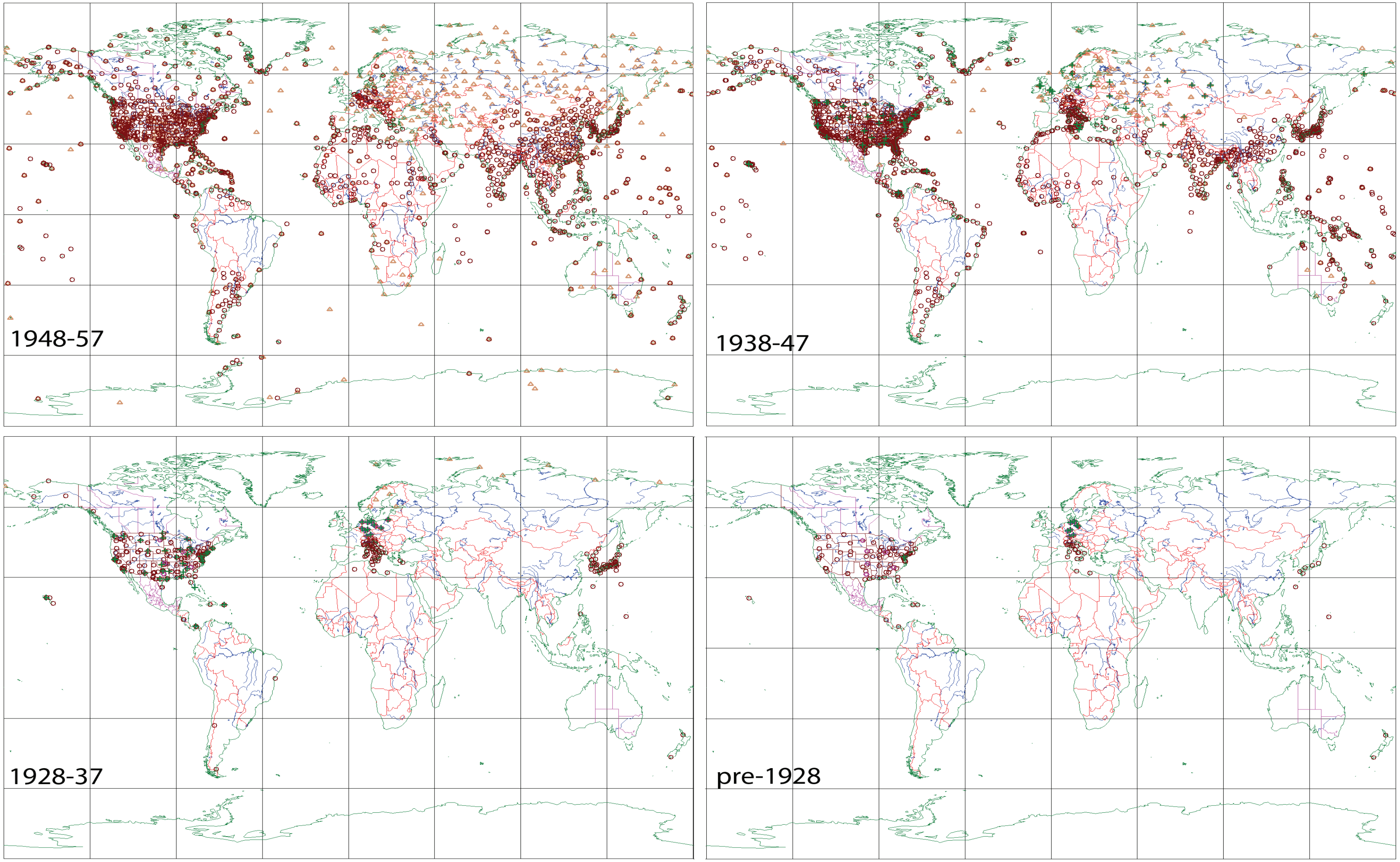

Figure 2: Maps showing the changing density of the CHUAN upper air station network over the decades 1948-57, 1938-47, 1928-37, and the pre-1928 period.

