Doctoral Thesis

Arctic climate from an upper level perspective arising from a new collection of historical upper air data

Author(s):
Grant, Andrea Nicole

Publication Date:
2008

Permanent Link:
https://doi.org/10.3929/ethz-a-005711885

Rights / License:
In Copyright - Non-Commercial Use Permitted
Arctic climate from an upper level perspective arising from a new collection of historical upper air data

A dissertation submitted to the
SWISS FEDERAL INSTITUTE OF TECHNOLOGY
ZURICH

for the degree of
Doctor of Sciences

presented by
ANDREA NICOLE GRANT
M.S., Materials Science, University of Minnesota
born 21 September 1972
citizen of the United States of America

accepted on the recommendation of
Prof. Dr. S. Brönnimann, examiner

Dr. M. Serreze, co-examiner

Prof. Dr. L. Haimberger, co-examiner

2008
On November 19, 1930 we had some “weather”, as some people say when they mean only the bad kind.

- H. Clark Fuller, Kite Flier, US Weather Bureau, Ellendale, North Dakota
Abstract

Arctic climate plays an important role in the global climate system and is often portrayed as the “canary in the coal mine” of anthropogenically induced climate change. The two major warm periods in the Arctic in the past century provide a pair of opportunities for understanding what drives Arctic climate on an interdecadal scale. Until now, only the surface instrumental record covered both warm periods. Here I present the development of a new upper air dataset and the application of this data to the temporal evolution of the vertical temperature structure in the Arctic.

Historical radiosonde data are known to suffer from inhomogeneities; data from the earliest part of the record (prior to the International Geophysical Year (IGY) in 1957-1958) has been considered even more suspect due to non-standardized launch times and early instrumentation development. In this work, data were gathered from a number of archives, including, in some cases, original paper records. A thorough cross-referencing of the sources resulted in the first comprehensive compilation of pre-IGY radiosonde data. This dataset was then validated using an innovative homogenization method which compares monthly mean data on a subset of levels to a statistically reconstructed reference series. Widespread uncorrected radiation and lag errors were found over Eurasia and inconsistent geopotential height units were used in the earliest years of the Former Soviet Union. The corrected data show large changes in height and temperature across Eurasia which resulted in more physically consistent fields (temperature has improved agreement with height). The corrections are important as they have a clear spatial structure that interferes with the planetary wave structure. The assimilation of uncorrected data into NCEP-NCAR reanalysis has led to a widespread warm bias in the reanalysis in the 1950s.

The new dataset was then used to question the validity of trend analysis in ERA-40 Reanalysis zonal mean temperature over the Arctic Ocean where no in situ upper air data are assimilated. Furthermore, the use of zonal means obscures important regional processes that affect Arctic climate. Examination of spatial and regional data provides a more complete understanding of the Arctic climate. Vertical time series were examined
for seven regions of the Arctic in all four seasons for evidence of the two warm periods recorded in the surface air temperature record. The two events (the 1920s to 1940s and 1990s to now) were found to have quite different seasonal and regional signals as well as vertical structure. Sparseness of the earliest data makes it difficult to conclusively identify the vertical structure of the early warm period, but some years were vertically coherent (warm at the surface and aloft) and some were not (warm at the surface only). The clear signal of the recent warm period throughout the troposphere is evident. The two warm periods are further distinguishable in that the early warm event was strongest in winter and almost exclusively seen in the Atlantic sector of the Arctic, while the recent warming is evident in all sectors of the Arctic and all seasons, although the strongest signal is in the Beaufort-Bering region in winter and spring. The recent warming is also growing more widespread and is better classified as a trend than an isolated warm event.

Reconstructed geopotential height fields show increased southerly advection from Europe into the Barents Sea area throughout the 1930s which is consistent with the continuously warm SAT in that region. Circulation in the 1940s weakened and grew disorganized, which is also consistent with the larger variability in temperature in the 1940s Atlantic Arctic. Sulfate levels in ice core records from Greenland and Svalbard provide independent validation of the anomalous circulation and raise the question of their own influence on the early warm event.

In conclusion, I describe several additional projects which could be undertaken using this new historical upper air dataset. This includes the validation of a new surface pressure based reanalysis which is already underway and the assimilation of this data into third generation reanalyses such as ERA-75. A more comprehensive validation of reanalysis products in the Arctic region could be studied as well as extending the climatology of the inversion layer statistics in the Arctic.
Zusammenfassung


Der neue Datensatz wurde verwendet, um Trendanalysen der zonal gemittelten Tem-

Rekonstruierte geopotentielle Höhenfelder zeigen in den 1930er Jahre erhöhte südliche Advektion von Europa Richtung Barentssee, was mit den kontinuierlich warmen Bodentemperaturen in jener Region konsistent ist. Die Zirkulation war in den 1940er Jahren abgeschwächt und wurde variabel, was ebenso mit der größeren Temperaturvariabilität in den 1940er Jahren in der Atlantischen Arktis im Einklang steht. Sulfat Werte in Eisbohrkernen von Grönland und Svalbard liefern unabhängige Verifizierungen dieser anomalen Zirkulation. Es bleibt jedoch unklar, ob diese selber auch einen Einfluss auf die frühere Warmphase gehabt haben.

In der Schlussfolgerung beschreibe ich mehrere zusätzliche Projekte, die mit Hilfe dieses neuen historischen Datensatzes unternommen werden könnten. Dies beinhaltet die Verifizierung von Reanalyse Daten, die auf Bodendruck-Daten basieren und bereits gerechnet wurden und die Assimilation dieser Daten in Reanalysen dritter Generation wie z.B. ERA-75. Dadurch könnte eine umfassendere Verifizierung der Reanalyse Produkte in der Arktischen Region studiert werden. Zudem könnte das langjährige Mittel der Inversionsschicht in der Arktis zeitlich ausgeweitet werden.
Chapter 1

Introduction

1.1 Motivation

In order to understand the potential impacts of anthropogenically induced climate change, a better understanding of the natural variability of the climate system is required. This is especially the case in the Arctic, where both observed natural variability and predicted trends are large. Furthermore, the Arctic hosts a number of sensitive ecosystems, such as extensive permafrost and sea-ice based hunting grounds, which are threatened by changing climate. In this work, I present the development of a new upper air dataset extending back to 1932 which allows analysis of 20th century Arctic climate from a three dimensional perspective. This casts a new light on the mechanisms which have played a role in the two major warm events of the past century.

The consensus from numerous coupled ocean-atmosphere models is that the Arctic region will show enhanced warming from increased greenhouse gases (a so-called “polar amplification” effect) (Räisänen, 2001), with warming of up to three times the global mean warming (Holland and Bitz, 2003). Paleoclimate records also show that, over the ice ages, tropical surface air temperature (SAT) changed by only a few degrees while polar surface air temperature changed by $10^\circ$C or more (Hartmann, 1994). While this enhanced response could potentially serve as an early warning of anthropogenic climate change, polar amplification has not been observed unanimously in analyses of 20th century data. Some have seen observational evidence of enhancement of global warming in the Arctic (Moritz et al., 2002; Johannessen et al., 2004; Overpeck et al., 1997), but Polyakov et al. (2002) have found the recent warming to be part of a low frequency oscillation with trends in Arctic SAT approximately equal to Northern Hemisphere SAT trends. Serreze and Francis (2006) address this discrepancy by discussing
two issues which confound the ability to see this signal clearly. First is that in the early stages of greenhouse gas warming, models exhibit a wide variety of regional patterns and magnitudes of warming which converge across the models as the anthropogenic forcing increases. Secondly, the natural low frequency variability in the Arctic which is driven by, for example, the Arctic Oscillation and the Pacific Decadal Oscillation obscures such an emerging climate change signal.

Polar regions play a critical role in driving some of the primary feedback processes in the ocean-atmosphere system, e.g., those involving planetary albedo, thermal inertia and the surface heat flux, as well as playing an important role in the thermohaline circulation of the oceans. The state of the Arctic has a considerable influence on global climate and vice versa. Disappearance of the permanent ice pack is one of the most popular examples of “tipping points” in the climate system, and recent evidence suggests there has been accelerated melting of Arctic sea ice, with the likely loss of permanent (i.e. summer) ice in the coming century (Overpeck et al., 2005; Meier et al., 2005). Climate in the Arctic region is influenced by many factors, for example: large scale atmospheric and oceanic circulation patterns such as the Arctic Oscillation and the Meridional Overturning Circulation, sea ice and cloud cover which both impact the surface energy budget on a local to regional scale, regional variability in the ocean-atmosphere heat flux, latitude limited insolation, and possible stratospheric influences.

A dramatic warm period in the Arctic occurred between the 1920s and 1940s, with the average surface air temperature reaching 1.7°C above normal (Polyakov et al., 2003). The cause of this warming and subsequent cooling is not well understood, although it is generally believed to be evidence of natural variability. Based on surface data, several mechanisms have been proposed for the early 20th century warming which can be grouped into circulation anomalies, natural variability, feedbacks, and changes in forcing. Changes in the shape and intensity of the Icelandic low have been noted which allowed both Greenland and Europe to experience warm winters after the mid 1920s (Rogers, 1985; Fu et al., 1999; Overland and Wang, 2005). Polyakov et al. (2003) found evidence of a low frequency oscillation in Arctic temperatures and sea ice extent in the Atlantic sector, while both Johannessen et al. (2004) and Wang et al. (2007) ascribed the early warm period to natural variability in their modeling studies. Evidence of a local pressure gradient off the northern coast of Norway, initiating a positive ocean-atmosphere feedback, was found by Bengtsson et al. (2004). Delworth and Knutson (2000) attributed the early warm period to a combination of natural variability and early anthropogenic greenhouse gas and sulfate forcing, while Overpeck et al. (1997) found a better match with decreased volcanic activity and increased solar activity and Soon (2005) found a striking concordance between solar activity and Arctic
surface air temperature.

Previous analyses have been limited to surface meteorological data as these are the only data that had been available for the first half of the 20th century. In this work, I present the development of a new upper air dataset which I then use to examine the vertical structure of the two major warm periods in the Arctic in the 20th century.

1.2 Outline

Chapter 2 (Grant et al., 2008b) describes the development of the upper air dataset used to examine the warmings. This is the first such comprehensive compilation of early upper air data ever attempted. The raw data sources are enumerated, and the process of cross referencing the stations and merging the duplicated and overlapping stations into a single dataset is discussed. The validation technique is presented along with numerous examples of typical problems encountered in the dataset. Finally, there is a brief discussion of the impact of the corrections that were identified and applied during the validation process; large geographical regions required significant correction in both temperature and geopotential height, affecting hemispheric-scale analyses.

Chapter 3 examines the vertical structure of the 20th century Arctic warm periods. Motivation is presented in Section 3.1 (Grant et al., 2008c), a comment on a recent Nature paper (Graversen et al., 2008) which discussed the vertical structure of the recent warming using zonal mean ERA-40 reanalysis. The problem of using of ERA-40 reanalysis for trend analysis in data sparse regions is addressed, and we suggest that a regional and seasonal perspective on the Arctic might be more informative. Section 3.2 (Grant et al., 2008a) then explores the vertical structure of the two warm periods in greater detail, including an analysis of reconstructed geopotential height fields which demonstrate anomalous circulation in the 1930s. Finally, a discussion of the impacts of anthropogenic aerosols on Arctic climate is presented; circumstantial evidence suggests these may have played a significant role in the early warming.

Chapter 4 summarizes the work presented, and discusses the further directions for this work. Potential uses of this newly developed dataset are discussed as well as the validation and further use of a new 20th century reanalysis which shows excellent agreement with the radiosonde data compiled here.

Appendix A (Ewen et al., 2008b) discusses a related validation project using the same technique as in chapter 2, but compared to reconstructions done by Tracy Ewen. Decisions were made jointly by Tracy Ewen and me.
Chapter 2

Validation of Historical Upper Air Data

A New Look at Radiosonde Data Prior to 1958

Andrea N. Grant¹, Stefan Brönnimann¹, Tracy Ewen¹, and Andrey Nagurny²

¹Institute for Atmospheric and Climate Science, ETH Zurich, Switzerland
²Arctic and Antarctic Research Institute, St. Petersburg, Russia

Submitted to Journal of Climate 17 March, 2008
Revised 21 August, 2008

Abstract

Historical radiosonde data are known to suffer from inhomogeneities. The first radiosonde intercomparison was made at Payerne, Switzerland in 1954, and a major international effort to standardize the network, including launch times, was made for the International Geophysical Year (IGY) in 1957-1958. Data from before this period, in some cases extending back as far as 1934, have been viewed with even more sus-
CHAPTER 2. VALIDATION OF HISTORICAL UPPER AIR DATA

These early data are scattered among numerous archives with a variety of station identifier schemes and quality control procedures, and some of the data has only recently been digitized from paper records. Here, the first systematic compilation of pre-IGY data is made, and a novel quality assessment technique is applied which reveals that much of the early data have uncorrected radiation and lag errors, especially in the former Soviet Union. Incorrect geopotential height units and problematic time stamps were also found. We propose corrections and present corrected hemispheric fields that show large changes and improved internal consistency in height and temperature across Eurasia compared with uncorrected data. The corrections are important, especially as they have a clear spatial structure that interferes with the planetary wave structure. These corrected data are useful for climate studies and considerably enhance the length and quality of the upper air record but may not be suitable for trend analysis. Assimilation of the uncorrected data has led to a widespread warm bias in NCEP-NCAR reanalysis in the 1950s.

2.1 Introduction

Radiosonde data is the primary tool by which we understand the vertical structure and circulation of the free atmosphere, both for forecasting purposes and climate studies. Unfortunately, this data record is fraught with problems. Although the apparently divergent temperature trends seen in the radiosonde and satellite records have, after over 15 years of debate, finally been harmonized (Sherwood et al., 2005; Santer et al., 2005; Mears and Wentz, 2005; Fu et al., 2004), many questions remain about the homogeneity of the radiosonde record (e.g., Lanzante et al., 2003; Free et al., 2002; Eskridge et al., 2003), limiting confidence in trend analysis.

Several efforts have been made to homogenize the radiosonde record (e.g., Free et al., 2005; Lanzante et al., 2003; Haimberger, 2007; Parker et al., 1997; Thorne et al., 2005). Homogenizing surface meteorological data generally relies on comparison with a suitable reference series. Such a reference is normally not available for upper air data, leading to a variety of techniques which can be broadly classified as statistical (e.g., stationarity of a time series (Lanzante et al., 2003)) or use of an unorthodox reference series (comparison with satellite data (Parker et al., 1997) or highly correlated distant stations (Thorne et al., 2005)). Intercomparison of different techniques do not routinely reveal the same breakpoints (Free et al., 2002), underscoring the difficulty of the problem. Most attempts have extended back only to 1958, the end of the International Geophysical Year (IGY) during which there was a dramatic expansion of the
worldwide radiosonde network as well as a standardization of launch times to 00Z and 12Z. Other homogenization efforts cover only the satellite period (back to 1979) (Parker et al., 1997). In most cases, pre-IGY data were simply too problematic or scarce to bother with, partly due to the irregular launch times. Lanzante et al. (2003) began by including data back to 1948, but in a number of cases were forced to discard the pre-IGY data.

Up to now, the pre-IGY data have never even been systematically compiled. The data are scattered among numerous archives, are cataloged via multiple station identifier schemes, and have been subjected to different quality control and data culling procedures. It was our hypothesis that some of the discarded earlier data may be usable after quality assessment and correction; therefore we attempted to compile a comprehensive archive of radiosonde data up to 1957. Much of the work involved detangling the overlapping archives to create a single dataset. These data are a valuable resource for studies of interannual climate variability, case studies, and as input to upcoming reanalysis and reconstruction projects. While trend studies based on the earlier data are problematic, having a better understanding of the errors and reliability of the earlier data can assist in decisions about the appropriateness of calculating trends.

This study applies a novel quality assessment technique to this new collection of pre-IGY data. The quality assessment process was developed during a project to digitize historical radiosonde data prior to 1948 (Brönnimann, 2003a). Monthly mean station data are compared to a statistically reconstructed reference series based on surface data and NCEP-NCAR Reanalysis (NNR, Kistler et al., 2001). Time series of anomalies from the reference can then be inspected for break points or variance which exceeds prespecified targets, and vertical profiles of the mean anomalies can be inspected for characteristic shapes of common errors. In most cases, the corrections are physically based and are corrected using independent information; this differs from many homogenization techniques which apply a statistical correction. Section 2.2 describes the data which were collected and assessed. Section 2.3 describes the quality assessment technique, with some sample stations being shown in Section 2.4. Overall data quality and findings are discussed in Section 2.5, including before and after figures of Northern Hemisphere station data and a brief discussion of widespread NNR biases due to assimilation of uncorrected data. Finally a conclusion summarizes the results in Section 2.6.
2.2 Data

Data from approximately 1500 radiosonde stations were collected from several archives: 1) the Integrated Global Radiosonde Archive (IGRA) (Durre et al., 2006) at the National Climatic Data Center (NCDC), 2) the United States Air Force Environmental Technical Applications Center tape deck 54 dataset (TD54) at the National Center for Atmospheric Research (NCAR), 3) NCDC’s tape deck 6201 compilation (TD-6201), 4) the Comprehensive Aerological Reference Data Set tape deck 542 archive (CARDS542) covering 1946 and 1947 (Eskridge et al., 1995) from NCAR, 5) data digitized internally within the working group of the authors at ETH (Brönnimann, 2003a; Brönnimann et al., 2005; Ewen et al., 2008b), 6) data from the Arctic region of the former Soviet Union which were recently digitized at the Arctic and Antarctic Research Institute (AARI), and 7) several miscellaneous sources such as Lindenberg, Germany and Payerne, Switzerland. Significant overlap existed among the archives, complicated by the fact that some were numbered according to the Weather Bureau Army Navy (WBAN) system and others according to the World Meteorological Organization (WMO) scheme; it was not initially clear which stations overlapped. Inconsistent station coordinates presented a further roadblock: for example, pairs of stations with identical WMO numbers but different coordinates were found, yet on further inspection they contained identical data. Other combinations also arose, where pairs of stations had different WMO numbers and same or different coordinates, with identical data in some cases and not in others. Different levels and timestamps were also present in soundings which were otherwise duplicates of each other.

