Modeling the 20th century Arctic Ocean/Sea ice system: Reconstruction of surface forcing

Frank Kauker,[1,2] Cornelia Köberle,[1] Rüdiger Gerdes,[1] and Michael Karcher[1,2]

[1] The ability to simulate the past variability of the sea ice-ocean system is of fundamental interest for the identification of key processes and the evaluation of scenarios of future developments. To achieve this goal atmospheric surface fields are reconstructed by statistical means for the period 1900 to 1997 and applied to a coupled sea ice-ocean model of the North Atlantic/Arctic Ocean. We devised a statistical model using a redundancy analysis to reconstruct the atmospheric fields. Several sets of predictor and predictand fields are used for reconstructions on different time scales. The predictor fields are instrumental records available as gridded or station data sets of sea level pressure and surface air temperature. The predictands are surface fields from the NCAR/NCEP reanalysis. Spatial patterns are selected by maximizing predictand variance during a "learning" period. The reliability of these patterns is tested in a validation period. The ensemble of reconstructions is checked for robustness by mutual comparison and an "optimal" reconstruction is selected. Results of the simulations with the sea ice-ocean model are compared with historical sea ice extent observations for the Arctic and Nordic Seas. The results obtained with the "optimal" reconstruction are shown to be highly consistent with these historical data. An analysis of simulated trends of the "early 20th century warming" and the recent warming in the Arctic complete the manuscript.

Citation: Kauker, F., C. Köberle, R. Gerdes, and M. Karcher (2008), Modeling the 20th century Arctic Ocean/Sea ice system: Reconstruction of surface forcing, *J. Geophys. Res.*, *113*, C09027, doi:10.1029/2006JC004023.

1. Introduction

[2] The distinct possibility of anthropogenic climate change makes it desirable to identify and attribute past long-term trends in the climate system. This approach complements efforts to estimate the future development of the climate system by running coupled climate models for future scenarios of increasing greenhouse gases and other anthropogenic radiatively active substances. The identification of long-term trends in the climate system is made difficult by the presence of energetic low-frequency natural variability [*von Storch et al.*, 2004]. Long simulations, covering thousand years or longer, with coupled climate models under preindustrial conditions are used to assess the natural variability [*Zorita et al.*, 2003]. The results reflect variability due to internal oscillations and due to prescribed forcing (e.g., fluctuations of solar radiation and volcanic aerosols). Trends can be compared statistically with those observed and, as far as model skill permits, anomalies of recent observed trends can be identified. This method has the disadvantage that because of the random nature of the phase in natural cycles in models and nature no direct comparison with observations is possible. Direct compari-

son with observations can be possible in strongly forced subsystems that do not show as large an internal variability as the coupled climate system. This is especially true for sea ice that adjusts on a time scale of a few years to any prescribed forcing [e.g., *Köberle and Gerdes*, 2003]. Thus for the ocean-sea ice subsystem, a different modeling approach has been pursued. Hindcast simulations forced with prescribed atmospheric surface fields yield model time series that can be directly compared with observations [e.g., *Hatun et al.*, 2005; *Kauker et al.*, 2003; *Polyakov et al.*, 2005]. The disadvantage of this approach is the shortness of available forcing data time series. Currently, data sets suitable for forcing ocean-sea ice hindcast, the NCEP and ECMWF reanalysis data, cover only 40 to 60 years, too short to resolve multi-decadal variability that is present in many quantities. A prominent example for such a multi-decadal signal in the coupled system of the high latitudes is the early 20th century warming [*Bengtsson et al.*, 2004].

[3] Here, we present a method to generate suitable forcing data for an ocean-sea ice model of the Atlantic and Arctic oceans. We combine long station data time series and existing two-dimensional reconstructions of SLP and surface air temperature with the NCEP reanalysis data in a statistical model. The basic idea is to up-scale local variability of station data and to down-scale large-scale variability of the gridded data sets to the regional-scale of the Arctic and North Atlantic area. Statistically up- and down-scaling techniques are frequently used in climate research [see, e.g., *Hanssen-Bauer et al.*, 2005]. Often a *Canonical*

---

[1] Climate System, Alfred Wegener Institute for Polar and Marine Research, Bremerhaven, Germany.
[2] O.A.Sys, Ocean Atmosphere System, Hamburg, Germany.

Correlation Analysis (CCA) or multi-variate regression is used to build the statistical model. The *Redundancy Analysis* employed here is less frequently used although the method is (at least theoretically) superior. The technique was previously applied for the reconstruction of atmospheric surface forcing fields for a coupled sea ice-ocean model of the Baltic Sea [*Kauker and Meier*, 2003; *Meier and Kauker*, 2003].

[4] Naturally, any such reconstruction has uncertainties and it is necessary to validate the reconstruction with independent data. Here, we employ sea ice extent observations that are compared with our model results. Sea ice extent has been reliably observed, for commercial purposes, for a long time in many regions of the Nordic Seas and the Arctic Ocean. With the exception of the reanalysis data, such sea ice extent data neither entered the reconstruction nor the data that we use to generate the reconstruction of atmospheric forcing fields. It can thus be regarded as independent data. Since sea ice extent is strongly forced by the atmosphere, there is a close relationship, although nonlinear, between the atmospheric fields and sea ice extent. Together, it represents a good measure for the success of the reconstruction.

[5] The paper focuses on the northern high latitudes where anticipated global change signals are large. The following section describes the available long-term atmospheric data sets that we considered for the reconstruction. The statistical method is described in the third section that also gives a critical assessment of the available long time series and the reconstructions based on these different data. The reconstructed forcing data are applied in section four where we attempt a validation of different data sets using historic sea ice extent data. We arrive at one reconstruction that has largest skill. The results are summarized in the last section.

2. Utilized Data Sets

2.1. Long-Term (Predictor) Data Sets

[6] The reconstruction relies on long time series that must extend as far back as one wishes to reconstruct the atmospheric forcing fields. There are various data sets with century long time-series and monthly temporal resolution available. We used here the gridded sea-level pressure (SLP) data set of *Trenberth and Paolino* [1980, 1981] (called MSLPG, hereafter), the sea-level pressure station data set of the International Arctic Research Center (IARC; http://www.iarc.uaf.edu courtesy I. Polyakov), the sea-level pressure station data set of the Swedish Meteorological and Hydrological Institute (SMHI) compiled by *Alexandersson et al.* [2000] and used in the Baltic Sea reconstruction [*Kauker and Meier*, 2003], the gridded 2m-meter air temperature (SAT) data set of the Arctic and Antarctic Research Institute (AARI) [*Alekseev*, 1999], the 2-meter air temperature station data set compiled by the AICSEX project http://www.nersc.no/AICSEX), and the 2-meter air temperature station data of IARC (courtesy I. Polyakov).

[7] Traditionally, these independent data are called *predictors* in the framework of redundancy analysis. Here, they are used to "predict" data on the regional Arctic scale. We used both gridded data sets and station data as predictors.

[8] The gridded data are usually constructed from station data and information from weather charts. Weather charts before the onset of numerical weather prediction are not only determined by physical reasoning but may also be
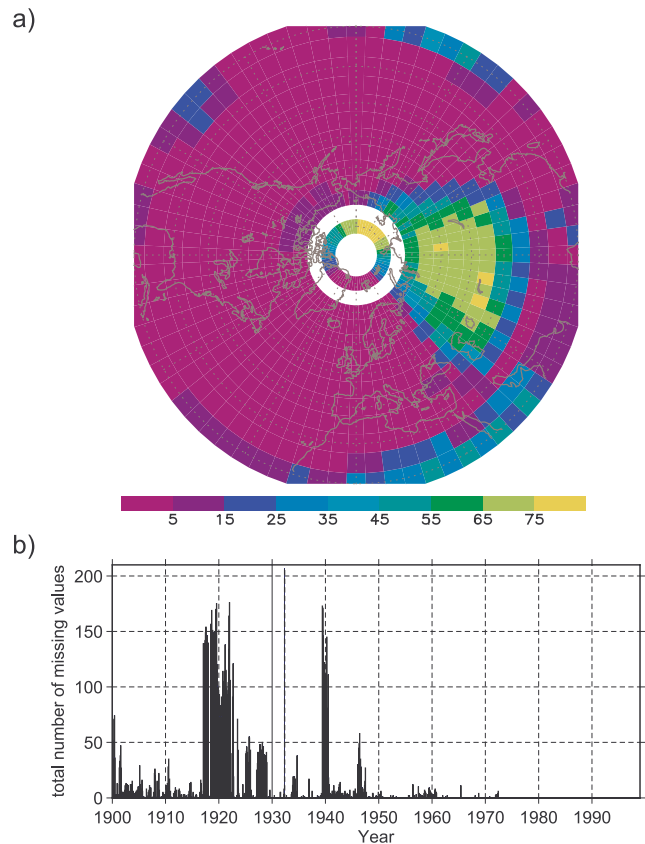