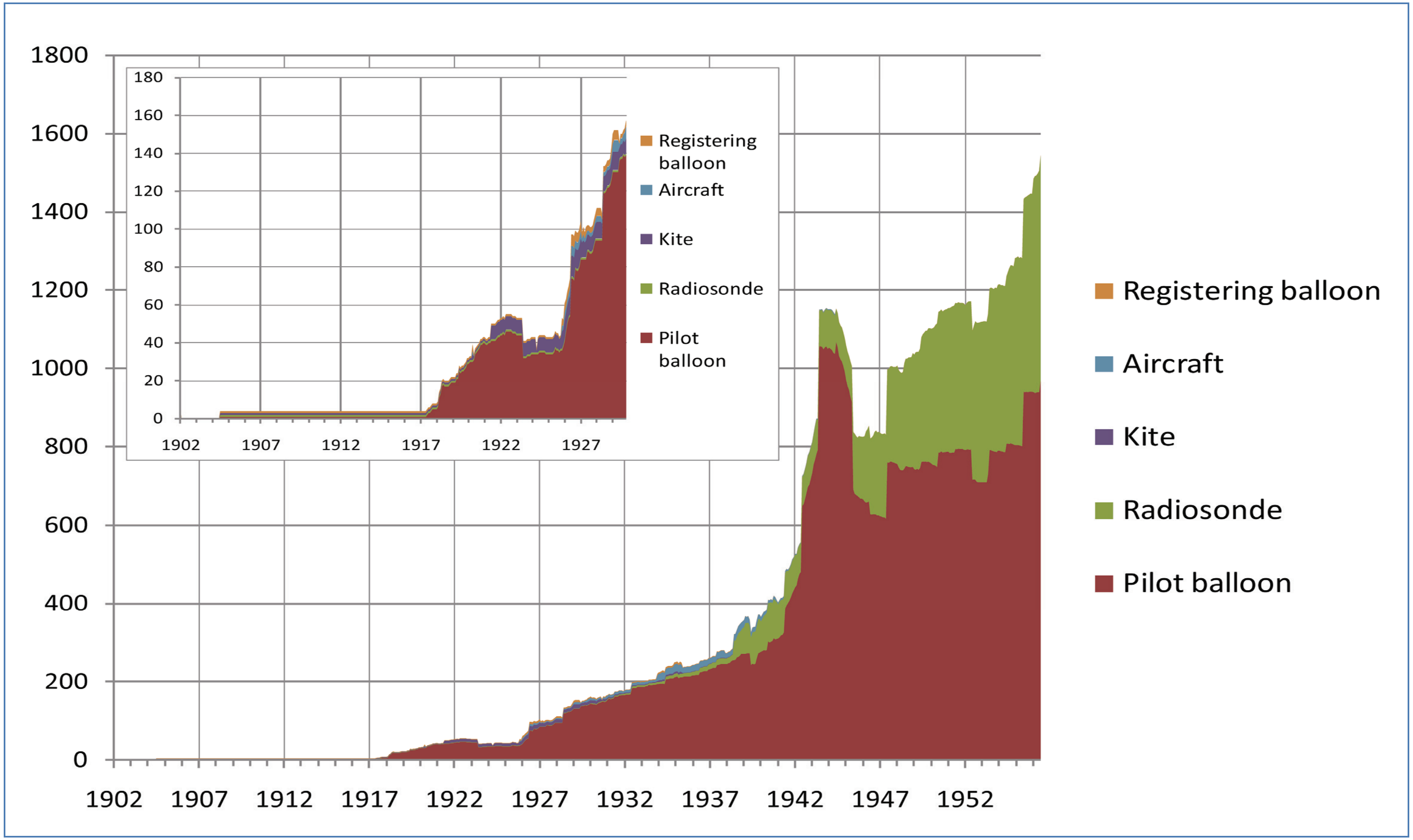

Figure 3: Number of available records versus time. Records that have multiple measurement platforms as well as records from identical stations (with respect to WMO station index or the National Climatic Data Center's Weather-Bureau-Army-Navy (WBAN) number) that have not yet been merged due to the lack of a quality assessment (see detailed station listing in online documentation, www.historicalupperair.org) are counted separately.