After identifying possible duplicates (identical station numbers or coordinates within 1° latitude and longitude), data were manually compared to identify duplicates. 879 unique stations were identified, shown in Figure 2.1. Because of systematic level deletion, duplicated timestamps, and overall data availability before 1958, sources were prioritized as follows (with records being completed with lower priority sources if additional time periods were available): 1) internally/AARI digitized, 2) TD54, 3) TD-6201, 4) CARDS, and 5) IGRA. IGRA was given a lower priority in the compilation process due to deletion of certain levels, deletion of early parts of some records, and duplicated timestamps at some stations. The stations were coded by radiosonde type so that the appropriate radiation and lag correction, if necessary, could be applied. Western European stations with no metadata were presumed to have used a Vaisala sonde, and Chinese stations were presumed to have used the Soviet Molchanov sonde. A fully cross referenced station list including all verdicts and corrections is available, along with the archive of monthly mean data, at http://www.historicalupperair.org
2.3 Quality Assessment

Data transmitted on the Global Telecommunications System (GTS) undergo routine quality control measures before being archived at NCDC ([Durre et al., 2006; Eskridge et al., 1995; Kalnay et al., 1996; Kistler et al., 2001]). Nevertheless, the record, especially in the earlier years, remains problematic (e.g., Lanzante et al., 2003). Having successfully assessed the quality of upper air data from 1939 to 1944 using a new statistical technique ([Brönnimann, 2003a]), we apply this method to the newly compiled pre-IGY data. The full details of the technique can be found in Brönnimann (2003b). An overview of the technique is presented next.
CHAPTER 2. VALIDATION OF HISTORICAL UPPER AIR DATA

2.3.1 Context and limitations

One of the main problems in homogenizing radiosonde data is a lack of neighboring stations, unlike surface data where suitable stations can normally be found. Various alternatives have been employed, such as self-homogenization techniques (Lanzante et al., 2003), comparison with satellite data (Parker et al., 1997), the use of “buddy” stations which show a high correlation with the station of interest (Thorne et al., 2005), and background fields from ERA-40 reanalysis (Haimberger, 2007).

In our approach, a statistically reconstructed monthly mean reference series was generated for both temperature and geopotential height at each station on a subset of pressure levels. Because of the potential errors involved in such a reconstruction, the quality assessment is limited to a small number of physically based errors for which all of the data for a given station (or large, contiguous temporal subsets, if necessary) can be corrected based on a single parameter. This is in contrast to the more statistical homogeneity approaches which identify errors in individual variables and heights and then adjust small sections of the time series independently of other levels and variables. Our approach limits the number of errors that can be addressed, but even with this conservative approach a large number of errors can be identified and corrected. Our approach also simplifies the identification of errors since we have an expectation of what the possible errors could be based on physical principles, knowledge about operational processing of the early data, and previous work with early upper air data (Brönnimann, 2003a; Ewen et al., 2008b).

The data were tested using classical statistical breakpoint detection tests (Alexandersson and Moberg, 1997; Lanzante, 1996), and we then take a conservative approach in accepting these breakpoints only when there is metadata supporting the adjustment. Prior to the IGY, such metadata could include the launch time change (to 00Z and 12Z) when it was likely that other operational changes were made, the IGY itself, or the change under consideration also occurring at other stations in the same country or network at the same time (such as the Soviet Union-wide change in reported units used for geopotential height in Jan, 1950). In practice this means that corrections apply for the entire pre-IGY timeseries or for large and contiguous subperiods (e.g., units correction from beginning-of-record to Dec 31, 1949). This conservative approach relies on an assumption of simplicity regarding operational processing; that is, that a station did not arbitrarily jump back and forth between processing routines from one sounding to the next. In some cases an endpoint was not identified: corrections were required up to the IGY, but due to the limited nature of this study, a specific endpoint beyond this cannot be specified. In summary, the errors that we identify are assumed to stem from opera-
tional or instrument calibration errors which are relatively constant for a period of time and then change, leading to corrections which apply for the entire time period under study, or large and contiguous sub-periods. The outcome improves spatial homogeneity in the early time period. While the temporal homogeneity is also improved, it may be the case that the corrected data are not suitable for trend analysis.

2.3.2 Reconstruction of the reference series

The statistically reconstructed reference series (hereafter: “reference” or “reference series”) was generated for each candidate series, that is, for each variable (temperature and height) at five pressure levels (850, 700, 500, 300, and 200 hPa) for each station. The reconstruction is based on a multiple linear regression of upper-air series (NNR interpolated to the station coordinates) as a function of surface data (temperature and pressure). The NNR series were regressed onto the surface predictors in the period 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 in the period 1948 to 1959 and estimates of skill (see below) were derived during this validation period. Note that the quality of NNR is presumably worse in the early years. This does not affect our reconstruction, but tends to lower the skill. Hence, the skill measure is conservative.

The surface predictors were defined for 11 regions (Fig. 2.1) and encompass the first 15 principal component time series of standardized surface air temperature anomalies and the first ten principal component time series of sea level pressure anomalies, where anomalies were defined based on the 1961-1990 mean seasonal cycles. Sea level pressure data were taken from HadSLP2 (Allan and Ansell, 2006); surface air temperature data were taken from the GISS analysis (Hansen et al., 1999) which was supplemented in the Arctic by data from Polyakov et al. (2003). Only nearly complete time series were chosen, and missing values in the calibration period were replaced by NNR anomalies at 925 hPa after standardization. This substitution is possible because the reconstruction is based on standardized temperature anomalies and tests show high correlation between standardized temperature anomalies at the surface and 925 hPa (see Brönnimann and Luterbacher, 2004).

This reconstruction technique relies on two assumptions. First, that there is longer term stationarity at a given location; i.e., that the relationship between the surface and upper air variables does not change on decadal scales. The second assumption is that NNR
itself is a suitable basis for a reference series. This latter assumption could be problematic as NNR has known inhomogeneities (e.g., Santer et al., 1999; Randel et al., 2000). A few sample reconstructions were performed using the post-1958 radiosonde data as the upper-air series; the comparison to the NNR-based reconstruction is shown in Figure 2.2. The reconstructions are virtually identical and have correlation coefficients ranging from 0.982 to 0.999; additional statistical metrics are listed in Table 2.1. Reconstructions based on post-IGY radiosonde data are not feasible for the majority of stations due to short records and missing data, but the excellent agreement with NNR-based reconstructions suggests our approach is acceptable.

Figure 2.2: Time series of the difference between the two reconstructions for North Platte, Nebraska, USA: those based on NNR based minus those based on post-1958 radiosonde data. Reconstruction differences are shown for temperature (left) and GPH (right) for 850, 700, 500, 300, and 200 hPa. Correlation coefficients and mean difference between the two reconstruction for each level and variable are given in the panels.
Table 2.1: Statistics comparing the reconstructions based on NNR to those based on the data. Correlation coefficients are given for the two reconstructions ($r$) and for NNR vs the data for the calibration period (1960-2000, $r_c$) and validation period (1948-1959, $r_v$) of the reconstruction. Difference between NNR and the data in the calibration period ($NNR - dat$) and between the two reconstructions ($\Delta_{recon}$) are also given.

<table>
<thead>
<tr>
<th></th>
<th>$r$</th>
<th>$\Delta_{recon}$</th>
<th>$r_v$</th>
<th>$r_c$</th>
<th>$NNR - dat$</th>
</tr>
</thead>
<tbody>
<tr>
<td>850 T</td>
<td>0.992</td>
<td>0.22</td>
<td>0.996</td>
<td>0.997</td>
<td>0.14</td>
</tr>
<tr>
<td>700 T</td>
<td>0.998</td>
<td>0.10</td>
<td>0.999</td>
<td>0.999</td>
<td>0.07</td>
</tr>
<tr>
<td>500 T</td>
<td>0.998</td>
<td>-0.15</td>
<td>0.998</td>
<td>0.999</td>
<td>-0.14</td>
</tr>
<tr>
<td>300 T</td>
<td>0.996</td>
<td>-0.20</td>
<td>0.996</td>
<td>0.998</td>
<td>-0.22</td>
</tr>
<tr>
<td>200 T</td>
<td>0.992</td>
<td>0.23</td>
<td>0.945</td>
<td>0.975</td>
<td>0.17</td>
</tr>
<tr>
<td>850 GPH</td>
<td>0.982</td>
<td>2.12</td>
<td>0.993</td>
<td>0.975</td>
<td>2.19</td>
</tr>
<tr>
<td>700 GPH</td>
<td>0.992</td>
<td>2.82</td>
<td>0.998</td>
<td>0.995</td>
<td>2.88</td>
</tr>
<tr>
<td>500 GPH</td>
<td>0.997</td>
<td>2.01</td>
<td>0.999</td>
<td>0.998</td>
<td>2.26</td>
</tr>
<tr>
<td>300 GPH</td>
<td>0.999</td>
<td>-1.32</td>
<td>0.998</td>
<td>0.999</td>
<td>-1.84</td>
</tr>
<tr>
<td>200 GPH</td>
<td>0.999</td>
<td>-2.34</td>
<td>0.998</td>
<td>0.999</td>
<td>-2.87</td>
</tr>
</tbody>
</table>

2.3.3 Data preprocessing

The radiosonde data were pre-processed in several steps. The data were checked for outliers using climatology, the lapse rate was checked, and a check for basic hydrostatic consistency was performed. Because the data were launched at inconsistent and varying times throughout the record, each sounding was then adjusted to the daily mean. These daily mean adjustments were based on a diurnal cycle climatology from NNR and were calculated for each station, height, and month of the year (Brönnimann, 2003a). The diurnal cycle climatology in NNR could have errors but was nonetheless the only choice. The diurnal cycle adjustments are very small in the free troposphere (~ 0.1°C and 4 gpm) and the physically based errors we identify are independent of this step. Furthermore, the daily mean values calculated from a variety of original launch times at a given station were not different when tested for statistical significance (not shown), further supporting the use of an NNR diurnal cycle climatology as a best first approach. After adjusting the soundings for launch time, they were averaged into daily and monthly means for comparison with the reconstructed series. Monthly means were calculated only if the data met the following criteria for any given level: at least 13 soundings present in the month, or no gaps longer than seven days (Brönnimann, 2003a). Only
temperature and geopotential height (GPH) were examined in this study.

2.3.4 Assessment basis

The data were then assessed on a station-by-station basis by comparing the monthly mean data and the monthly mean reference series at the subset of pressure levels (850, 700, 500, 300, and 200 hPa). The difference between the two (monthly mean data – reference) is termed the bias. For each station a suite of diagnostic plots was created: mean bias as a function of pressure for both annual and seasonal means, time series of the bias at each level, scatter plots of the raw data at each level versus the neighboring level, and plots of the monthly mean real-valued data as a function of pressure.

As noted above, the technique employed here is based on the a priori assumption that there is a small number of possible errors which can be reliably identified in this manner. “Theoretical” examples of these possible errors are shown in Figure 2.3, where data from a high quality station (which required no correction) has had an artificial error added to a summer and a winter profile. The error was generated by applying the correction code to an individual sounding. The radiation and lag correction follows the general framework in Väisälä (1941, 1949) and Raunio (1950) of determining the insolation at a given height, time of day, and day of year for the station, and assumes a fixed lapse rate of 5m/s (see also Brönnimann, 2003a); the correction differs for each country only in the two parameters of lag time constant and the pressure dependence factor. This more generic correction, rather than a highly detailed instrument specific correction is more appropriate for the early data. The pressure correction is based on the physical principle of a measurement being recorded at the wrong pressure and is functionally similar to the lag correction.

The panels in Figure 2.3 show the difference between the data with the error and the original (correct) data, in order to provide a clean example of the vertical profile of these common radiosonde errors. Each of them has unique and distinct features: a units problem (geodynamic meters instead of geopotential meters, Fig. 2.3a) has an error only in height, and is of extremely large magnitude at higher altitudes. Radiation and lag (Fig. 2.3b) shows increasing bias with altitude for both temperature and GPH, and is less than about 3°C and 100 m at the tropopause. Pressure errors (Fig. 2.3c) have a similar shape in the mid-troposphere (both increasing with height) but have two notable differences from radiation and lag: 1) in all seasons, the temperature error drops to zero at the tropopause when the lapse rate changes sign, and 2) the magnitude of the height error can reach any magnitude (and is generally substantially larger than
radiation and lag). The constant temperature offset (Fig. 2.3d) shows a constant bias with altitude in temperature and an increasing, but small magnitude, bias in height as altitude increases (although any magnitude is possible for exceedingly large temperature offsets). We bundled radiation and lag together because in our experience, they are either both applied or neither has been applied in an operational manner. As the lag error is approximately constant while the radiation error changes with the time of day and season, the seasonal cycle of the errors can give further information on this. Obviously, it is possible for a station to have more than one of these errors at the same time, none of them, or other errors not addressed in the current study.

2.3.5 Reconstruction skill

In order to assess the station quality with confidence, at least five monthly means were required, and the reconstruction was required to have reasonable skill. The skill metric determined in the validation period was the reduction of error (RE) statistic (Cook et al., 1994):

\[
RE = 1 - \frac{\sum_t (x_{\text{rec}} - x_{\text{obs}})^2}{\sum_t (x_{\text{null}} - x_{\text{obs}})^2}
\]

where \(t\) is time, \(x_{\text{rec}}\) is the reconstructed value, \(x_{\text{obs}}\) is the observed value, and \(x_{\text{null}}\) is a null hypothesis or no-knowledge prediction (e.g. constant, climatology, random, persistence, etc.). In our case, since we reconstruct anomalies, the chosen null hypothesis is a zero anomaly (i.e., the mean annual cycle in the calibration period). RE can range from \(-\infty\) to 1, where \(RE = 1\) means a perfect reconstruction, \(0 < RE < 1\) means there is predictive skill in the reconstruction, and \(RE = 0\) means the reconstruction is no better than the input NNR climatology (the “no-knowledge prediction”). A random number with the correct variance would yield an RE of \(-1\). RE is preferred over correlation or explained variance, as the latter two do not account for a bias in the reconstruction, while RE does.

RE above 0.5 was chosen as a cutoff for reasonable skill in this work; this choice is somewhat arbitrary. An RE of 0.5 corresponds roughly to an explained variance of 50% or a correlation of 0.7. It was generally the case, however, that reconstructions with skill between 0 and 0.5 still captured climatic features reasonably well (that is, they were still well correlated); the lower skill tended to be reflected in variability that was too small compared to NNR or the data themselves. For this reason, the low skill reconstructions were given less emphasis, but were still consulted during quality assessment decisions.
CHAPTER 2. VALIDATION OF HISTORICAL UPPER AIR DATA

Figure 2.3: Artificial errors added to a summer (black o, Jul. 12, 1949, 16Z) and winter (grey x, Jan. 13, 1951, 15Z) profile from Frankfurt, Germany; plots show the difference between the data with the error and the original (correct) data for: a) units, b) radiation and lag, c) constant temperature, and d) pressure error. Each error has a characteristic vertical profile and unique features which allows it to be distinguished from the other errors.

2.3.6 Statistics

The target quality for bias with respect to the reference was $\pm 0.75 ^\circ C$ for temperature at all levels and $\pm 15 - 30$ gpm (increasing with height in the atmosphere and for our subset of levels) (Brönnimann, 2003b). The target precision was $1.6 ^\circ C$ and 30-80 gpm for
temperature and height, respectively (that is, 90% of the data must lie within these limits). The targets were chosen based on the ultimate application of the post-assessment data, in this case, climate variability studies; they were chosen such as to allow identification of impact-relevant climate anomalies on a monthly-to-multiannual time scale (e.g., droughts, severe winters, monsoon changes, El Niño related anomalies, etc.).