Figure 1. The number of missing values at each grid box for the period 1900 to 1997 (a) and the total number of missing monthly values as a function of time (b) for the *Trenberth and Paolino* [1980] (MSLPG) data.

influenced by prejudices. For instance, during the beginning of the last century it was common sense among meteorologists that the Arctic Ocean lies beneath a strong high-pressure system. *Jones* [1987] could show that this led to a bias of about 4 to 6 hPa in the weather charts. This is an example of inhomogeneity in gridded data sets. In general, it holds that gridded data sets are much more difficult to assess than are station data.

[9] On the other hand, station data may also be inhomogeneous (e.g., showing abrupt changes) "if the instrument (or the observer) changes, the site is moved, or recording practices are changed" [*von Storch and Zwiers*, 1998]. By cross-checking reconstructions based on gridded data sets and station data sets one may hope to be able to identify systematic errors in these data sets.

[10] All these data sets have gaps. In principle, data gaps can be handled by the statistical reconstruction method, but causes uncontrollable uncertainties in the reconstructed variables. We allowed no more than 120 missing monthly values per station or grid box in the period 1900 to 1997. Allowing no more than 120 missing data is a compromise between too few stations on the one hand (if less missing data are allowed) and too gappy data sets on the other hand (if more missing data are allowed). Stations or grid boxes with more than 120 missing values are discarded.

[11] The number of data gaps of the selected stations or grid boxes is illustrated Figure 1 that depicts the number of
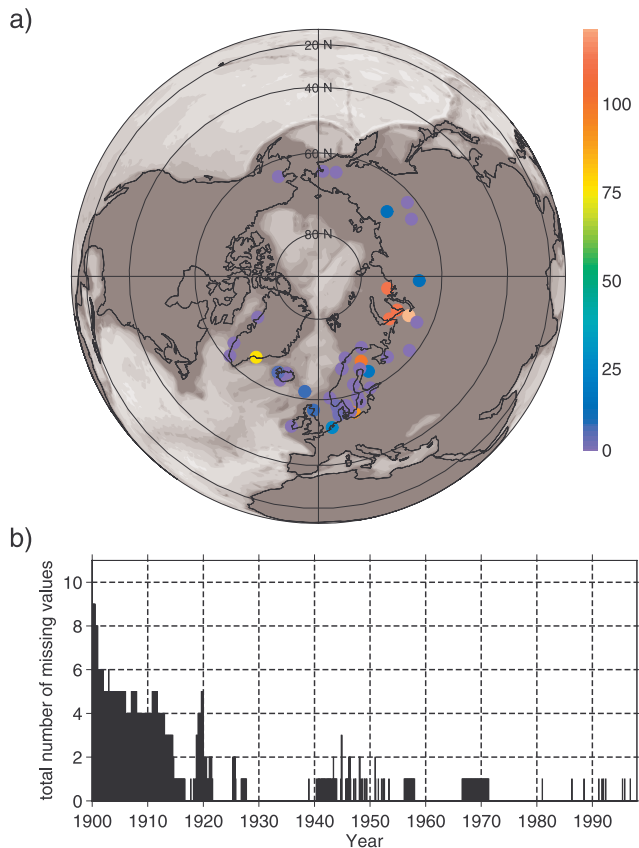
a)

b)

Figure 2. The number of missing values at each station for the period 1900 to 1997 (a) and the total number of missing monthly values as a function of time (b) of the combined data set of SLP of IARC and SMHI.

missing values at each grid box for the period 1900 to 1997 and the total number of data gaps at each point in time for the MSLPG data. The longitude-latitude grid consists of 72×15 boxes ranging from $0°E$ to $355°E$ and from $15°N$ to $85°N$. 216 out of 1084 grid boxes are discarded because of too many missing values (the complete latitude circles $15°N$, $75°N$, and $85°N$). Except $15°N$, $75°N$, and $85°N$ grid boxes with the highest number of gaps are located over Siberia and at the $80°N$ latitude (Figure 1a). Time periods with the highest number of data gaps are the periods of both World Wars and the years following World War I (Figure 1b).

[12] To get the largest benefit from the SLP station data we merged the data from IARC and SMHI. In total the IARC data set contains 133 stations and the SHMI data set 20 stations (mostly over Northern Europe). However, only 48 stations are considered for the reconstruction because of too many missing values at most of the stations. In the following we will refer to this data set as IARC/SMHI. The location and the number of data gaps can be seen in Figure 2a. Unfortunately, the final data set contains only one station over North America. Temporally, the data coverage of the selected stations is very high. It is only at the beginning of the last century that a larger number of stations (up to 9) contain considerable gaps.

[13] The SAT data set of the AARI is organized on a $10° \times 5°$ longitude-latitude grid covering the whole

Northern Hemisphere. The data set was compiled from 1486 meteorological stations in the Northern Hemisphere, including land- and drifting-stations from the Arctic. North of $20°N$, all grid boxes are taken into account (total number 580, Figure 3a). The total number of missing values at each point in time is lowest during the 1960s and 1970s and highest during the 1980s and 1990s (even higher than during the early 20th century).

[14] Additionally, we merged the SAT station data sets of the AICSEX project and from IARC. The AICSEX data set contains 1373 stations and the IARC data set 133 stations. However, only 122 stations fulfill our criteria of no more than 240 missing monthly values (Figure 4a). The number of missing values at a point in time is low (about or lower 10) during almost the whole century except for the 1990s (Figure 4b).

2.2. Spatial High-Resolution Data Set

[15] The standard forcing of the $1° \times 1°$ coupled sea ice-ocean model NAOSIM used here consists of daily NCEP/NCAR reanalysis data [Kalnay et al., 1996] from 1948 to 1997. The variables which will be used are SLP, wind stress (calculated from the 10m winds), SAT, dew-point temperature, cloudiness and scalar winds. The model domain encompasses the North Atlantic north of $20°N$, the Nordic Seas, and the Arctic Ocean. Daily NCEP/NCAR reanalysis
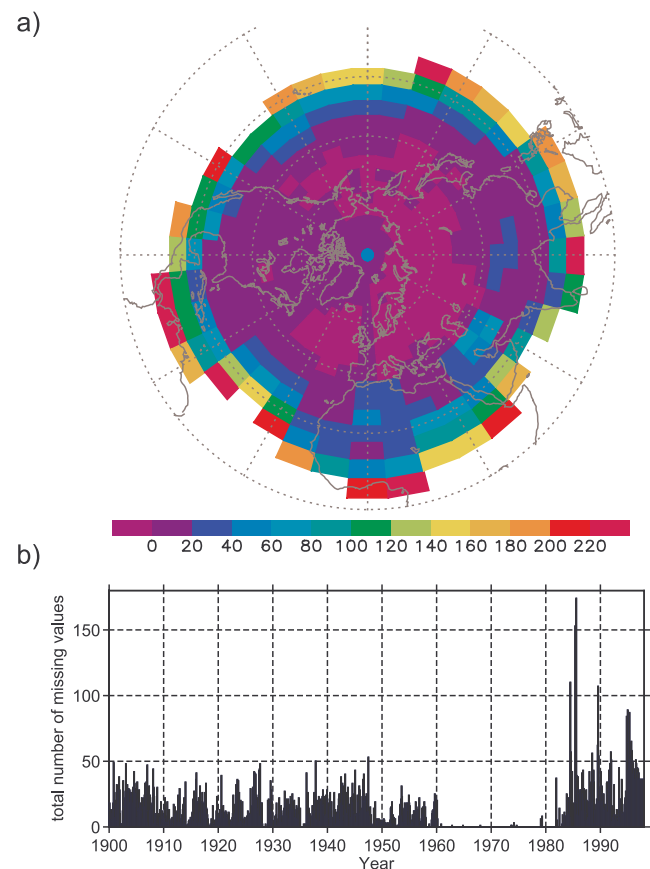
a)

b)