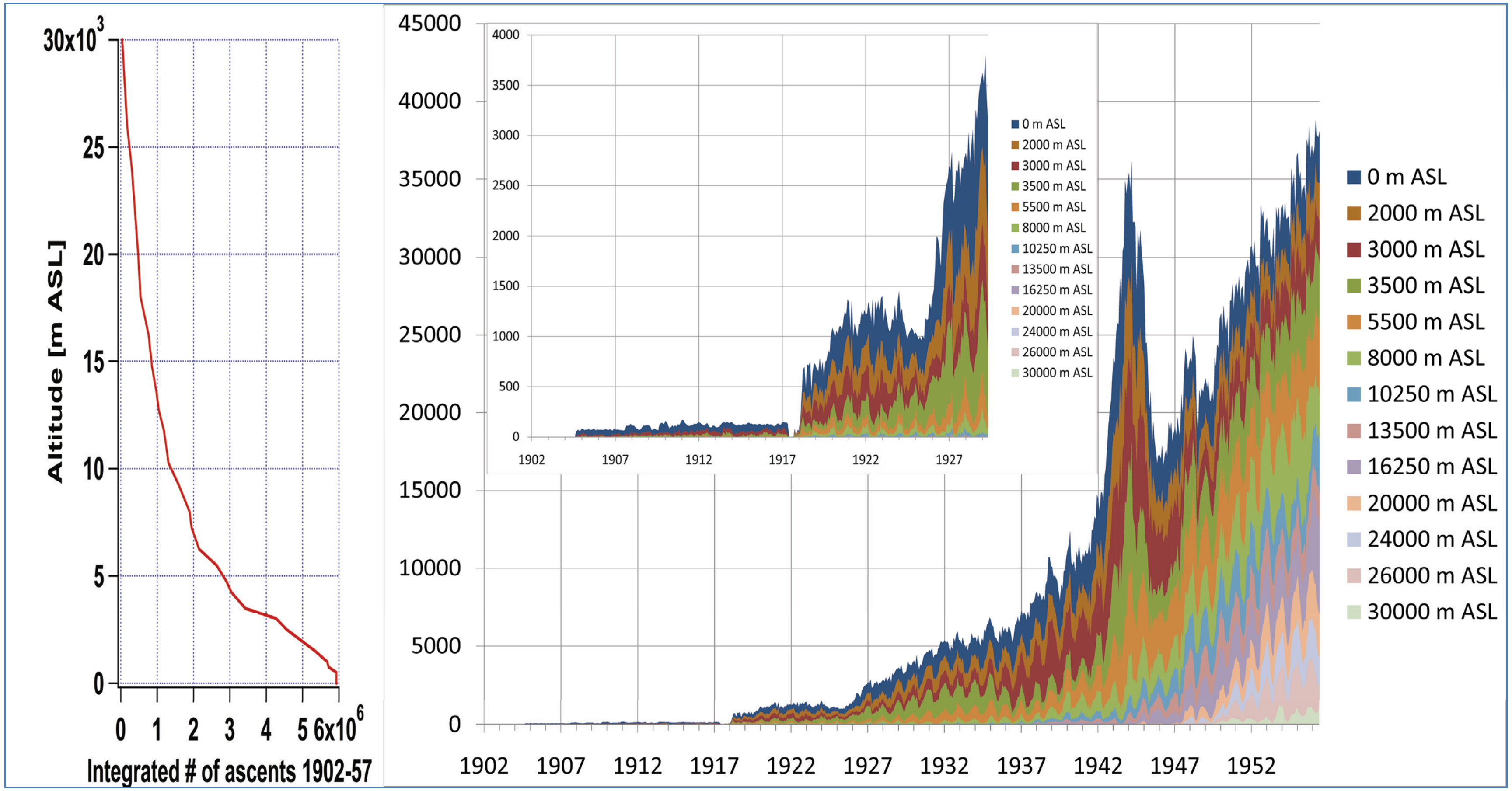

Figure 4: Left panel: Vertical profile of the total number of daytime (06-18 LT) ascents integrated over the period 1902-56. Right panel: Number of daytime ascents reaching different altitude levels above the mean sea level (ASL) versus time.

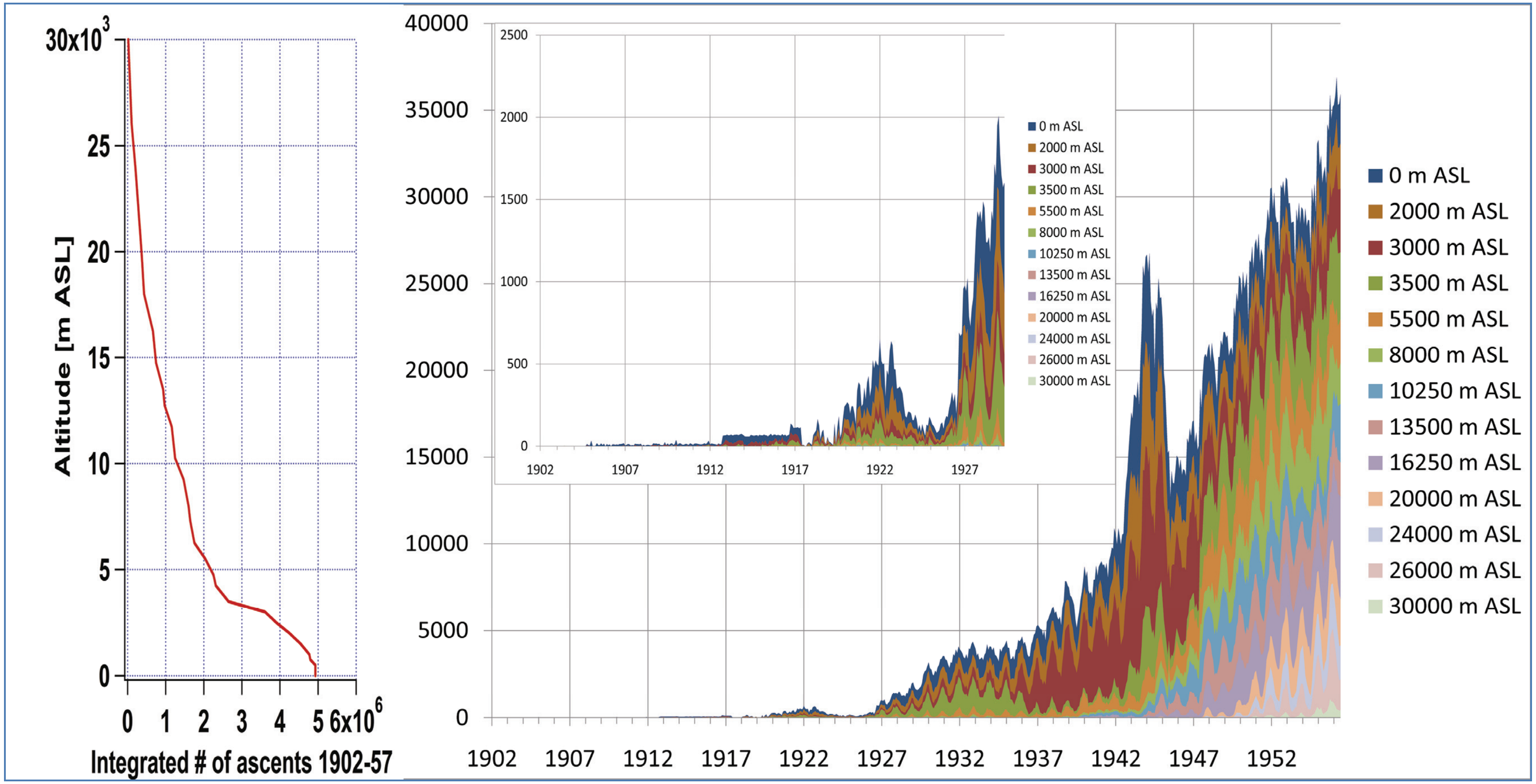

Figure 5: Left panel: Vertical profile of the total number of nighttime (18-06 LT) ascents integrated over the period 1902-56. Right panel: Number of nighttime ascents reaching different altitude levels above the mean sea level (ASL) versus time.