Statistical significance of the bias and the variance (too much variability) were calculated. The bias was determined significant if it exceeded twice the standard error $SE_{o-t}$, where

$$SE_{o-t} = \sqrt{\frac{s_{o-r}^2 + s_{r-t}^2}{n}}$$

and $s_{o-r}^2$ is the variance of the observations with respect to the reconstructions, and $s_{r-t}^2$ is the variance of the reconstructions with respect to the true values and is estimated from the variance of the model residuals in the calibration period, and $n$ is the number of months with data. The variance was tested according to $s_{o-r}^2 \leq s_{targ}^2 + s_{r-t}^2$ where $s_{targ}^2$ is the predefined target variance (1.6°C or 30-80 gpm) and $s_{o-r}^2$ was calculated from the observations and reconstructions. Because this variance is only an estimation that was sometimes based on a small sample, we used the lower 95% confidence limit of the estimated variance based on a $\chi^2$ distribution (Brönnimann, 2003a). The standard normal homogeneity test (Alexandersson and Moberg, 1997) was applied to the data and the reference series, while the self-homogeneity test developed by Lanzante (1996) was applied both to the time series of the raw data and of the bias. It should be noted that the statistics presented here were an aid to judgement, providing a metric for acceptable bias, but that the quality assessment process is primarily based on the vertical structure of the bias rather than minute details of the statistical tests.

### 2.3.7 Decisions

The final verdict for each station was based on these statistical tests combined with some expert judgement. For example, clusters of homogeneity breakpoints at the same time at different levels and for both temperature and GPH presented a stronger case for an inhomogeneity than an isolated result from one of the tests. This approach was taken because this quality assessment method is primarily for judging data on a station-by-station basis (that is, looking at the quality of the station as a whole) rather than a technique to statistically homogenize individual data points within a station time series. Initial verdicts fell into three categories: accept the station as is, correct the station, or reject the station. Stations which were deemed acceptable (after correction, if necessary) were further given a flag indicating higher or lower quality. Higher quality
indicated the final data had nearly zero bias at all levels, while the lower quality showed good agreement but with a small, non-zero bias (less than the target) in one or more levels.

In order to assess the quality of a station time series, the reconstruction had to have reasonably good skill, that is, above our RE threshold of 0.5. This left a large number of initially unassessed stations where the RE was below 0.5 at all or most levels for both temperature and height. Examination of the time series of the reconstruction compared to NNR from 1948 to 1957 showed that in the vast majority of cases, RE between 0 and 0.5 coincided with situations where the reconstruction clearly captured the climatic fluctuations in the time series but had variability that was too small. In this case, the reconstruction, although of low skill, was used in assessing the quality of the station time series. The three verdict categories (accept good quality, accept lower quality, or reject) were also applied to stations where the reconstruction skill was too low to make a definitive verdict. The poor skill should be noted and these stations should be rejected for applications requiring only very high quality data. Corrections to these stations were specified only if independent information existed (e.g., if a network wide radiation and lag correction was relevant) but no constant temperature adjustments were made. A small number of stations were totally unassessed due to insufficient data for assessment (less than 5 monthly means) or insufficient input data to reconstruct the reference series.

### 2.3.8 Corrections

With this technique, the diagnosis of problems is separate from the correction in most cases. Rather than performing statistical homogeneity adjustments, the majority of corrections were performed if the bias was characteristic of a problem with a known physical basis, such as uncorrected or undercorrected radiation and lag errors, pressure sensor errors, temperature sensor offsets, or units problems (e.g., geodynamic rather than geopotential meters). The corrections involved three stages: 1) the errors were preliminarily diagnosed via their characteristic vertical profile of the bias (see Fig. 2.3), 2) the corrections were applied based on independent information about the data (e.g., radiation correction as given in Teweles and Finger (1960)) so independence from the NNR-based reconstructions is maintained (and thus independence from NNR and its inhomogeneities) and 3) checking whether the application of the correction removed the bias in both temperature and GPH. The third criteria was critical—in some cases, the original bias appeared to be of one type, but the correction failed to resolve it. The
failure in some cases meant a new bias appeared (e.g., temperature was “fixed” but then GPH had a bias), or there was a remaining bias with an irregular profile. The only correction that used information from the reference series and was thus dependent on NNR (via the reconstructions) was the constant temperature correction. For many applications the correction to NNR as a standard is desirable, and it is consistent within our framework (e.g., the use of NNR climatologies to adjust the diurnal cycle and to present anomalies). For other applications it may be more desirable to correct the series with respect to another standard. 

Although Teweles and Finger (1960) refer to “solar radiation temperature correction used by various countries” (emphasis added), data from the former Soviet Union all showed signs of uncorrected radiation and lag (RL) errors, causing some confusion about whether the published corrections had been applied operationally. Nagurny (1998) also apply RL corrections to all former Soviet Union data before 1957, confirming the diagnosis of uncorrected RL errors. The specific ending date for the RL corrections was determined for each station from the diagnostic plots. After correction, some stations still had significant upper level warm anomalies up to and including 1954, suggesting that the earlier instruments might require an even stronger correction than that published in Teweles and Finger (1960), although the exact value of that larger correction remains undetermined. In general it seems that the use of country to identify radiosonde type is an acceptable approach for the early data for two reasons; during this time, the radiosonde was generally being developed at a governmental level for national or regional use rather than as a commercial product, and the radiation and lag correction we use follows a generic framework (see Section 2.3.4). The missing meta information clearly is a problem and the corrections have to be reassessed once more information becomes available. 

In all cases, decisions and correction amounts were based on the aggregate monthly mean but were applied to individual soundings which were reprocessed into new monthly means. Corrections were made to the temperature and then geopotential height was recalculated using the hydrostatic equation (Brönnimann, 2003a).

### 2.4 Example Stations

The mean bias of both temperature and height as a function of pressure is the centerpiece of the quality assessment technique, summarizing the overall quality of the station. Figure 2.4 shows four example stations with correction type: a) accepted as-
is (Caribou, Maine, USA), b) radiation and lag correction (Dikson Island, Russia), c) constant temperature offset (Wernigerode, Germany), and d) rejected due to large and unidentifiable errors (Warsaw, Poland). For the two stations which were adjusted (b and c), both before and after are shown. The error bars about the data are $\pm 1SE_{\omega-r}$ (of the data) and the error bars around the zero line are $\pm 1SE_{r-t}$ (of the reconstruction).

The skill of the reference series was low for some stations and variables. Levels with good skill are shown as circles (before correction) and stars (after correction) with heavy, solid lines. The poor skill levels are plotted with a dashed line; low skill levels were given no or very little weight in determining the quality of the data, depending on the availability of other, high-skill levels. However, they can be useful in determining the overall consistency of a correction, especially those that were applied network-wide.

The typical data problems have characteristic vertical structure in the bias as seen in Figure 2.3. For Caribou, Maine (Fig. 2.4a), the station which was accepted as-is, the bias is very close to zero at all levels for both variables. In the radiation and lag example (Diskon Island, Russia, Figure 2.4b, cf. Fig. 2.3b), the temperature and GPH biases increase with height as shown in the grey line. After the radiation and lag correction (black lines with stars), the data fall into line with the reconstruction. Constant temperature offsets were specified for stations which show the same temperature offset at all levels: a constant temperature bias with height, and a slightly increasing GPH bias with height (e.g, Wernigerode, Germany, Figure 2.4c, cf. Fig. 2.3c). In most cases, it is unclear if the data are biased or the reconstruction is; this adjustment reflects the fact that the apparent disagreement between the data and the reconstruction is resolved with a constant temperature adjustment. For this reason they are termed adjustments rather than corrections, as they adjust the radiosonde data to the NNR 1960-2000 climatology used in the reconstructions. The adjustment was considered mandatory if the offset is both significant (i.e., more than $2SE_{\omega-t}$) and larger than the target, and optional if it was significant but less than the target.

Some stations were found to have unidentifiable and uncorrectable errors, e.g., positive temperature bias and negative GPH bias (or vice versa) or biases which changed erratically in time. In these cases, correcting one variable would make the other worse for part or all of the record (e.g., Warsaw, Poland, Figure 2.4d). If these uncorrected errors were larger than the target ($\pm 0.75^\circ$C and $\pm 15 - 30$ gpm), the station was rejected; if they were within these limits, the station was accepted but coded as lower quality. In the case of Warsaw, the large error in the bias plot is clearly reminiscent of a pressure sensor error (cf. Fig. 2.3d). However, the conflicting temperature and height errors (positive temperature bias with negative height bias at low levels), and a worsening of
some levels and variables after applying the proposed correction, led to a decision to reject the entire station. This decision was also made because the pressure correction requires good skill in the upper troposphere to fine tune the magnitude of the pressure error (an iterative process); the low skill at higher altitudes precludes this adjustment at this time.

About one third of the stations could not be assessed with complete confidence due to poor skill in the reconstruction (below 0.5 at most or all levels in both temperature and GPH). Because the reconstruction does still capture some information about the climate, the vertical bias plots were examined and the unassessed stations were categorized as accept or reject, although the lack of reconstruction skill should be kept in mind. Some of these stations with poor skill were specified as needing a radiation and lag correction if they were part of a network-wide correction, such as in the former Soviet Union, and the low-skill reconstruction plots showed improvement after the correction, confirming their usefulness. No constant temperature adjustments were specified for these stations.

In Figure 2.5, time series for the examples shown in Figure 2.4 are presented for both temperature and height on the subset of pressure levels used for quality assessment. In the case of the rejected station, the decision was based on the large and conflicting errors in temperature and pressure. In particular, 700 hPa has a large positive temperature bias but a small positive GPH bias alternating with no GPH bias, while 500 hPa shows a positive temperature bias but nearly zero GPH bias, and 200 hPa shows no temperature bias and a large positive GPH bias except for some months in 1952. The poor skill in the temperature reconstruction complicates the picture. The accepted stations, on the other hand, show biases which are internally physically consistent and relatively stable over time.

Figure 2.6 shows a time series of bias at 500 hPa for Yekaterinburg (formerly Sverdlovsk), Russia, a station with a units problem. When upper air measurements were first developed, heights were normally reported in geodynamic meters rather than geopotential meters. The difference between geodynamic and geopotential meters increases with height and can be hundreds of meters in the stratosphere yet is very difficult to detect in time series of real-valued data, such as in Figure 2.6a. In a time series of anomalies from the reconstruction (Figure 2.6b), the problem is easily detected.

Time series of the number of corrected stations (as a fraction of the number of stations) and the total number of stations are shown in Figure 2.7. Note that in the time series of corrected stations, individual stations may be represented more than once at any timestep if they had multiple corrections, and that the number of stations is presented on
Figure 2.4: Bias of the data relative to the reconstruction as a function of pressure for temperature (left) and GPH (right); data are shown for beginning of record to the end of 1957 or the end of record, if before 1957. Grey lines show before correction and black lines show after correction. Dashed lines indicate that the reconstruction did not meet the predefined metric in quality. Error bars about the data are $\pm 1SE_{o-r}$ and those about the zero line are $\pm 1SE_{r-t}$ of the reconstruction. Examples are shown for: a) station requiring no correction (Caribou, Maine, USA), b) radiation and lag correction (Dikson Island, Russia), c) constant temperature adjustment (Wernigerode, Germany), and d) a rejected station (Warsaw, Poland).
Figure 2.5: Time series of temperature (left within pair) and height (right within pair) relative to the reconstruction at 850, 700, 500, 300, and 200 hPa. Grey lines show before correction and black lines show after correction. Dashed lines indicate that the reconstruction did not meet statistical significance in quality. Examples are shown for the same stations as in Fig. 2.4: a) station requiring no correction (Caribou, Maine, USA), b) radiation and lag correction (Dikson Island, Russia), c) constant temperature adjustment (Wernigerode, Germany), and d) a rejected station (Warsaw, Poland).
CHAPTER 2. VALIDATION OF HISTORICAL UPPER AIR DATA

Figure 2.6: Time series of (a) raw data and (b) anomaly from the reconstruction for Yekaterinburg (formerly Sverdlovsk), Russia at 500 hPa, before (grey) and after (black) units correction. A log scale. The sparsity of data in the early years causes the correction time series to appear jumpy before 1940. In some cases the ending time for the correction was clear (e.g. the network wide radiation and lag correction in the former Soviet Union and the units problems where there was a clear step-function change in the bias time series). In other cases the correction was applied to the entire time period; this is especially the case with the constant temperature adjustments. Because this study focused on pre-IGY data, this led to an unspecified ending time for many corrections. Corrections were identified and applied only up to the end of 1957 (or earlier, if specified).

Figure 2.7: Number of stations which required a correction as a fraction of the number of stations with data for each year (top), and the number of stations (bottom).
These examples cover simple, easily corrected errors in the data which have nonetheless eluded detection using standard quality control procedures but that we were able to identify by examining anomalies from a reconstructed reference series. A map showing station data before and after correction, differences, and required correction type is shown in Figure 2.8, revealing how widespread these errors are. Units and radiation and lag were generally confined to the former Soviet Union, while the constant temperature corrections reveal no underlying pattern. This is discussed in more detail in Section 2.5.

During the crosscheck of duplicated stations in the original data collected for this study, some stations were identified with incorrect timestamps. In one example, Aktjubinsk, Kazakhstan, the data from the TD-54 source had two soundings per day (03Z and 15Z), while the same station in the IGRA database had four soundings per day (03Z, 05Z, 16Z, and 17Z). On closer inspection, the two additional soundings in IGRA (05Z and 17Z) were found to be duplicates of the other two in IGRA but with a timestamp two hours later, consistent with the station being at UTC+3. A complete list of such stations was not made, as the decision to prioritize other sources over IGRA had already been made and the problem was found only in the early part of the record. Where noted, this information is added to the cross-referenced station list. Table 2.2 presents a summary of statistics describing the correction types both globally and by network.

### 2.5 Large Scale Impacts of the Corrections

The widespread geographical extent of the corrections has implications for hemispheric scale analysis figures because the extent of the errors is on the same scale as large scale atmospheric features such as planetary waves. Station data for 1956 showing 500 hPa temperature and height are presented in Figure 2.8. The original and corrected data are shown as well as the difference between them; data are shown as anomalies from a 1961-1990 NNR climatology for each station and the correction type is coded by the symbol in the difference plots. The corrected data are colder and the heights are lower across the whole of Eurasia, reflecting the network wide radiation and lag corrections. Constant temperature adjustments are more isolated and do not exhibit such coherency in sign and scale. The temperature and height fields match each other better, and thus are more physically consistent, after correction. The spatial structure of the correction is important, especially spatially large networks such as the former Soviet Union which introduce errors of the spatial scale of planetary waves. The unit corrections (necessary only on data prior to 1950 in most cases) caused a raising of the
layer which is larger than the magnitude of the radiation and lag correction, leading to a situation where temperatures decreased but heights increased after correction before 1950.
Table 2.2: Stations requiring correction by sonde type as a proxy for network (see station list for complete details). The former Soviet Union is labeled “SU”; “SU strong” are former Soviet Union stations requiring a stronger-than-published correction up to and including 1954.

<table>
<thead>
<tr>
<th></th>
<th>Vaisala</th>
<th>US</th>
<th>SU</th>
<th>SU strong</th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td>Radiation &amp; lag</td>
<td>31</td>
<td>3</td>
<td>177</td>
<td>26</td>
<td>237</td>
</tr>
<tr>
<td>GPH units</td>
<td>3</td>
<td>0</td>
<td>64</td>
<td>18</td>
<td>85</td>
</tr>
<tr>
<td>Constant Temperature (bias: +,-)</td>
<td>29 (18,11)</td>
<td>23 (15,8)</td>
<td>39 (32,7)</td>
<td>5 (4,1)</td>
<td>96 (69,27)</td>
</tr>
<tr>
<td>Optional Const. Temp. (bias: +,-)</td>
<td>7 (7,0)</td>
<td>42 (7,35)</td>
<td>45 (45,0)</td>
<td>4 (4,0)</td>
<td>98 (63,35)</td>
</tr>
<tr>
<td>no correction</td>
<td>106</td>
<td>279</td>
<td>55</td>
<td>0</td>
<td>440</td>
</tr>
<tr>
<td>Accepted, high quality</td>
<td>109</td>
<td>296</td>
<td>141</td>
<td>26</td>
<td>572</td>
</tr>
<tr>
<td>Accepted, lower quality</td>
<td>66</td>
<td>50</td>
<td>99</td>
<td>0</td>
<td>215</td>
</tr>
<tr>
<td>Rejected</td>
<td>24</td>
<td>9</td>
<td>28</td>
<td>0</td>
<td>61</td>
</tr>
<tr>
<td>Unassessed (no data or recon)</td>
<td>20</td>
<td>11</td>
<td>0</td>
<td>0</td>
<td>31</td>
</tr>
<tr>
<td>Total</td>
<td>219</td>
<td>366</td>
<td>268</td>
<td>26</td>
<td>879</td>
</tr>
</tbody>
</table>
Figure 2.8: 1956 annual mean anomaly of 500hPa temperature (left) and GPH (right) from original data (top), corrected data (middle), and difference (corrected-original; bottom). Anomalies are from a 1961-1990 NNR climatology. In the difference temperature (lower left), radiation and lag corrections are indicated with a plus, constant temperature corrections are indicated with right and left facing triangles for positive and negative biases, respectively. White circles indicate a station with no (or rejected) data. In the GPH difference (lower right panel), units corrections are indicated by upward pointing triangles. The radiation and lag and constant temperature corrections affect both temperature and height but the symbols are separated for clarity.
A considerable fraction of the data assessed in this study was assimilated into NNR. Examination of real valued annual mean data compared to NNR reveals that NNR has a warm bias over much of Eurasia in the 1950s. Figure 2.9 shows annual mean NNR minus annual mean station data in 1956 for the Northern Hemisphere using both the original and the corrected data. NNR has a very slight cold bias over Eurasia compared to the uncorrected data, but the errors are less coherent in sign and location and suggest that NNR is well constrained by the assimilated (erroneous) data. When compared against the corrected data, however, NNR has a coherent warm and high bias over Eurasia. This hemispheric scale bias would have an impact on analyses of NNR prior to 1958, particularly on principal component analyses of large-scale fields such as teleconnection studies. NNR prior to 1958 should be treated with some caution. Note that radiosondes are not the only source of error in NNR, and that, conversely, the new radiosonde product is by no means error free. Furthermore, the topography of the Tibetan Plateau is clearly revealed by the large swath of negative height biases in both the before and after figures. It is well known that NNR has some difficulty over significant terrain features, in particular showing a cold and low bias over the Himalayas (e.g., Xie et al., 2008). Furthermore, the reconstructions in the area consistently had low skill, hampering our ability to stringently assess the quality of stations in this region.

