Editorial Manager(tm) for Climate Dynamics
Manuscript Draft

Manuscript Number:

Title: On the vertical structure of climate variability and change over the Arctic during the past 100 years

Article Type: Original Article

Keywords: Arctic climate; temperature trend; historical data

Corresponding Author: Stefan Bronnimann

Corresponding Author's Institution:

First Author: Stefan Bronnimann

Order of Authors: Stefan Bronnimann;Andrea Grant;Gilbert Compo;Tracy Ewen;Thomas Griesser;Andreas Fischer;Martin Schraner;Alexander Stickler

Abstract: Understanding the vertical structure of tropospheric temperature variability and trends in the Arctic is important for diagnosing the processes behind these variations. Because the Arctic climate undergoes strong changes on annual to multidecadal scales, long data sets of the thermal structure and atmospheric circulation over the Arctic are desirable. However, studies have shown that even recent reanalysis data sets might not always be suitable for this purpose. Here, we systematically compare the thermal structure of the Arctic troposphere using chemistry-climate model simulations and nine observation-based, globally and vertically resolved data sets, in total, spanning more than 120 yrs. We analyze the temperature variability on time scales ranging from daily to multi-decadal trends. Comparisons on the daily scale between reanalysis data and historical upper-air observations were performed for Spitsbergen, Norway for historical extreme seasons (cold winter 1911/1912, warm winter 1944/1945). Excellent agreement was found at upper levels. Near the ground, however, systematic differences occurred. The extreme anomalies during these two winter seasons could be related to persistent atmospheric circulation anomalies. Both historical reanalysis and statistical reconstructions revealed similar patterns of circulation. On the interannual time scale, the correlation between all data sets is high, but there are systematic differences between the data sets in terms of absolute values as well as in terms of the magnitude of variability. With respect to overlapping 20-yr averages and trends in the vertical structure, the data sets also agree relatively well, although differences are more pronounced than for the interannual scale.
Together, the data sets show that, for some Arctic regions, the mean temperature during the last two decades is unprecedented. Importantly, the vertical structures of the early and late twentieth century warming periods are noticeably different. Also, the vertical structure of the recent trend reveals a clear amplification at the surface in autumn to spring. None of the data sets shows a clear sign of a systematic trend in interannual variability. Simulations with a chemistry-climate model qualitatively reproduce the vertical structure of a warming trend in the early twentieth century and recent period. Analysed together and in combination with climate models, the new long data sets presented in this paper will allow further insights into the processes underlying Arctic climate variability.

Suggested Reviewers: Kevin Wood
Kevin.R.Wood@noaa.gov

Expert in historical Arctic data

Takashi Yamanouchi
National Institute of Polar Research, The Graduate University for Advanced Studies, Tokyo
yamanou@nipr.ac.jp
Expert in Arctic climate

Cecilia Bitz
bitz@atmos.washington.edu

Valentin Semenov
vsemenov@ifm-geomar.de

Manuscript
Click here to download Manuscript: Broennimann.doc Click here to view linked References

On the vertical structure of climate variability and change over the Arctic during the past 100 years

Stefan Brönnimann[1,2], Andrea N. Grant[1], Gilbert P. Compo[3,4], Tracy Ewen[5], Thomas Griesser[1], Andreas M. Fischer[5], Martin Schraner[1], Alexander Stickler[1,2],

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

[2] *Oeschger Centre and Institute of Geography, University of Bern, Switzerland*

[3] *Climate Diagnostics Center, CIRES, University of Colorado, Boulder, USA*

[4] *Physical Sciences Division, Earth System Research Laboratory, NOAA, Boulder, USA*

[5] *Department of Geography, University of Zurich,, Switzerland*

[6] *MeteoSwiss, Zurich, Switzerland*

Corresponding author:

Stefan Brönnimann

Oeschger Centre for Climate Change Research and Institute of Geography

University of Bern

Hallerstr. 12

CH-3012 Bern, Switzerland

e-mail: broennimann@env.ethz.ch

Phone: ++41 031 631 88 85

Fax: ++41 031 631 85 11

Abstract

Understanding the vertical structure of tropospheric temperature variability and trends in the Arctic is important for diagnosing the processes behind these variations. Because the Arctic climate undergoes strong changes on annual to multidecadal scales, long data sets of the thermal structure and atmospheric circulation over the Arctic are desirable. However, studies have shown that even recent reanalysis data sets might not always be suitable for this purpose. Here, we systematically compare the thermal structure of the Arctic troposphere using chemistry-climate model simulations and nine observation-based, globally and vertically resolved data sets, in total, spanning more than 120 yrs. We analyze the temperature variability on time scales ranging from daily to multi-decadal trends.

Comparisons on the daily scale between reanalysis data and historical upper-air observations were performed for Spitsbergen, Norway for historical extreme seasons (cold winter 1911/1912, warm winter 1944/1945). Excellent agreement was found at upper levels. Near the ground, however, systematic differences occurred. The extreme anomalies during these two winter seasons could be related to persistent atmospheric circulation anomalies. Both historical reanalysis and statistical reconstructions revealed similar patterns of circulation. On the interannual time scale, the correlation between all data sets is high, but there are systematic differences between the data sets in terms of absolute values as well as in terms of the magnitude of variability. With respect to overlapping 20-yr averages and trends in the vertical structure, the data sets also agree relatively well, although differences are more pronounced than for the interannual scale.

Together, the data sets show that, for some Arctic regions, the mean temperature during the last two decades is unprecedented. Importantly, the vertical structures of the early and late twentieth century warming periods are noticeably different. Also, the vertical structure of the recent trend reveals a clear amplification at the surface in autumn to spring. None of the data sets shows a clear sign of a systematic trend in interannual variability. Simulations with a chemistry-climate model qualitatively reproduce the vertical structure of a warming trend in the early twentieth century and recent period. Analysed together and in combination with climate models, the new long data sets presented in this paper will allow further insights into the processes underlying Arctic climate variability.

1. Introduction

The Arctic is one of the focus regions of climate change studies. On the one hand, the Arctic is a primary heat sink of the global climate system and thus globally relevant. Because of its complex surface cover, a large range of physical, chemical, and radiative processes act on its climate, including atmospheric and oceanic circulation, processes on the ocean-atmosphere-sea ice interface, processes involving the land-surface and cryosphere, clouds, and aerosols. On the other hand, the Arctic is warming much more rapidly than the globe, indicating that there might be interactions and feedback involving these processes. In fact, current global warming is amplified considerably in the Arctic (Symon et al. 2004, Serreze et al. 2006, Solomon et al. 2007). Future climate is also expected to warm much more rapidly in the Arctic than elsewhere (Meehl et al. 2007) and there is also growing concern of irreversible and abrupt changes occurring in the Arctic system (Holland et al. 2008).

The processes behind the recent Arctic warming amplification, however, are still not fully understood (Serreze et al. 2009, Screen and Simmonds 2010). Although diminishing sea ice has been clearly identified as one major contributor, the roles of moisture, clouds, aerosols, the land surface, and atmospheric circulation are debated (e.g., Chapin et al. 2005, Garret and Zhao 2006, Lubin and Vogelmann 2005, Francis and Hunter 2007, Schweiger et al. 2008, Graversen et al. 2008, Simmonds and Keay 2009, Chylek et al. 2009, Bekryaev et al. 2010). In this context it is interesting to note that the Arctic previously underwent a period of pronounced warming in the early twentieth century (e.g., Polyakov et al. 2003, Bengtsson et al. 2004, Overland et al. 2004, Johannessen et al. 2004, Wang et al. 2007, Kauker et al. 2008, Wood and Overland 2010, Wood et al. 2010) which was also accompanied by a global (in particular, North Atlantic) warming period (the Early Twentieth Century Warming, hereafter termed ETCW, see also Brönnimann 2009). Comparing the two periods might reveal important details about the dynamics of Arctic climate processes.

One key to a better understanding of Arctic climate changes lies in the vertical thermal structure, which might give clues to the underlying processes (Graversen et al. 2008, Serreze et al. 2009, Screen and Simmonds, 2010). For instance, based on results from climate model simulations, the vertical structure of Arctic warming seems to be an ideal fingerprint for anthropogenic changes. According to models, in contrast to many other regions of the globe, the Arctic should warm fastest near the ground (Meehl et al., 2007). This is because latent heat release during condensation in the upper troposphere does not play a large role in Arctic

warming, and at the same time, the lower troposphere is often stably stratified, particularly during winter (e.g., Persson et al., 2004). Hence warming processes and amplification mechanisms that involve sea ice or the surface energy budget primarily affect the surface layers. The energy consumed and released during the melting and freezing of ice further modulates the heating of the surface layers and its seasonal cycle, tending to stabilize the atmosphere in spring and summer (during the melting season) and destabilize in autumn and winter.

In contrast, atmospheric circulation changes (which may be anthropogenic or naturally caused) will more likely affect a larger column, perhaps the entire lower and middle troposphere. In fact, Graversen et al. (2008) found that the recent (1979 to 2001) zonally-averaged Arctic warming during the warm season peaked at about 700 hPa and concluded that trends in northward energy transport explained a substantial part of the trends. However, the data set used in that study (ERA-40 reanalysis, Uppala et al. 2005) has known problems in tropospheric temperatures over the Arctic (Bromwich and Wang 2005) and other data sets do not support the findings (Thorne, 2008, Grant et al. 2008b, Bitz and Fu, 2008). Therefore, before addressing changes in the vertical thermal structure, the data quality and suitability needs to be discussed.

The main goal of this study is to assess the quality of various observational data sets containing global vertical temperature data with respect to their representation of Arctic climate. In order of period covered, these are: The Twentieth Century Reanalysis (20CR), two statistical reconstructions (REC1, REC2), upper-air observations (CHUAN), NCEP/NCAR Reanalysis (NNR), ERA-40 reanalysis, JRA25, and ERA-Interim (Table 1).

We analyze variability and trends in these data sets in comparison with model simulations. Although tentative conclusions can be drawn from the ensemble of partly or fully independent data sets, none of the data sets alone is sufficient for addressing long-term trends in the vertical structure. Ultimately, data sets should be created that are suitable for detection and attribution studies of Arctic climate change which previously have been based on the spatial pattern of the Arctic warming near the ground (Gillet et al. 2008).

Section 2 gives a description of the data used. Section 3 shows results in the form of time-height plots. In Section 4 we discuss trends in temperature at levels throughout the troposphere and lower stratosphere in all data sets and compare these trends with interannual variability and with climate model simulations. Conclusions are drawn in Section 5.

2. Data

In this paper we compare nine different data sets (see Table 1) with respect to the temperature variability in the troposphere and lower stratosphere over the Arctic. They comprise observation-based data sets, statistical reconstructions, reanalyses, and an ensemble simulation with a chemistry-climate model. Note that more than these nine data sets could be used. For instance, even within the group of station upper-air data sets, several different products could be compared (e.g., RATPAC (Free et al. 2005), RAOBCORE (Haimberger 2007), and the Iteratively Homogenized radiosonde data (Sherwood et al. 2008)). Also, gridded radiosonde data products could be used (HadAT (Thorne et al. 2005), RICH (Haimberger 2007)). Further, there are other reanalyses (e.g., MERRA, 20CR Version 1, NCEP/DOE Reanalysis 2) and different satellite data sets, at least for the sub-Arctic, could be used. Rather than comparing all available data sets, we focus on those that have a long record as well as on some of the widely used reanalyses data sets.

a. Observations

As a primary source for our comparisons, we use observational datasets, keeping in mind that measurements and averages based on them contain errors. To represent the near-surface air temperature, we use the gridbox anomaly dataset of CRUTEM3v (Brohan et al. 2006). For all higher levels of the atmosphere we use the temperature observations from a combination of radiosonde, kite, and aircraft-based measurement contained in the Comprehensive Historical Upper Air Network (CHUAN, Stickler et al. 2010, Grant et al. 2009a, Brönnimann 2003).

The upper air data were quality assessed following Grant et al. (2009). Corrections were applied up to the end of 1957. For the period from 1958 to present, we used data from the Integrated Global Radiosonde Archive (IGRA, Durre et al. 2006) with RAOBCORE version 1.4 corrections applied (Haimberger 2007).

The European Arctic is believed to be a critical region for our understanding of Arctic climate processes (see Bengtsson et al. 2004). As part of the assessment of the available data sets for this region, several case studies were performed for Spitsbergen, Norway. This region of the Arctic exhibits particularly high temperature variability (see also Grant et al. 2009a). Moreover, data coverage is favorable for Spitsbergen. We chose extremely warm or cold winters for this comparison and present historical upper-air data from tethered balloons and kites from Advents

Bay and Ebeltofthamn, 1911-1913, as well as radiosonde data from Nordaustlandet from 1944-1945, respectively, both performed by German observers.

The Ebeltofthamn data were originally published by Rempp and Wagner (1916), Wegener (1916) and Wegener and Robitzsch (1916a,b). The balloons often did not reach very high altitude, however, during the 22 months of measurements, 80 profiles reached an altitude of 1500 m asl (approximately 850 hPa). Additionally, the vertical levels reported were changed during the period such that a detailed study of inversion layers is not possible with the later data.

The data from 1944/1945 are from the German war operation "Haudegen" led by Wilhelm Dege (Selinger, 2001). In total 132 radiosonde ascents were performed between November 1944 and June 1945. Pilot balloon observations were also made (until Sep. 1945, when the station was finally uncovered, making this the last German unit to surrender), but not used in this project. We digitized temperature and geopotential height (GPH) from the radiosonde measurements on standard pressure levels as given in Dege (1960). The source does not mention whether radiation and lag error corrections were made. Since the data were published in 1960, we assume that these errors were in fact corrected. We also tested the possible bias from using uncorrected data (approach according to Brönnimann, 2003) and found that it might lie between -1 and +0.3 °C (the average over all levels considered here is -0.33 °C). For the winter period (a focus of this paper), when the radiation errors are small, the bias is even smaller.

Note that both the tethered balloon data and the radiosonde data have various sources of uncertainties. These might be particularly large in the harsh Arctic environment. Unfortunately, we have no estimation of the error for these specific Arctic sites. In an accompanying paper (Brönnimann et al. 2010b), we attempt to quantify the error for early ship-based upper air data measured with kites and radiosondes. Here, random errors might be of a similar magnitude of about 1-1.2 °C, in addition to the biases such as those sketched above.

b. Reconstructions

Temperature fields for the period 1880-1957 were taken from a statistical reconstruction (Griesser et al. 2010). The reconstructions are based on a principal component regression. The predictors used were historical surface data from station observations (temperature), gridded SLP, as well as (after 1918) upper-air data (temperature, GPH or pressure, and winds). The

predictands used were GPH 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. This reconstruction is termed REC1. In addition to temperature and GPH, there is also a reconstruction of global 3 km winds (*u* and *v* components) that was performed in the exact same manner and with the same predictor data (see also Stickler et al. 2010); it is also termed REC1 hereafter.