Figure 3. The number of missing values at each grid point for the period 1900 to 1997 (a) and the total number of missing monthly values as a function of time (b) for the AARI data set.
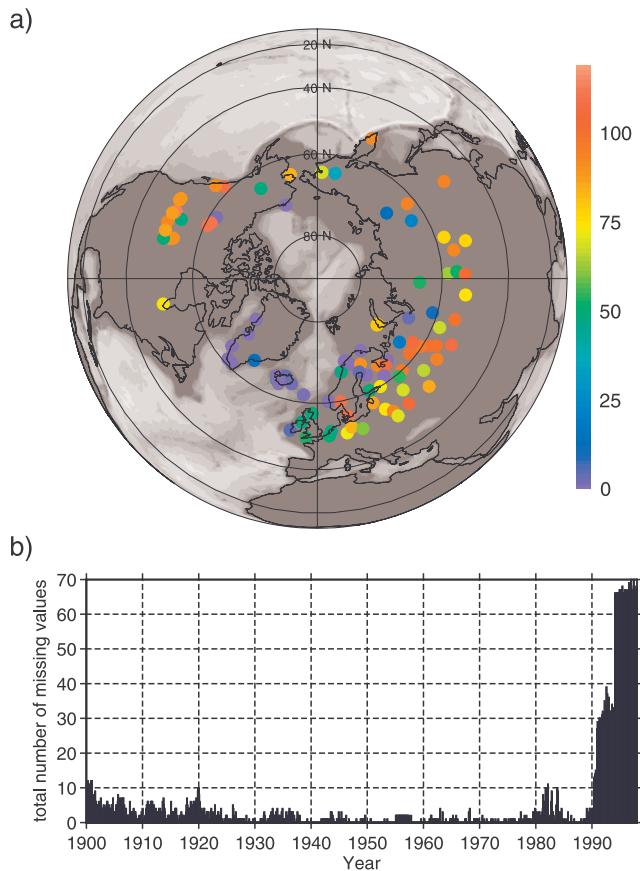
a)

b)

Figure 4. The number of missing values at each station for the period 1900 to 1997 (a) and the total number of missing monthly values as a function of time (b) for the combined surface air temperature (SAT) data set of AICSEX and IARC.

surface variables are interpolated onto the (rotated) $1° \times 1°$ spherical grid.

[16] The objective of the reconstruction is to produce forcing data on the model grid for the whole 20th century. This will be achieved by establishing a statistical link between the interpolated NCEP/NCAR data and the long-term (predictor) data sets (chap. 2.1) in an overlapping period. These spatial high-resolution data are called predictands.

3. Reconstruction

3.1. Method

[17] The statistical model linking the NCEP/NCAR reanalysis data and the selected long time series data is based on the redundancy analysis. Details of the redundancy analysis can be found in *Kauker and Meier* [2003] (and references therein) where the method has been used to generate atmospheric forcing data for a 100 year simulation of the Baltic ocean-sea ice system. A less mathematical description is given in the Appendix. The redundancy analysis yields pairs of patterns of the predictor and the predictand in which the predictand pattern is optimized to represented the highest possible variance in the fitting period. Frequently used techniques to identify pairs of

patterns are the *Canonical Correlation Analysis* (CCA) [*Hotelling*, 1936; *von Storch and Zwiers*, 1998] and the *Maximum Covariance Analysis* (MCA, often (misleadingly) called Singular Value Decomposition) [*Wallace et al.*, 1992; *von Storch and Zwiers*, 1998]. While the CCA maximizes the correlation between the corresponding pattern coefficients the MCA maximizes the (cross-) covariance or the co-variability. However, the optimization of the link between the predictor and the predictand is non-symmetric because the objective is to maximize the variance of the predictand that can be represented. Properties of the predictor patterns, such as the amount of variance they represent, are irrelevant to the problem. The redundancy analysis technique directly addresses this problem by identifying patterns that are strongly linked through a regression model. Patterns are selected by maximizing predictand variance. This technique was developed and applied in the early 1970s by *Tyler* [1982] in field of econometrics.

[18] From the five decades of NCEP data we used three decades for the fitting of the statistical model and 2 decades for validation. Most robust results were obtained when the years 1958 to 1987 were used for model fitting while the two periods 1948 to 1957 and 1988 to 1997 were left for validation. A different partition of the fitting and validation period results in less skill (Brier skill score) in the validation.

3.2. Selection of Predictors and Forcing Variables

[19] There are several data sets containing long-term time series of atmospheric variables that could be used to reconstruct the necessary forcing fields for an ocean-sea ice model. Different combinations of these data sets are possible. A reconstruction for different time scales (daily, monthly, seasonal etc.) can be attempted. Furthermore, it is an open question which forcing variables are the most important to describe the long-term development in ocean and sea ice. The following section describes some of the decisions that we made in arriving at the final reconstructed atmospheric forcing data set.

[20] In the first attempt we decided to reconstruct all relevant forcing variables, i.e., wind stress, SAT, dew-point temperature, scalar wind, cloudiness, and precipitation using the MSLPG data set as predictor (Rec I, see Table 1). Using the two decades 1948 to 1957 and 1988 to 1997 for validation, we checked the skill of the reconstruction and its dependence on time scale by comparing the proportion of explained variances (Brier-based score) in the validation period. Highest skill was found when we filtered the data with a running-mean of 37 months. However, with Rec I we got reasonable skills only for the wind stress and SAT reconstructions.

[21] The lack of skill in many forcing variables forced us to limit the reconstruction to wind stress (calculated from the NCEP 10m-wind), SLP, and SAT, the forcing fields for which we achieved the highest skills. We use monthly climatology for all other forcing variables. This has been chosen in accordance with the *Arctic Ocean Model Inter-comparison Project* (AOMIP, see http://fish.cims.nyu.edu/project_aomip/experiments/coordinated_alysis/overview.html) where daily SLP and SAT NCEP/NCAR fields for 1948 to 2003 are used. If SLP is reconstructed the wind stress is calculated from SLP also according to the AOMIP proto-col. Further on we decided to reconstruct monthly fields.

Table 1. Overview of the Reconstructions[a]

Reconstruction	Predictor Data Set(s)	Reconstructed Time Scales	Reconstructed Variables
Rec I	MSLPG	running mean 37 months	STRESS, SAT, DEW, SCAWND, CLD, PRECIP
Rec II	IARC/SMHI SLP, AARI SAT	monthly	SLP, SAT
Rec III	IARC/SMHI SLP, AICSEX/IARC SAT	monthly	SLP, SAT
Rec IV	IARC/SMHI SLP, AARI SAT	monthly	STRESS, SAT

Reconstruction	Restored Daily Variability	Climatological Variables
Rec I	–	–
Rec II	LP, SAT	DEW, CLD, PRECIP
Rec III	SLP, SAT	DEW, CLD, PRECIP
Rec IV	STRESS, SAT	DEW, CLD, PRECIP

[a]Acronyms not mentioned in the main text are DEW (dew point temperature at 2m), STRESS (surface wind stress), SCAWND (scalar wind at 10m), CLD (cloud cover), and PRECIP (precipitation).

Daily variability is reinstalled by adding the intra-monthly variability of a fixed year. This procedure is similar to that used by *Röske et al.* [2006] to construct a climatological atmospheric surface atlas including daily variability. Like Röske, we selected the year 1982.

[22] In Rec II (Table 1) we reconstructed monthly SLP using the combined IARC and SMHI monthly SLP station data (see Figure 2). Monthly SAT was reconstructed using the monthly gridded SAT of the AARI as predictor.