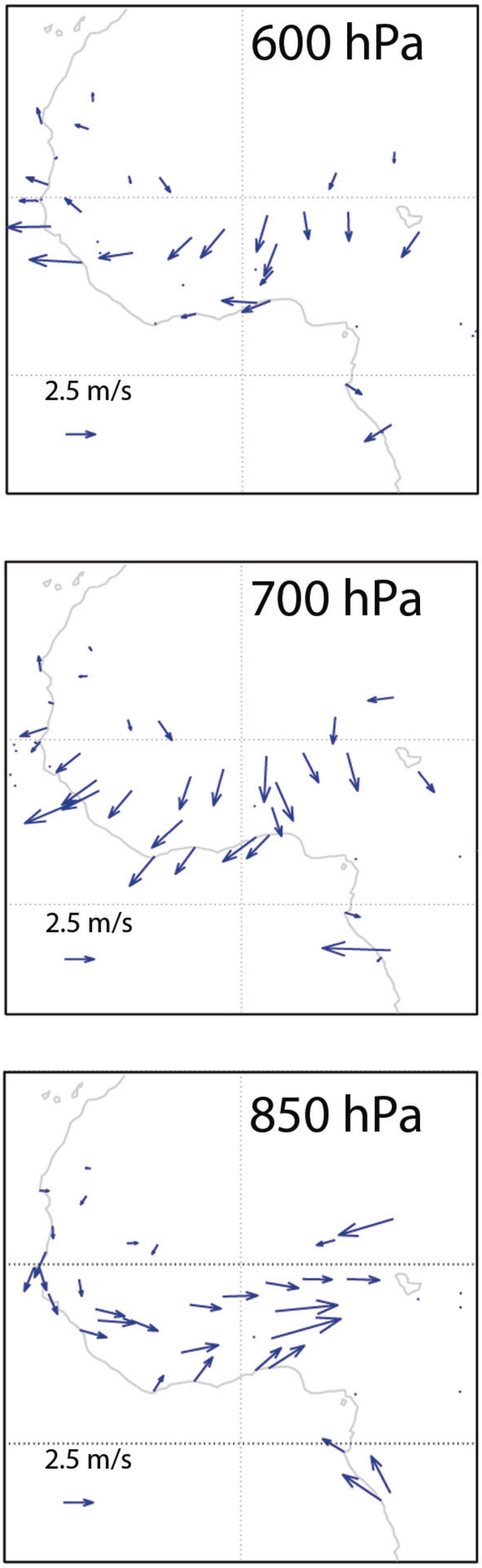

Figure 6: Fields of the JJA mean difference of wind data from the NCEP-NCAR 50-Year Reanalysis relative to wind observations from pilot balloons and in West and Central Africa for the years 1948-1957. The mean difference has been calculated from average U and V wind differences between reanalysis winds and single wind observations (pilot balloon ascents) from CHUAN in a time window of max. 2 h around the reanalysis times (00, 06, 12, 18 UTC).