A second reconstruction, REC2, was recently performed in order to avoid the strong limitations of constraining stationary patterns (large-scale empirical orthogonal functions) and thus stationary teleconnections (see Brönnimann et al. 2010a, for details). The approach of REC2 is similar to REC1 except that it was performed grid column by grid column (rather than with hemispheric fields) using only predictors in the "cone of influence" of that grid column (defined as within 1200-1500 km of the location, thus avoiding calibration by means of a possible negatively correlated series). This alleviates the need for stationary patterns, at the expense of a sparse data set. REC2 covers the period 1918-1957. In some figures we show after 1957 the calibrated reconstruction using the predictor network from 1957 for comparison with reanalysis data (REC2-cal., note that gaps in the predictors after 1957 are filled with data extracted from ERA-40, see Brönnimann et al. 2010a for details).

c. Reanalysis data sets

We use five different reanalysis data sets in this study, one of which covers the ETCW in the Arctic. The Twentieth Century Reanalysis version 2 (20CR) is a global 3-dimensional atmospheric dataset that reaches back to 1871 (Compo et al. 2011). It is based on an assimilation of surface observations of synoptic pressure and on HadISST (Rayner et al. 2003) monthly sea surface temperature and sea ice distributions as boundary conditions. Assimilation is performed using an Ensemble Kalman filter with first guess fields generated by a 2008 experimental version of the US National Centers for Environmental Prediction Global Forecast System atmosphere/land model (NCEP/GFS) at a spatial resolution of T62, with 56 ensemble members. Thus 20CR not only provides 6-hourly global analyses (ensemble mean) but also their uncertainty (the ensemble standard deviation). Details and validation results are given in Compo et al. (2011).

Other reanalysis data sets used include NCEP/NCAR (NNR hereafter) from 1948 to 2009 (Kistler et al., 2001), ERA-40 from 1958 to 2002 (Uppala et al., 2005), JRA-25 from 1979 to

2007 (Onogi et al., 2007), and ERA-Interim from 1989 to 2007. These four reanalysis data sets, although widely used, may not all be suitable for all purposes. Errors and inconsistencies in the assimilation system or in the data assimilated can lead to inhomogeneities and errors. Errors relevant for the Arctic include a warm bias in NNR over the former Soviet Union in 1948–1957 due to uncorrected radiation errors in the radiosonde data (Grant et al. 2009a). In the case of ERA-40, problems with satellite radiance assimilation over the ice-covered Arctic Ocean are documented (Bromwich and Wang 2005, Uppala et al. 2005), which could lead to spurious trends (e.g., Thorne 2008, Grant et al. 2008).

d. Model simulations

Observed, reconstructed, and analysed vertical profiles are compared with results from the Chemistry-Climate Model SOCOL (Schraner et al. 2008). We used an ensemble of nine simulations performed in an "all forcings" set-up starting in 1901 and ending in 1999. The model was constrained with monthly varying SSTs (HadISST; Rayner et al. 2003), sea ice, land-surface conditions, stratospheric aerosols, solar variability, surface concentrations of greenhouse gases and ozone depleting substances, emissions of short lived species, and the Quasi-Biennial Oscillation (QBO) in the stratosphere (Fischer et al. 2008). SOCOL is a combination of the middle atmosphere version of ECHAM4 (Manzini and McFarlane 1998) coupled to the chemistry-transport model MEZON (Egorova et al. 2003). It is a spectral model with T30 horizontal truncation and 39 vertical levels with a model top at 0.01 hPa. The model simulations used here participated in an intercomparison within the framework of the CLIVAR "Climate of the 20th Century (C20C)" Project. Results showed a reasonable performance for selected 20th century climate events (Scaife et al. 2009); most notably, the simulations qualitatively (although not quantitatively) reproduced the trend in the North Atlantic Oscillation between the 1960s and the 1990s.

3. Analysis procedure

a. Day-to-day variability

Before turning to interannual and decadal variability, we compared the historical observations (CHUAN) and 20CR at the level of individual profiles for the case of Spitsbergen, Norway. The selected periods include the winters 1911/1912 and 1944/1945, which are among the coldest and warmest winters of the 20th century, respectively. To facilitate comparison we subtracted a common climatology from each data set. We used NNR data for this purpose,

namely a climatology of daily mean values as a function of the day of year that is given and recommended on the website of NOAA/PSD and refers to the period 1968-1996 (note that for the comparisons of the interannual variability, where more data sets than NNR are involved, we use 1961-1990 as a reference). These data also were subsampled and interpolated to the location and time of the ascents.

b. Seasonal-regional averages

A special focus of this paper is on seasonal and regional averages, i.e., various regions of the Arctic and different seasons are analyzed separately. Averaging is straight forward for gridded data, but more difficult (both with respect to time and space) for the observations. A glance at the upper-air observational data coverage for temperature is shown in Fig. 1, which gives the number of observations per level, season, and year for the sector [10°W-100°E, 60°N-90°N]. The number of observations was very high in the 1970s and 1980s and has decreased back to the level of the early 1950s since then. The number also decreases very rapidly (by more than an order of magnitude) particularly before the 1950s. Also, due to the different reporting some of today's standard pressure levels are missing. In the early years, measurements were not always performed daily. There were many gaps, and valid monthly means (for CHUAN monthly means require 13 days of observations or no gap longer than one week) can hardly ever be calculated, although there would be more data.

Therefore, in order to better exploit all observations, the following procedure was used. The Arctic was divided into 54 equal area grid cells, as shown in Fig. 2, and time was subdivided into weeks. The scale of both the grid cells (approx. 800 km x 800 km) and the seven day blocks were chosen as representative of the intraseasonal large scale in order to maximize the information contained in the spatially and temporally sparse measurements. Anomalies of individual soundings were calculated relative to a 1961-1990 monthly climatology from NNR for each location and then averaged within the equal area grid cells and seven day blocks. The mean values per grid cell and week were then aggregated into sectors and seasons.

The four seasons were defined as the 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). The overlaps (1 March, 1 December) are necessary for obtaining an integer number of weeks (thirteen) for averaging. Rather than dividing the Arctic into three or four large sectors, the sectors were defined as combinations of 4-6 neighbouring grid cells with good in-situ data

coverage (Fig. 2), thus further increasing the density of observations. Seasonal-regional means were then calculated from the grid cell averages if 50% of the grid cells in a region and 50% of the weeks in the season had data. Still, averages for the period prior to 1948 will have a different quality than those from the period later on and should be treated with caution. It is therefore important to compare this data set with other available data sets. For the gridded data sets we simply averaged the region for the sectors as shown in Fig. 2 and used climatological seasons rather than to subsample all data sets to the exact times and locations of the observations. Note that for this reason, the comparability between observations and other data sets is limited (i.e., the difference also contains the error of representativity, as will be discussed), but on the other hand the interpretation of the result is more robust as they represent the mean value for a sector (whereas the sub-sampling would "transfer" uncertainties in the observational data, such as those due to changes in locations, over to the other data sets).

Seven regions could be defined by the procedure described above, resulting in 28 combinations of seasons and regions. We have analyzed all combinations and report on these results at the beginning of Section 4b, but for brevity's sake we show figures for only four sectors, each for one season. We chose to show (1) the European Arctic in winter, (2) Western Siberia in spring, (3) the Bering Strait region in summer and (4) the Baffin region in autumn (see Fig. 2 for definition). These combinations capture different characteristics of Arctic climate. Moreover, combinations (1), (3), and (4) correspond to regions and seasons with a large variability in sea ice. Note that the Arctic Ocean is underrepresented and land areas are overrepresented in this selection.

We show seven levels, namely 1000, 850, 700, 500, 400, 300, and 200 hPa. For the observation-based data set, surface air temperature (from CRUTem3v, Brohan et al. 2006) is shown rather than 1000 hPa from CHUAN, which is often extrapolated or not reported. Due to irregular reporting, the 925 and 600 hPa levels were omitted in the CHUAN averages. Similarly to the day-to-day variability, we analyze the regional-seasonal averages in the form of anomalies. For this purpose, the mean annual cycle from the years 1961-1990 was subtracted. All analyses were performed using both NNR and ERA-40 as a common reference as well as using each data set as a self-reference (which however only works for long data sets). Due to the documented errors in the vertical temperature structure in the Arctic in ERA-40 (Bromwich and Wang 2005) we show mainly the analyses with NNR as a common reference.

c. 20 year means and trends

In order to address lower frequency variability in Arctic climate, we analyzed moving 20-yr averages and moving 20-yr trends for the seasonal and regional averages defined above. We also analyzed interannual variability within the moving 20-yr windows in order to address long-term changes in interannual variability. The size of the window (20-yr) reflects the fact that Arctic temperature is known to show variability on this time scale (e.g., Polyakov et al. 2003, Overland et al. 2004). Low-frequency processes might exist which have an effect on Arctic climate. It also reflects the length of the shortest data set. The analyses are then performed with 10-yr overlapping windows (i.e., 20 yr windows moving in steps of 10 years). Longer lasting trends thus appear as a sequence of trend profiles with the same sign or as difference between 20-yr averages.

Several starting and ending years of data sets lie in the years 7-9 of a decade (NNR, ERA-40, JRA-25 and ERA-Interim start in 1948, 1957, 1979 and 1989, respectively, REC1 and SOCOL end in 1957 and 1999, most other data sets between 2007 to 2009). Therefore, to fully exploit the lengths of the data sets we chose the intervals 1908-1927, 1918-1937, …, 1988-2007. Not more than five missing seasons are allowed, neither the first nor the last 2 years can be missing. Trends were calculated using least squares regression.

4. Results and discussion

a. Day-to-day variability during cold and warm winters in the early decades

The analysis of a very warm and a very cold winter in Spitsbergen may illustrate the potential role of atmospheric circulation anomalies and show similarities and differences in temperature and circulation depicted in the different data sets. Figure 3 shows a direct comparison between temperatures profiles from tethered balloons and kites and reanalysis data from 1911-1913; expressed as anomalies from the NNR climatology (1968-1996). The observations often show strongest anomalies near the ground, which may be a real feature or due to an inaccurate depiction of the surface layer in the reference (NNR). The observations for the cold winter show relatively shallow surface inversions (<200 m), and sometimes inversions at higher levels (200-1100 m asl). The profiles from 20CR (Fig. 3) are much warmer near the ground, particularly in winter and particularly during cold days, which might be partly due to a likely bias in specifying sea ice, leading to anomalous heat flux (Compo et al. 2011). However, other factors (i.e., specific local conditions, interpolation, time mismatch, etc.) might also contribute.

Despite these systematic differences, we find relatively good correlations of the anomalies on a day-to-day scale (Table 2). At the surface, correlations are low (around 0.4), but above around (1000 m asl) we find anomaly correlations of 0.6 to 0.8. Single warm or cold profiles are well reproduced.

Figure 4 shows the same analysis for the winter 1944/1945. Here, radiosondes were used rather than tethered balloons, covering the entire troposphere and even lower stratosphere (not shown). The results are similar to those for the 1911/1912 case. Again, the warm bias at the surface in 20CR is clearly visible, and a cold tropopause bias appears. But otherwise the agreement is excellent, with anomaly correlations (Table 3) between 0.7 and 0.9 in the lower and middle troposphere. For GPH, anomaly correlations exceed 0.9 from 850 to 300 hPa. Hence, both data sets seem to capture similar features of day-to-day variability. Strong positive temperature anomalies of 10 °C or more appear simultaneously in both data sets.

In view of the errors in the historical upper-air data, the interpolation procedure, and the possible effect of the time mismatch (0-3 hours) the correlations in both episodes are considered to be very good. For further discussion and a more quantitative assessment that includes different sources of error in the upper-air data as well as the ensemble spread in 20CR see Brönnimann et al. (2010b).

One expects that the extreme temperature anomalies in both cases are related to anomalous atmospheric circulation. Interestingly, in both cases, the temperature anomalies show a high persistence. The cold anomalies in the winter 1911/12 lasted from late December 1911 to mid-February 1912, with only a few anomalously warm profiles. The warm period in 1944/1945 lasted from the start of the measurements in November 1944 to mid-March 1945, interrupted only by two cold spells in mid-winter. Consequently, we analyzed the seasonally averaged circulation for the two winters (December through February). Figure 5 shows the wind fields for 700 hPa (20CR) and 3 km (REC1) respectively. The two fields are in very good agreement for the case of the winter 1911/1912. Spitsbergen was in a northwesterly flow and likely often received air from the high Arctic. The westerlies were shifted to lower latitudes and the northern North Atlantic was under a high-pressure influence. In fact, 20CR shows frequent blocking in this region. The low temperatures in Spitsbergen are thus in agreement with the wind fields.

In contrast to the situation of 1911/12, during the warm winter 1944/45, southwesterly advection dominated, transporting warm air masses to Spitsbergen. The westerlies are shifted northwards in both data sets. However, there are slight differences between the two data sets in that the southerly component was stronger in 20CR than in REC1.

The analyses of these two winters fit well with the decadal scale circulation anomalies sketched by Grant et al. (2009b) for cold and warm decades before and during the ETCW in the Arctic. The cases show that circulation variability is similar in the historical data sets; in fact, the agreement for geopotential height is even better than for temperature (Table 3). Hence uncertainties in atmospheric circulation are not the main contributor to uncertainties in temperature.

b. Interannual variability

Before turning in more detail to interannual variability in the four selected seasonal-regional averages, we briefly highlight the main features of the analyses of all 28 seasonal-regional averages in the CHUAN data. Winter exhibits the largest interannual variability, spring and autumn are less variable, while summer has the smallest variability. Winter also has a number of multi-annual clusters of warm (or cold) years which is not seen in summer. The vertical structure of individual years is more coherent in autumn than in other seasons. Another feature is an anti-correlation of the temperature anomalies in the troposphere and lower stratosphere (the seemingly random anomalies at 300 hPa in winter and 200 hPa in summer are due to the variable height of the tropopause). A stretch of cold winters in the 1950s at all levels in the Bering Strait region combined with a warm upper troposphere in summer in the same years is unusual, as is the vertically coherent warm period in the European Arctic in the early 1970s winters when the rest of the Arctic was quite cold.

For the four selected seasonal-regional averages, the interannual variability in all data sets is shown in the form of time-height cross-sections in Figs. 6-9. We use NNR as a common reference here for all data sets. Note that, in several cases, we have combined more than one data set in one panel for ease of presentation.

In terms of differences between data sets, several features are worth noting in these figures. Firstly, there are differences between the data sets in terms of absolute values. As noted earlier, 20CR is warmer than NNR and in fact warmer than all other data sets near the ground, and it is colder than all others near the tropopause. ERA-40 and ERA-Interim are also warmer at the

surface than NNR. Both ERA datasets have prescribed fractional sea ice concentration in a grid box, while NNR has prescribed either 100% or 0% only. Note that in the case of 20CR part of this difference can be attributed to a likely error in the specification of sea-ice in 20CR (Compo et al. 2011). However, other factors might also contribute. The stable arctic boundary layer is arguably not well suited for comparing reanalysis data sets. Small differences in the representation of the boundary layer, orography, and interpolation to pressure levels might have large effects on such a comparison.