[23] Figure 5 depicts the leading mode (26% described variance of the dependent data, see Table 2) which resembles the pattern associated with the North Atlantic Oscillation (NAO) [*van Loon and Rogers*, 1978; *Hurrel*, 1995]. Note the correspondence of the station SLP predictor data and the dependent data, the predictand. This correspondence holds also for higher-order modes (the modes are sorted in accordance to the explained variance of the predictand pattern) as we illustrate by the example of the 6th mode (Figure 6). The triple structure of this mode strongly resembles the *EAST Atlantic Jet* teleconnection pattern (see e.g., http://www.cpc.ncep.noaa.gov/data/teledoc/eajet.html).

[24] The time series of the redundancy modes are obtained by projecting the predictor and predictand fields onto the redundancy modes (see equations (A3) and (A4) in the Appendix and Figure 7). In the overlapping period 1948 to 1997 the strong coherence of both time series can be seen. The corresponding correlation coefficients are given in Table 2.

[25] Alternatively to the calculation of the wind stress from the reconstructed SLP we reconstructed the wind stress directly with the help of the IARC/SMHI predictor. The reconstruction of the wind stress gives no further insights into the methodology and we refrain from a detailed discussion and list only numerical values of explained variances of the predictand and correlation coefficients in Table 3.

[26] After calculating the redundancy modes and the corresponding time series the reconstruction is built using equation (A6). To assess the skill of the reconstruction the explained variances described locally are calculated with respect to the NCEP/NCAR reanalysis (Figure 8). The region of high locally explained variances (80% to 90%) encompasses the Nordic Seas, the Irminger Sea, part of the Labrador Sea, and the Barents Sea. This is the case in both, the fitting period and the validation period and underlines the robustness of the reconstruction. Most parts of the Arctic show explained variances of about 50%, except parts of the Beaufort Sea where explained variances close to zero are obtained. The regions of vanishing explained variances in the Arctic are located where no station data are available.

[27] For the SAT reconstruction of Rec II the AARI data set is utilized. The second redundancy mode (17% explained variance, see Table 4) is shown in Figure 9. Both patterns show a positive signal over almost the whole Arctic, the Labrador Sea, Davis Strait, and Baffin Bay and a weak negative signal over the Nordic Seas. The corresponding time series (Figure 10) contains large positive values in the 1930, the 1940s, and the 1990s. While the first mode of this reconstruction is connected with the NAO (not shown) the second mode is very likely connected to the Arctic warming in the first half of the last century as can be seen in Figure 10. Although the predictor and predictand 60-month running-mean filtered time series are highly correlated from about 1956 to the end of the time series, they show large discrepancies for the earlier validation period 1948 to 1955. This may be an indication of a lesser accuracy of the early years of NCEP/NCAR reanalysis. This has been described as due to the upper air network (rawinsonde) which was gradually developing from 1948 to 1957 [*Kistler et al.*, 2001].

[28] The variance described locally by the reconstruction in both the fitting and the validation period (Figure 11) are much smaller then for the previously discussed SLP reconstruction (compare Figure 8). Highest values of up to 80% can be found next to Franz-Joseph Land and the Labrador Sea for the fitting period. Typical values for the Arctic are 50% to 60%. For the validation period these values are reduced by about 10% to 20% (Figure 11, right).

[29] The AARI data set used as a predictor in Rec II and Rec IV has the usual disadvantages of gridded data sets. Amounts and quality of data are different for each grid cell, i.e., the data are not homogeneous in time. It is anticipated that the uncertainty is largest for the years before about
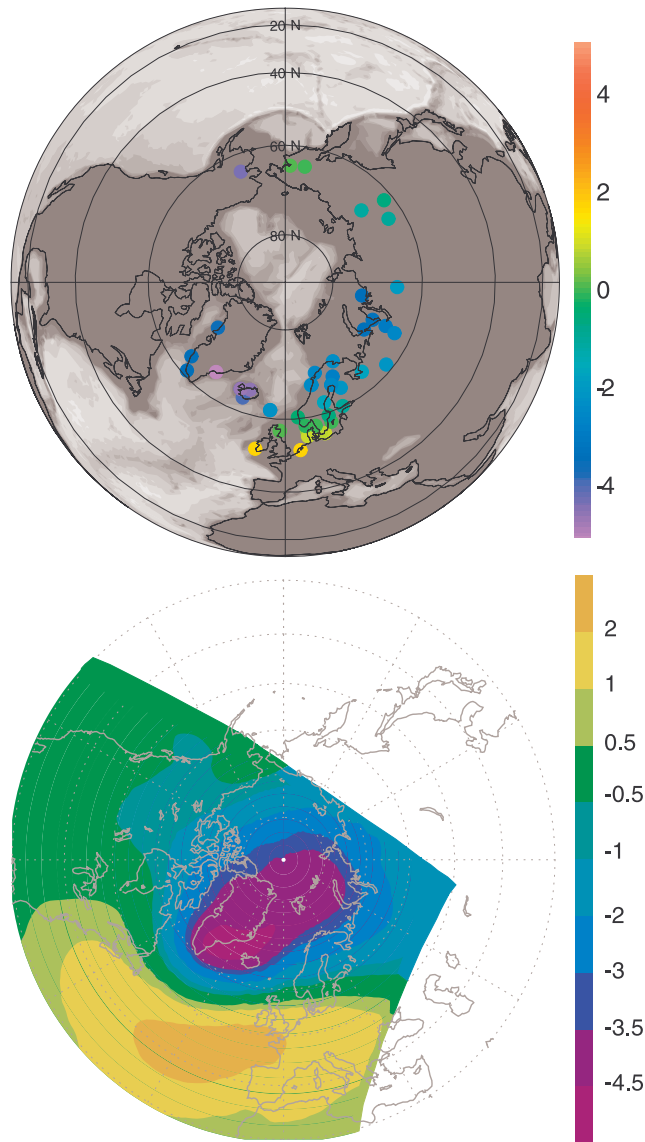
Figure 5. The first redundancy mode of the SLP reconstruction with the IARC/SMHI predictor. The upper top shows the predictor SLP pattern [*hPa*]. The predictand pattern is shown in the bottom [*hPa*] in a polar-stereographic projection to allow better comparison with the predictor data. White areas on the predictand pattern are located outside of the model domain.
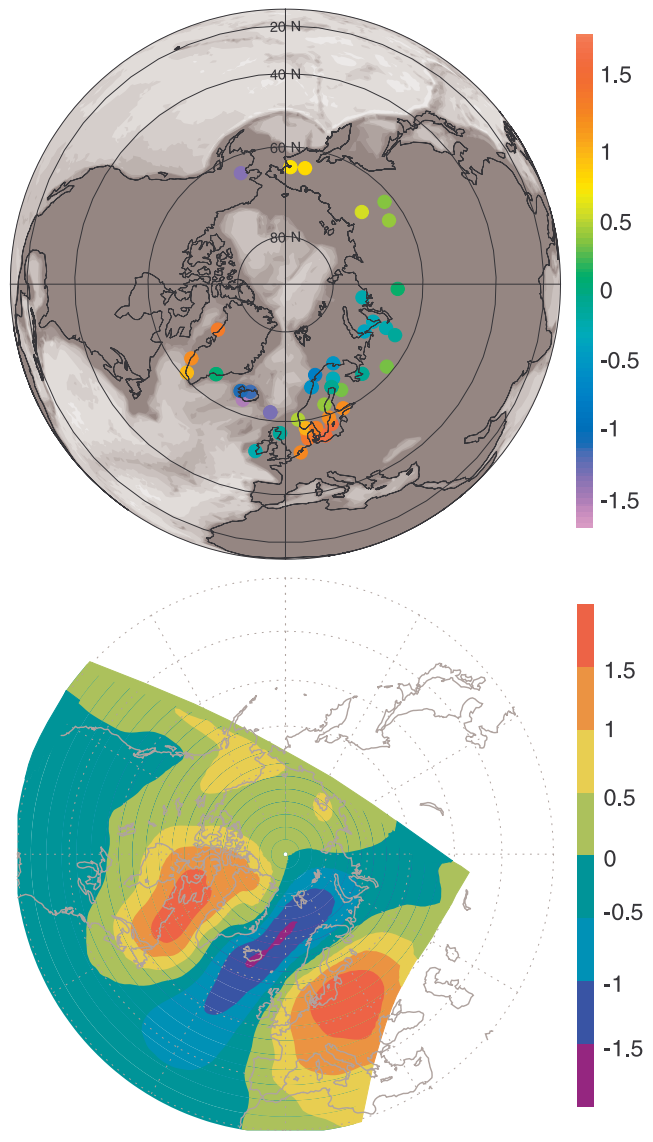
Figure 6. As in Figure 5 but for the 6th redundancy mode.