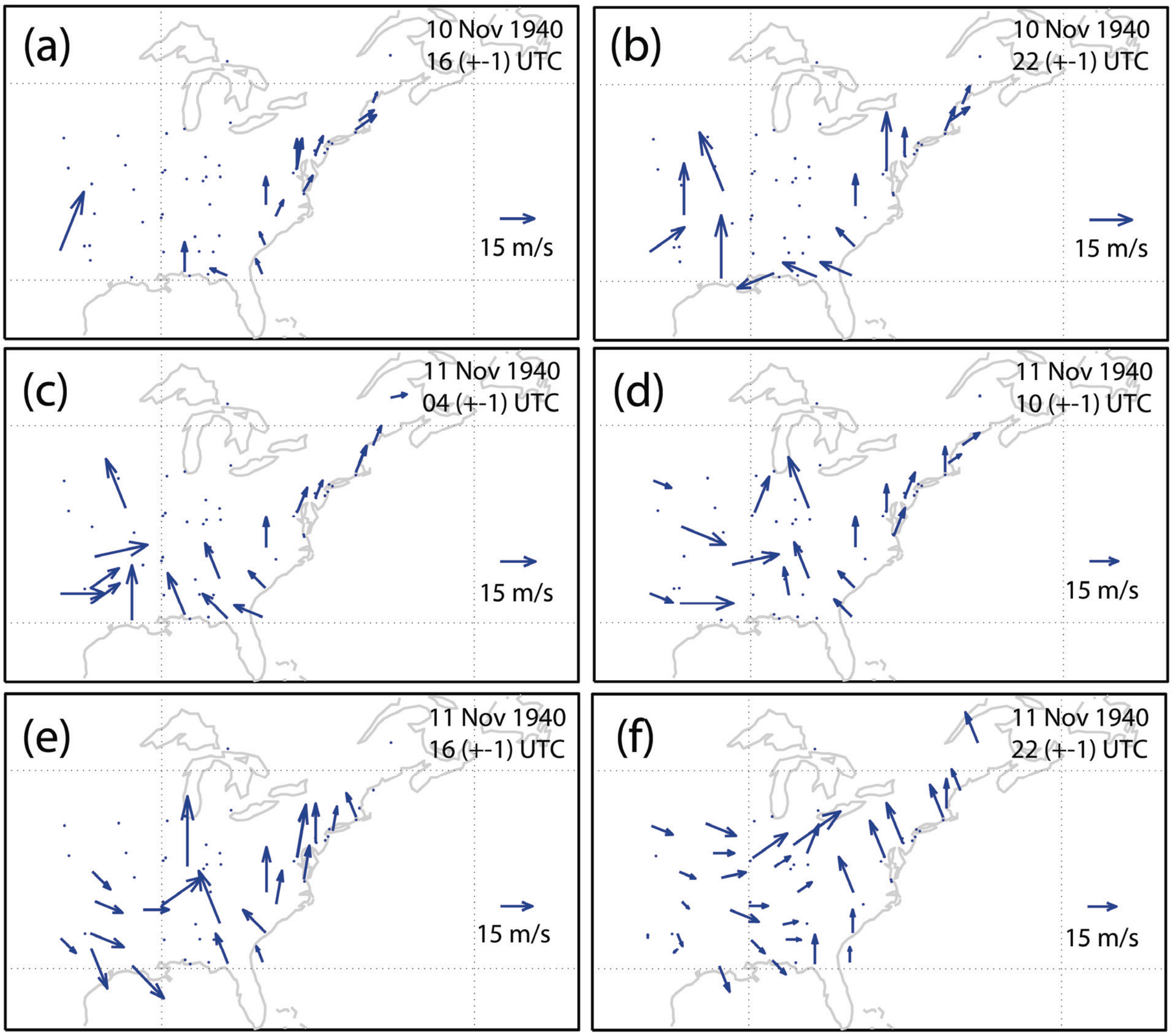

Figure 7: Wind fields at 1000 m MSL from pilot balloon measurements for 10 and 11 November 1940, the days of the explosive development of the so-called "Armistice Day Storm", for the eastern part of North America (30-50°N, 100-65°W). Wind fields are shown in 6-hourly intervals in panels (a)-(f). Because the ascent times are slightly different at some stations, observations taken inside an interval of (±1 h) of the main observation times (04, 10, 16, 22 UTC) have been taken into account. Blue dots symbolize stations for which no measurement data are available for the time indicated. The arrow on the right side shows the scale length for 15 m/s.

Figure captions:

Figure 1: Maps showing the global distribution of all CHUAN upper air stations (1902-57). For clarity, measurement platforms are presented separately in the four panels.

Figure 2: Maps showing the changing density of the CHUAN upper air station network over the decades 1948-57, 1938-47, 1928-37, and the pre-1928 period.

Figure 3: Number of available records versus time. Records that have multiple measurement platforms as well as records from identical stations (with respect to WMO station index or the National Climatic Data Center's Weather-Bureau-Army-Navy (WBAN) number) that have not yet been merged due to the lack of a quality assessment (see detailed station listing in online documentation, www.historicalupperair.org) are counted separately.

Figure 4: Left panel: Vertical profile of the total number of daytime (06-18 LT) ascents integrated over the period 1902-56. Right panel: Number of daytime ascents reaching different altitude levels above the mean sea level (ASL) versus time.

Figure 5: Left panel: Vertical profile of the total number of nighttime (18-06 LT) ascents integrated over the period 1902-56. Right panel: Number of nighttime ascents reaching different altitude levels above the mean sea level (ASL) versus time.

Figure 6: Fields of the JJA mean difference of wind data from the NCEP-NCAR 50-Year Reanalysis relative to wind observations from pilot balloons and in West and Central Africa for the years 1948-1957. The mean difference has been calculated from average U and V wind differences between reanalysis winds and single wind observations (pilot balloon ascents) from CHUAN in a time window of max. 2 h around the reanalysis times (00, 06, 12, 18 UTC).

Figure 7: Wind fields at 1000 m MSL from pilot balloon measurements for 10 and 11 November 1940, the days of the explosive development of the so-called "Armistice Day Storm", for the eastern part of North America (30-50°N, 100-65°W). Wind fields are shown in 6-hourly intervals in panels (a)-(f). Because the ascent times are slightly different at some stations, observations taken inside an interval of (±1 h) of the main observation times (04, 10, 16, 22 UTC) have been taken into account. Blue dots symbolize stations for which no measurement data are available for the time indicated. The arrow on the right side shows the scale length for 15 m/s.

Table 1. Number of soundings for different measurement platforms and time periods.

Platform	pre-1928	1928-37	1938-47	1948-57
Aircraft	1419	15363	7392	0
Kite	29850	28501	495	0
Pilot balloon / registering balloon	240200	1061328	3794542	5101760
Radiosonde	0	1884	147099	2319339

ONLINE SUPPLEMENT: EVALUATION, ADJUSTMENT AND QUALITY ASSESSMENT.

GPH and T data.

In CHUAN, large parts of the data, including the complete radiosonde data as well as most of the wind data (all pilot balloon data on geometrical altitude levels, >98% of all wind data) have been evaluated. For the GPH and temperature (T) data, the quality assessment/quality control (QA/QC) process starts with a range check relative to a reference climatology to flag implausible values. This climatology has been extracted from the NCEP-NCAR 50-Year Reanalysis and interpolated in time between the standard times of observation, yielding a climatology of GPH, T as well as the corresponding vertical gradients for any given time of day, month, level, location, and variable. In addition, a plausibility screening has been performed including checks for consistency with hydrostatic balance.