Secondly, the amount of variability varies greatly between the data sets. CHUAN shows a relatively high variability in the early years that contrasts with that in later years. This variability is likely an artifact of the sampling. A variance adjustment could be applied in this case (here we stick to simple processing steps that do not go beyond averaging). Other data sets such as REC1 or REC2 show very little variability. Again, this is understandable from the data sets, which in essence are based on linear regression and thus underestimate the variance by construction. The 20CR, in turn, seems to show a similar amount of variability in the earlier period as in later periods or in other reanalysis data sets. This suggests that the amount of variability is well captured in the 20CR. This is interesting as the simulations with the chemistry-climate model SOCOL show clearly less variability in Arctic surface temperatures prior to 1950 than afterwards (we take up this point again in Sect. 4c) which we interpret as the consequence of the almost constant sea ice used as a boundary condition in the first half of the century. The same sea-ice data were provided to 20CR, but obviously did not have this adverse effect.

Several features of interannual variability appear in all data sets, e.g., the cold winters in 1940-1942 in the European Arctic that were likely caused by related to the El Niño event at that time (Brönnimann et al. 2004). Other prominent features are the cold anomalies in spring in Western Siberia in the 1960s and the warm early 1990s in almost all seasons and sectors. The warm anomalies in the NNR in the upper troposphere over Western Siberia in spring in the late 1940s and early 1950s (Fig. 6) likely are an artifact of problems in data processing. Approximately 30 stations in the former Soviet Union have a suspected undercorrected radiation and lag error during that time, which is corrected in CHUAN but not in NNR (Grant et al. 2009a).

The interannual variability is well captured in most data sets. As an example, Table 4 shows the correlation between 20CR and CHUAN in the European Arctic in all seasons. Correlations are between 0.75 and 0.93 for winter and spring throughout the lower and middle troposphere.

Correlations decrease at the tropopause level (most likely due to varying tropopause height), and they are considerably smaller for the spring and summer seasons. For this analysis, the entire length of the series was used.

In order to better understand the differences between the historical data sets, we analyze correlations for the 1930-1957 period for winter (Dec.-Feb.) for all combinations of the four historical data sets (Table 5). In general, correlations are high, but there are slight differences. REC1 shows the lowest correlation with observations (CHUAN) as well as with all other data sets (REC2, 20CR). In the troposphere, the highest correlations are found for 20CR. REC2 shows slightly lower correlations with CHUAN than 20CR but shows the highest correlations of all data sets in the stratosphere.

This result is confirmed when extending the comparison also to the later reanalysis data sets in Table 6. Here we show the correlation of monthly anomalies (again for the European Arctic) between all gridded data sets for different time periods (note that in the latter case anomalies are based on a self-climatology in order to avoid effects of seasonally dependent biases). With the exception of REC1, all correlations up to 500 hPa are >0.83 in the 1918-1957 period and >0.93 thereafter. The main result of this comparison is again that the data sets agree well among each other. However, REC1 performs somewhat worse than the other data sets, 20CR performs well in the troposphere but not the stratosphere, while REC2 performs well in the stratosphere. Note that the data sets are not fully independent.

c. Bi-decadal means and trends

Figures 10-13 show vertical structures of temperature trends in overlapping 20-yr periods for different data sets for the four seasonal-regional averages. Trends are not consistent through time, space, and season. Positive trends alternate with negative trends, though it is visually apparent that positive trends dominate in the troposphere compared to negative trends in the lower stratosphere. The data sets once again agree relatively well, but there are differences in magnitude. The agreement is promising as some of the data sets are fully or partly independent.

One common feature of most or all seasons is that the warming is especially strong in the 1978-1997 period, which is due to a step-like warming in the late 1980s, at least in some of the sectors. For the European Arctic and Siberia in winter and spring, possible explanations include the overlay of a pronounced global warming episode with a sequence of an El Niño/Southern Oscillation cycle in 1986-1989 (an El Niño event with low Arctic temperatures followed by a

La Niña event with high temperatures in the Arctic), followed by a volcanically perturbed period (Pinatubo eruption in 1991, accompanied by a winter warming). A strong trend in the North Atlantic Oscillation between the 1960s and the 1990s (which itself could be partly due to the same causes) also might have contributed. The role of other influences (e.g., solar variability, ozone depletion, tropospheric aerosols) remains to be examined in more detail.

In autumn (Baffin region, Fig. 13), the pattern is somewhat different. Positive trends throughout the troposphere are very large in the past 20 years, a feature that is clearly unprecedented in the past 100 years.

Another common feature in all data sets and seasons is the cooling trend in 1948-1967, following the peak of the ETCW. In NNR over Siberia in spring, the cooling is particularly pronounced which is most likely due to the warm bias in the first half of that period. However, other data sets also show a consistent cooling.

Concerning the vertical structure, all recent warming trends except for summer (Bering Strait region) are strongest near ground. For the ETCW the vertical structure of the trend is data set dependent.

The chemistry-climate model simulations (ensemble mean) qualitatively reproduce the two-step warming, i.e., a phase of warming in the first half of the 20th century and a recent warming, but the rates of the two warmings are smaller than in observations, which is to be expected due to ensemble averaging (see also Figure 10 in Fischer et al., 2008). Shorter features such as the cooling in 1948-1967 are not reproduced. Concerning the vertical structure, the model suggests that the recent warming is strongest near the ground in all seasons, similar to the observation-based data sets (note that the 1988-2007 period cannot be shown for the model because model simulations stop in 1999).

The 20-yr trends show large differences from one time window to the next. In order to focus on the multidecadal changes, we compare 20-yr averages for these different time windows in Figs. 14 and 15. Here the data are expressed with respect to self-climatologies of the period 1961-1990 in order to remove biases (consequently, JRA-25 and ERA-Interim cannot be shown). Individual data sets are shown with symbols. First we focus on a comparison between the warm periods during the ETCW (1918-1937; only 20CR and REC1 are available for this period) and during the recent warming (1988-2007, CHUAN, 20CR, NNR), respectively (Fig. 14). The profiles are well constrained in the recent period, while there are relatively large differences

during the ETCW. However, despite these differences a change in the profile shape appears, with a stronger warming near the ground than aloft in 1988-2007 in all data sets and seasons (except for 20CR in spring) and an opposite tendency in 1918-1937 (though with exceptions).

In Fig. 15 we show all 20-yr periods in all data sets. Although there are differences between the observation-based data sets) as outlined above, we also plot the average of all available observation-based data sets (solid lines) to visualise the common features. Note that the averaged profile should be treated with caution as it contains data sets (e.g., ERA-40) that are known to have errors. Most or all observation-based data have issues in the Arctic that may affect the trends in the vertical structure. Despite these issues, an average profile may help for visualization. For ease of presentation we show only the spread of the data sets in Figure 15 (a "full" version of the figure with each data set shown in as a different symbol is given in the supplementary material).

The spread in the ensemble of observation-based data sets for the early twentieth century is affected by arguably artificial trends in 20CR. Most notably, 20CR shows much higher anomalies than the other data sets in the lower stratosphere in autumn to spring and the opposite near the ground in summer. Figure 15 shows that the average for the last 20-yr period (1988-2007) not only differs from the 1918-1937 period, but from all other periods. The magnitude of the warm anomaly is unprecedented in all seasons and regions. Although there are differences between the data sets, the recent warming surpasses that of the 1918-1937 and 1928-1947 periods even in the European sector in winter (where the ETCW was strongest) and even more so in the other seasons and regions. The average over all data sets (solid line) is outside the spread (bars) of any period in all seasons up to 850 hPa (in winter) or higher. Also, a very clear vertical structure appears, with the strongest warm anomalies near the ground. This is different from the ETCW, which rather seems to show a vertically coherent warming. However, there are differences between the data sets in the early decades, many of the 20-yr profiles do not agree between the data sets, resulting in a large spread.

The climate model simulations agree fairly well again with respect to the two phases of warming and the vertical structure of the anomalies except for the stratospheric warming in the early twentieth century, which is likely overestimated in 20CR and hence affects the averaged profile. The model shows a too cold lower troposphere in the first half of the twentieth century in winter. The variability between 20-yr averages is much smaller, which is expected since the model simulations represent an ensemble mean over 9 simulations.

The vertical structure of the variability in the 20-yr averages follows that of interannual variability relatively closely (Fig. 16). The main features here are the maxima of variability at the surface (except summer) and in the lower stratosphere and the minimum at 300 hPa (ca. 9 km asl) above the tropopause in all seasons. Also noteworthy is the fact that in summer, variability is smaller near the ground than at 850 hPa, which might be due to melting sea ice. The largest variability is found in spring. In winter, there is an additional maximum of variability at 700 hPa, which might indicate an effect of circulation variability. We find no clear evidence for a change in variability, although we note that the two most recent periods have above average variability in spring and autumn (note, however, that 20 yrs are not sufficient to give a good estimation of the standard deviation).The climate model simulations produce a similar vertical pattern of variability, with strongest near-surface variability in autumn and spring. The model shows a clear increase in variability with time, which we interpret as an artifact of the sea-ice boundary condition used. In the first half of the twentieth century, sea-ice concentration in the HadISST data set that was used as boundary condition exhibits far too little variability (Rayner et al. 2003).

5. Conclusions

The vertical thermal structure of the Arctic troposphere is studied in nine observation-based data sets and a set of model simulations spanning the past 20-140 years. Comparisons on the day-to-day scale between historical reanalysis data and upper-air observations for Spitsbergen, Norway for a very cold winter (1911/1912) and a very warm winter (1944/1945) reveal a generally good agreement, but systematic differences near the surface and in the stratosphere were found. Excellent agreement is found for tropospheric circulation patterns. The temperature anomalies during these two anomalous periods appear to be consistent with the observed persistent atmospheric circulation anomalies.

The correlation between all data sets on the interannual time scale is also high, but there are systematic differences in term of absolute values as well as in terms of the magnitude of variability. Good agreement was also found for trends within overlapping 20-yr periods, while for 20-yr averages a larger spread between data products was found.

Note that almost all of the data sets have known problems in the Arctic. ERA-40 has problems with satellite radiance assimilation over the ice-covered Arctic Ocean (Bromwich and Wang 2005, Uppala et al. 2005), as discussed above. Bromwich et al. (2007) performed an assessment

for the main reanalyses ERA-40, NNR, and JRA-25 in the polar regions and discussed differences in the data sets (addressing also clouds and cyclones). Lüpkes et al. (2010) compared ERA-Interim data with ship-based observations and found problems in the specification of sea-ice in ERA-Interim. NNR has a warm bias over the former Soviet Union in 1948–1957 due to uncorrected radiation errors in the radiosonde data (Grant et al. 2009a). Deficiencies are also found for the long data sets discussed in this paper. 20CR has a warm bias due to misspecification of sea-ice (Compo et al. 2011) and a cold bias in the stratosphere. Upper-air observations may have remaining instrumental biases, especially in the early years. Furthermore, their coverage leads to a "coastal bias" of regional averages. Finally, both reconstructions (REC1, REC2) have not been validated for trend analysis. By construction, they are only able to reproduce trends whose spatial or vertical patterns can be represented by the patterns of interannual variability.

It should be stressed, however, that all these data sets are extremely valuable for studying Arctic climate variability. Knowing the shortcomings and differences, information can be gained even on trends from analysing all data sets individually and by combining the results.

Analyzed together, the data sets show that the last two decades appear unprecedented in terms of the magnitude of the warm anomaly in the lower troposphere. The rate of warming between the early 1980s and present is also outstanding. The vertical structure of the trend shows a clear amplification of the recent trend at the surface in autumn to spring. The ETCW does not seem to show a clear surface amplification, but the data are more uncertain and no firm conclusion can be drawn. No trend in interannual variability is found. Simulations with a chemistry-climate model qualitatively reproduce both the ETCW and the recent warming, including their vertical structure. Although tentative conclusions can be drawn from the ensemble of partly or fully independent data sets, none of the data sets alone is sufficient for addressing long-term trends in tropospheric or stratospheric temperature. In a synopsis and in combination with climate models, however, the new long data sets will allow further insights into the processes underlying Arctic climate variability. Ultimately, new data sets should be created that are suitable for trend analysis, but more work needs to be done, including further recovery of historical upper-air data from the Arctic. The new upper-air data presented in this paper will be incorporated into the CHUAN data set (Stickler et al. 2010).

Acknowledgment. This work was supported by the Swiss National Science Foundation through the project "Past climate variability from an upper-level perspective" as well as through the National Competence Centre in Research (NCCR) Climate. The Twentieth Century Reanalysis Project (20CR) used resources of the National Energy Research Scientific Computing Center and of the National Center for Computational Sciences at Oak Ridge National Laboratory, which are supported by the Office of Science of the U.S. Department of Energy under Contract No. DE-AC02-05CH11231 and Contract No. DE-AC05-00OR22725, respectively. Support for the 20CR dataset is provided by the U.S. Department of Energy, Office of Science Innovative and Novel Computational Impact on Theory and Experiment (DOE INCITE) program, and Office of Biological and Environmental Research (BER), and by the National Oceanic and Atmospheric Administration Climate Program Office. Constructive comments by P. Sardeshmukh of the University of Colorado CIRES/CDC and NOAA/ESRL/PSD on an earlier version of this manuscript are greatly appreciated.

References

Bekryaev RV. Polyakov IV, Alexeev VA (2010) Role of Polar Amplification in Long-Term Surface Air Temperature Variations and Modern Arctic Warming. J Clim 23:3888–3906

Bengtsson L, Semenov V, Johannessen O (2004) The early twentieth-century warming in the Arctic–a possible mechanism. J Clim 17:4045–4057

Bitz CM, Fu Q (2008) Arctic warming aloft is data set dependent. Nature 455:E3-E4

Brohan P, Kennedy J, Harris I, Tett S, Jones P (2006) Uncertainty estimates in regional and global observed temperature changes: A new data set from 1850. J Geophys Res 111: D12106. doi:10.1029/2005JD006548.

Bromwich D, Wang S (2005) Evaluation of the NCEP-NCAR and ECMWF 15-and 40-yr reanalyses using rawindsonde data from two independent Arctic field experiments. Mon Wea Rev 133:3562–3578

Bromwich DH, Fogt RL, Hodges KI, Walsh JE (2007) A tropospheric assessment of the ERA-40, NCEP, and JRA-25 global reanalyses in the polar regions. J Geophys Res 112:D10111, doi:10.1029/2006JD007859

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

Brönnimann S (2009) Early twentieth-century warming. Nature Geosc 2:735-736.

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

Brönnimann S, Griesser T, Stickler, A (2010a) A gridded monthly upper-air data set back to 1918. Clim Dyn doi:10.1007/s00382-010-0940-x (online first)

Brönnimann S, Compo G.P, Allan R, Adam W, Spadin R (2010b) Early ship-based upper-air data and comparison with the Twentieth Century Reanalysis. Clim Past Disc 6:2423-2453

Compo GP and co-authors (2011) The Twentieth Century Reanalysis Project. Q J R Meteorol Soc (submitted)

Chapin FR and co-authors (2005) Role of land-surface changes in Arctic summer warming. Science 310:657–660

Chylek P, Folland CK, Lesins G, Dubey MK, Wang M (2009) Arctic air temperature change amplification and the Atlantic multidecadal oscillation. Geophys Res Lett 36:doi:10.1029/2009GL038777

Dege W (1960) Wissenschafliche Beobachtungen auf dem Nordostland von Spitzbergen. Berichte des Deutschen Wetterdienstes Nr. 72 (Bd. 10), Offenbach a M

Durre I, Vose RS, Wuertz DB (2006) Overview of the Integrated Global Radiosonde Archive. J Clim 19:53-68

Egorova T, Rozanov E, Zubov VA, Karol IL (2003) Model for investigating ozone trends (MEZON). Atmos Ocean Phys 39:277-292

Fischer AM and co-authors (2008b) Interannual-to-Decadal Variability of the Stratosphere during the 20th Century: Ensemble Simulations with a Chemistry-Climate Model. Atmos Chem Phys 8:14371-14418.