2.6 Conclusion

Worldwide radiosonde data from prior to 1958 were compiled and reevaluated using a novel quality assessment technique which creates a statistically reconstructed reference series. Errors which otherwise prove difficult to detect are identified more clearly when the data are compared to such a reference series. Pervasive errors were found in some networks over the Eurasian continent affecting both temperature and height. In particular, radiation and lag errors were common; these errors, which affect both temperature and height, are easily corrected using published corrections (Teweles and Finger, 1960). Unit errors in the geopotential height were also common. These simple errors have a large impact on resulting analysis of the spatial field as they affect data over a geographical area that is on the same scale as atmospheric phenomena such as planetary waves which can be analyzed through modes of variability (such as the Arctic Oscillation/North Atlantic Oscillation or Pacific North America pattern). Some stations were also found to have a constant temperature offset relative to the reconstruction which must be corrected if the data are to be merged with NNR for other purposes. The corrected data show improved internal consistency between temperature and height.
Figure 2.9: 1956 annual mean NNR minus raw (top) and corrected (bottom) temperature (left) and GPH (right) at 500 hPa. NNR appears to have a very slight cool bias relative to the uncorrected data, but generally shows good agreement with some random errors. NNR has a widespread warm and high bias relative to the corrected data due to the assimilation of uncorrected data. Note the persistent low bias over the Tibetan Plateau where NNR has difficulty with the terrain.
Of 879 unique stations, 787 were of sufficient quality after quality assessment and correction (if necessary) to be kept, while 92 stations were rejected due to unresolvable errors or data sparsity. Monthly mean data are available in an archive along with a complete index of stations (cross referenced to several source archives) and the verdicts.

Examination of NNR revealed that the reanalysis closely follows the assimilated data leading to a widespread warm and high bias in the 1950s due to assimilation of large amounts of uncorrected data. Caution should be used in interpreting hemispheric scale analysis of NNR prior to 1958.

The homogenization of historical radiosonde data is a difficult process which has been tackled many times. While errors undoubtedly remain in the data, the corrections presented here are one step in the process of increasing the usefulness of the earlier data and providing the climate community with a new comprehensive dataset.

**Acknowledgments**

We thank the anonymous reviewers for helpful comments which resulted in a much improved paper. This work was supported by the Swiss National Science Foundation. The TD54 and CARDS542 datasets were kindly provided by Joey Comeaux at NCAR, and the TD6201 dataset by Joe Elms at NCDC. Lindenberg, Germany data were provided by the German Meteorological Service, the Payerne data by MeteoSwiss, and the Ilmala data by the Finnish Meteorological Institute.
Chapter 3

Vertical Structure of 20th Century Arctic Warmings

This chapter utilizes the new data from Chapter 2 to examine the vertical structure of the two main warming events in the Arctic in the 20th Century. Section 3.1 (Grant et al., 2008c) motivates the analysis by comparing zonal mean trends in ERA-40 reanalysis versus radiosonde data, a response to a recent Nature paper (Graversen et al., 2008) which examined the vertical structure of the recent Arctic warming in ERA-40 reanalysis. Section 3.2 (Grant et al., 2008a) explores the regional and seasonal nature of the two warm periods in the upper air temperature record. Anomalous circulation during the early warming is examined in reconstructed geopotential height, and a possible cause of the early warming involving anthropogenic aerosols is presented.
3.1 Reanalysis over data sparse regions

Recent Arctic warming vertical structure contested

A. N. Grant¹, S. Brönнимann¹, L. Haimberger²

¹Institute for Atmospheric and Climate Science, ETH Zurich, Universitàtstrasse 16, CH-8092 Zurich, Switzerland
²Department of Meteorology and Geophysics, University of Vienna, Althanstraße 14, A-1090 Vienna, Austria


The vertical structure of the recent Arctic warming contains information about the processes governing Arctic climate trends. Graversen et al. (2008) argue, on the basis of ERA-40 reanalysis data (Uppala et al., 2005), that a distinct maximum in 1979-2001 warm-season (April-October) Arctic temperature trends appears around 3 km above ground. Here we show that this is due to the heterogeneous nature of the data source, which incorporates information from satellites and radiosondes. Radiosonde data alone suggest the warming was strongest near ground.

Graversen et al. (2008) claim that the warm-season temperature trend has a maximum at around 700 hPa, polewards of 75°N, and argue that anomalous heat advection from more southerly latitudes is important. However, the ERA-40 reanalysis may not be suitable for trend analysis as it incorporates information from different observing systems such as satellite and radiosonde, which might be inconsistent, in particular with respect to trends (Fu et al., 2004; Mears and Wentz, 2005). Radiosonde measurements provide vertically resolved temperature profiles in the troposphere, whereas satellites provide information on a weighted average over a thick layer. Furthermore, the ERA-40 assimilation system extrapolates information from data-rich to data-sparse areas,
which is less reliable than observations. The ERA-40 reanalysis in the polar region has not been sufficiently validated by in situ observations and documented problems with satellite radiance assimilations over the Arctic Ocean (Uppala et al., 2005; Bromwich and Wang, 2005) could lead to spurious trends.

A map of warm-season trends at 700 hPa (the peak level of the polar warming trend in Graversen et al. (2008)) from ERA-40 and radiosonde observations (Durre et al., 2006; Haimberger, 2007) confirms that the enhanced warming signal lies mostly in areas with no radiosonde data coverage (Fig. 3.1a). This is particularly so polewards of 75°N, where the trend appears strongest in Graversen et al. (2008). Moreover, the few radiosonde data available near or polewards of 75°N show modest trends. To illustrate the effects on the vertical structure of the trend, we calculated zonally averaged vertical temperature trends from (1) ERA-40 reanalysis data, (2) such data subsampled to locations where radiosonde information is available (that is, where ERA-40 is best constrained) and (3) from only radiosonde data. The trend in the reanalysis (Fig. 3.1b) is a reproduction of Fig. 4a in Graversen et al. (2008) and exhibits a maximum at 700 hPa, polewards of 75°N. Subsampling the ERA-40 reanalysis (Fig. 3.1c) reveals clearly different trends, and calculating trends directly from radiosondes alters matters even further (Fig. 3.1d). The result is independent of the methods used to homogenize the radiosonde data (unadjusted, RAOBCORE v1.4 (Haimberger, 2007) and RICH (Haimberger et al., 2008); not shown). The radiosonde data (note that some regions are not well covered and some levels are missing because of inconsistent reporting) have their strongest trend near the ground, not above the boundary layer as in the full reanalysis. This is important because boundary layer processes are much more locally driven and simultaneously not well represented in a reanalysis. The same result is found when analysing a subregion with relatively even radiosonde coverage (inset, Fig. 3.1a), and during the remainder of the year (not shown\(^1\)).

Arctic climate is controlled by processes operating on scales from local to global, including transport effects; forcings such as greenhouse gases, aerosols and clouds; and feedbacks such as the well-known sea-ice-albedo feedback. The temperature profile can be a clue to the underlying processes, but to disentangle the contributions to Arctic temperature trends fully, vertical temperature structures should be addressed in a regionally and seasonally resolved manner. Furthermore, the large interannual variability in the Arctic, coupled with the sensitivity of trends to both end points and season definitions, suggests care should be taken in interpreting trends over short periods.

\(^1\)See appendix to this section in the thesis.
In conclusion, some features of the temperature trends calculated in Graversen et al. (2008) reflect possible inhomogeneities or artefacts in the ERA-40 reanalysis rather than true climate signals, as they appear not to be supported by observations. ERA-40 reanalysis is a valuable tool in calculating circulation effects, especially on a sub-decadal basis, but inhomogeneities and gaps in the global observing system tend to make trends from reanalyses unreliable, particularly in data-sparse regions.
Figure 3.1: Vertical structure of Arctic temperature trends for April to October, 1979-2001. Trends were calculated from seasonally averaged monthly anomalies using least squares regression (not more than one missing month per season allowed, not more than five missing seasons in 1979-2001, neither first nor last two years can be missing). (a) Trend field at 700 hPa from ERA-40 (Uppala et al., 2005) and from radiosonde data (Durre et al., 2006; Haimberger, 2007) (circles) with 75°N latitude circle indicated by the thin solid line. (b) Trends of zonal mean temperature as a function of latitude and altitude from ERA-40. (c) Same as (b), but from ERA-40 subsampled to the locations and times where radiosonde data is available (anomalies zonally averaged in equal-area latitude bands). (d) Same as (c) but for radiosonde data. Inset in (a) average trend profiles of the region 58°N-82°N, 100°W-25°E for full ERA-40 (solid line), subsampled ERA-40 (dashed) and radiosonde data (dotted).
Appendix

Arctic trends in winter showed similar disparities in the full reanalysis versus subsampled reanalysis and radiosondes, as seen in Figure 3.2, which shows trends for the winter half of the year (Nov to Mar) for 1979-2001 (specifically, Nov 1978 to Mar 2001). The deviation of the full reanalysis is especially apparent at the tropopause over the pole, where the trend was strongly affected by the extreme outlier months of February and March 2001. The trends in Graversen et al. (2008) were calculated as the linear regression of the individual months rather than as a trend of seasonal means, magnifying the impact of outliers.

The spatial pattern of warming appears quite different in winter compared to summer, although in both cases the strongest trend is over the area north of the Canadian Archipelago and Greenland. The assimilation of HIRS (High resolution Infrared Radiation Sounder) satellite radiances into ERA-40 introduced a cold bias to the lower tropospheric temperatures over ice-covered oceans for the years 1979-1996 (Bromwich and Wang, 2005; Uppala et al., 2005); the removal of the bias starting in 1997 introduced a known inhomogeneity over the ice-covered oceans. That the strongest trends in both winter and summer in the full ERA-40 reanalysis are over the area of the Arctic Ocean with the most consistent (and thickest) coverage of sea ice supports the likelihood that the anomalous trend is spurious.

While surface and 700 hPa temperatures might not be expected to have the same trends, a casual comparison of Figures 3.1 and 3.2 with surface air temperature (SAT) trends for January and July from 1980–1999 in Kuzmina et al. (2008) shows some similarities. All four SAT datasets in their Figure 4 show a band of cooling stretching from Central Russia (60°-90°E) and the Kara Sea eastward along the coast and then south into Alaska and northwestern Canada. A similar feature appears in Figure 3.2. In summer, the 700 hPa trends also show similarity to the SAT trends in showing warming nearly everywhere except a small region around central Asia (Turkmenistan and Tajikistan). It is especially interesting to note that the ERA-40 July SAT trends (Kuzmina et al. (2008) Figure 4b, far right) do not show warming over the Arctic Ocean, as the reanalysis here is based on assimilated sea surface temperatures which are not contaminated by the HIRS inhomogeneity.
Figure 3.2: As in Figure 3.1, but for winters 1979-2001 (November 1978 - March, 2001)
3.2 Arctic warmings: An upper level perspective

The 20th century Arctic warm events from an upper level perspective

Andrea N. Grant, Stefan Brönnimann, Thomas Griesser, Tracy Ewen

Institute for Atmospheric and Climate Science, ETH Zurich, Switzerland

In preparation for Journal of Climate

Abstract

Newly validated radiosonde and aircraft data extending back to 1932 were examined for evidence of the warm event that occurred in Arctic surface air temperatures in the 1930s and 1940s and the ongoing Arctic warm period. The strongest signal of the early warm period was found in wintertime over Europe and western Russia. Sparsity of data in the early part of the record make it difficult to conclusively report the vertical structure of the warming, but the results suggest that, in the early warm period, some years showed vertical coherence and others did not. The above normal temperature signal was most congruent in the 1930s and began to disintegrate in the 1940s. The recent warm period, in contrast, was most strongly seen in the region around the Beaufort Sea and Bering Strait in winter and spring and exhibited coherent warming in the troposphere although the warming was strongest at the ground. Some warming was seen in all regions and seasons in the recent period and the warming is becoming more widespread; that is, that signal is a trend rather than an isolated warm event. Circulation based on reconstructed geopotential height showed excellent agreement with above normal warm air advection into the Atlantic sector of the Arctic in the 1930s which had previously been seen in sea level pressure fields. The zonal circulation is even more enhanced at 700 hPa and shows development of both strengthened cyclonic circula-
CHAPTER 3. VERTICAL STRUCTURE OF ARCTIC WARMINGS

3.2.1 Introduction

Recent decades have seen an increase in tropospheric temperatures across the globe and a simultaneous decrease in stratospheric temperatures (e.g. Thompson and Solomon, 2005). This coupling of trends is considered to be a signature reaction to a change in greenhouse gas forcing, which is highly likely to be due to anthropogenic greenhouse gas emissions (Solomon et al., 2007). From the 1920s to the 1940s, the Arctic experienced a dramatic warm event which is generally accepted as an episode of natural variability (Johannessen et al., 2004; Wang et al., 2007). The recent warming, since the 1990s, is seen as a combination of natural variability and anthropogenic forcing. Investigating the vertical structure of these two events would elucidate their mechanisms (radiative versus dynamic) and also clarify if the two events are related, both of which would aid Arctic climate modelers.

Climate trends have different signatures (often called “fingerprints”) in spatial and vertical extent depending on the forcing driving them. Models can be used to describe the spatial and vertical structure expected for a given, isolated forcing. Observational data are then analyzed with respect to detecting these fingerprints, which can lead to the attribution of a trend to a forcing. In the current study, the vertical structure of the Arctic warming is proposed to potentially serve as a means to separate forcings or to identify feedback processes.

Greenhouse gases and solar variability, while both having an effect on global mean surface temperature, can be distinguished via their impact on the stratosphere: modeled response to CO$_2$ shows a warming surface and troposphere (the warming increases with height most strongly in the tropics with a weakening gradient poleward) and a cooling stratosphere while solar variability has been modeled to have a similar response in tropospheric temperatures and a very weak warming of the northern hemisphere stratosphere and cooling of the southern hemisphere stratosphere in zonal and annual
means (e.g. Joshi et al., 2003; Hansen et al., 1997). A decrease in stratospheric ozone has a similar pattern in temperature response as an increase in CO₂ but with much larger magnitude of changes in temperature and stronger gradients in vertical temperature structure (Joshi et al., 2003). Volcanic eruptions produce a transient (1-2 years) cooling of the surface and warming of stratospheric temperatures from aerosols injected into the stratosphere, although secondary dynamical effects cause a warming of northern hemisphere winter surface temperatures (Robock, 2000). Tropospheric aerosols which scatter shortwave radiation (such as sulfates) have a cooling effect in the global mean on the surface while shortwave-absorbing aerosols in the troposphere such as black carbon have a cooling effect at the surface (reduced downwelling shortwave) and a warming effect in the aerosol layer (absorption of shortwave) (e.g. Haywood and Boucher, 2000). Because black carbon can interact with the climate system via numerous pathways, it can play an especially important role in regional climate change which is still a matter of controversy (Ramanathan and Carmichael, 2008). In polar regions and on glaciers, black carbon deposited on the surface can have a net warming effect (enhanced surface absorption of downwelling shortwave from lowered surface albedo overwhelms the cooling from reduced insolation) (McConnell et al., 2007), while the increased longwave emissions from clouds in the presence of sulfates overwhelms the shortwave effects in polar night, producing a net warming at the surface (Garrett and Zhao, 2006; Lubin and Vogelmann, 2006). All tropospheric aerosols have impacts on regional scales as the aerosols are relatively short lived (1-2 weeks). Furthermore, feedbacks operating in the climate system can complicate the picture. The well know ice-albedo feedback has a temperature signature at the surface from changing ice coverage, while changes to cloud coverage from either aerosol indirect effects or height of saturation levels (warmer air being less cloudy if all other things are held constant) would affect the surface via altered shortwave and longwave energy but can also have stratospheric effects because of these same changes in reflected shortwave and emitted longwave radiation (e.g. Fueglistaler and Fu, 2006). Finally, if forcings act to change atmospheric dynamics, or dynamics change through natural variability, other mechanisms come into play. Changes in advection which are coherent throughout the troposphere would cause vertically coherent changes in temperature.