The data were additionally screened for breakpoints using classical statistical breakpoint detection tests (Alexandersson 1986; Lanzante 1996). A conservative approach was chosen by accepting these breakpoints only when metadata supporting the adjustment were present (the most frequent case was the point of merging of data from different sources). In practice this means that corrections apply for the entire pre-IGY time series or for large and contiguous subperiods. The quality of all p/GPH and T data has been further assessed using statistically reconstructed reference series (see Brönnimann 2003, Ewen et al. 2008a) that are based only on surface data and thus independent from the upper air data. The reference series was generated for each candidate series, that is, for each variable (T and GPH) at five p levels (850, 700, 500, 300, and 200 hPa) for each station. The reconstruction is based on a multiple linear regression of upper-air series (NCEP-NCAR 50-Year Reanalysis interpolated to the station coordinates) as a function of surface data (T and p). The reanalysis series were regressed onto the surface predictors for the period from 1960 to 2000 (the model calibration period) to create the statistical model. This model was then applied to the surface predictors in the pre-IGY period to reconstruct a reference series for each variable, height, and station. The model was validated for the period 1948–59 and estimates of skill were derived during this validation period. Vertical profiles as well as time series (see Fig. 6 of Grant et al. 2009a) of the resulting differences of T and GPH with respect to the reconstructions were then analyzed. Depending on the shape of the vertical profiles of the errors of both T and GPH, the errors were attributed to unit errors, lag and radiation errors, T offsets, p offsets, or combinations. Fig. S1 shows a schematic illustration of the different cases. Large unattributable errors led to rejection, otherwise an adjustment was attempted of either the whole series or parts of the series.

If possible, we use physics based adjustments that are independent of the steps described above. For instance, we applied known radiation-error characteristics of a specific sonde (see Brönnimann 2003b). Additional information on typical error characteristics comes from other studies listed in Table 1 of Brönnimann (2003b) as well as from WMO international radiosonde intercomparisons (OMI 1951, OMM 1952). In some cases, physical parameters (such as lag coefficients) needed to be estimated, which was done by adjusting the parameter in such a way that the error in both the T and GPH profiles would vanish when averaged over the given data source and station network. Only the adjustments for T and p offsets rely on the estimated error profile (and hence depend on the reference series). Pressure offsets are difficult to correct and often led to rejection. Note that all adjustments affect T (or p) first, with GPH being recalculated accordingly. The method is described in detail in Brönnimann (2003b), Ewen et al.

(2008a) and Grant et al. (2009a). The versions of the dataset before and after these adjustments are termed “R” (“raw”) and “C” (“corrected”), respectively (note that “C” does not contain series that were rejected during the procedure).

Finally, for further processing to daily and monthly means, the values have been adjusted to the daily mean based on a climatology of the diurnal cycle from NCEP-NCAR reanalysis (see Brönnimann 2003b). This is necessary as there was no standard ascent time. Versions of the different adjustment steps (radiation and lag correction, diurnal cycle correction) are available for the fully time-resolved dataset (single ascents). Monthly means are generally provided as an “R” version as well as with all adjustments applied to the data (“C” version). For one of the source datasets (UA_39_44_V1.1) an additional version is available in which gaps in the data of some stations have been supplemented with data from surrounding stations (“S” version).

Wind data.

For the wind data in CHUAN, the evaluation is based on the comparison with statistically reconstructed U and V fields (using the same reconstruction method with surface pressure and upper air data as described in Griesser et al. 2009, calibrated with the ERA-40 reanalysis). We performed t tests on the significance (s={90%,95%}) of the deviation of the difference between measured and statistically reconstructed U and V winds from 0 at an altitude of 3 km MSL. An altitude of 3 km MSL was chosen because there are still enough pilot balloons that reached this altitude and at the same time this level is clearly located above the ABL for most of the stations, which otherwise could introduce inconsistencies between observed and reconstructed series. The station records were then automatically classified into four categories with respect to their quality: accepted, accepted but further tests recommended (termed suspicious, with two subcategories), rejected and unassessed (see online documentation for detailed classification criteria).

The target for the “C” version of the wind data was set to an absolute bias of < 1.5 m/s with respect to the reconstructions and a precision of the monthly means of < 2.5 m/s (determined as 1σ of the difference measurement-reconstruction) for the U and the V wind. Note that assessing the bias in this way is difficult in cases where the reanalysis itself is strongly biased. However, this should be true only for regions from which only small numbers of observations have been assimilated into the reanalyses (large parts of the oceans, some parts of the tropics and of the southern hemisphere). As an example (see above), we have found systematic differences between the NCEP-NCAR 50-Year Reanalysis and the pilot balloon wind data in the West African Monsoon region (stations 641-677) contained in CHUAN in the years 1948-57 (Fig. 6). The respective observations have not been assimilated into the reanalysis. Figure S2 displays the probability density functions (PDFs) of the mean bias (over the whole time for each series) of the U and V wind as well as of the precision. The quality criteria applied are indicated by the grey vertical lines.