Francis JA, Hunter E (2007) Changes in the fabric of the Arctic's greenhouse blanket. Environ Res Lett 2:doi:10.1088/1748–9326/2/4/045011

Free M, Seidel DJ, Angell J, Lanzante JK, Durre I, Peterson TC (2005) Radiosonde Atmospheric Temperature Products for Assessing Climate (RATPAC): A new data set of large-area anomaly time series. J Geophys Res 110:D22101. doi:10.1029/2005JD006169

Garrett T, Zhao C (2006) Increased Arctic cloud longwave emissivity associated with pollution from mid-latitudes. Nature 440:787–789

Gillett NP, Stone DA, Stott PA, Nozawa T, Karpechko AY, Hegerl GC, Wehner MF, Jones PD (2008) Attribution of polar warming to human influence. Nature Geosc 1:750-754, doi:10.1038/ngeo338

Grant A, Brönnimann S, Haimberger L (2008) Recent Arctic warming vertical structure contested. Nature 455:E2-E3. doi:10.1038/nature07257

Grant A, Brönnimann S, Ewen T, Nagurny A (2009a) A new look at radiosonde data prior to 1958. J Clim 22:3232-3247

Grant AN, Brönnimann S, Ewen T, Griesser T, Stickler A (2009b) The early twentieth century warm period in the European Arctic. Met Z 18:425-432

Graversen R, Mauritsen T, Tjernström M, Källèn E, Svensson G (2008) Vertical struture of recent Arctic warming. Nature 451:53–56

Griesser T, Brönnimann S, Grant A, Ewen T, Stickler A, Comeaux J (2010) Reconstruction of global monthly upper-level temperature and geopotential height fields back to 1880. J Clim 23:5590-5609

Haimberger L (2007) Homogenization of radiosonde temperature time series using innovation statistics. J Clim 20: 1377–1403

Holland MM, Bitz CM, Tremblay B, Bailey DA (2008) The role of natural versus forced change in future rapid summer Arctic ice loss, In Arctic Sea Ice Decline: Observations, Projections, Mechanisms, and Implications (ed. DeWeaver ET, Bitz CM, Tremblay LB), Geophys Monogr Ser 180:133-150, AGU, Washington, D. C.

Kauker F, Köberle C, Gerdes R, Karcher M (2008) Modeling the 20th century Arctic Ocean/Sea Ice system: Reconstruction of surface forcing. J Geophys Res 113:C09027

Kistler R and co-authors (2001) The NCEP-NCAR 50-year reanalysis: monthly means CD-ROM and documentation. B Amer Meteorol Soc 82:247-267

Johannessen OR and co-authors (2004) Arctic climate change: observed and modelled temperature and sea-ice variability. Tellus 56A:328–341

Lubin D, Vogelmann A (2006) A climatologically significant aerosol longwave indirect effect in the Arctic. Nature 439:453–456

Lüpkes C, Vihma T, Jakobson E, König‐Langlo G, Tetzlaff A (2010) Meteorological observations from ship cruises during summer to the central Arctic: A comparison with reanalysis data. Geophys Res Lett 37:L09810, doi:10.1029/2010GL042724

Manzini E, McFarlane N (1998) The effect of varying the source spectrum of a gravity wave parameterization in a middle atmosphere general circulation model. J Geophys Res 103:31523-31539

Meehl GA and co-authors (2007) Global Climate Projections. In Climate Change 2007: The Physical Science Basis. Contribution of Working Group I to the Fourth Assessment Report

of the Intergovernmental Panel on Climate Change (ed. Solomon S et al.). Cambridge University Press, Cambridge, United Kingdom and New York, NY, USA, pp. 747-845

McConnell J R and co-authors (2007) 20th-century industrial black carbon emissions altered Arctic climate forcing. Science 317:1381–1384

Onogi K and co-authors (2007) The JRA-25 Reanalysis. J Met Soc Jap 85:369-432

Overland J, Spillane M, Percival D, Wang M, Mofjeld H (2004) Seasonal and regional variation of Pan-Arctic surface air temperature over the instrumental record. J Clim 17:3263–3282

Persson P O G, Fairall C W, Andreas E L, Guest P S, Perovich D K (2002) Measurements near the Atmospheric Surface Flux Group tower at SHEBA: Near-surface conditions and surface energy budget. J. Geophys. Res. 107: 8045, doi:10.1029/2000JC000705

Polyakov I, Bekryaev R, Alekseev G, Bhatt U, Colony R, Johnson M, Maskshtas A, Walsh D (2003) Variability and trends of air temperature and pressure in the maritime Arctic, 1875-2000. J Clim 16:2067–2077

Rayner NA, Parker DE, Horton EB, Folland CK, Alexander LV, Rowell DP, Kent EC, Kaplan A (2003) Global analyses of sea surface temperature, sea ice, and night marine air temperature since the late Nineteenth Century. J Geophys Res 108:4407, doi:10.1029/2002JD002670

Rempp G, Wagner A (1916) Die Temperaturverhältnisse über Spitzbergen in the Adventbai 1911/12. Veröffentlichungen des Deutschen Observatoriums Ebeltofthafen-Spitzbergen 1, 27 pp.

Scaife AA and co-authors (2009) The CLIVAR C20C Project: Selected 20th century climate events. Clim Dyn 33:603-614,

Schraner M R and co-authors (2008) Technical Note: Chemistry-climate model SOCOL: version 2.0 with improved transport and chemistry/microphysics schemes. Atmos Chem Phys 8:5957-5974

Schweiger AJ, Lindsay RW, Vavrus S, Francis JA (2008) Relationships between Arctic sea ice and clouds during autumn. J Clim 21:4799–4810

Screen JA, Simmonds I (2010) The central role of diminishing sea ice in recent Arctic temperature amplification. Nature 464:1334–1337

Selinger F (2001) Von „Nanok“ bis „Eismitte“ - Meteorologische Unternehmungen in der Arktis 1940-1945. Schriften des Deutschen Schiffahrtsmuseums

Serreze MC, Francis JA (2006) The Arctic amplification debate. Clim Change 76:241–264

Serreze MC, Barrett AP, Stroeve JC, Kindig DN, Holland MM (2009) The emergence of surface-based Arctic amplification. The Cryosphere 3:11-19

Sherwood SC, Meyer CL, Allen RJ (2008) Robust tropospheric warming revealed by iteratively homogenized radiosonde data. J Clim 21:5336–5350

Simmonds I, Keay K (2009) Extraordinary September Arctic sea ice reductions and their relationships with storm behavior over 1979–2008. Geophys Res Lett 36:L19715, doi:10.1029/2009GL039810

Solomon S and co-authors (eds) (2007) Climate Change 2007: The Physical Science Basis. Contribution of Working Group I to the Fourth Assessment Report of the Intergovernmental Panel on Climate Change. Cambridge University Press, Cambridge, United Kingdom and New York, NY, USA

Stickler A and co-authors (2010) The Comprehensive Historical Upper Air Network (CHUAN). B Amer Meteorol Soc 91:741-751. doi: 10.1175/2009BAMS2852.1

Symon C, Arris L, Heal B (eds) (2004) Arctic Climate Impact Assessment. Cambridge Univ. Press

Thorne P (2008)Arctic tropospheric warming amplification? Nature 455:E1-E2, doi:10.1038/nature07256

Thorne PW, Parker DE, Tett SFB, Jones PD, McCarthy M, Coleman H, Brohan P (2005) Revisiting Radiosonde Upper-Air Temperatures From 1958 to 2002. J Geophys Res 110:D18105, doi:10.1029/2004JD005753

Uppala SM and co-authors (2005) The ERA-40 re-analysis. Q J R Meteorol Soc 131:2961-3012

Wegener K (1916) Die Technik der Drachen- und Ballonaufstiege im Winter 1912/13 zu Ebeltofthafen (Spitzbergen). Veröffentlichungen des Deutschen Observatoriums Ebeltofthafen-Spitzbergen, Heft 2, 3-9.

Wegener K, Robitzsch M (1916a) Ergebnisse der Pilot-Visierungen während der Überwinterung 1912/13. Veröffentlichungen des Deutschen Observatoriums Ebeltofthafen-Spitzbergen 3, 18 pp

Wegener K, Robitzsch M (1916b) Ergebnisse der Fessel-Aufstiege während der Überwinterung 1912/13. Veröffentlichungen des Deutschen Observatoriums Ebeltofthafen-Spitzbergen 4, 21 pp

Wang M, Overland J, Kattsov V, Walsh J, Zhang X, Pavlova T (2007) Intrinsic versus forced variation in coupled climate model simulations over the Arctic during the twentieth century. J Clim 20:1093–1107

Wood KR, Overland JE (2010) Early 20th century Arctic warming in retrospect. Int J Climatol 30:1269-1279, doi: 10.1002/joc.1973

Wood KR, Overland JE, Jónsson T, Smoliak BV (2010) Air temperature variations on the Atlantic‐Arctic boundary since 1802. Geophys Res Lett 37:L17708, doi:10.1029/2010GL044176

Table 1. Upper-air data sets used in this study. Note that time period, time resolution, and spatial resolution represent the form in which the data sets were used in this study, not the original resolutions and time periods. UA = upper-air observations, SLP = sea-level pressure, SAT = surface air temperature, SST = sea-surface temperature

#	Data set	Abbr.	Period	Type	Input	Time resolution	Spatial resolution	Reference
1	Comprehensive historical upper-air network	CHUAN	1905-2006	Observations	-	State	>3000 stations globally	Stickler et al. 2010 Grant et al. 2009 Brönnimann 2003
2	Reconstructions	REC1	1880-1957	Statistical reconstructions	UA, SLP, SAT*	monthly mean	2.5°	Griesser et al. 2010
3	Reconstructions	REC2	1918-2001	Statistical reconstructions	UA, SLP, SAT*	monthly mean	2.5°	Brönnimann et al. 2010
4	Twentieth century reanalysis, vers. 2	20CR	1871-2008	Data assimilation (Ensemble Kalman Filter, NCEP/GFS model)	SLP, monthly SST	6-hourly	2°	Compo et al. 2011
5	NCEP/NCAR reanalysis	NNR	1948-2009	Data assimilation (Statistical Interpolation, NCEP/MRF model)	All	Daily	2.5°	Kistler et al. 2001
6	European reanalysis	ERA-40	1957-2002	Data assimilation (3D Var, IFS model)	All	Monthly	2.5°	Uppala et al. 2005
7	Japanese reanalysis	JRA-25	1978-2008	Data assimilation (3D-Var, JMA model)	All	Monthly	2.5°	Onogi et al. 2007
8	European reanalysis	ERA-Interim	1989-2009	Data assimilation (4D Var, IFS model)	All	Monthly	1.5°	-
9	Chemistry-climate model	SOCOL	1900-1999	"all forcings" simulations	SST, sea ice, other forcings	Monthly	3.75°	Fischer et al. 2008

* ERA-40 was used for calibration

Table 2. Coefficients of correlation (r) between anomalies of observed temperature (individual ascents minus a daily NNR climatology, 1968-1996, linearly interpolated to the observations) and reanalysed temperature from 20CR (closest standard time, linearly interpolated to the observations) as a function of altitude for Spitsbergen, 1911-1913 (n gives the number of observation pairs).

altitude (m asl)	200	500	1000	1500	2000	2500	3000	3500
n	78	165	125	80	39	19	11	7
r	0.405	0.397	0.514	0.600	0.585	0.686	0.846	0.925

Table 3. Coefficients of correlation (r) between anomalies of observed temperature and GPH (individual ascents minus an NNR climatology for the closest standard time, linearly interpolated to the observations) and reanalysed temperature and GPH from 20CR (closest standard time, linearly interpolated to the observations) as a function of pressure for Spitsbergen, 1944-1945 (n gives the number of observation pairs). Anomalies are constructed as in Table 2.

	pressure level	1000 hPa	850 hPa	700 hPa	600 hPa	500 hPa	400 hPa	300 hPa	200 hPa
n	GPH	130	132	132	132	132	131	120	105
	temp	95	132	132	132	132	131	120	106
r	GPH	0.601	0.942	0.943	0.940	0.937	0.931	0.925	0.658
	temp	0.729	0.813	0.875	0.877	0.863	0.800	0.515	0.442

Table 4: Correlations (r) between seasonal mean temperatures of 20CR and CHUAN for the European Arctic for different levels (note that SAT from CRUTem3v is used instead of CHUAN 1000 hPa temperature). “n” gives the number of seasonal means used for the analysis.

		1000 hPa	850 hPa	700 hPa	500 hPa	400 hPa	300 hPa	200 hPa
n	DJF	97	54	66	62	51	59	59
	MAM	97	54	64	61	50	61	61
	JJA	97	53	64	61	51	60	60
	SON	96	54	65	62	51	59	59
r	DJF	0.852	0.885	0.828	0.754	0.714	0.472	0.472
	MAM	0.873	0.870	0.757	0.706	0.557	0.074	0.074
	JJA	0.907	0.805	0.657	0.368	0.518	0.271	0.271
	SON	0.931	0.918	0.818	0.883	0.900	0.292	0.292

Table 5: Correlations between Dec.-Feb. mean temperatures for the European Arctic for different levels in 20CR, CHUAN, REC1 and REC2 for the period 1930-1957.

Comparison	Period	850 hPa	700 hPa	500 hPa	300 hPa	200 hPa
CHUAN-20CR	1930-1957	0.981	0.795	0.700	0.124	0.174
CHUAN-REC2	1930-1957	0.912	0.714	0.702	0.742	0.806
CHUAN-REC1	1930-1957	0.714	0.619	0.543	0.729	0.650
REC1-REC2	1930-1957	0.762	0.772	0.721	0.619	0.675
REC1-20CR	1930-1957	0.825	0.808	0.752	0.553	0.013
20CR-REC2	1930-1957	0.905	0.915	0.845	0.368	0.070

Table 6: Correlations of monthly temperature anomalies (with respect to the period 1961-1990 in each data set; ERA-40 was used for REC1 and REC2) for the European Arctic between different gridded data sets.