A study by Graversen et al. (2008) looked at the vertical structure of the recent (1979 to 2001) Arctic warming in the ERA-40 reanalysis trends; they found the peak of the warming to be at about 700 hPa in summer for zonally averaged data². Furthermore, they found a relationship between the 500 hPa temperature field and northward energy

²More fundamental questions about the quality of reanalysis in data sparse regions were addressed in Section 3.1.
transport (across the 60N circle) which they claimed explained a “substantial” part of the Arctic temperature trends (in fact only $\sim 10\%$). Arctic climate, however, is influenced by regional processes and circulation patterns and looking at zonal means obscures these mechanisms. It is important to examine regional and seasonal features of the two major warm periods in the Arctic in the 20th century in addition to their vertical structures.

### 3.2.2 Previously Proposed Mechanisms

#### The early warm period

Instrumental records show a strong jump in Arctic surface air temperature (SAT) from gridded CRUTEM3v data (Brohan et al., 2006) in 1920 followed by a warm period in the 1930s and 1940s, as seen in Figure 3.3. This warm period is found in the entire northern hemisphere, although it is has the largest magnitude in the Arctic. Investigation of proxy data suggests that this warming was actually the tail end of a gradual increase which began around 1850 (Overpeck et al., 1997), although this early feature is not apparent in Figure 3.3, perhaps due to the limited spatial extent and changing spatial coverage of the earliest parts of the instrumental record or the difference between proxy and instrumental records. Based on surface data, several mechanisms have been proposed for the early 20th century warming which can be broadly grouped into four themes: circulation anomalies, natural variability, feedbacks, and changes in forcing.

#### Circulation anomalies

Three studies have examined sea level pressure (SLP) during the warm period and all found unusual warm air advection. Rogers (1985) found that from 1900 to 1925, there was a high frequency of years with a strengthened Icelandic low causing increased warm air advection into Europe (and thus warmer SAT) accompanied by cool temperatures in Greenland, a classic strengthening of the zonal circulation. From 1925 to 1944, that pattern was rarely seen, instead one of two cases was present: a) when the Icelandic low was strong (deep), it was farther southwest and was rotated, allowing warm air advection into both Europe and Greenland (that is, the air advecting onto Greenland came from around 65N rather than 70-75N while Europe experienced the classic increased zonal circulation), and b) when the Icelandic low was weak, it’s position was variable and more polar highs crossed
Europe. Fu et al. (1999) split the early warm period into three eras: a) 1903-1918—before the large temperature jump—which had an enhanced North Atlantic Oscillation (NAO, or zonal circulation), b) 1919-1929—during the jump—with a further strengthening of the circulation, and c) 1930-1940—after the jump—with a weakening of both the Icelandic low and the Azores high and a disorganization of the pressure anomalies. Overland and Wang (2005) found a trough of low pressure extending from northern Quebec to Iceland for the winters of 1928-1935 which changed the circulation around Greenland. Hanssen-Bauer and Forland (1998), on the other hand, created a multiple linear regression model of SLP (on Svalbard and northern Norway) and temperatures on Svalbard which represents the link between regional circulation and temperature on Svalbard. They found that this model cannot reproduce the early warm period, leading them to conclude that other factors, for example sea ice and sea surface temperatures (SST), must have played a role.

Natural variability

Polyakov et al. (2003) found evidence of a low frequency oscillation in Arctic temperatures and sea ice extent in the Atlantic sector, although Johannessen et al. (2004) argued against such a recurring cycle due to the dif-
ferent spatial patterns of the two warm periods. Johannessen et al. (2004) ascribed the early warm period to natural variability in their modeling study, as did Wang et al. (2007).

Feedbacks

Bengtsson et al. (2004) presented a hypothesis based on a local pressure gradient between the northern coast of mainland Norway and Svalbard which induced anomalous advection of warm water into the Barents Sea, thus reducing sea ice and increasing surface heat flux and SAT. Hanssen-Bauer and Førland (1998) also discuss the possibility of feedbacks due to a retreat of the sea ice edge as playing a role in the early warming.

Changes in forcing

Delworth and Knutson (2000) attributed the global early warm period to a combination of natural variability and early anthropogenic greenhouse gas and sulfate forcing, while Overpeck et al. (1997) found a better match with decreased volcanic activity and increased solar activity. Soon (2005) found a striking concordance between solar activity and Arctic SAT, although Bengtsson et al. (2004) points out that the evidence of long term changes in solar irradiance are still controversial.

The recent warm period

While the early warming is generally accepted to be an episode of natural variability, the consensus on the recent warming is that it involves anthropogenic influences, in particular well mixed greenhouse gases. Modeling studies are unable to reproduce the recent Arctic warming without including anthropogenic forcing (Johannessen et al., 2004; IDAG, 2005), although a universal “fingerprint” of greenhouse gas induced warming in Arctic SAT is not yet clear (Serreze et al., 2000), at least not in the emerging phase of such a warming (Serreze and Francis, 2006). Mechanisms for the recent warming include an increase in downwelling longwave radiation due to an increase in clouds and water vapor which is driving the change in ice edge position (Francis and Hunter, 2006). Changes in advection due to strongly positive values of the Arctic Oscillation (AO) since 1989 and a change in the Pacific North America (PNA) and Pacific Decadal
Oscillation (PDO) indices since the late 1970s have also been noted (e.g. Turner et al., 2007; Overland et al., 2002). The AO has returned to more neutral values in the past several years, while the Arctic has remained warm, ruling out the positive AO as the sole cause of the warming. The cluster of positive AO years were likely still an important part of the recent warming due to initiation of a sea-ice feedback following the export of large amounts of thick, multi-year ice during the positive AO years (Rigor and Wallace, 2004; Stroeve et al., 2005). Changes to the land surface in Arctic regions is playing a role in high latitude warming by altering surface albedo due to changing vegetation and duration of the snowpack (Chapin et al., 2005). It is likely that the increasing greenhouse gas concentrations will operate both via a direct radiative effect and through changes in dynamical mechanisms, e.g. possibly increasing the likelihood of positive AO states (Kuzmina et al., 2005).

3.2.3 Data

Recently digitized and validated radiosonde data (Grant et al., 2008b) allow a view of the temporal evolution of the vertical temperature structure over the Arctic. The upper air data presented in this study help present a clearer picture of how the vertical structure of the early warm period differs from the recent warming in the Arctic, allowing the two warmings to be distinguished more conclusively. Note that in this paper the term “warming” is used to indicated a warm period rather than a warming trend.

We note that while it is tempting to calculate trends when presented with a time series, care must be taken in interpreting the results. Linear regressions are notoriously sensitive to endpoints (e.g. Lanzante, 1996) and can be foiled by both large interannual variability and the presence of multi-year autocorrelation. Being overly reliant on trends can also obscure the details of what is happening. In this work we instead choose to explore the time series themselves rather than trends.

The upper air data in this study extends back to 1932 at the earliest, thus we focus on the 1930s and 1940s which are the heart of the early warm period. Data coverage is very uneven prior to 1948 and is primarily concentrated in Eurasia. For the recent warming we focus on the 1990s and 2000s when widespread evidence of warming was apparent, although hints of the warming can be seen in the 1980s. SAT data is from CRUTem3v (Brohan et al., 2006), and SLP is from HadSLP2 (Allan and Ansell, 2006).

Geopotential height fields were reconstructed using a principal component regression model (Griesser et al., in preparation) similar as in earlier work (Brönnimann and Luterbacher, 2004). The predictors were historical surface and upper air temperature and
pressure time series, the predictands were geopotential height and temperature fields from ERA-40 reanalysis (Uppala et al., 2005). The model was calibrated in the period 1958-2001 and optimized using split sample validations within that period (as in Brönnimann and Luterbacher, 2004).

### 3.2.4 Regional and seasonal means

In this analysis, the Arctic was divided into the 54 equal area gridcells shown in Figure 3.4. The northernmost cells were stretched slightly to accommodate isolated stations. Anomalies of individual soundings were calculated relative to a 1961-1990 monthly climatology from NCEP3-NCAR4 Reanalysis (NNR, Kistler et al., 2001). Climatologies were also calculated using ERA-40 reanalysis and the data itself and were virtually identical. Due to the short nature of many of the records, using a self-climatology resulted in much more data being thrown out due to lack of climatology; thus an NNR climatology was chosen for consistency with other datasets used in this study. The anomalies of the individual soundings were then regridded into the equal area gridcells and 7 day blocks. The scale of both the gridcells and the 7 day blocks were chosen as representative of the synoptic scale: the cells are about 800km on a side and were the largest area (and longest time) that could be seen to be realistically represented by a single point. The four seasons were defined as the thirteen week periods of 1 December to 1 March (winter), 1 March to 31 May (spring), 1 June to 31 August (summer) and 1 September to 1 December (autumn). Seasonal-regional means were calculated from the gridcell averages if 50% of the gridcells in a region and 50% of the weeks in the season had data. Time-height datapoints that did not meet this criteria are still plotted but are hatched. The failure to meet the criteria was more likely to be due to missing data rather than rapid changes in the station network, therefore they could still be comparable to neighboring points although clearly not as reliable in representing the larger region and season. All discussion is based on non-hatched time-height datapoints.

Different combinations of gridcells within 60-degree-longitude wedges were combined to inspect the spatial coherency of the temperature structure within the 60 degree wedges (e.g., for 0° to 60°E, we examined [1, 25, 43]; [1, 2, 25, 43]; [1, 2, 35, 36, 43]; [2, 3, 26, 43, 44]; [3, 4, 27, 44]; and [1, 2, 3, 25, 26, 43, 44]). With the exception of the area near the Bering Strait, the different possible combinations of gridcells within a 60 degree wedge gave virtually the same picture in the vertical time series. Gridcell

---

3National Centers for Environmental Prediction, part of the U.S. National Oceanic and Atmospheric Administration

4U.S. National Center for Atmospheric Research
combinations with the best overall data coverage were chosen as representative, and the presence of above normal temperatures at the surface and in the free troposphere was noted for the four classically defined seasons (DJF, MAM, JJA, SON) and the four decades of the 1930s, 1940s, 1990s, and 2000s (along with other striking features in the time series). These representative timeseries for all four seasons are shown in Figures 3.5 to 3.8. The levels shown are: SAT (from CRUTem3v (Brohan et al., 2006)), 850, 700, 500, 400, 300, and 200 hPa. 925 and 600 hPa were omitted due to irregular reporting. For clarity of identification, the regions have been named as follows: Europe (0°-60°E), central Russia (60°-120°E), east Siberia (120°-150°E), Bering (150°-210°E), Beaufort (210°-240°E), North America (240°-300°E), and Greenland (300°-360°E). The results are, by definition, difficult to summarize, but the major cold and warm features are concisely noted on a map in Figure 3.9.

Figure 3.4: Map showing the radiosonde stations in the Arctic used in this study along with the equal area gridcells used for regional averaging.
Figure 3.5: Time series of radiosonde temperature anomalies (relative to 1961-1990 NNR climatology) as a function of height for winter (DJF). The vertical structure and timing of the signal was generally robust to the gridcells used; therefore, the gridcell combination with the best data coverage was selected (shaded cells on map). Hatched time-height squares did not meet the 50% of gridcells or 50% of weeks in the season criteria and may not be representative of the region and season.
CHAPTER 3. VERTICAL STRUCTURE OF ARCTIC WARMINGS

Figure 3.6: As in Figure 3.5, but for spring (MAM).
Figure 3.7: As in Figure 3.5, but for summer (JJA).
Figure 3.8: As in Figure 3.5, but for autumn (SON).
Figure 3.9: Summary of the major warm and cold features in each region, based on Figures 3.5 to 3.8.
CHAPTER 3. VERTICAL STRUCTURE OF ARCTIC WARMINGS

3.2.5 Results

General features

The nature of interannual variability in the Arctic has several interesting features in the vertical time series figures. Winter has the largest interannual variability, as expected. Spring and autumn have similar variability, already much reduced from winter, while summer has the smallest variability. Winter also has a number of multi-annual clusters of warm (or cold) years whereas summer shows a tendency towards interannual variability. Greenland always has less variability than the other regions; this sector is primarily coastal stations which would likely be more moderated by an oceanic climate. Autumn has a larger incidence of vertically coherent individual years than the other seasons, pointing to the importance of advection in this season.

Several additional strong features are apparent in the time series which, in some cases, might be remaining problems with the data or the climatology. The anti-correlation of the troposphere-stratosphere temperatures is apparent in the line of seemingly random anomalies at 300 hPa (winter) and 200 hPa (summer) which stands out to the viewer. This is due to the variable height of the tropopause itself, leading to that pressure level being within the troposphere in some months or seasons and in the stratosphere in others, and thus being coherent with the layers below or above, respectively. It could also be due to the use of NNR as a climatology which can have the wrong tropopause height. The stretch of vertically coherent cold winters in 1950s North America—Beaufort combined with warm upper troposphere in summer in the same years is unusual, as is the vertically coherent warm period in Europe in the early 1970s winters when the rest of the Arctic was quite cold.

The apparent cold signal in the upper troposphere and stratosphere over Central Russia and East Siberia in 1950s summers could be related to the radiation and lag correction or the choice of climatology. The extreme cold in central Russia in the early 70s summers followed by a similar signal in North America in the later part of the decade might suggest a see-saw between these two regions, although the cold signal is apparent in all four seasons in central Russia, perhaps pointing to a change in instrumentation or operational processing algorithms.

The Early Warm Period

The SAT record of the early warm period is non-monotonic (Fig. 3.3): there is a jump in 1920 followed by a warm period with two peaks (the two warmest winters were 1937
and 1944) separated by unusually low temperatures that were related to the 1939-1942 El Niño (Brönnimann et al., 2004). This El Niño resulted in anomalously low SAT in Europe and high SAT in Alaska in winter coupled with high temperatures in the lower stratosphere over Eurasia. These features are clearly visible throughout the troposphere in Fig. 3.5 (the cold signal in the Europe wedge and the warm signal in the Beaufort wedge), including the weaker El Niño signal over the Alaskan region. The striking vertical coherence of the El Niño signal is also apparent, reaching to the tropopause over Europe. In the remaining discussion, the El Niño years are excluded from the description of the features in the early warm period.

The early warm period was primarily confined to winter in the area between 60W and 120E, encompassing both Europe and Greenland. Normally Greenland and Europe are out of phase because of the NAO (i.e., they are on opposite sides of the Icelandic low), but the early warm period was a time when both areas were warm due to a slight shift in the shape of the Icelandic low which allowed warm air advection into Europe coupled with more easterly (rather than northerly) advection over Greenland (Rogers, 1985). The warmer temperatures were clearly present in the surface data in winter in the 1930s. In the 1940s, the warming disappeared over Greenland and weakened over Europe and strengthened over central Russia. There are two factors that point to a dynamical cause of the warm anomaly: one is the spatial see-saw in temperatures apparent in the three sub-periods (warm Europe—cold Beaufort in the 1930s, then cold Europe—warm Beaufort during the El Niño, then a return to warm Europe-central Russia—cold Beaufort in the 1940s) and the other is the eastward shift of the warm signal after the El Niño.

In autumn, individual years were warm throughout the entire Arctic in both decades with a cluster of warm years in the 1940s over central Russia. No warming was seen in summer except over Greenland. Spring also showed little warming—individual years at the surface in North America and central Russia—although the El Niño is still evident. The scattered nature of the individual warm years in spring and summer could point to normal interannual variability combined with season-to-season autocorrelation (e.g., lingering from warm winters) rather than an independent signal of warming.

In winter over the area influenced by the North Atlantic (0°-60°E), some years display a vertically coherent signal (e.g., 33-34 and 40-42) while other years do not (e.g., 37-39). The sparseness of the data precludes drawing conclusions about the vertical structure of regionally averaged data.
The Recent Warm Period

The recent warm period was far more widespread in space and some signal appeared in all regions and all four seasons with the single exception of winter in the Bering Strait. It also appears to be a trend rather than an isolated warm period. To some degree, it presents a picture which is consistent with normal interannual variability within the context of overall background warming.

The strongest signals were warmings in Beaufort—North America in winter, Bering in spring, and North America in autumn. The North American and Bering signals were overall strongest at the surface but were clearly apparent throughout the troposphere, while the Beaufort signal was quite strong at both the surface and 850 hPa but was also apparent throughout the troposphere. In the other seasons and regions, the warming was generally seen as individual years and appeared at only the surface in some areas and in the full column in others. Vertically coherent individual years are especially noticeable in both spring and autumn. East Siberia shows coherent vertical structure in summer and the summer trend is strong at 850 hPa in several other regions. This feature superficially resembles the mid-tropospheric peak in warming in Graversen et al. (2008) but could be related to the different spatial coverage and time sampling of the radiosonde data and the SAT (a gridded product which, furthermore, was not subsampled to stations and times where radiosonde data was present). Furthermore, the feature arises from the data rather than an assimilation problem which caused a step function in ERA-40 reanalysis over the Arctic Ocean basin (Bromwich and Wang, 2005).

The Bering Strait has a different seasonal pattern in the recent warming: there is no signal in winter but it is apparent in spring-summer-autumn, and is remarkably strong in spring, coinciding with the extensive changes in sea ice area in recent years.