Additionally, time series of the measurements, the reconstructions, their difference and ratio have been checked visually for inhomogeneities and consistency of the variances whenever the ratio measurement/reconstruction was found to be < 0.52, but > 0, or > 1.9 for either the U or the V wind (pointing at unit problems), or whenever the ratios for the U and V wind differed in their sign and > 0.5 in their value. Stations which were classified rejected or unassessed have been discarded in the corrected (“C”) version of the dataset. No adjustments have been applied in this case.

Caveats.

Although the data have undergone a number of important adjustments, CHUAN is undoubtedly much less mature than similar datasets for the modern period. Given the greater limitations in temporal and spatial sampling compared to the latter, this implies substantially more uncertainty. The adjustment strategy used to account for changes in instruments and practices is, partly due to the lack of important metadata, far less complex than that used for modern radiosonde and satellite data corrections.

Even though we have attempted to ensure that an adequate number of measurements goes into the calculation of the monthly means (see above), there may remain some temporal sampling bias issues. Given the lack of sophistication of the historical measurement systems there might still be a significant "fair weather" bias for the pilot balloon wind data, as well as a strong preference for day versus night observations. Therefore the data, in their current form, are of limited value for assessing long-term changes and trends. However, this does not preclude these data for use in such endeavors in the future, after additional work, and used in conjunction with other data and metadata.

Finally, we want to point out that this dataset is in its infancy compared to similar upper-air datasets from more modern times. The latter, derived from more sophisticated measurements systems, with more complete temporal and spatial coverage, have required decades of evaluation and adjustment - a continuing process - as compared to CHUAN which is just at the starting point.

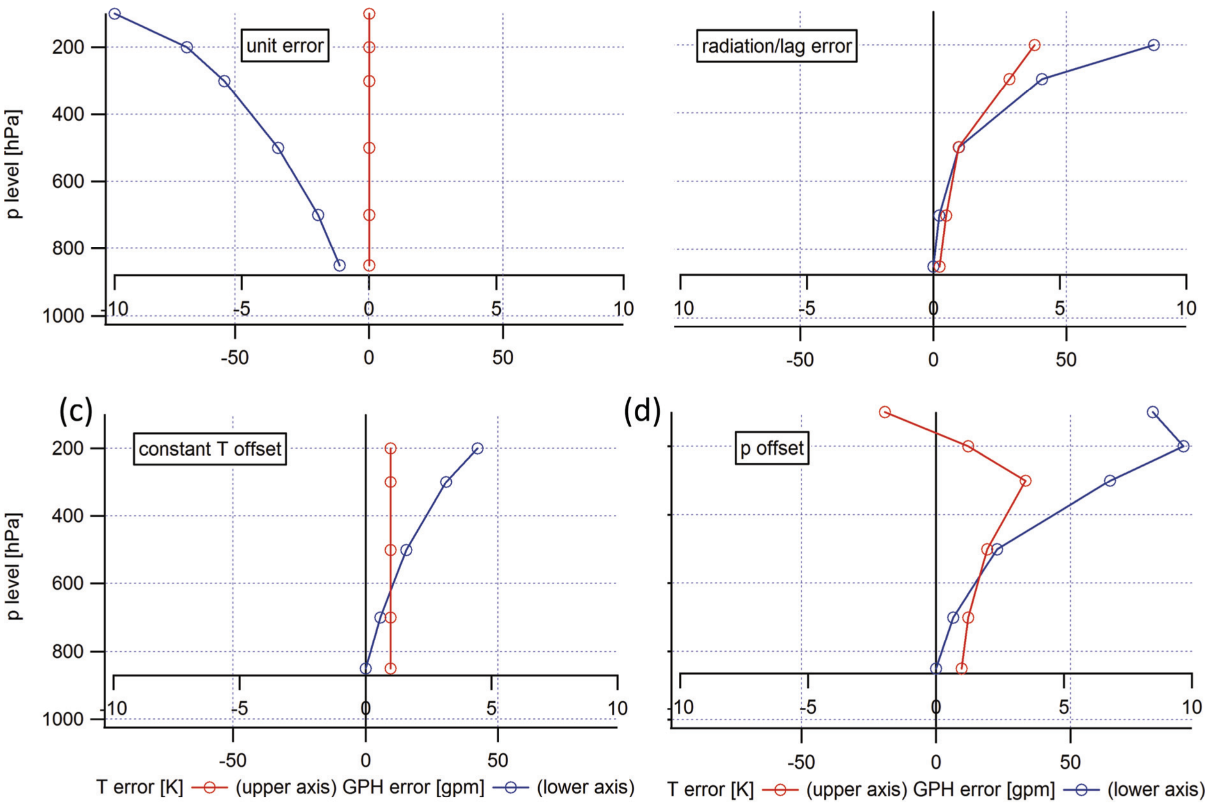

Figure S1: Schematic showing the 4 different error cases identified in the context of the radiosonde QA/QC process and used for the adjustment/rejection of the data. For each station, the vertical profile of the differences between the observations and statistically reconstructed GPH and T fields was analyzed. Depending on the form of the profile, 4 cases could be distinguished: (a) Unit error with growing difference in GPH with altitude. (b) Typical profiles indicative of a radiation (and possible lag) error. They are characterized by a T offset increasing with altitude even beyond the tropopause level (see e.g. Durre et al. 2002). According to the hydrostatic equation, this leads to a strongly growing GPH offset with altitude. (c) Constant T offset. In this case, the GPH offset grows only logarithmically with altitude. (d) p offset. T and GPH differences growing with altitude up to the tropopause level, followed by a sharp swing to negative T errors and declining GPH errors point to a pressure offset in the data.