Comparison	Period	850 hPa	700 hPa	500 hPa	300 hPa	200 hPa
20CR-REC1	1918-1957	0.784	0.781	0.744	0.368	0.227
20CR-REC2	1918-1957	0.873	0.873	0.836	0.421	0.246
REC1-REC2	1918-1957	0.850	0.854	0.833	0.641	0.586
20CR-NNR	1948-1957	0.934	0.939	0.928	0.686	0.464
REC1-NNR	1948-1957	0.834	0.848	0.828	0.721	0.720
REC2-NNR	1948-1957	0.939	0.951	0.951	0.897	0.791
20CR-NNR	1958-2001	0.941	0.961	0.947	0.637	0.369
20CR-ERA40	1958-2001	0.947	0.959	0.932	0.658	0.369
ERA40-NNR	1958-2001	0.986	0.987	0.978	0.904	0.985

Figures

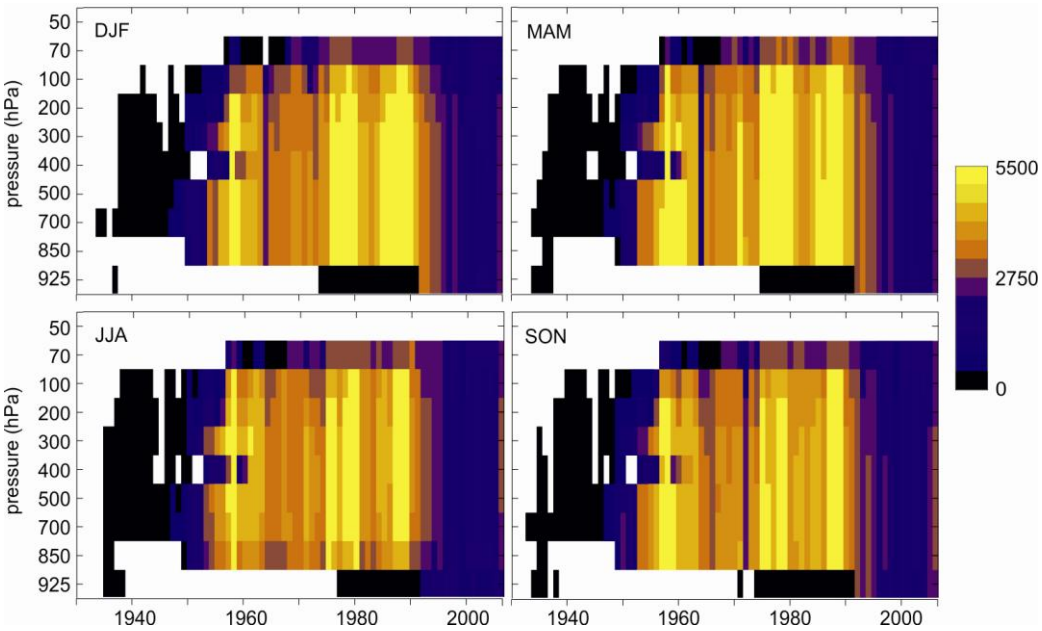

Fig. 1: Number of upper-air observations available for the sector [10°W-100°E, 60°N-90°N] for different seasons.

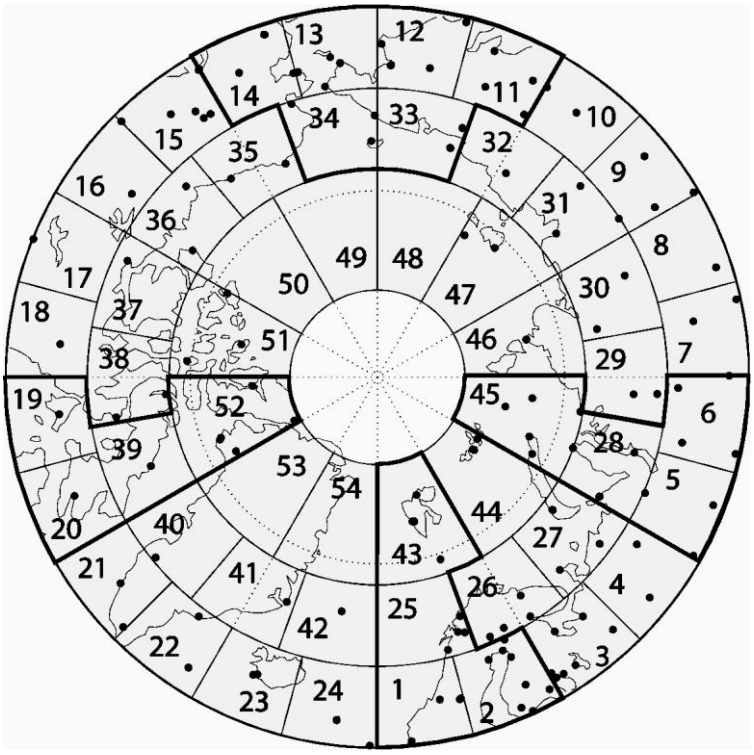

Fig. 2: Map showing the radiosonde stations in the Arctic used in this study along with the equal area grid cells used for regional averaging and the four regions for which analyses are presented, namely the European Arctic (cells 1, 2, 25, and 43), Siberia (5, 6, 28, 45), the Bering Strait region (11, 12, 13, 14, 33, 34) and the Baffin region (19, 20, 39, 52).

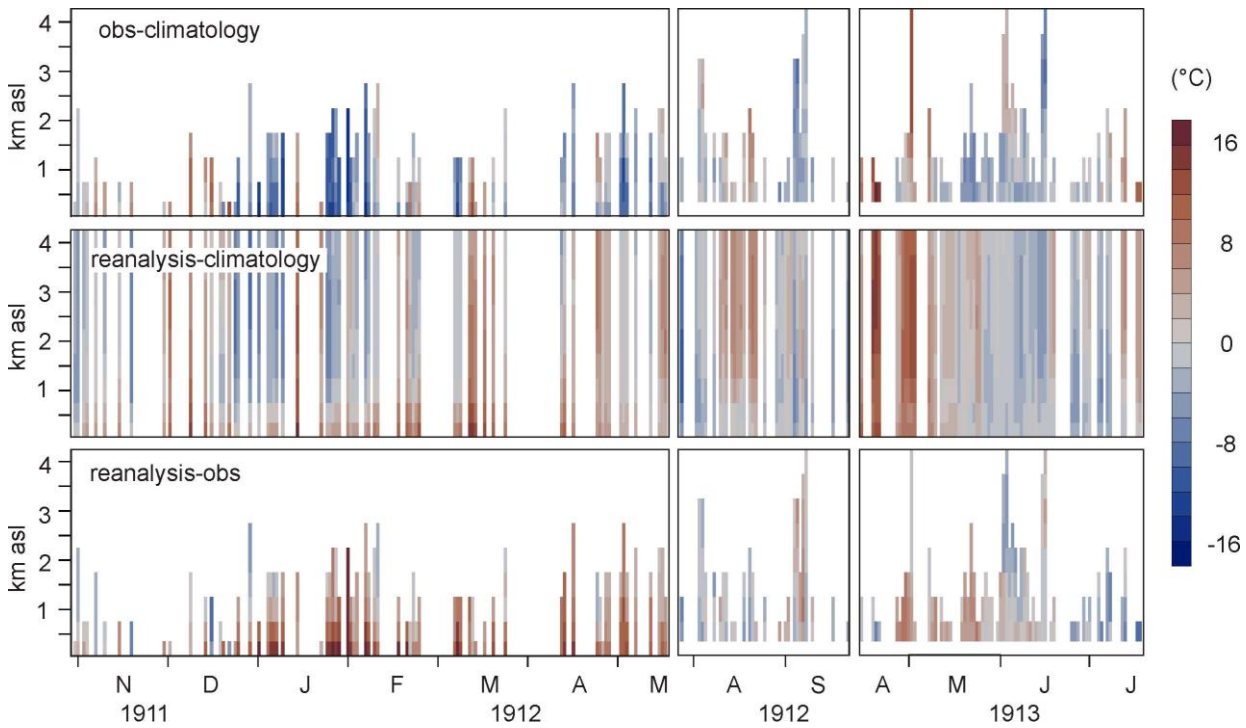

Fig. 3: Anomalies of daily temperature profiles (as a function of altitude above msl) from Spitsbergen, Nov. 1911-May 1912, Jul. 1912-Sep. 1912, Apr. 1913-Jul. 1913, from observations (top), 20CR (middle, both with respect to a 1968-1996 climatology from NNR), and their difference (bottom).

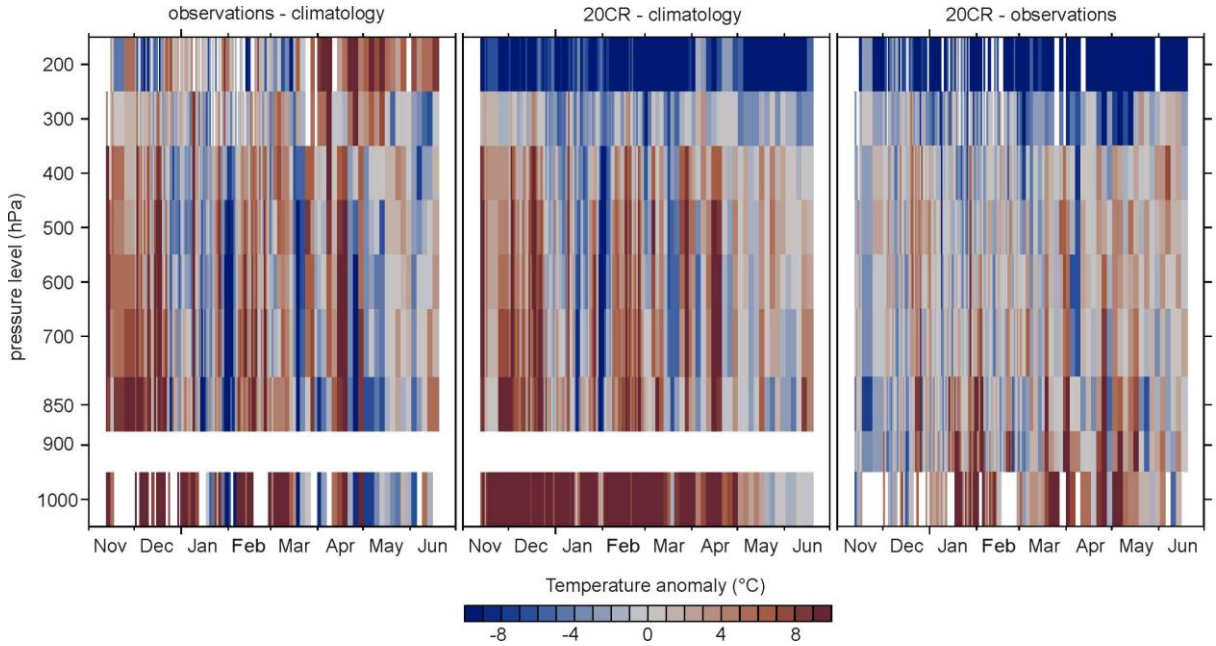

Fig. 4: Anomalies of daily temperature profiles (as a function of pressure) from Spitsbergen, 1944-1945, from observations (left), 20CR (middle), and their difference (right). Anomalies are constructed as in Fig. 4. Because of differences in reporting (925 hPa in NNR, 900 hPa in observations and 20CR), no climatology and hence no anomalies are available for 925 hPa.

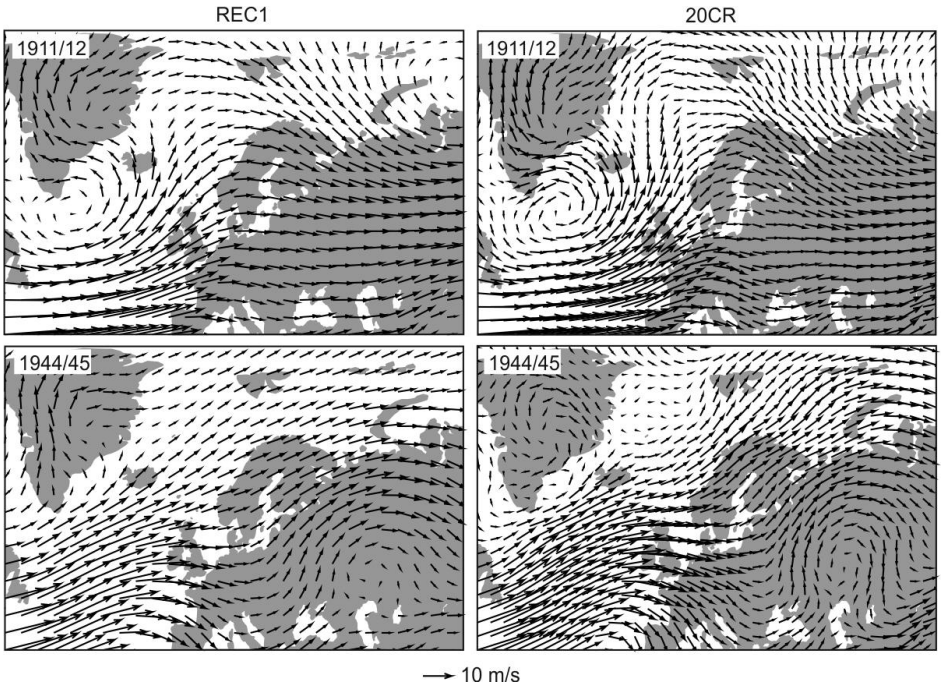

Fig. 5: Averaged wind field in Dec.-Feb. 1911-1912 (top) and Dec.-Feb. 1944-1945 (bottom) in in REC1 at 3 km asl (left) and in 20CR at 700 hPa (right).

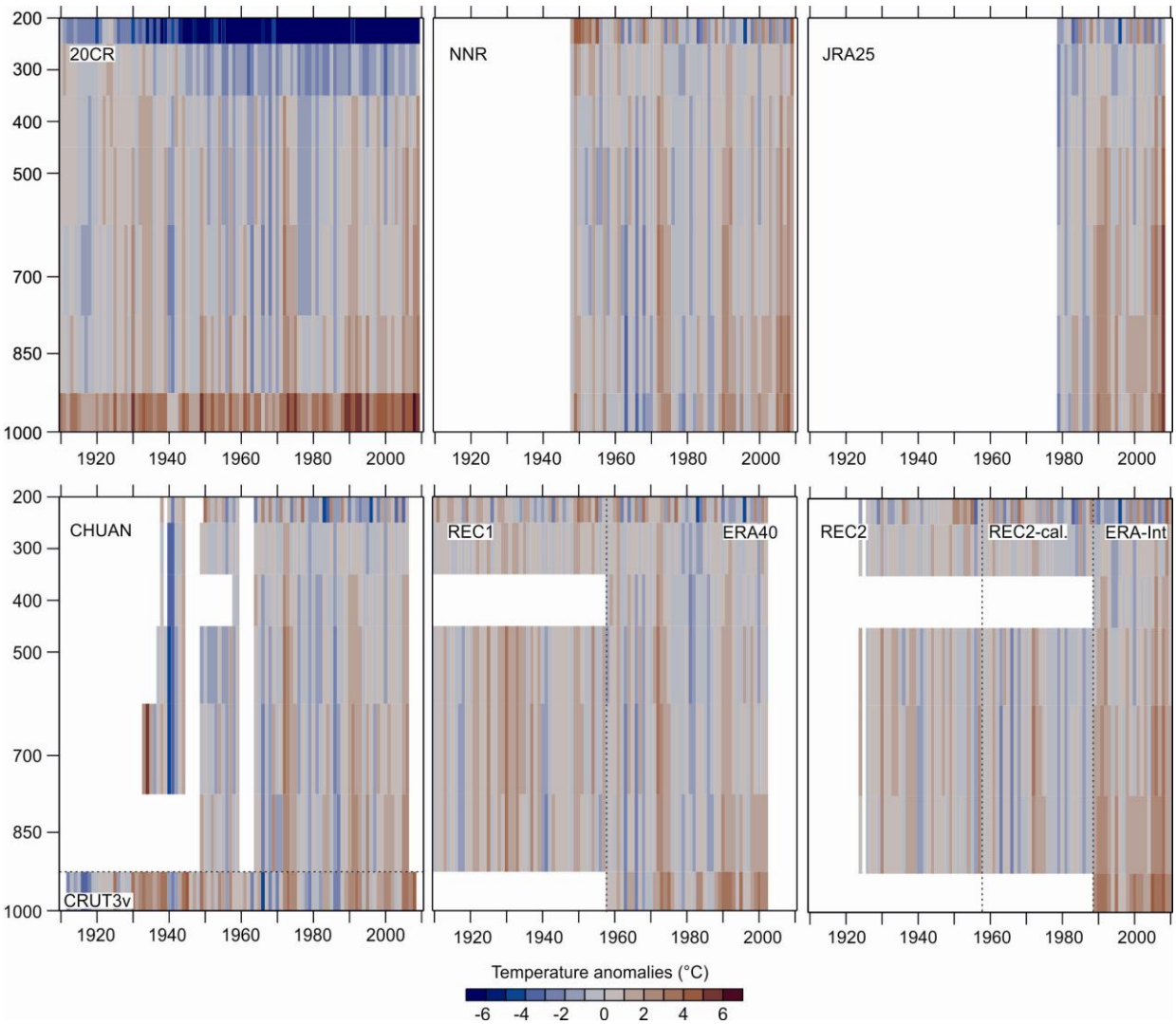

Fig. 6: Time-height cross-section of seasonal mean temperature anomalies as a function of pressure and time for different data sets for the European Arctic (see Fig. 2) in winter. All anomalies are with respect to NNR (1961-1990) except CRUTEM3v (self-climatology, see Brohan et al. 2006). Note that for visualisation purposes, non-overlapping data sets have been combined in some cases, indicated by dashed lines). Between the end of the reconstruction period of REC2 (1957) and the start of ERA-Interim (1989) we show the calibration period of REC2.