3.2.6 Discussion

Circulation anomalies in the 1930s and 1940s

Reconstructed geopotential height (GPH) fields provide new information on the circulation during the early warm period. Decadal mean SLP (from HadSLP2 Allan and Ansell, 2006) and 700 hPa GPH height anomalies (relative to a 1961-1990 climatology) are shown in Figure 3.10 (1939-1942 is shown as a separate figure due to the strong El Niño). While SLP showed enhanced advection into Greenland from less northerly latitudes during the 1930s (that is, air coming from ~ 65N rather than 75N), the 700
hPa fields show additional southerly advection directly over Greenland as the isobars around the Icelandic low shifted. The progressive strengthening of zonal circulation is apparent in the 1920s, in agreement with Rogers (1985); Fu et al. (1999) and the generally positive values of the AO/NAO index in the first half of that decade. In the 1930s, the Icelandic low remained strong and a high pressure anomaly developed over Europe at 700 hPa, both of which funneled warm air from the Atlantic and west coast of Europe directly into the Greenland and Barents Seas. During the El Niño, the Icelandic low weakened remarkably, bringing the well known cold temperatures to Europe and the Scandinavian Arctic (Brønnimann et al., 2004). Anomalous circulation in the 1940s was far less coherent at 700 hPa, in agreement with the disintegration of the signal in SLP (Rogers, 1985; Fu et al., 1999).

There is, of course, interannual variability within the 1930s circulation, but the presence of both the cyclonic circulation over southern Greenland—Iceland and the anti-cyclonic circulation over Europe is present in most of the years, as seen in Figure 3.11. The winter of 1936-1937 is particularly interesting, as this was one of the two warmest winters on record in the early warm period as well as being a winter with a remarkably low value of the PNA index (Ewen et al., 2008a) which is clearly evident in the weakening of the Aleutian low. This circulation brought extreme cold surface temperatures into western parts of mainland US and the central Canadian Arctic and warm temperatures into the Alaskan region, although they are not clearly evident in Figure 3.5 due to the spatial averaging.

We suggest that the early warm period was brought about through a chain of events: in 1920, the AO moved into a strongly positive phase for several years, initiating enhanced zonal circulation and warm air advection from Europe into the European Arctic and perhaps the export of some older, thicker sea ice out of the Arctic (Rigor and Wallace, 2004; Vinje, 2001). The AO then returned to neutral or negative values, but the anomalously strong advection remained in the 1930s, coupled with an additional anti-cyclonic circulation cell over Europe, as seen in reconstructions of geopotential height in Figure 3.10. Evidence of the enhanced circulation can also be found in ice core records from Svalbard and Greenland (Fig. 3.12), which show increasing amounts of non-sea-salt sulfate aerosols from anthropogenic sources (as opposed to volcanic sources) in the 1930s and 1940s, although the timeline of the Svalbard icecore should not be inspected on a scale lower than five years due to possible summer melting at that site. Furthermore, sulfate aerosols themselves have a warming effect on Arctic surface air temperatures (Garrett and Zhao, 2006; Lubin and Vogelmann, 2006), further enhancing the warming as discussed below. The entire chain of events may possibly have been initiated due to the first anthropogenic aerosols (both sulfate and black car-
Figure 3.10: 700 hPa geopotential height (top in pair) and SLP (bottom in pair) anomalies (relative to 1961-1990) for decadal mean winters (DJF): 1900-1909 (top left pair), 1910-1919 (top center pair), 1920-1929 (top right pair), 1930-1938 (lower left pair), 1939-1942 El Niño (lower center pair), and 1943-1949 (lower right pair). Heavy solid line is the 0 contour, thin solid lines are positive anomalies and dashed lines are negative anomalies. Contour spacing is 1 hPa for SLP and 10 m for 700 hPa height. SLP data is from HadSLP2 (Allan and Ansell, 2006)
Figure 3.11: 700 hPa geopotential height anomalies (relative to 1961-1990) for individual winters (DJF) from 1930 to 1938. Heavy solid line is the 0 contour, thin solid lines are positive anomalies and dashed lines are negative anomalies. Contour spacing is 20 m. The majority of years have a high pressure anomaly over Europe and a low pressure anomaly over Greenland—Iceland, a very strong feature in the decadal mean. Note also the extremely strong negative PNA in 1937.
bon) which began arriving in the Arctic (Greenland and Svalbard) toward the end of the 1800s (Goto-Azuma and Koerner, 2001; Isaksson et al., 2001; McConnell et al., 2007). This coincided with an upturn in Arctic temperature as recorded in proxy data (Overpeck et al., 1997) which, for the first time since $\sim 1600$, rose above the natural variability starting in the late 1800s.

The role of aerosols in the early warm event

Anthropogenic aerosols are normally credited with having a cooling effect on surface temperatures, as indeed they do in temperate regions (e.g. Jones et al., 2007; Ramanathan et al., 2001). In non-polar regions, changes to incoming solar radiation often dominate other changes to the surface energy budget. In the polar night, however, longwave radiation overwhelmingly determines the surface energy budget of the Arctic. Even in the sunlit half of the year the high surface albedo causes longwave radiation to continue to have a significant impact. For example, clouds have a net warming effect on Arctic SAT throughout the entire year, except for a few weeks at the height of summer. And while, overall, pollution such as sulfates have been well documented as cooling the surface (e.g. Jones et al., 2007; Ramanathan et al., 2001), it is important not to neglect the regional responses which can differ significantly from the global mean.

The climatic effects aerosols which have been transported to the Arctic have recently been documented for sulfates (Garrett and Zhao, 2006; Lubin and Vogelmann, 2006) and black carbon (McConnell et al., 2007), raising an interesting possibility of a previously unsuspected anthropogenic cause of the early warm period. Recent experimental studies established that when low, thin water clouds are present in the Arctic, the addition of sulfate aerosols increases the longwave emissivity of the cloud by 3 to 5 W/m$^2$, warming the surface by up to 1.5°C (Garrett and Zhao, 2006; Lubin and Vogelmann, 2006). Although the microphysical state of Arctic clouds is not well sampled, observational studies have found liquid water clouds predominate in summer and that winter clouds are generally mixed phase (Intrieri et al., 2002) and the two studies showing the enhanced longwave emissivity were both observational, confirming that the overlap of pollution and susceptible clouds exists. Arctic haze is generally observed in winter and spring. In addition, the presence of aerosols directly alters the longwave emissivity of the atmosphere apart from (and in addition to) the cloud emissivity effects, adding an additional 4 W/m$^2$ (Ritter et al., 2005). Direct heating of the aerosol layer (typically at around 2km and containing both sulfates and black carbon) in spring has also been observed (Treffeisen et al., 2005), accompanied by surface cooling (Quinn et al., 2008; Treffeisen et al., 2005). Finally, black carbon has been estimated to induce a direct sur-
face heating of around 3 W/m$^2$ during summers from 1906 to 1910 (McConnell et al., 2007).

Proxy records of Arctic temperatures reveal that the early warming was actually the tail of a longer warming that rose above the natural variability of the past 400 years starting around 1850 (Overpeck et al., 1997). Interestingly, ice core records from Greenland and Svalbard, shown in Figure 3.12, demonstrate a rise in non-sea-salt sulfate pollution at around the same time (Goto-Azuma and Koerner, 2001; Isaksson et al., 2001), which has been attributed to increases in anthropogenic emissions from Europe and North America (Goto-Azuma and Koerner, 2001; Moore et al., 2006).

Knowing that aerosols warm the Arctic in winter and that they were present in the beginning of the 20th century raises the question of the role played by the influx of anthropogenic pollution in the early warming. It is plausible that the aerosols: 1) triggered the warming by initiating a transition into a warm phase of the decadal-scale natural variability endemic to the Arctic, 2) enhanced a natural warming that was already underway, or 3) both triggered and enhanced or sustained a warming. Furthermore, there are two points in time at which the aerosol role could have been enhanced: 1) at the first arrival of aerosols into a previously pristine Arctic (starting around 1850 with this effect peaking by 1910) and 2) another flood of aerosols in the 1930s and 1940s due to anomalous advection which is seen in reconstructions of geopotential height (see above). The highly positive AO seen in the first half of the 1920s likely caused above average ice export (and lower ice concentrations were observed during that decade in the Nordic Seas (Vinje, 2001)) further enhancing the warming already in place. Aerosols primarily warm the surface in winter and the aerosol layer in spring. Although the data is sparse, the upper air data in Section 3.2 do show more warming at the surface than aloft, perhaps supporting this theory. Other years have vertically coherent temperatures which are consistent with changes in advection. The recent warming has been evident at both the surface and the free troposphere well above the aerosol layer, indicating that other processes are involved although aerosols could still play a role. The stronger signal of the recent warming in the Bering side of the Arctic could be related to the influx of anthropogenic pollution from Asia in the recent decades (Herber et al., 2002; Koch and Hansen, 2005), although Stohl (2006) concluded Asian black carbon was unlikely to reach the lower troposphere in the Arctic.
Figure 3.12: Ice core records showing sulfate levels. For Greenland records, raw data are shown along with smoothed emission curves for North America (solid smooth curve, left hand emissions scale on right side of figure) and North America + Europe (dashed smooth curve, far right scale). Greenland data reproduced from *Goto-Azuma and Koerner* (2001); Svalbard data reproduced from *Isaksson et al.* (2001)
3.2.7 Summary

Considered from the perspective of the troposphere, the two warm periods in the Arctic over the past century appear quite different due to their different vertical structure as well as their spatial and seasonal patterns. The three dimensional view presented here allows the warm periods to be distinguished from each other in a way that was not possible using only SAT, and further counters the theory of a low frequency oscillation found by Polyakov et al. (2003).

The early warm period was, for the most part, confined to wintertime Europe where it was consistently strong at the surface and intermittent aloft. The recent warm period is spatially widespread and has some signal in all regions and seasons and appears to be spreading; that is, it is a trend rather than an isolated warm period. It is strongest near the ground but is also widely present in the free troposphere. The strongest signals are in the Beaufort-North America area in winter and the Bering Strait in spring. Reconstructed geopotential heights showed advection anomalies in the 1930s which were consistent with the anomalous temperatures in the early warm period as well as the record of non-sea salt sulfate (presumably of anthropogenic origin as there were no volcanic eruptions in that period) in Arctic ice cores. A sequence of events was presented for the early warming: strengthening of zonal circulation in the 1920s followed by continued strong zonal circulation and the development of anomalous anticyclonic circulation over Europe in the 1930s which advected warm air into the Scandinavian Arctic. Although the early warm event appears to be dynamically driven at heart, this advection also brought sulfate pollution which likely further enhanced the warming. The warming effect of the aerosols may also have been a trigger mechanism as early as 1850.

The different characteristics of the warm periods can also serve as an aid in climate model validation, as the appearance of these features would suggest the model has the appropriate dynamical or radiative forcing to bring about each of the two warm periods.

Acknowledgements

This work was supported by the Swiss National Science Foundation. We thank Andrey Nagurny at the Arctic and Antarctic Research Institute for the Russian radiosonde data. The TD54 and CARDS542 datasets were kindly provided by Joey Comeaux at NCAR, and the TD6201 dataset by Joe Elms at NCDC.
Chapter 4

Conclusion

4.1 Summary

A more complete understanding of 20th century Arctic warming events requires analyzing them from an upper level perspective. This work presented a new dataset which was created for this purpose and analyzed the signal of the two major Arctic warm periods in the 20th century as seen in the new dataset.

First, a new collection of radiosonde data from 1932 to 1958 was compiled and validated. This is the first time such a comprehensive dataset from the early part of the record has been created, and extends the record of atmospheric temperatures and heights from 48 years to 75 years, an addition of over 25 years.

Data was drawn from a number of international sources and in some cases was digitized directly from paper records. The stations were comprehensively cross referenced for both time and location, and a single set of merged stations was created. The data was then validated using a novel technique which compares monthly mean data on a subset of levels with a reconstructed reference series. This creates an anomaly time series which is particularly powerful at revealing characteristic errors of radiosondes. Certain networks were found to have errors in reported units and uncorrected radiation and lag errors.

Coherent corrections covered much of Eurasia. Prior to 1948, the temperatures were lowered to correct for radiation and lag errors, but the corrections to geopotential height due to radiation and lag were overwhelmed by the units correction, resulting in an overall lifting of the layer. After 1948 the corrections were of the same sign (lowering temperatures and lowering heights). The corrected figures of temperature and height show improved physical consistency (that is, the structure in temperature at a given pressure
level more closely matches the structures in the height field). The corrections covered such a widespread area that hemispheric-scale field analyses using radiosonde data (or products derived from it, such as reanalysis) should be reconsidered in light of the corrected data.

The new data were then examined for information about the two major warm periods in the Arctic in the 20th century: 1) the warm period from the 1920s through the 1940s and 2) the warming from the late 1980s to now. The events were looked at on three axes: the vertical structure, the spatial (regional) extent, and the seasonal strength of the events. The two warm periods appear quite different: while the early warming was most concentrated in the European sector of the Arctic and was strongest in winter, the recent warming has been far more widespread and has appeared, to some degree, in all seasons and regions; it is also a trend rather than an isolated warm period. The strongest signal of the recent warming is in winter in the Beaufort Sea and North America and in spring in the area around the Bering Strait. The sparsity of data in the early period makes it difficult to conclusively state the vertical structure of that warm period: there are individual years with warming aloft in the Atlantic Arctic while the warming is temporally coherent at the surface. The recent warming, in contrast, has considerable signal aloft, although the trend of warming from 1979-2001 is strongest at the ground, contrary to the results of Graversen et al. (2008) who found the strongest warming trend in ERA-40 reanalysis to be at about 700 hPa.

Reconstructed geopotential heights from 1900 to 1950 showed anomalously strong advection from Europe and the North Atlantic into the Atlantic Arctic in the 1930s, in agreement with both the warm temperatures and the increased levels of anthropogenic sulfates seen in ice core records from Svalbard and Greenland. The circulation weakened and shifted in the 1940s, also in agreement with the somewhat more disorganized nature and eastward shift of the warm signal in that decade and the lower sulfate levels in the ice core records. The presence of the sulfate aerosols may have enhanced the early warm event as well.

4.2 Future Work

With the development of this new, comprehensive upper air dataset extending back to the 1930s, several interesting projects could be pursued.

In collaboration with Gilbert Compo, our data is being used to validate a new reanalysis
CHAPTER 4. CONCLUSION

64

from NOAA\(^1\) and CIRES-ESRL\(^2\) based entirely on surface pressure back to 1900 (Compo et al., 2006). The reanalysis shows excellent agreement with data at midlatitude and Arctic stations, as seen in the correlation of observed and analyzed 700 hPa temperature anomalies at Ilmala, Finland and Lindenberg Germany and 500 hPa geopotential height anomalies at Ilmala (Figure 4.1). After validation is complete, a case study of the two warmest winters (1937 and 1944) would help us understand both the continuity of the early warm period as well as its similarities to and differences from the recent warming. The case study could be extended by examining similar winters from long control runs of general circulation models which can be identified based on surface air temperature patterns, providing a better match than the use of zonal means due to the regional structure in the early warm period. New reconstructions of the first half of the twentieth century (Griesser et al., in preparation) used in Section 3.2 could also be used in the case study. The new data can also be assimilated into third generation reanalyses such as ERA-75.

![Figure 4.1](image)

Figure 4.1: Scatterplot of observed versus analyzed temperature anomaly at 700 hPa for Lindenberg, Germany (left) and Ilmala, Finland (center), and observed versus analyzed geopotential height anomaly at 500 hPa for Ilmala (right). Figures courtesy of Compo (2008).

With each new generation of models, climate can be simulated with increasingly higher resolution. Validating such high performance models requires datasets which extend as far back in time as possible and have the highest quality. The significantly extended upper air dataset presented in this work could be a valuable resource to climate modelers in a variety of projects such as model intercomparisons, investigations of variability in long term control runs and both long term forced runs and shorter transient studies.

\(^1\)U.S. National Oceanic and Atmospheric Administration

\(^2\)Cooperative Institute for Research in Environmental Sciences and the Earth System Research Laboratory, both in the U.S.
Another pertinent question that could be addressed is the quality of current reanalysis products over data sparse regions such as the Arctic. Known problems assimilating satellite data over the Arctic Ocean have led to discontinuities in tropospheric temperatures in ERA-40 (Bromwich and Wang, 2005), and an examination of NCEP-NCAR reanalysis showed significant problems with the seasonal cycle in surface air temperature (Makshtas et al., 2007). Preliminary investigation of the vertical structure of six-hourly NCEP-NCAR reanalysis compared to individual radiosonde soundings suggests that there are some problems, many of which can be traced to a surface bias which affects the entire profile. Both ERA-40 and NCEP-NCAR reanalysis could be compared to individual soundings over land (from the new dataset described earlier) and over the ocean (from the North Pole Drifting Stations (Fetterer and Radionov, 2000) after validation).

An extension of Arctic inversion characteristics, such as those done by Serreze et al. (1992) would also be valuable in furthering our understanding of the variability of inversions and how they change in the long term and during warm periods. Dramatic changes in the surface energy budget of the Arctic will occur under any climate change and will certainly impact the stability of the boundary layer.
Appendix A

Monthly Weather Review Data

The paper included below covers a related validation project. As second author, I performed the validation and statistical calculations on the monthly mean data. Decisions about corrections and rejections were made jointly by Tracy Ewen and me.