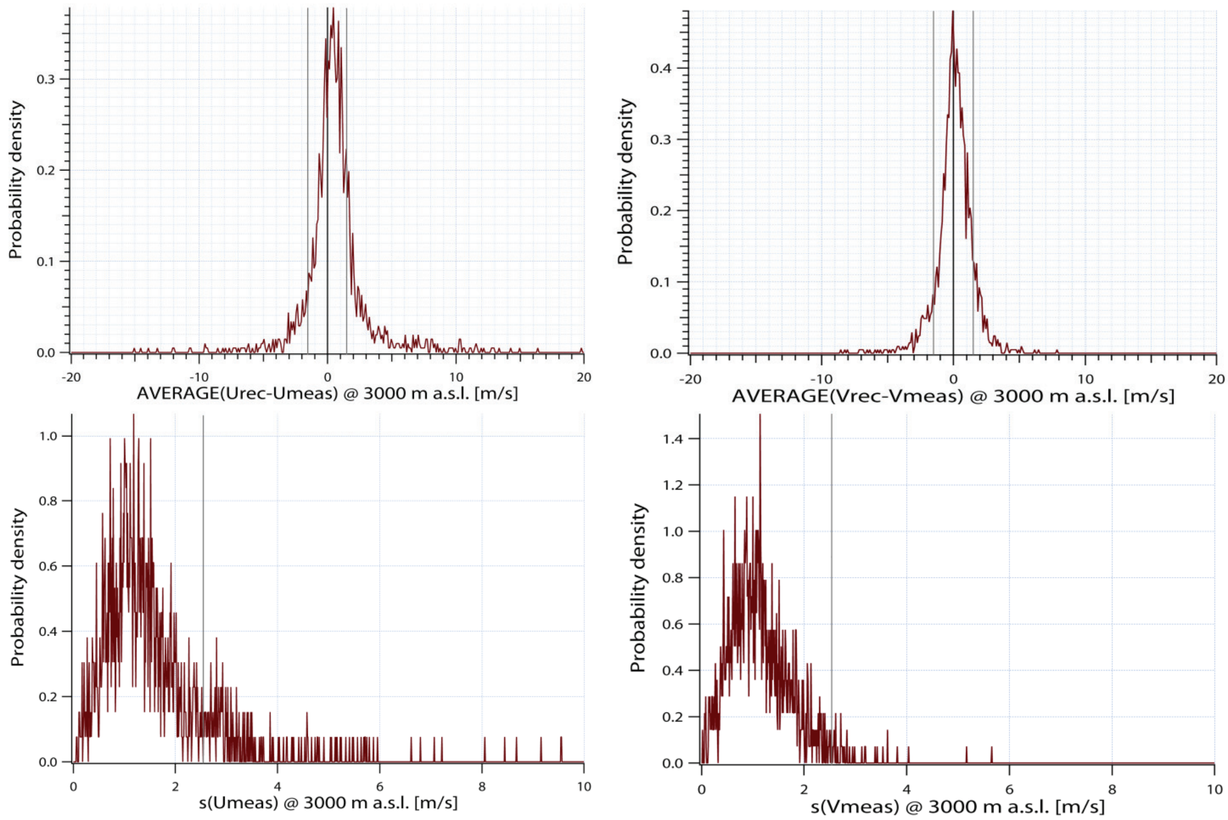

Figure S2: Quality assessment of the wind data. Panel (a) shows the PDF of the temporal mean bias between the observations and the statistical reconstructions for all records for the zonal wind (U). The grey vertical lines show the quality criterion chosen as an upper limit for the bias of the record to be included in the C version of the dataset (1.5 m/s). Panel (b) is the same for the V wind (V). Panel (c) displays the PDF of the precision of all records defined as 1σ of the time series of the differences reconstruction-observation. Again, the grey vertical line marks the quality criterion used for the C version of the dataset (2.5 m/s). Panel (d) is the same as (c) for the V wind.

Figure S1: Schematic showing the 4 different error cases identified in the context of the radiosonde QA/QC process and used for the adjustment/rejection of the data. For each station, the vertical profile of the differences between the observations and statistically reconstructed GPH and T fields was analyzed. Depending on the form of the profile, 4 cases could be distinguished: (a) Unit error with growing difference in GPH with altitude. (b) Typical profiles indicative of a radiation (and possible lag) error. They are characterized by a T offset increasing with altitude even beyond the tropopause level (see e.g. Durre et al. 2002). According to the hydrostatic equation, this leads to a strongly growing GPH offset with altitude. (c) Constant T offset. In this case, the GPH offset grows only logarithmically with altitude. (d) p offset. T and GPH differences growing with altitude up to the tropopause level, followed by a sharp swing to negative T errors and declining GPH errors point to a pressure offset in the data.

Figure S2: Quality assessment of the wind data. Panel (a) shows the PDF of the temporal mean bias between the observations and the statistical reconstructions for all records for the zonal wind (U). The grey vertical lines show the quality criterion chosen as an upper limit for the bias of the record to be included in the C version of the dataset (1.5 m/s). Panel (b) is the same for the V wind (V). Panel (c) displays the PDF of the precision of all records defined as 1σ of the time series of the differences reconstruction-observation. Again, the grey vertical line marks the quality criterion used for the C version of the dataset (2.5 m/s). Panel (d) is the same as (c) for the V wind.