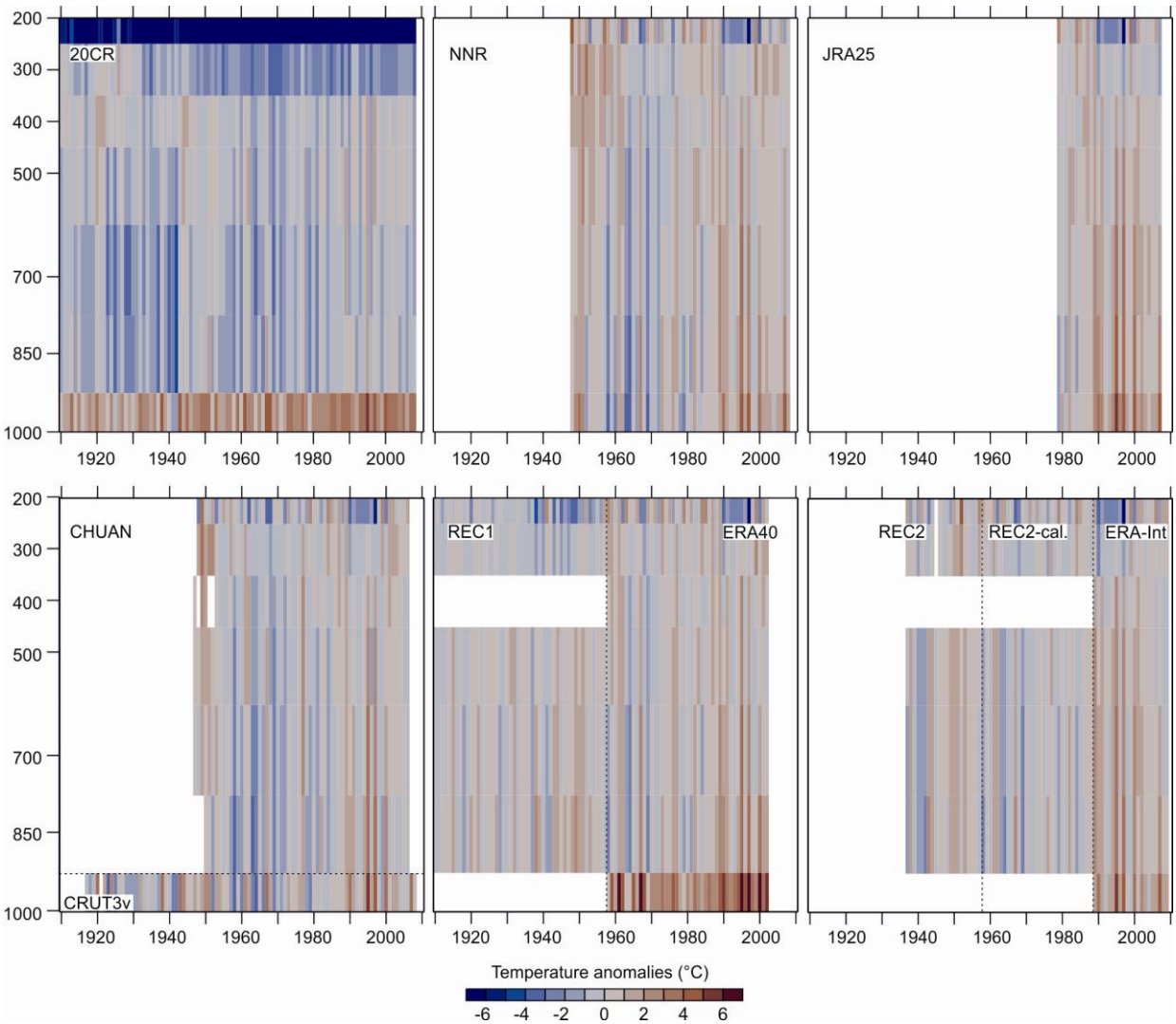

Fig. 7: Same as Fig. 6 for Western Siberia (see Fig. 2) in spring.

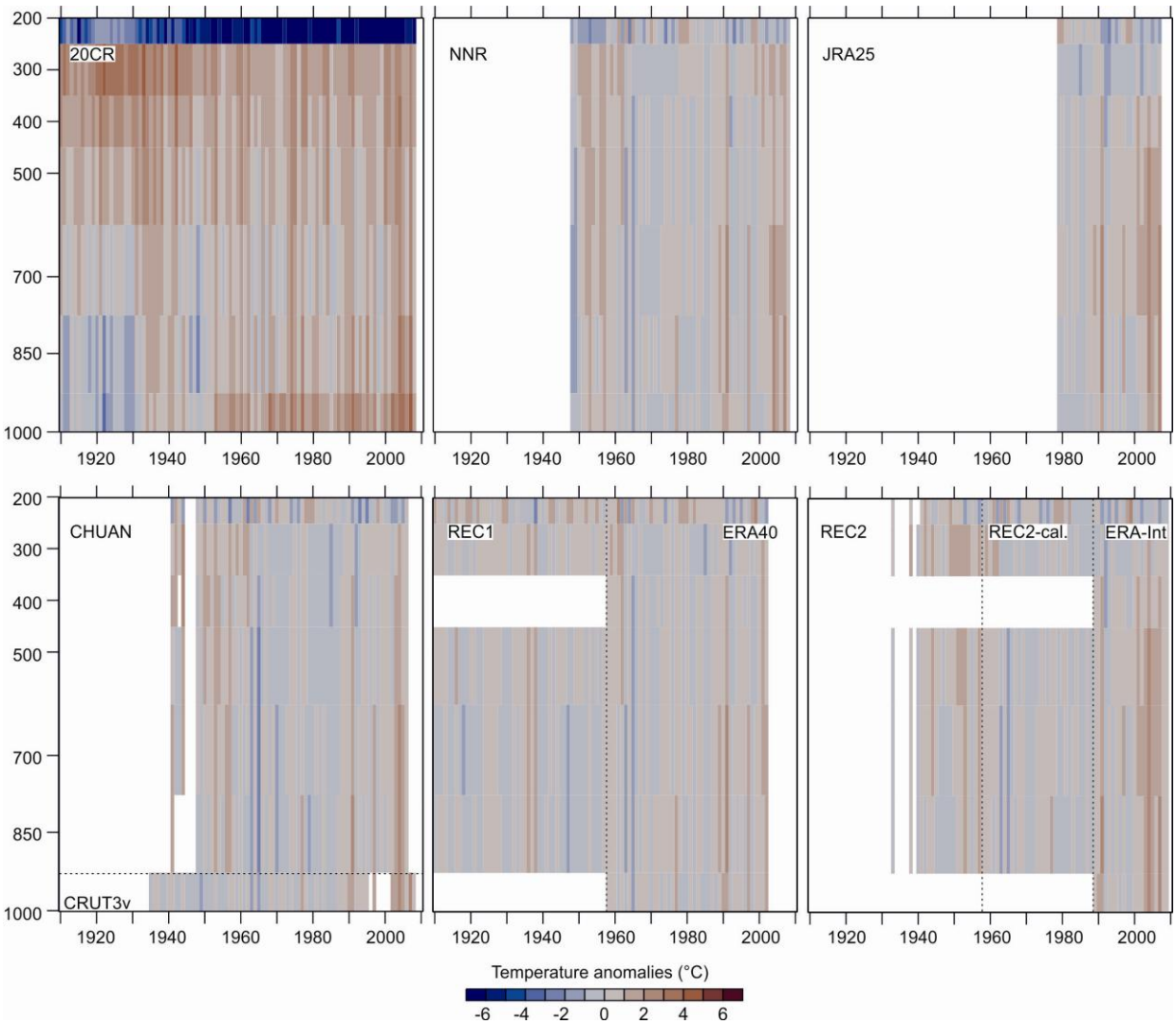

Fig. 8: Same as Fig. 6 for the Bering Strait region (see Fig. 2) in summer.

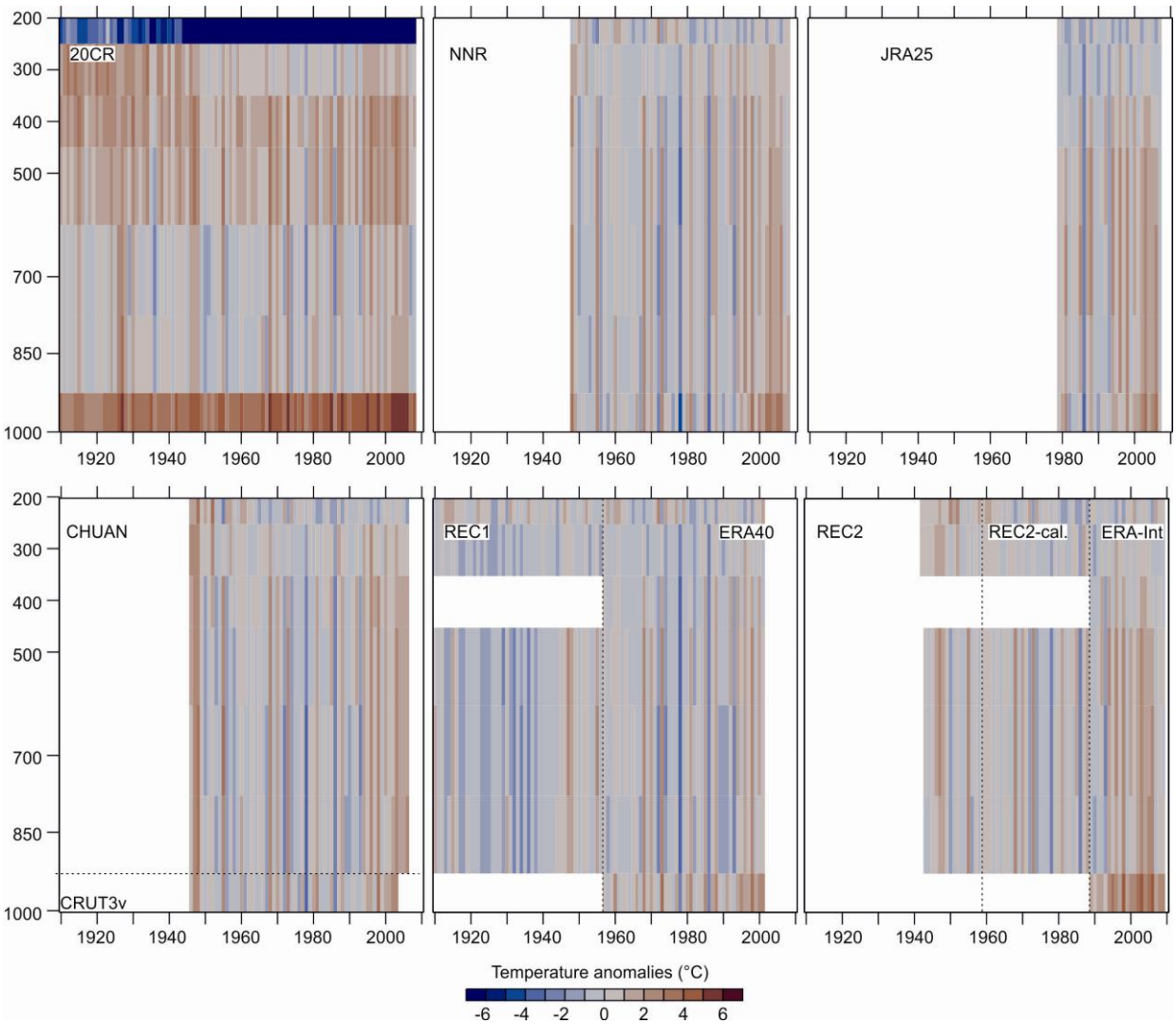

Fig. 9: Same as Fig. 6 for the Baffin region (see Fig. 2) in autumn.