A monthly upper-air data set for North America back to 1922 from the Monthly Weather Review

Tracy Ewen, Andrea N. Grant, and Stefan Brönnimann

Institute for Atmospheric and Climate Science, ETH Zurich, Switzerland


On November 19, 1930 we had some “weather”, as some people say when they mean only the bad kind.

- H. Clark Fuller, Kite Flier, US Weather Bureau, Ellendale, North Dakota
Abstract

Upper-air observations with kites, aircraft, and radiosondes were performed in the United States operationally since the 1890s. In this paper, we present a re-evaluation of newly digitized monthly mean values from the Monthly Weather Review back to 1922. Data from 46 US weather stations are presented with a focus on early kite and aircraft observations during the 1922-1938 period. A quality assessment of the data, based on reconstructed reference series is carried out and the quality of the monthly mean data is found to be quite high. Anomalies of upper-level temperatures from the re-evaluated station data are compared to surface fields for selected periods and show consistent spatial patterns.

A.1 Introduction

Changes and variations in atmospheric circulation during the early part of the twentieth century are difficult to study without access to upper-air data. Currently, reanalyzed data sets are available back to 1948 (NCEP/NCAR Reanalysis, Kalnay et al., 1996; Kistler et al., 2001) and 1957 (ERA-40 Reanalysis, Uppala et al., 2005). Few upper-level meteorological data sets based on observations extend further back than 1948 (see for example, Brönnimann, 2003a), although operational upper-air observations were performed since the 1890s. Our understanding of atmospheric circulation, especially upper level fields extending into the higher troposphere during this period is therefore limited by the amount of upper-air data that is available in digital format. Such data would be very valuable for studying climate extremes in North America during the early part of the 20th century, for example, the Dust Bowl droughts in the 1930s. The processes involved in this drought, which affected the southern United States and the Great Plains regions, cannot be fully understood with available surface data. Data from higher levels in the atmosphere are required in order to fully understand the dynamical processes that govern large-scale circulation and the corresponding effects on surface climate over North America. The addition of newly re-evaluated upper-air historical data to current climate data sets is an essential component if we are to improve our understanding of mechanisms involved in these anomalies during the early 20th century.

In this paper, we present a monthly upper-air data set for North America that extends back to 1922. The data have been digitized from past issues of the Monthly Weather Review. The new data set includes monthly means of temperature and pressure data from kite, aircraft and radiosonde observations during the 1922-1947 period for 136
North American stations. A considerable fraction of the 1939-1944 data, mainly radiosonde data converted to standard pressure levels, were already presented elsewhere (Brönnimann, 2003a). Here we focus on the 1922-1938 data, which mainly consist of temperature from kite and aircraft ascents given on fixed altitude levels. In total, data from 46 stations are available for this period. We present analyses of select stations, including the longest kite record at Ellendale, North Dakota (a US Weather Bureau station). The quality of the data is assessed using statistically reconstructed reference series and an analysis of anomaly maps for selected months is carried out.

The outline of this paper is as follows. A historical overview of kite and aircraft measurements and associated errors in the United States are presented in Section A.2. In Section A.3 the data published in the Monthly Weather Review are introduced. The digitizing and correction procedure is discussed in Section A.4. Section A.5 outlines the quality assessment of the data and corrections to the historical series are discussed. This is followed by results, which include an analysis of selected anomalies over the 1922-1938 period, in Section A.6 and conclusions in Section A.7.

### A.2 Instruments, observations and errors

The US Weather Bureau started upper-air observations in the late 1890s (Marvin, 1895). During the 1910s, a network of kite stations was built up, which included the station at Mount Weather, Virginia. At Ellendale kite flights started on December 17, 1918 (Gregg, 1918) and were intermittently recorded until 1922. After this almost daily soundings were performed through to 1933.

A daily kite sounding would usually be made based on favorable morning pilot balloon readings. If winds were too light, lighter than \( \approx 12 \text{ m.p.h.} \), a kite could not be lifted off the ground as kites were quite heavy. A kite sounding would also not be made if weather conditions were not favorable or if winds were potentially strong enough to destroy the kite. As H. Clark Fuller wrote about conditions at the Ellendale station, “You might wonder if we ever had too much wind for flying kites. Yes, indeed we did, not very often but more than a few times. My record mentions October 27 and 28, 1931 as two days together on which there were no flights for that reason. There were other times when blizzard winds were too strong. On January 19, 1928 my note says it was the windiest day for over a year, all day long 40 to 50 m.p.h. and twice 52 m.p.h.” Thunderstorms also posed a serious danger, and kites would not be launched if such systems were approaching. Other possible reasons for no readings included a “breakaway”, in which
turbulent conditions exceeded the recommended pull on the line causing the wire to snap, ideal conditions for the formation of snow and ice on the kites and wires, which might cause the kites to fall. Although bad weather often prevented kite flights, at the Ellendale station staff sometimes found it necessary to ski into work in order to carry out the daily sounding (Fuller, 1976).

If the conditions were favorable for a launch, kites were first chosen according to the strength of the winds present, which had been determined earlier in the morning from the pilot balloon soundings. A smaller kite was generally selected for stronger winds and larger kites for weaker winds. All kites at the Ellendale station were Marvin-Hargrave box kites with spruce supports and cloth sails (see Figure A.1a). Three types of sails were available, two white and one black. Once the kite was chosen, a Marvin meteorograph was baselined by comparing it to the other instrumentation in the vented shelter before being mounted on the kite. The meteorograph was attached to the kite so that the air would flow freely through a tube containing the temperature and hydrograph elements. The kite was fastened to piano wire (≈ 10 miles in total) which was unreeled from a drum and comprised the reel assembly, located in the reel house. At the Ellendale station, an electric motor was installed in July 1918 and was used to pay out the wire for the kite flight. Generally more than one kite was launched as the weight of the wire was usually too heavy for a single kite to support.

After each kite observation, all daily data were telegraphed to Washington D.C. where they would be used for 36 hour forecasts. An early example of aerial photography was developed by H. Clark Fuller at the Ellendale station, with the invention of the “aerocam” which was mounted on a kite. An aerial photograph taken with the “aerocam” is shown in Figure A.1b.

Several deficiencies led to the discontinued use of upper-air kite soundings: kite wires were often difficult to see and posed an increasing hazard to aviators; wind data were unreliable when observations were taken under frequently changing winds, including storms, as they took too long to adjust (Samuels, 1926) and in general, kite observations also required too much time, effort and manpower (Geddes, 1939). One of the biggest deficiencies was that the mean altitude attained by a kite was usually around 2500 m. The maximum altitude rarely exceeded 5000 m, and observations from higher altitudes were needed. As of 1925, the record height reached by a sounding kite was ≈ 9000 m (Talman, 1931).

The Weather Bureau had been experimenting with aircraft observations since 1919 by attaching meteorographs to airplanes, observations called “APOBS” (Huschke, 1959). From 1925 on, aircraft observations were performed routinely at several stations. Dur-
ing the early 1930s, aircraft observations became popular as they were generally easier to carry out than kite observations, with a single flight taking approximately 1 hour compared to 4-5 hours for a single kite sounding. Pilots could record both altitude and visibility during the flight as well, which added to the observational record. In January 1938 several stations began to use radiometeorographs, replacing aircraft observations, which recorded both temperature and pressure on fixed altitude levels. While planes continued to be used for this purpose until 1943, their use was greatly curtailed afterwards. Airplane observations had several weaknesses: the average altitude reached was \( \approx 5000 \) m, data could not be evaluated until the observation was completed and observations could not be taken in stormy weather, as with kites, although an airplane could withstand worse weather than a kite could.

Several errors relating to reading the meteorograph, used for both kite and aircraft soundings, have been recorded (Clayton, 1904). These include, but are not limited to: instrumental errors (changes in placement of the recording pens, although these were usually fixed by calibrating with a standard before each flight), errors in reading from the output data on the meteorograph (the time error related to the use of different pens on the meteorograph and error in reading charts arising from a thickening of the recorded trace due to vibrations), errors due to topography and local microclimate, and
errors due to observations being limited to certain weather conditions. Despite problems with aircraft soundings, aircraft data were often used to validate radiosondes and this was an important step in improvements to the early radiosonde network (e.g. Diamond et al., 1938). As airplanes became faster and flew to higher altitudes however, associated errors due to friction and pressure effects from the aircraft became too large and eventually the observations made by radiosondes were better than aircraft measurements (Scherhag, 1948). However, the eventual inability of kite and aircraft meteorographs to achieve high altitudes, operate in all weather, and provide data in real-time helped foster the development for the radio transmission of upper-air data and the use of radiosondes. In addition, flights were often life threatening due to lack of oxygen at the high altitudes reached and several pilots were killed in the 1930s (Hughes and Gedzelman, 1995).

Radiosondes were introduced by the US Weather Bureau in 1938 and in 1939, a network was established. The data from 1939 on as well as more detailed descriptions of the instruments, can be found in (Brönnimann, 2003a). In the following, we focus on the aircraft and kite data up to 1938.

### A.3 Upper-air data published in the Monthly Weather Review

From 1922 on, the data from the upper-air network were published in the form of monthly mean values in the Monthly Weather Review. For the 1922-1938 period, data from 46 individual stations were published and the location for each is shown in Figure A.2. These data include 6 kite stations, 39 aircraft stations and 9 radiosonde stations (8 of which overlap with aircraft stations). Kite observations of temperature, relative humidity and winds were reported from six stations (Broken Arrow, Drexel, Due West, Ellendale, Groesbeck, Royal Center) from January 1922 until observations ended between 1931 and 1933 (except Drexel, Nebraska which ended in 1926). Ellendale was the last station with recorded kite data. The last record in the MWR was for the month of April in 1933, and two months later, the station was closed. Observations from US Naval air stations, including Pensacola, San Diego, Washington and Hampton Roads were also reported in the Monthly Weather Review between 1926 and 1930. Starting in July 1931 Aircraft observations are reported from Chicago, Cleveland, Omaha and Dallas (see Figure A.2), followed in 1932 by Atlanta, Norfolk (Naval air station), Boston (observations were made by MIT) and in 1933 by Pembina, North Dakota. In July 1934,
16 new stations were added which included Weather Bureau, US Naval and US Army aircraft stations. By 1937, the Weather Bureau aircraft observations were reported from 30 different stations.

Figure A.3 shows the total number of digitized monthly mean values (all levels) over the 1922-1938 period for each observation type - kite, aircraft or radiosonde (radiometerograph). Additionally, in Figure A.4, the number of monthly mean values for Ellendale, North Dakota (kite, station 255) and Washington, D.C. (aircraft followed by radiosonde observations, station 259) is shown for each altitude level. Observations taken with radiometerographs did not start at Washington, D.C. until June 1938 and heights above 5000 m are not recorded in the MWR until August 1938, thus only 4 months of observations above this height are shown for the period considered here (for later radiosonde data see Brönnimann, 2003a).

The monthly mean values for temperature (°C, recorded to the nearest 0.1°C) and humidity are given on fixed altitude levels for the 1922-1936 period and were reported in the “Free-air Conditions” section of the MWR, which subsequently changed to the “Aerological Observation” section in 1928. Starting in January 1937, mean barometric pressures (hPa, recorded to the nearest hPa) are available along with equivalent potential temperatures on fixed altitude levels. We have thus digitized monthly temperatures up to and including December 1936 and temperature and pressure for the 1937-38 period. Observations are recorded on altitude levels from the surface, every 500 m up to 3000 m and then every 1000 m thereafter, with the exception of kite observations before February 1929, which are recorded from the surface, every 250 m up to 1500 m and then every 500 m thereafter. Most of the records do not go above 5000, until the introduction of the radiosonde network in 1938, when some soundings reached a maximum level of 20,000 m. Due to the short pressure record, our analysis of the digitized

Figure A.2: Station map showing the location of the 46 stations. Kite stations - grey boxes, aircraft stations - black triangles, radionsonde stations - grey circles.
APPENDIX A. MONTHLY WEATHER REVIEW DATA

Figure A.3: Number of kite, aircraft and radiosonde observations between 1922-38

Data will focus exclusively on temperature.

From the archives, we have digitized the monthly means as daily observations were not printed. The number of observations per month that was used to calculate the mean is not recorded prior to January 1935. After this, the number of observations used for each monthly mean is listed and all stations which had more than 15 observations per month are used for the final data product. Prior to this, we have no information about the number of observations used for the final monthly mean, so have included all the

Figure A.4: The number of observations for stations 255 (Ellendale, North Dakota) and 259 (Washington, D.C.)
available data. An example which highlights this lack of information was recorded in the “Aerological Section” in September 1928, and L.T. Samuels writes, “The largest deviation (from the mean) is found at Due West but, owing to the relatively few kite observations possible at that station during the month, a comparison was made between the resultant winds as indicated by the morning pilot-balloon observations and the normals...” (Samuels, 1928).

A.4 Digitizing and correction procedure

The digitizing and correction procedure followed is the same as in Brönnimann et al. (2006). A detailed description of the digitizing procedure can be found elsewhere Brönnimann et al. (2006); here we give a brief summary. The data were digitized using speech recognition or key typing. The data underwent a plausibility screening consisting of range checks for absolute values, vertical gradients, and the hydrostatic balance. Corrections were then made to account for changes in the time of day of observations. As in Brönnimann (2003a), we adjusted all data to daily means by adding the climatological difference in a given variable between the time of day of observation and the daily mean (calculated for each location, level, variable, and calendar month from NCEP/NCAR data, 1968-1996). Although the diurnal cycle in the reanalysis is model generated and may deviate from the true diurnal cycle (in fact, due to the many interpolation steps in the process of obtaining this climatology, the diurnal cycle is most likely underestimated), the deviation from the true diurnal cycle is likely smaller than inconsistencies that would arise when comparing or merging data obtained at different times of the day.

The start times for many of the observations changed throughout the 1922-1938 period. Kite observations were usually made in the morning, after the pilot balloon soundings, and took somewhere between 4-5 hours for the entire ascent and descent to be made. However, there is no mention of the exact time of kite ascents in the archives during the early years of our record, until September 1932, when it states that Ellendale kite observations are made \( \approx 9 \) am, seventy-fifth meridian (EST) time. Starting in March 1929, pilot balloon soundings are noted as being made at 7 am EST. According to Gregg (1921), morning kite soundings were favored by the forecasters in Washington and also preferred by the observers as there was a “general lack of convection” in the early morning. We have assumed that all kite soundings were carried out at 9 am EST, unless otherwise noted. Starting in July 1932 the time for Weather Bureau airplane observations is listed as 5 am (EST) and Navy observations near 7 am (EST).
Other corrections that could apply for kite and aircraft observations such as radiation and lag errors are not well known and difficult to apply due to a lack of information concerning changes in instrumentation and measurement details. Although the corrections have been applied for daily radiosonde observations by Brönnimann (2003a), they were found to be unnecessary at tropospheric levels. We have therefore not applied any further corrections.

A.5 Quality assessment

The target accuracy (bias) and precision (90% spread) for the monthly mean temperatures are chosen to be $\pm 0.5^\circ\text{C}$ and $\pm 1.5^\circ\text{C}$, respectively. A data set that is demonstrated to be within these limits can be useful for climatological applications (such as analyzing extremes and interannual variability), though probably not for trend analyses. In this Section we describe the tests used in order to assure this quality.

Reconstruction of reference series

In order to test the quality of our upper air series, a reference series is needed. No neighboring or high altitude surface station series are available for this time period, so we use statistically reconstructed reference series for each variable (level) and station location. Surface data are used as predictors for the statistical model.

To reconstruct reference series for the 1922-1938 period, statistical models need to be calibrated in a period for which both the station data and the predictor data are available. As the upper-air series do not continue into the present, they were supplemented with reanalysis data for the calibration. We used NCEP/NCAR reanalysis for the 1948-2003 period, interpolated to the station location and levels. Despite possible inaccuracies or inhomogeneities (e.g. Santer et al., 1999; Randel et al., 2000), this is the best currently available long term global meteorological data set for this purpose and the data quality for North America is quite good. As a test, we performed the same reconstructions using ERA-40 data (discussed below).

As predictors we used surface air temperature data taken from the NASA-GISS database (Hansen et al., 1999). For each upper-air station, three surface stations within a 30 km range from each upper-air station were chosen and any gaps in the surface data were filled with anomaly data from neighboring stations. Sea-level pressure (SLP) and horizontal pressure gradients (Allan and Ansell, 2006, HadSLP2) are used as additional
Appendix A. Monthly Weather Review Data

The quality of the reference series was addressed using the reduction of error statistics (RE, Cook et al., 1994). Values of $RE$ can be between $-\infty$ and 1 (perfect reconstructions). $RE > 0$, determined in an independent period, is normally considered an indication that the model has predictive skill.