Table 2. Explained Variances of the Predictand Patterns and the Correlation of the Predictor and Predictand Time Series in the Fitting and Validation Period for the Leading Modes of the SLP Reconstruction Using the IARC/SMHI Data Set[a]

Mode	σ_{exp} [%]	r fit	r val
1	26	0.98	0.95
2	14	0.98	0.96
3	11	0.90	84
4	9	0.877	0.86
5	6	0.91	0.85
6 to 8	≤4	≤0.87	≤0.77

[a]Eight modes are used for this reconstruction.

1930, especially for the off-shore areas of the central Arctic. To estimate the uncertainties related to this inhomogeneity, we set up an alternative SAT reconstruction based on the AARI data set excluding all "off-shore" grid boxes in the central Arctic. Following an analysis of Alekseev et al. (G.V. Alekseev et al., Regional and seasonal features of two periods of greatest warming in the Arctic in the 1920–1940s and 1980–1990s, unpublished manuscript, 2008.) we calculated the mean SAT over the ocean northward of 62°N for the whole domain, for the Atlantic region (90°W to 90°E), for the Pacific region (90°E to 90°W), and for the Greenland/Iceland region (90°W to 0°E) (Figure 12). The major differences between both AARI based reconstructions are located in the Pacific region where the root-mean-square deviation (rmsd) amounts to 0.28°C (rmsd = 0.11, 0.04, 0.05°C for the whole domain, the Atlantic region, and the Greenland/Iceland region, respectively). Surprisingly, the deviation at the Pacific region is largest for the both warm periods (exceeding 0.4°C about 1940 and in the 1990s) and the rmsd is even slightly lower prior to 1930 (0.24°C) than
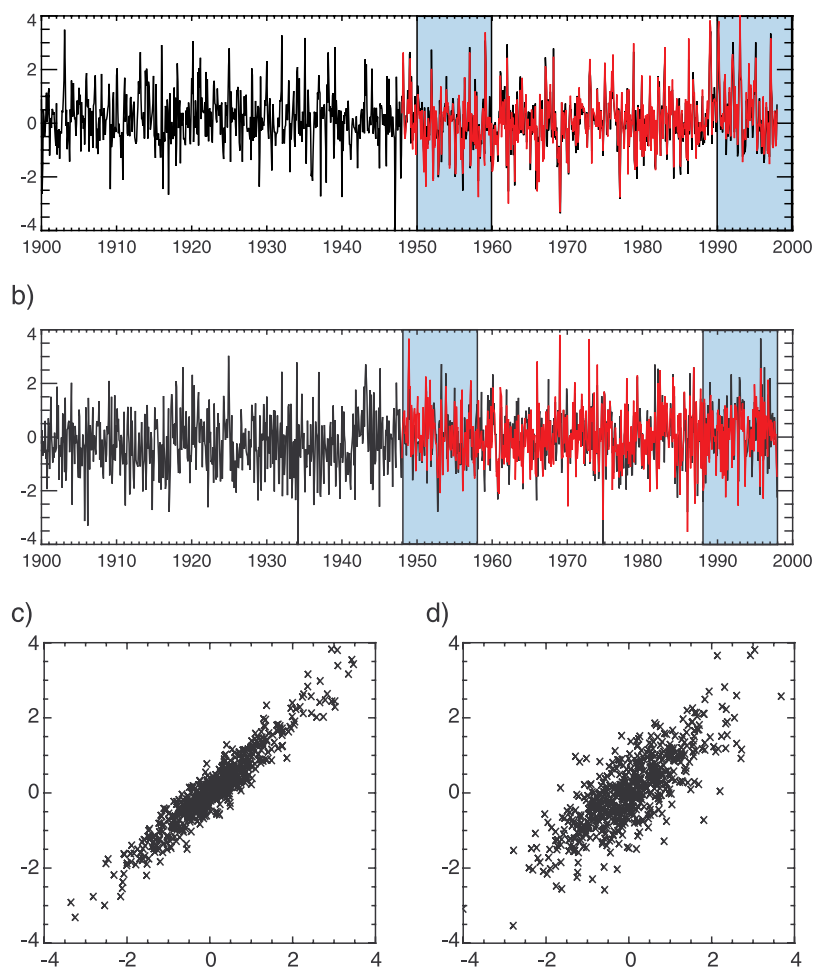
Figure 7. Monthly time series of the first (a) and sixth (b) and scatter plots of the first Figures 7c and sixth 7d redundancy modes of the SLP reconstruction based on the IARC/SMHI predictor data set. The solid line in Figures 7a and 7b shows the projection $\vec{X}^T \vec{b}_j$ for 1900 to 1997. The red line in Figures 7a and 7b is the projection $\vec{Y}^T \vec{a}_j$ for the period 1948 to 1997, i.e., the period where NCEP/NCAR data exist. In Figures 7a and 7b the validation periods 1948 to 1957 and 1988 to 1997 are shaded light blue. In Figures 7c and 7d the abscissa is displaying the projection $\vec{X}^T \vec{b}_j$ and the ordinate the projection $\vec{Y}^T \vec{a}_j$ for 1948 to 1997. See the Appendix for the nomenclature of the projections.

after 1930 (0.28°C). Compared to the SAT from the NCEP reanalysis (1948–1997, see Figure 12) the AARI reconstruction based on all data performs much better than the reconstruction in which no "off-shore" central Arctic where used. We therefore will use the full AARI data set in the following reconstructions.

[30] We set up also a reconstruction based on the AICSEX/IARC station data set (Rec III, see Table 1). In contrast to the second redundancy mode of Rec II (compare Figure 9) the second mode of Rec III (15% described variance, see Table 5 and Figure 13) shows less pronounced positive anomalies in the central Arctic but stronger positive anomalies in the Kara Sea. The corresponding time series (Figure 14), although having some similarity with the corresponding time series of Rec II (compare Figure 10), shows a less pronounced positive anomaly in the 1930s and 1940s and in the 1990s. The predictor and predictand time series are less correlated than in Rec II. Especially the positive anomaly in the 1990s is not captured by Rec III, possibly a consequence of reduced

data coverage during that period (compare Figure 4). In general, the correlation coefficients in the fitting and validation of all redundancy modes are lower than in Rec II (compare Tables 5 and 4).

[31] Comparing the correlation coefficients of the AARI SAT (Rec II) and the AICSEX/IARC (Rec III) redundancy modes we conclude that the AARI SAT reconstruction is superior to the AICSEX/IARC reconstruction. However, in certain areas this does not hold. The locally explained variances (Figure 15) in both the fitting and the validation period are almost everywhere lower than in Rec II (compare Figure 11) except in the vicinity of Iceland where six stations enter the calculation (see Figure 4).

[32] The performance of Rec III is also depicted in Figure 12. The largest discrepancies between reconstructions II and III can be found for the 1930s/1940s and the 1990s in the Pacific area (Figure 12c). There, Rec III using the AICSEX/IARC data shows only weak anomalies while Rec II using the AARI data shows anomalies reaching up to

Table 3. Explained Variances of the Predictand Patterns and the Correlation of the Predictor and Predictand Time Series in the Fitting and Validation Period for the Leading Modes of the Wind Stress Reconstruction Using the IARC/SMHI Data Set[a]

Mode	σ_{exp} [%]	r fit	r val
1	29	96	89
2	21	0.94	0.87
3	12	0.94	0.88
4	8	0.76	0.61
5	5	0.75	0.63
6 to 8	≤3	≤0.66	≤0.63

[a]Eight modes are used for this reconstruction.

Table 4. The Explained Variances of the Predictand Patterns and the Correlation of the Predictor and Predictand Time Series in the Fitting and Validation Period for the Leading Modes of the SAT Reconstruction Using the AARI Data Set[a]

Mode	σ_{exp}, %	r fit	r val
1	29	0.94	0.84
2	17	0.89	0.78
3	10	0.89	0.74
4	8	0.88	76
5	5	0.78	0.58
6 to 17	≤4	≤0.78	≤0.60

[a]Seventeen modes are used for this reconstruction.