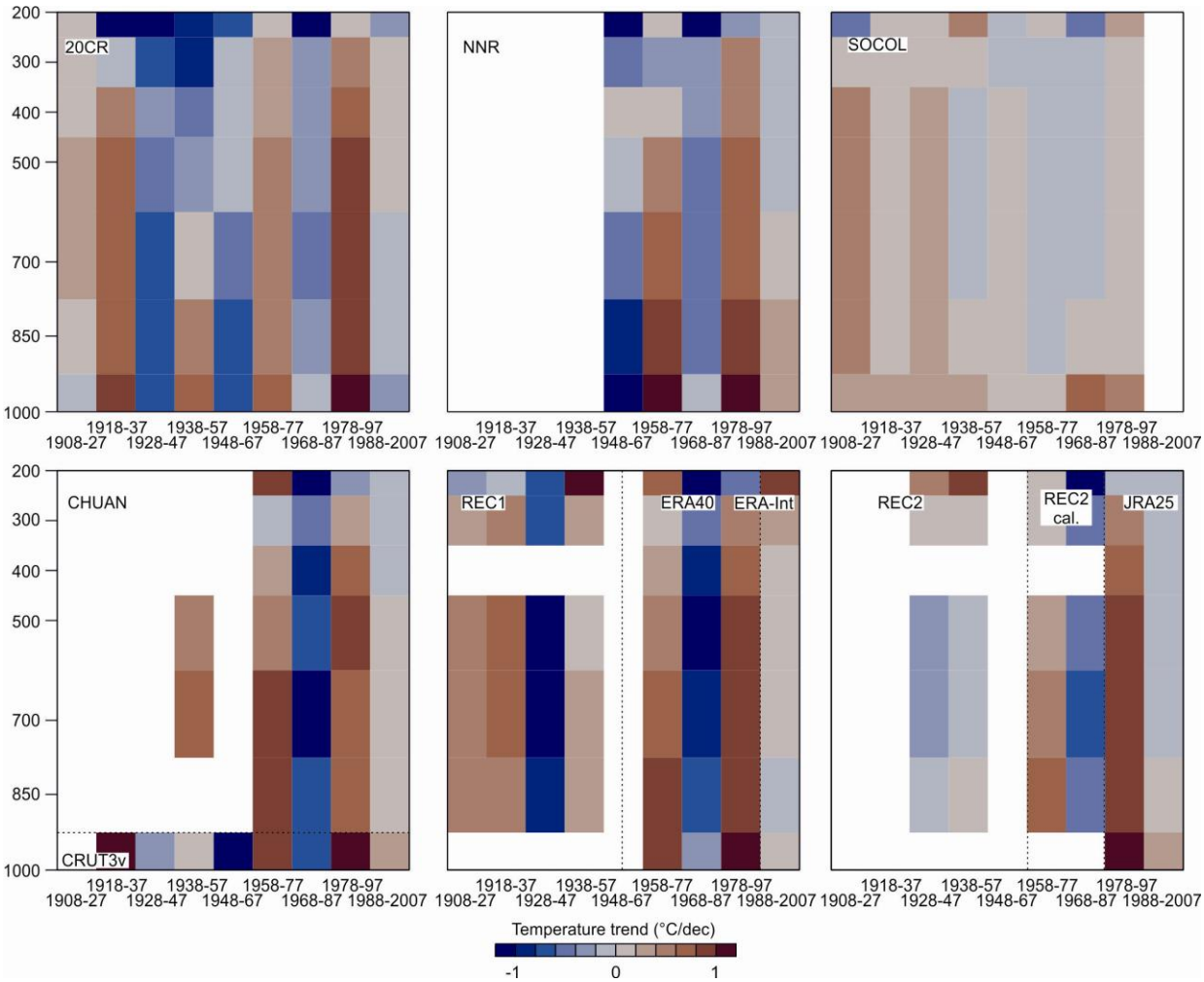

Fig. 10: Trend in seasonally-averaged temperature profiles over 20-yr periods as a function of pressure and time period for different data sets for the European Arctic (see Fig. 2) in winter. Note that for visualisation purposes, non-overlapping data sets have been combined in some cases, indicated by dashed lines). Between the end of the reconstruction period of REC2 (1957) and the start of ERA-Interim (1989) we show the calibration period of REC2.

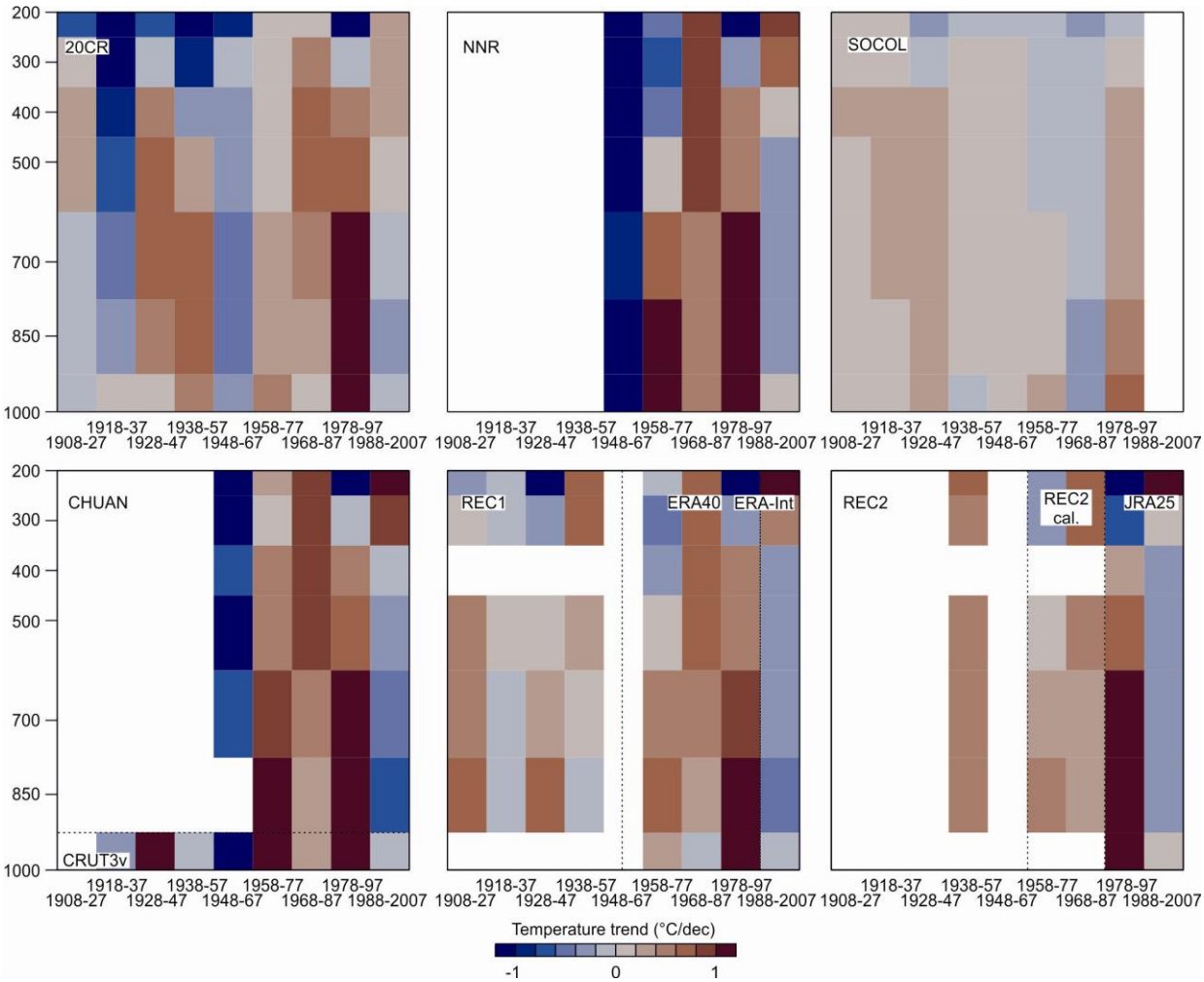

Fig. 11: Same as Fig. 10 for Western Siberia (see Fig. 2) in spring.

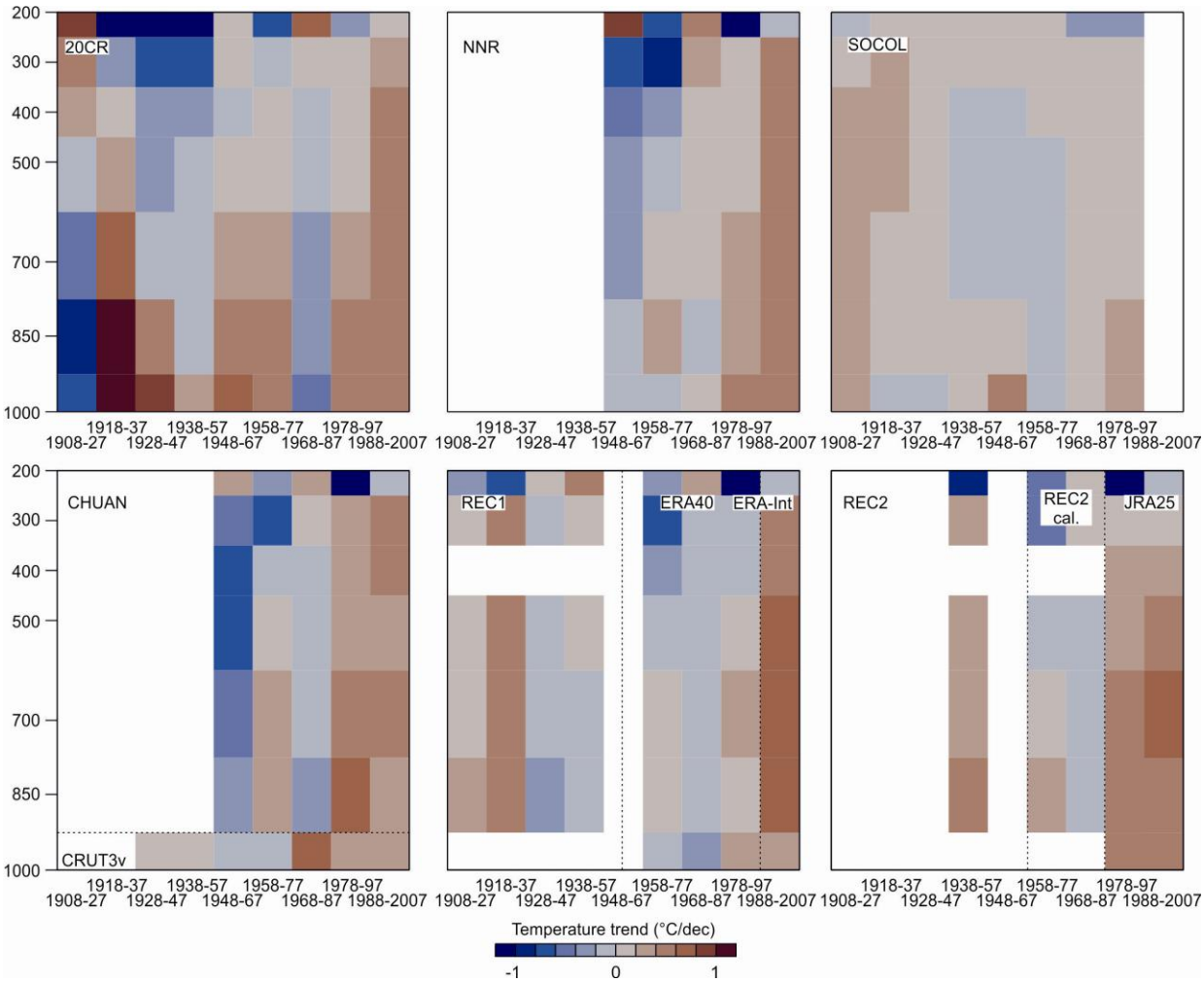

Fig. 12: Same as Fig. 10 for the Bering Strait region (see Fig. 2) in summer.

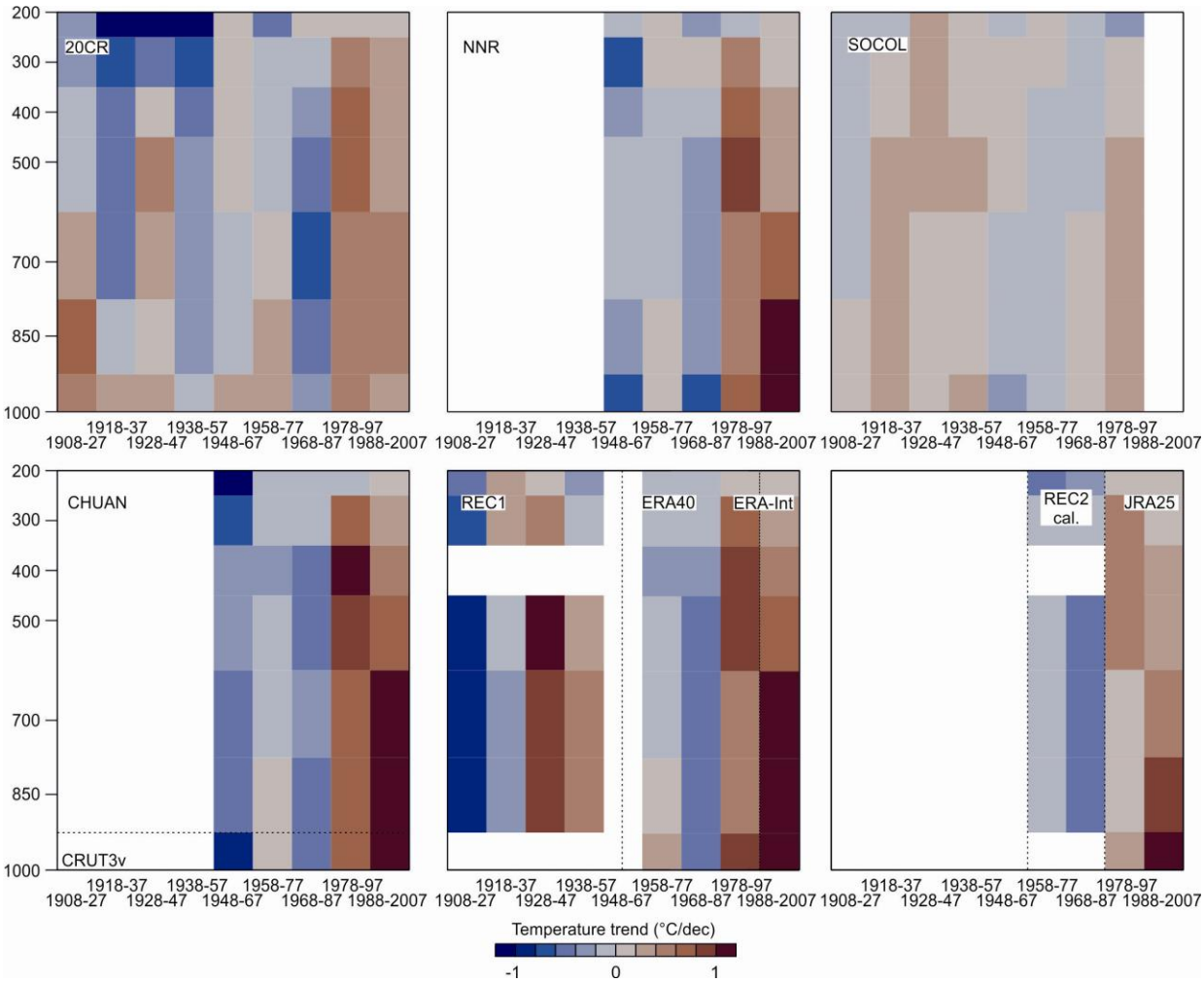

Fig. 13: Same as Fig. 10 for the Baffin region (see Fig. 2) in autumn.