**Test statistics**

The historical upper-air data were compared to the reconstructed reference series; we first performed a visual test and then analyzed and tested the bias (t-test) and the variance (Chi-square test). Additionally, we performed two homogeneity tests (Alexandersson and Moberg, 1997; Lanzante, 1996). Tests were only performed if $RE > 0.5$ and the number of non-missing data points was at least 5. In addition, the boundary layer (normally all levels up to 2000 m) was discarded as the reanalysis is not expected to be a reasonable substitute for the upper-air series in this level. If all criteria were fulfilled and a temperature offset was found, a correction was calculated as the mean bias above 2000 m for all levels where the first set of controls passed ($RE > 0.5$, $n_{obs} > 5$, significance test of the bias). A temperature adjustment was then made whenever the target precision of $\pm 0.5^\circ C$ was exceeded. It is important to keep in mind that often not all of the criteria are met and the final decision is somewhat subjective as it is also based on a thorough visual assessment of the data. For example, the boundary layer approximation of 2000 m may be shifted to 1500 m, depending on the station, whether or not there are many observations at that level and the $RE$ values.

Although this proved to be a powerful approach for assessing the quality of historical data series, there are additional fundamental problems besides those already mentioned. Errors may be introduced by interpolating the reanalysis to station locations and in the vertical in order to get temperatures on altitude levels. Additionally, there is an underlying assumption that the relationship between the predictors and predictand during the reconstructed period is the same as the calibration period. This is not necessarily true, especially given that our historical data spans the Dust Bowl period during the 1930s, when air temperatures were warmer than normal. Because the predictors are all taken from the surface, there is reduced skill (smaller $RE$ values) in the reconstruction when moving to higher levels in the atmosphere. In this study however, our data is mostly under 6000 m, so is not as affected by this reduction in skill with height. As a final product, if a temperature correction is made based on the reconstructed reference series, it is important to keep in mind that the new data series is no
longer independent from the data sets used for the reconstruction (as predictors and predictand). Finally, this is a novel test for historical series as often reference series are not available. There are however additional tests that could be performed to assess the quality of the data and make correction decisions. Tests for homogeneity of each series are also carried out and are discussed in the following section.

In addition, reconstructions for each historical series (as above), using ERA40 reanalysis from 1957-2001 were carried out and the quality from both reconstructed series was evaluated. For all stations, the reconstructed reference series and corresponding statistics, including plots of the mean bias for each level, using either NCEP/NCAR reanalysis or ERA40 reanalysis were compared, and very little difference between the reconstructions was found. All reconstructed reference series discussed in the following results section use NCEP/NCAR reanalysis for the reconstruction.

A.6 Results

Analysis of selected series

In this section, results from the quality assessment are discussed together with an analysis of temperature anomalies over North America for selected seasons over the 1922-1938 period. We highlight two stations from our data set - Ellendale, North Dakota as it is the longest kite record, and Washington, D.C., which also covers a long period, from 1926-1938.

Figures A.5 shows the mean difference between the observations and the reconstructed reference series, together with the standard errors of the reconstructed mean anomalies (grey bars around the zero lines) and the standard errors of the mean difference (confidence intervals with whiskers) for four stations. Ellendale, North Dakota (a) and Broken Arrow, Oklahoma (b), Washington, D.C. (c), and Cheyenne, Wyoming (d) are shown.

Both kite stations reveal a similar difference profile, namely a much cooler surface and boundary layer in the observations than the reconstructions. Up to 2000 m, observations become increasingly warmer than the reconstructions and between 2000 m and 3000 m, become cooler again. This pattern reverses several times, producing a typical 'zig-zag' profile that was found in all kite stations. Above 4000 m the observations are warmer than the reconstructions until there is little gradient between 4500 m and 5000 m. This profile shape is a result of the NCEP/NCAR reanalysis; in the case of Ellen-
dale, these differences can be seen in the winter-time temperature profiles (Figure A.6, discussed below). All $RE$ values between 2000 m and 5000 m are quite high for all 46 stations; for example, $RE$ values at Ellendale are $0.87 > RE > 0.74$ and Washington, $0.91 > RE > 0.80$. $RE$ values are higher closer to the surface and decrease in height as all predictors used for the reconstruction are surface variables.

The original profile (before correction) for Washington, D.C. is shown as a grey profile, together with a constant temperature correction (black profile). The correction, $T = 0.57^\circ$ is calculated as the average of the mean difference for all levels above 2000 m. Figure A.5d shows the difference profile for Cheyenne, a high altitude site (1873 m a.s.l) which reveals the large difference between the observations and reconstructions in the boundary layer. The actual temperature profiles (not shown) reveal that the reconstructions match the reanalysis climatology quite well and underestimate the cold temperatures observed in all seasons, although more so in summer and fall than in winter.

Altogether, all stations except six, out of the total 46 stations that were assessed according to our criteria outline above, passed with no temperature adjustments. Only one station, Dallas, could not be assessed as $RE < 0.5$ for all levels above 2000 m.

In order to better understand the typical profile characteristics of the kite data, seasonal profiles (DJF, MAM, JJA, and SON) from Ellendale are shown in Figure A.6. The original data are shown in black, the reconstructed data in grey and reanalysis (climatology) is shown as a light grey curve (no hatch marks). Differences in the boundary layer between the observations and reconstructions are quite clear and in all seasons the reconstructions and reanalysis are too warm. The pronounced A.5 5a) is clearly dominated by differences in the reconstruction and observations in winter. In winter, the temperature inversion is poorly represented in the reconstructions. Both reanalysis and reconstructed series show a weakened inversion layer in winter-time as compared to the observations. The inversion which is already forming in the fall, is not apparent in either the reconstruction or reanalysis at all. The inversion at Ellendale is typically caused by southerly winds being cooled at the surface as they move to higher latitudes; this cooling produces stable atmospheric conditions and therefore does not extend to the upper levels (Samuels, 1926). The presence of this inversion illustrates the value of having upper-air data available for this region.

In general, both the reanalysis and reconstruction show less seasonal variability than the observations for Ellendale. Similar profiles for Washington show that both the reanalysis and reconstruction are very similar to the observational profile over the seasons, which is also clear in the difference profile in Fig A.5c.
Figure A.7 shows time series of observed and reconstructed anomalies for 3000 m at Ellendale and Washington. The agreement between the reconstructed and observed temperature anomalies is very good; both time series have similar variability and the reconstructions follow the extreme anomalies in the observations quite well. Gaps in the data can be seen for the Washington station for all levels, with the longest apparent break between December 1933 and February 1934. Large anomalies during the 1930s correspond to large surface temperature anomalies during the Dust Bowl period and are discussed below.
Figure A.6: Mean temperature profiles for Ellendale, North Dakota over the seasons. Black line shows observations, grey line with hatches shows the reconstruction and light grey line (no hatches) show 1961-90 climatology from NCEP/NCAR reanalysis.

**Homogeneity tests**

We applied the standard normal homogeneity test (Alexandersson and Moberg, 1997) for each station, using our data series and the corresponding reconstruction as the (only) neighboring series. We also carried out an additional homogeneity test (Lanzante, 1996) for each station. This test was designed for use on upper air time series where an appropriate neighboring station is generally not available. Normally it would be run on the time series of the raw data, however we also opted to run it on the time series of the difference between the observations and reconstructions. For the major-
ity of the stations, all tests revealed no significant coherent pattern of break points or trends between the station and reconstructed reference series at all levels. Although break points may appear due to station relocation or instrument changes, we found no significant breaks corresponding to known shifts or changes in the metadata available. For example, the Washington Naval air station began recording pressure data in January 1937 and no significant shifts were found at this time. Break points were found however in November 1934 for all levels, using the standard normal homogeneity test. As with several other stations, based on this break point, we further split the time series up into before and after the break point. By separating the series, larger offsets in the difference profile after the split were found. Since we have no physical basis for making such an adjustment, only the constant correction (discussed above) was applied to the entire series. Homogeneity tests, while informative, are only one of a suite of tools available to an experienced analyst; we found it important to keep in mind the importance of not over analyzing a short time series (Wunsch, 1999).

We have further investigated changes in instrumentation as discussed in the MWR, and checked gaps in time series which may have been affected. For example, radiometerograph records were introduced at several stations in January 1938, although no break
points in the station time series were found at this time. For the Ellendale station, no clear break points were found. Altogether, although the homogeneity tests revealed break points which were used for further investigation of the time series, no adjustments were justified or carried out.

North American temperature anomalies; 1922-1938

In order to investigate North American temperature anomalies over our historical period (1922-1938), two warm summers during the Dust Bowl period in the 1930s are investigated. Figures A.8 and A.9 show results for summer (JJA) surface air temperature (Rayner et al., 2003; Jones and Moberg, 2003, HadCRUT2v) over North America for the summer of 1932 (hereafter, S1932) and summer of 1936 (hereafter, S1936), together with temperature anomalies from our corrected station series for 2000 m and 4000 m levels. The anomalies from the station series are calculated using 1961-90 NCEP/NCAR reanalysis as climatology.

There is good correspondence between the surface temperature fields and the corrected station data for both summers. The temperature patterns at 2000 m and 4000 m anomalies for S1932, show relatively good correspondence with the surface pattern however there are some discrepancies, most prominently at stations in San Diego and Atlanta. The station network in summer 1932 is not sufficient to capture the gradient across the Rocky Mountains. The station anomalies for the S1936 however, capture many of the small scale features such as the trough in temperature coming down into Washington state from Canada, with negative anomalies in Seattle and positive anomalies in Spokane. Similarly, the cold anomalies in eastern Canada in the surface field are found at upper-levels at Sault Ste. Marie and the east coast of the US, at Boston. At Honolulu the cold Pacific feature is also well captured. 2000 m anomalies at El Paso are slightly negative, which is slightly further south in the surface field, down into Mexico and Baja, however it may be realistic as it is very close to this boundary shown in the surface field. Overall, the magnitude of the anomalies is quite close to those for the surface field, with strong anomalies over the central US.

Winter-time temperature anomalies are shown for 1925/26 and 1936/37 in Figs. A.9 and A.10, respectively. The two winters represent extreme positive and negative values of a PNA index that was reconstructed from independent (pilot balloon) data (see Ewen et al., 2008a). The winter 1925-1926 is an early year in our data set and shows our six kite stations. The corrected 2000 m anomalies for all kite stations show a similar pattern as the surface temperature anomalies, with positive anomalies in the central US...
Figure A.8: Station temperature anomalies for 4000 m (top) and 2000 m (middle) heights from the summer of 1932 (JJA). Surface temperature (JJA) is also shown (bottom) from HadCRUT2v.
Figure A.9: Station temperature anomalies for 4000 m (top) and 2000 m (middle) heights from the summer of 1936 (JJA). Surface temperature (JJA) is also shown (bottom) from HadCRUT2v.
up to Ellendale, and a slightly weaker anomalies at Groesbeck and Due West, South Carolina. At 4000 m, both Groesbeck and Due West show larger negative anomalies with respect to both 2000 m and surface anomalies. The anomaly at Drexel Nebraska is relatively small. Note that this station showed a ‘zig-zag’ pattern in the difference profile, but a correction was not warranted based on the available information. Note also that Royal Center, Indiana, was corrected by $T = 1.13^\circ$.

The winter 1936/37 case with the strong negative PNA index shows remarkably good agreement with the surface anomaly field, as did the summer anomalies in S1936. There is a distinct negative PNA pattern shown in the surface anomaly field across North America, with the coldest anomalies centered around the Rocky mountains, in Alberta, British Columbia and Montana. Cold anomalies extend into the central US and down into California, and warm anomalies over the Great Lakes and eastern US. The 2000 m station anomalies show very good agreement with the surface field, with the exception of Spokane. This station had the largest temperature correction out of all the six corrected stations ($T = -1.67^\circ$). The seasonal profile plots for Spokane (not shown) reveal that the correction greatly improves the warm bias in all seasons, with the exception of winter, which might still be too warm after the correction.

A.7 Conclusions

A new historical upper-air data set has been digitized and re-evaluated from archives of the Monthly Weather Review for the 1922-1947 period. A quality assessment, based on reconstructed reference series was carried out and temperature corrections were applied to six stations. Warm summers during the Dust Bowl period in the 1930s, together with winter-time anomalies over the historical period were selected and corrected station anomalies at 2000 m and 4000 m were compared to surface temperature fields.

This analysis shows that our validation procedure works quite well. From the six stations that required temperature corrections, the station anomaly plots show good agreement between the surface temperature anomalies and the upper levels for most stations. The data match better in the later years, and 1936 summer anomalies and winter 1936-1937, during a strong positive PNA year, show extremely good agreement at all stations, except one (Spokane), and levels.

It is important that additional historical upper-air data, which extend back to the early 20th century, and which may be found in various archives, be digitized and re-evaluated. We have demonstrated that the quality of this early Monthly Weather Review data,
Figure A.10: Station temperature anomalies for 4000 m (top) and 2000 m (middle) heights from the winter of 1925/26 (DJF). Surface temperature (DJF) is also shown (bottom) from HadCRUT2v.
Figure A.11: Station temperature anomalies for 4000 m (top) and 2000 m (middle) heights from the winter of 1936/37 (DJF). Surface temperature (DJF) is also shown (bottom) from HadCRUT2v.
which include data from early kite and aircraft stations, are very high. Our technique to assess the quality is also novel and could be applied to other historical data series and made available to the climate community. Contributions of these data sets to future reconstructions of upper level fields and reanalysis projects would be extremely beneficial and help us to better understand mechanisms related to early 20th climate anomalies and extremes.

Acknowledgements

We would like to thank Sandra Wiche from the National Weather Service for providing us with many helpful historical documents and photos from the Ellendale station during H. Clark Fuller’s time spent as a kite flier for the U.S. Weather Bureau there. We also gratefully acknowledge funding awarded to SB from the Swiss National Science Foundation.
Bibliography


Lanzante, J., 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, 1996.


Väisälä, V., *Der Strahlungsfehler der finnischen Radiosonde*, Mitteilungen des Meteorologischen Instituts der Universität Helsinki No. 47, 1941.


Acknowledgements

I would like to thank Stefan Brönnimann for help and guidance throughout my entire doctoral work. I also thank Tracy Ewen for going above and beyond the call of duty as a group member, coauthor, and most importantly, friend. I thank Atsumu Ohmura for insightful and amusing discussions on many topics.

I am indebted to Annie, Bill, Bob, Valerie, Vicki, and Angela: without your help, I truly could not have done this. I thank Mike and Steve for endless discussions about all things, polar meteorology and otherwise, and for the initial motivation to study climate. I thank my parents for their endless support and enthusiasm for each new adventure I undertake. Finally I express my deep and abiding love and gratitude for my constant companion and comforter, Jesus Christ.
Curriculum Vitae

Andrea Nicole Grant
Seebahnstrasse 121, 8003 Zurich, Switzerland
Citizen of the United States of America

Education

Sc.D. Environmental Science, ETH Zurich, 2008 (expected)
Advisor: Stefan Brönnimann

M.S. Materials Science, University of Minnesota, 2002
Advisors: Robert Cook and Chris Palmstrøm

B.S. Physics, San Francisco State University, 1996, *cum laude*
Advisor: Roger Bland

Research Experience

Research Assistant, ETH Zurich, 2005–2008
Compilation and validation of historical radiosonde data. Analysis of 20th Century
Arctic climate from an upper level perspective.

Staff Scientist, Mount Washington Observatory, New Hampshire, 2003–2005
Digitization and validation of historical meteorological data.

Science Technician, RPSC, Amundsen-Scott South Pole Station, 2000–2001
Responsible for routine operations on six experiments in a remote location. Repaired
hardware and software malfunctions in extreme weather conditions with limited re-
resources.
Research Assistant, University of Minnesota, 1998–2000
Designed and constructed instrumentation to improve flux measurements in Molecular Beam Epitaxy.

Developed hardware and software for second generation semiconductor reliability test instrumentation.

Research Assistant, San Francisco State University, 1994–1996
Developed hardware and software to measure bulk water temperature.

Teaching Experience
Teaching Assistant, Field Course, ETH Zurich, Summer 2007
Teaching Assistant, Numerical modelling of weather and climate, ETH Zurich, Summer 2007
Teaching Assistant, Bachelor Student Seminar, ETH Zurich, Winter 2006–2007
Teaching Assistant, Introduction to Materials Science Lab, University of Minnesota, Fall 1999

Publications
Grant, A; Brönnimann, S; Haimberger, L: Recent Arctic warming strongest near ground. Nature, in press.
Grant, A; Brönnimann, S; Ewen, T: A New Look at Radiosonde Data Prior to 1958. Journal of Climate, in submission.
Brönnimann, S; Annis, J; Dann, W; Ewen, T; Grant, A; Griesser, T; Krahenmann,
S; Mohr, C; Scherer, M; Vogler, C: A guide for digitising manuscript climate data. Climate of the Past, 2, 137–144, 2006.


Conferences and Workshops

NCCR Summer School, Switzerland, August, 2007
IPY Sea Ice Summer School, Svalbard, July, 2007
CLASSIC Spring School, Sweden, April, 2007
EGU General Assembly, Vienna, April, 2007
AMS Annual Meeting, Texas, January, 2007
NCCR Summer School, Switzerland, August, 2006
EGU General Assembly, Vienna, April, 2006
NCCR Summer School, Switzerland, August, 2005

Honors and Awards

Joachim de Giacomi Research Grant, 2007
Outstanding Teaching Assistant, 1999-2000
Graduate Research Fellowship, University of Minnesota, 1999-2000
CIC WISE Travel Grant, 1999
BP Graduate Fellowship, 1998
Physics Department Honored Graduating Student, 1996
Physics and Astronomy Club Undergraduate Research Scholarship, 1996