1.5°C. In the Atlantic region, both reconstructions are mostly coherent except for the 1910s and the 1960s where Rec II exhibits much lower temperatures than Rec III. The anomalies in the Greenland/Iceland area are dominated in both reconstructions by a strong increase of the temperature during the 1920s.

[33] On the basis of the analysis presented so far it is hardly possible to decide which of the two SAT reconstructions (Rec II or Rec III) is superior in the early 20th century though the AARI reconstruction is clearly superior in the fitting and validation period. In the following section we shall test the reconstructions regarding their ability to reproduce observed sea ice variability.

4.   20th Century Ocean-Sea Ice Simulation

4.1.   Model Description

[34] Here, we apply the reconstructed atmospheric forcing data in a hindcast simulation with the $1° \times 1°$ version of AWI's NAOSIM (North Atlantic-Arctic Ocean-Sea Ice Models) hierarchy. The procedure is the same as in the AOMIP hindcast calculations for the second half of the 20th century that are discussed in a number of publications [e.g., *Köberle and Gerdes*, 2007; *Johnson et al.*, 2007]. A detailed model description can be found in the work of *Köberle and Gerdes* [2003]. The model has 19 unevenly spaced levels in the vertical. The model domain contains the Arctic Ocean,
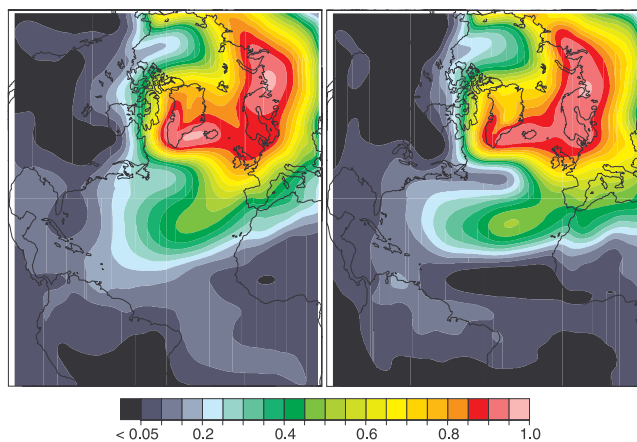

Figure 8. The locally described variances in the fitting (left) and the validation (right) periods for the monthly SLP reconstruction based on the IARC/SLP station data.
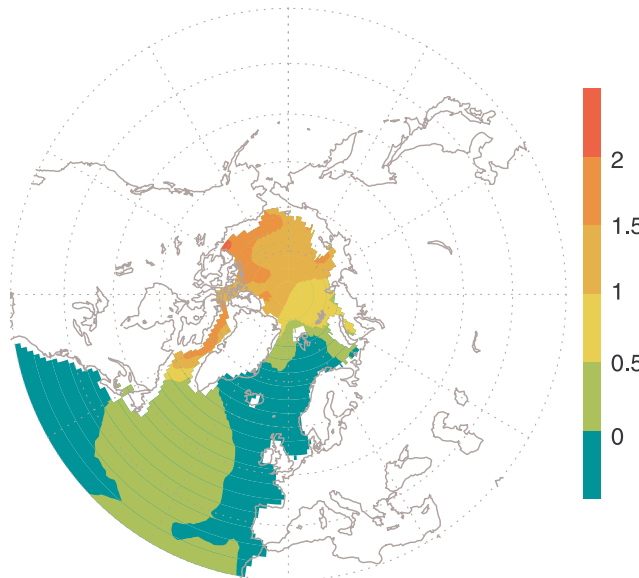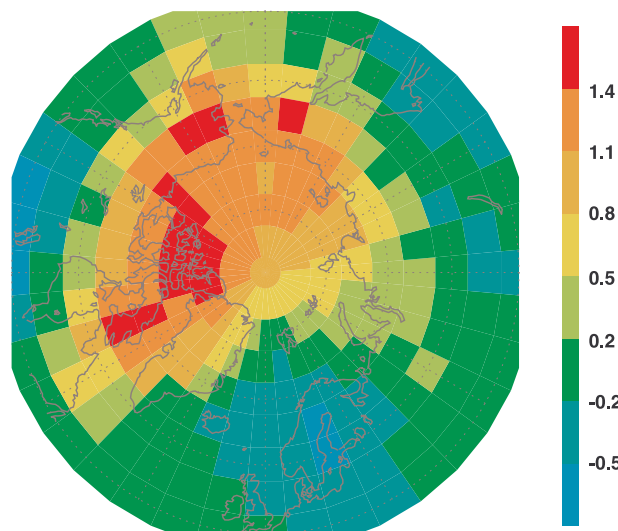

Figure 9. The second redundancy mode of the SAT reconstruction with the AARI SAT predictor. The top shows the predictor SAT pattern [°C]. The predictand pattern is shown in the bottom [°C] in a polar-stereographic projection to allow better comparison with the predictor data. White areas over ocean in the predictand pattern are located outside the model domain.
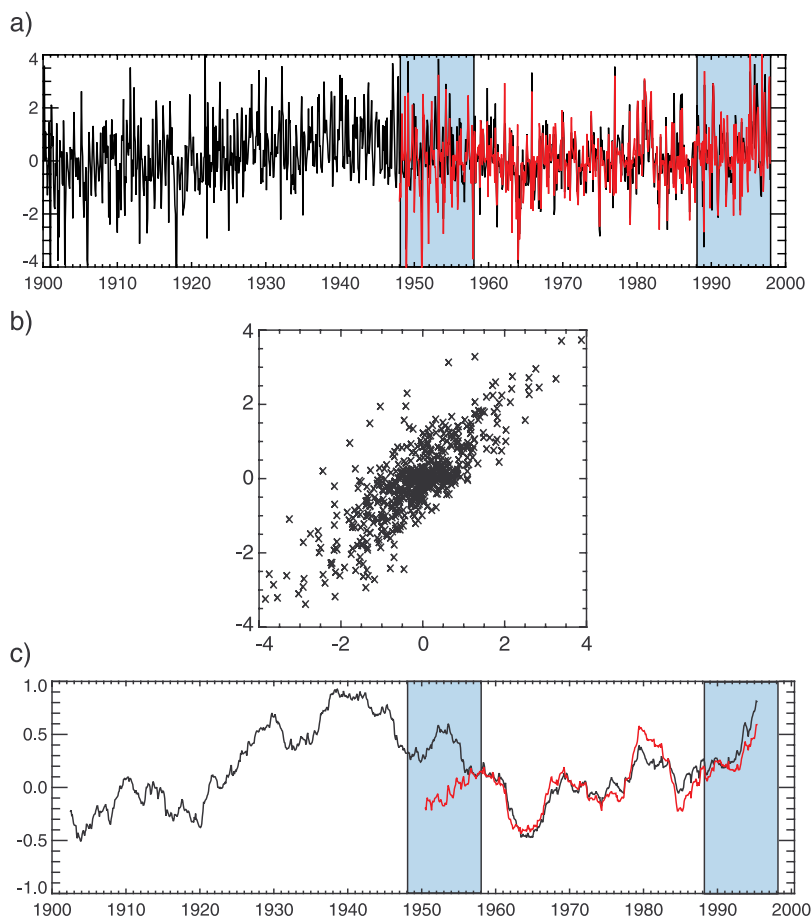
Figure 10. Time series of the second redundancy modes of the SAT reconstruction based on the AARI data set (a), the same time series filtered with a 60-month running mean (c), and a scatter plot of the unfiltered data (b). The solid lines in Figures 10a and 10c show the projection $\vec{X}^T\vec{b}_j$ for 1900 to 1997. The red line is the projection $\vec{Y}^T\vec{a}_j$ for the period 1948 to 1997, i.e., the period where NCEP/NCAR data exist. In Figures 10a and 10c the validation periods 1948 to 1957 and 1988 to 1997 are shaded light blue. In Figure 10b the abscissa is displaying the projection $\vec{X}^T\vec{b}_j$ and the ordinate the projection $\vec{Y}^T\vec{a}_j$ for 1948 to 1997. See the Appendix for the nomenclature of the projections.

the Nordic Seas and the Atlantic north of approximately 20°S. The model is formulated on a spherical grid that is rotated such that geographical 30°W meridian becomes the equator of the grid while the Pole is situated at 60°E on the geographical equator. At the southern boundary an open boundary condition has been implemented following *Stevens* [1991], allowing the outflow of tracers and the radiation of waves. The other boundaries are treated as closed walls.