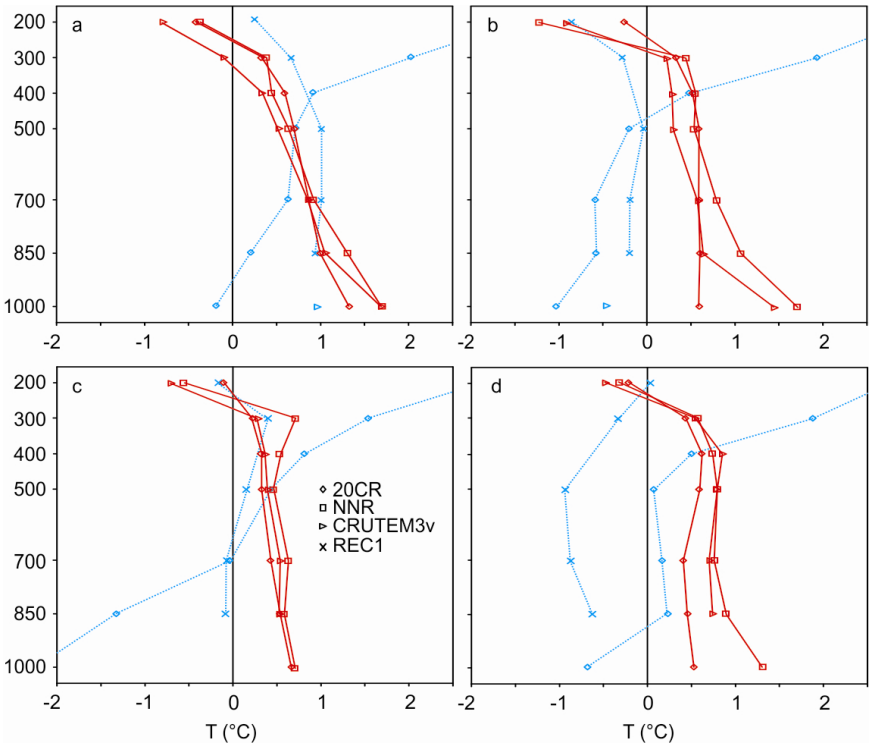

Fig. 14: Temperature anomaly averages (relative to self-climatologies 1961-90) in two 20-yr windows for different data sets for different seasonal-regional averages (a = European Arctic in winter, b = Western Siberia in spring, c = Bering Strait in summer, d = Baffin region in autumn). Blue symbols and dashed lines denot 1918-1937, red lines and symbols denote 1988-2007.

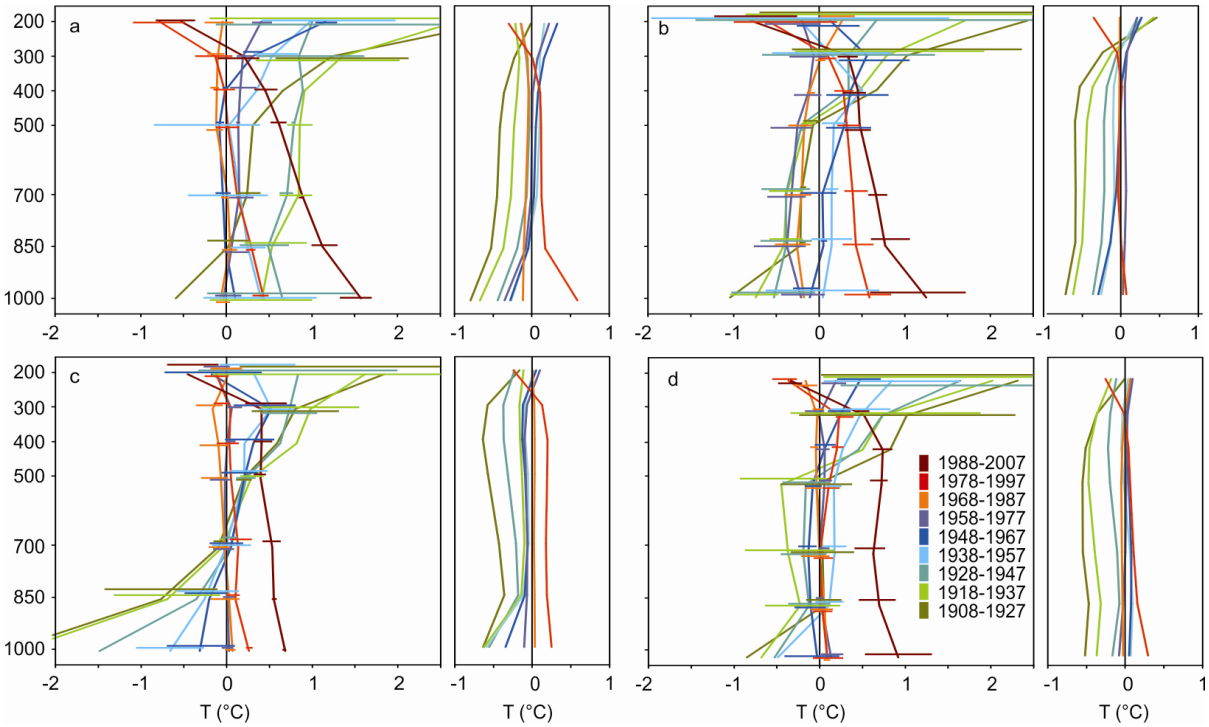

Fig. 15: Temperature anomaly averages (relative to self-climatologies 1961-90) in 20-yr windows for different data sets (left) and model simulations (right) for different seasonal-regional averages (a = European Arctic in winter, b = Western Siberia in spring, c = Bering Strait in summer, d = Baffin region in autumn). The solid line gives the mean value of all observation based data sets, the horizontal bars (slightly displaced in the vertical for better visualization) indicate the spread. A full version of this figure (including symbols for each data set) is given in the electronic supplement).

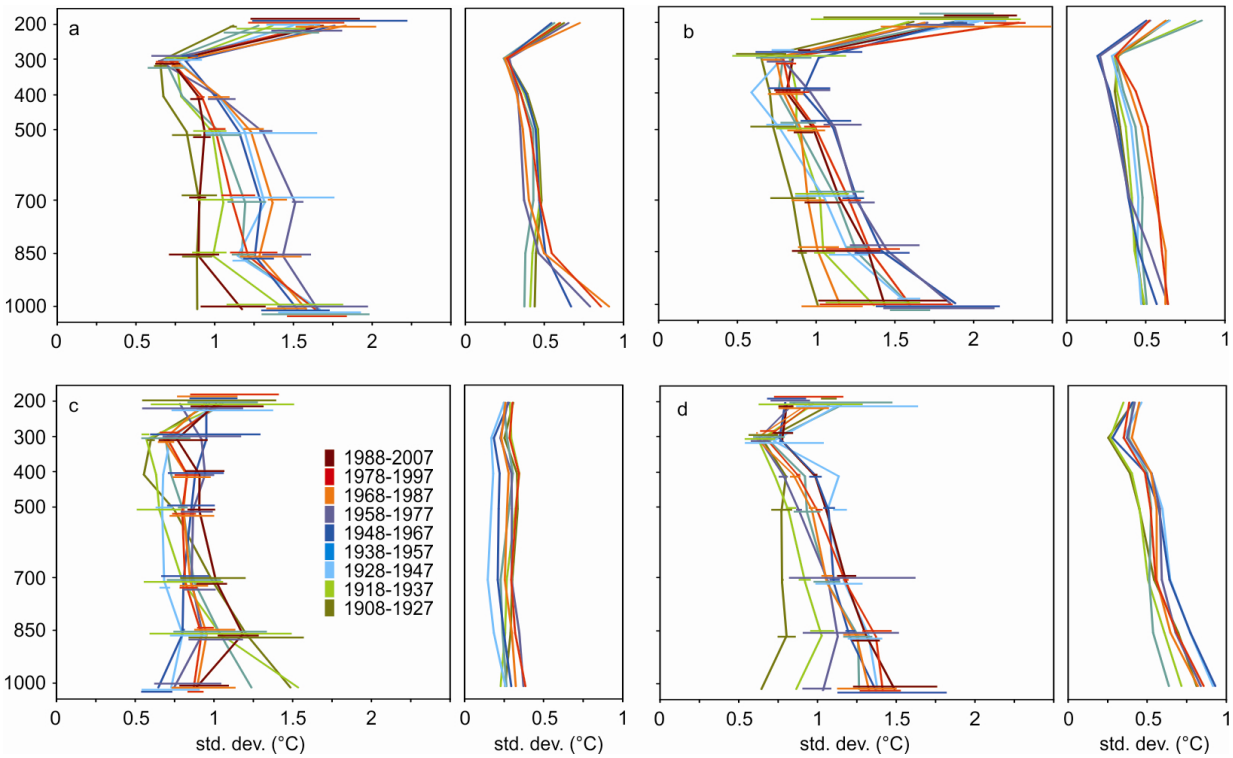

Fig. 16: Standard deviation (right) of seasonal temperature averages in 20-yr windows for different data sets (left) and model simulations (right) for different seasonal-regional averages (a = European Arctic in winter, b = Western Siberia in spring, c = Bering Strait in summer, d = Baffin region in autumn). The solid line gives the mean value of all observation based data sets, the horizontal bars (slightly displaced in the vertical for better visualization) indicate the spread. A full version of this figure (including symbols for each data set) is given in the electronic supplement).

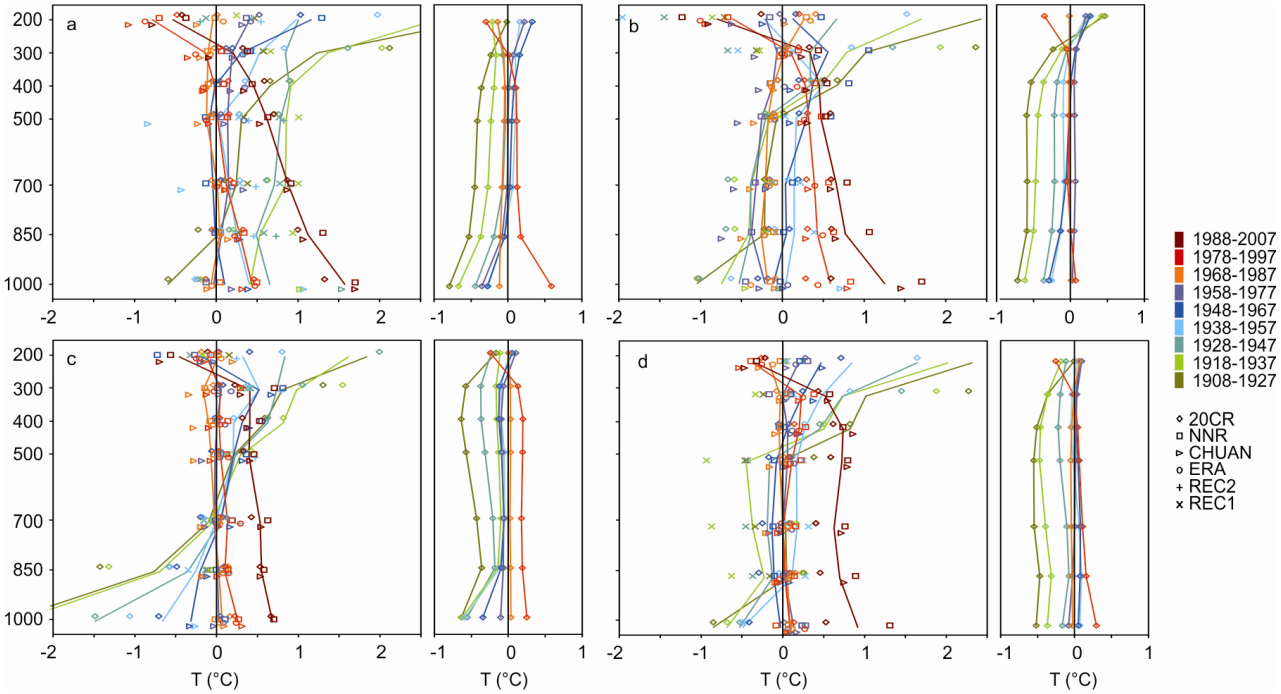

Fig. S1: Temperature anomaly averages (relative to self-climatologies 1961-90) in 20-yr windows for different data sets (left) and model simulations (right) for different seasonal-regional averages (a = European Arctic in winter, b = Western Siberia in spring, c = Bering Strait in summer, d = Baffin region in autumn). Note that symbols are slightly displaced in the vertical for better visualization. The solid line gives the mean value of all observation based data sets (see also Fig. 15)

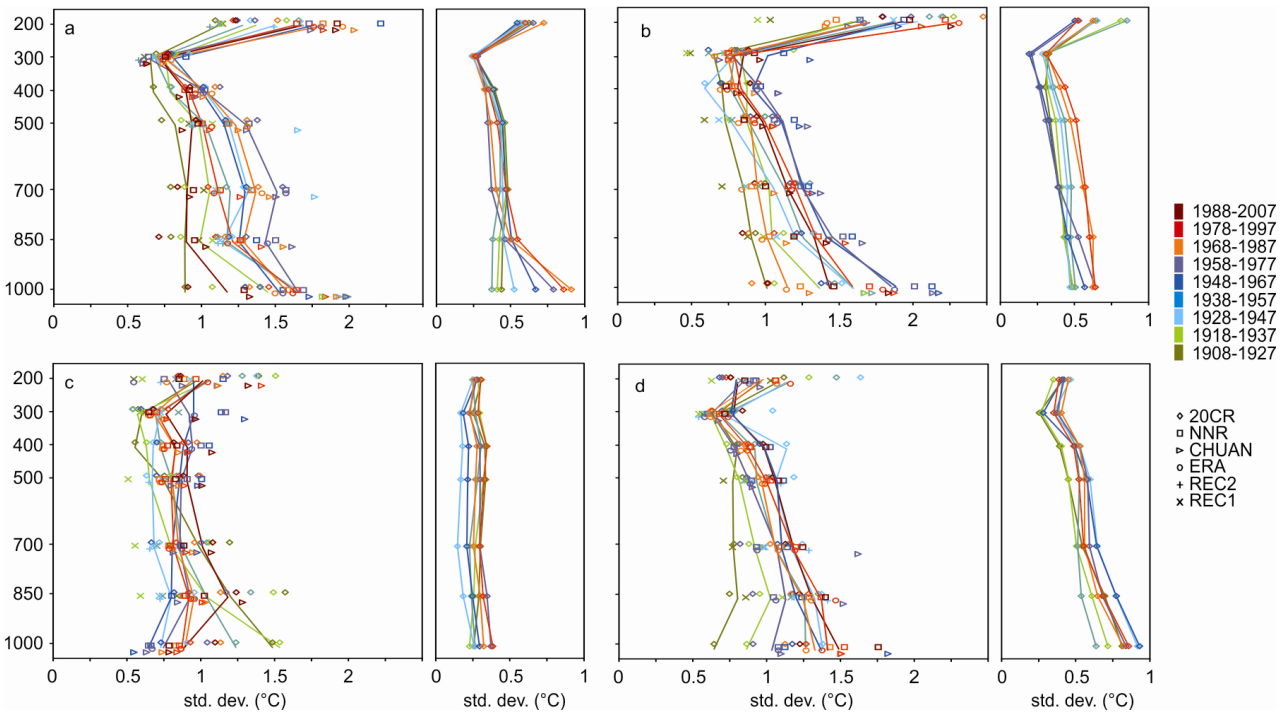

Fig. S2: Standard deviation (right) of seasonal temperature averages in 20-yr windows for different data sets (left) and model simulations (right) for different seasonal-regional averages (a = European Arctic in winter, b = Western Siberia in spring, c = Bering Strait in summer, d = Baffin region in autumn). Note that symbols are slightly displaced in the vertical for better visualization. The solid line gives the mean value of all observation based data sets (see also Fig. 16).