[35] A dynamic-thermodynamic sea ice model with a viscous-plastic rheology [*Hibler*, 1979] is coupled to the ocean model. The prognostic variables of the sea ice model are ice thickness, snow thickness, ice concentration, and ice drift. Snow and ice thicknesses are mean quantities over a grid box. The thermodynamic evolution of the ice is described by an energy balance of the ocean mixed layer following *Parkinson and Washington* [1979]. Freezing and melting are calculated by solving the energy budget equation for a single ice layer with a snow layer. The surface heat flux is calculated from standard bulk formula using

prescribed atmospheric data and sea surface temperature predicted by the ocean model. The sea ice model is formulated on the ocean model grid and uses the same time step. The models are coupled following the procedure devised by *Hibler and Bryan* [1987].

4.2. Comparison With Historical Sea Ice Extent Data

[36] We will restrict the analysis of the model simulations to the sea ice extent for which historical observations exist. Sea ice extent is here defined as the area within the 15% sea-ice concentration margin. Widely used observational data sets are *Chapman and Walsh* [1993] of the sea ice extent of the Northern Hemisphere and the Arctic/Barents Sea sea ice extent data of *Zakharov* [1997] which includes Russian data not used by *Chapman and Walsh* [1993]. Both data sets show differences from each other that illustrate the uncertainties in these historical estimates [*Johannessen et al.*, 2004]. In the Atlantic-European Sector (Greenland and Barents Sea) actual observations for April to August enter these data sets for the whole period after 1900. In the
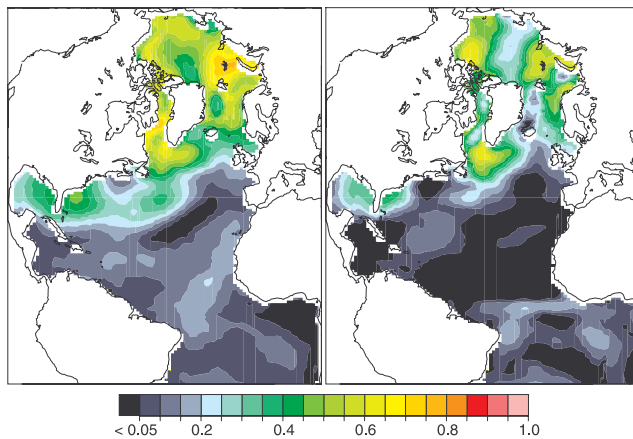
Figure 11. The locally described variances in the fitting (left) and the validation (right) periods of the monthly SAT reconstruction based on the AARI data set.

Siberian Sector (Kara, Laptev, East Siberian, and western Chukchi Seas) only observations of August sea ice extent are available and those only after 1924. After the late 1950s, all seasons have been sampled. Thus the annual mean values in these data sets before the late 1950s are themselves reconstructed by statistical means using a functional relationship between the Atlantic-European Sector and the Siberian Sector ice extents established during the period after the late 1950s. We refer to *Johannessen et al.* [2004]

for a detailed discussion on the historical sea ice extent observations and the method used to build annual means.

[37] Figure 16a depicts the annual sea ice extent filtered by a 3-year running-mean of the Arctic Ocean, the Barents Sea, and the Greenland Sea (hereafter called the *Northern Icy Ocean* following the notation of *Alekseev et al.* [1999]) for Rec I to Rec IV, a hindcast forced in accordance with the AOMIP protocol, the *Zakharov* [1997] data, and satellite derived ice extent from the *Goddard Space Flight Center* (GSFC) for 1979 to 2000 [*Cavalieri et al.*, 2003]. Satellite data, the Zakharov data, and model results for all four reconstructions agree well in the period after 1979 (Figure 16b). While smaller deviations exist, all main features of interannual variability are reproduced. Note that the years 1988 to 1997 are not used for the building of the reconstructions, i.e., the performance in these years is an independent test. Larger deviations exist between 1900 and the mid-1920s as well as between 1950 and 1970. In addition, Rec I contains large deviations for the 1940s. Table 6 lists correlation coefficients and described variances for the whole 20th century and for 1979 to 1997. Rec I has the lowest skill for the whole period and will be not considered further. Rec II to IV's skills are very close. For 1900 to 1924 Rec IV fits the Zakharov data best and Rec III worst.

[38] While Rec II to Rec IV yield similar results from about 1945 to 1970 the reconstructions show large deviations from Zakharov's data. A reason for this behavior is unclear but might be connected to the absence of the eastern Chukchi and Beaufort Seas in Zakharov's data.

[39] According to the criteria of correlation and explained variance compared with satellite observations, the AOMIP simulation performs slightly worse than the best simulations
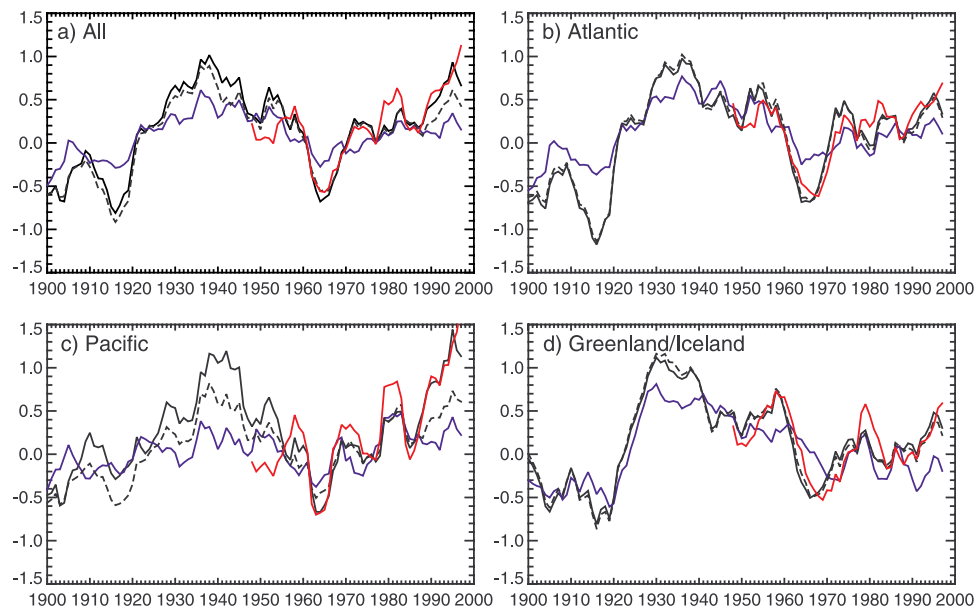

Figure 12. A comparison of the reconstructed SAT based on the AARI data set (solid line), based on the AARI data set without off-shore central Arctic data (dashed line), based on the AICSEX/IARC station data set (blue line), and the SAT from the NCEP reanalysis (red line) averaged over ocean areas northward of 62°N and filtered with a 60-month running mean. (a) The whole area, (b) the Atlantic region (90°W to 90°E), (c) the Pacific region (90°E to 90°W), and (d) the Greenland/Iceland region (90°W to 0°E).

Table 5. Explained Variances of the Predictand Patterns and the Correlation of the Predictor and Predictand Time Series in the Fitting and Validation Period for the Leading Modes of the SAT Reconstruction Using the AICSEX/IARC SAT Station Data Set[a]

Mode	σ_{exp} [%]	r fit	r val
1	26	0.92	0.74
2	15	0.82	0.57
3	10	0.75	0.70
4	8	0.71	0.48
5	6	0.57	0.41
6 to 17	≤ 4	≤ 0.67	≤ 0.40

[a]Seventeen modes are used for this reconstruction.

with reconstructed forcing data. This is surprising as one could suppose the AOMIP forcing data being closer to the actual conditions than the reconstruction. However, the reconstruction methodology ensures that only variability which is found in both, the predictor and the predictand data sets, is incorporated in the reconstruction. The temporal variability is given by the temporal variability of the predictor fields (compare equation (A6)). This filters out possible spurious variability of the reanalysis and results in higher skills of Rec II and Rec IV.

[40] Although the differences are moderate, the comparison with Zakharov's data yields that Rec II and Rec IV are superior to Rec III. Rec II has a slightly higher skill for 1979 to 1997 while Rec IV has higher skill for the whole 20th century. We will continue the discussion only for Rec II and Rec IV.

[41] Observations of August sea ice extent based on aircraft and ship observations exist for the Russian Arctic shelf seas from the mid 1920s onward. Following the analysis of *Polyakov et al.* [2003] we calculated the August sea ice extent for the East Siberian Sea (as the best resolved shelf sea in the model) for the AOMIP hindcast and for Rec II and Rec IV. Salient features of the Polyakov et al. time series are negative anomalies in the early years of the 20th century, a positive anomaly in the late 1920s, a decline of sea ice cover from then until around 1960s, and a dramatic decline and recovery of the sea ice over around 1990 (Figure 17). The results using reconstructed forcing fields reproduce the Polyakov et al. time series very faithfully since the mid-1920s. The outstanding 1990s event is reproduced by all three simulations. Rec II and Rec IV underestimate the event somewhat while the AOMIP simulation overestimates the event compared to Polyakov et al. The largest discrepancies compared to Polyakov et al. are found before the mid 1920s. During this period, the Polyakov et al. time series were composed on the basis of "occasional" ship observations [see *Polyakov et al.*, 2003].

[42] *Vinje* [2001] published April sea-ice extent in the Nordic Seas for 1860 to 1998. The area he referred to as the Nordic Seas comprises the Greenland, Iceland, Norwegian, Barents, and western Kara seas bounded by 30°W−70°E, and 80°N. The data are taken from ship logs as well as from satellite data for the most recent period (see the Appendix of *Vinje* [2001] for a detailed description of the data). We calculated the April sea-ice extent for the Nordic Seas area for the reconstructions Rev II and Rev IV.

[43] The trend of the ice extent in the Nordic Seas amounts in the *Vinje* [2001] data to about 400,000 km^2 over the 20th century. Rec II underestimates this trend strongly (about 100,000 km^2) while Rec IV simulates a trend of about 300,000 km^2 for 1900 to 1997. According to this performance of Rec II and Rec IV in the Nordic Seas, Rec IV is clearly superior to Rec II.

[44] The simulation forced with Rev IV (Figure 18) reproduces the ice extent maximum in the late 1960s as well as the minimum in the early 1990s and the maximum in the late 1920s. The simulation shows a minimum of ice extent in the mid 1940s which cannot be verified with observations because only for the years 1942 and 1949 data exist.

[45] Recently *Divine and Dick* [2006] published time series of ice edge anomalies spanning the period 1750–2002 for the geographical domain 30°W to 70°E similar to the domain used by *Vinje* [2001]. Their data are essentially
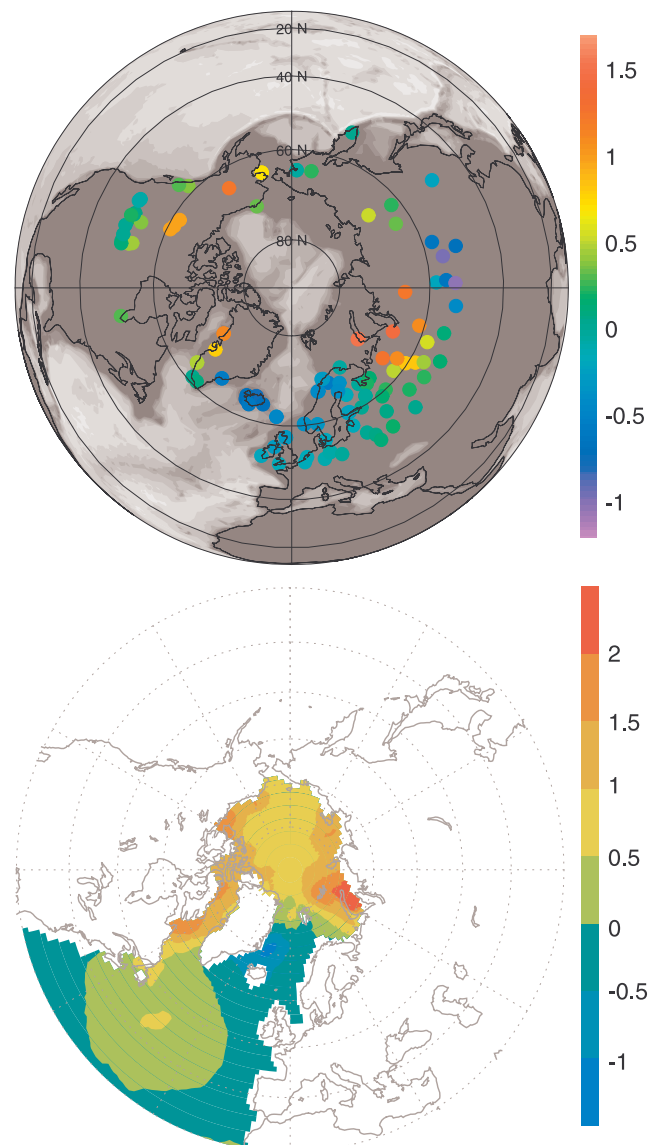

Figure 13. As in Figure 9 but for the SAT reconstruction based on the AICSEX/IARC station data.
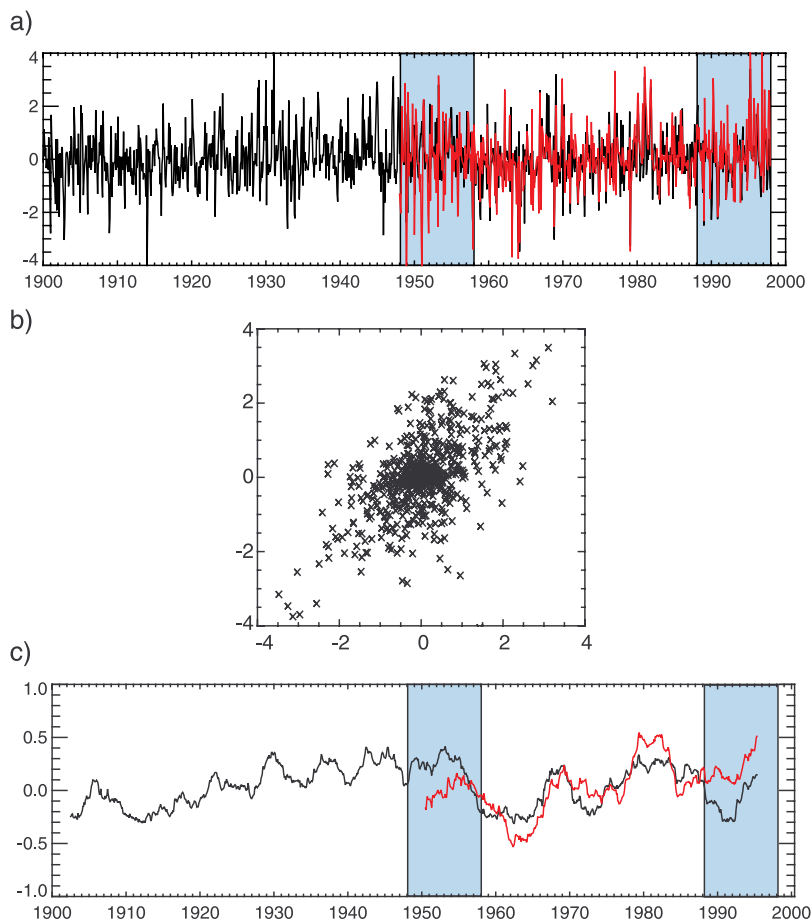
Figure 14. As in Figure 10 but for the SAT reconstruction based on the AICSEX/IARC station data.

the same as those used by Vinje (see http://nsidc.org/data/g02169.html). We compare ice-extent anomalies of Rec IV with the ice edge anomalies by *Divine and Dick* given for the whole domain for April, June, and August (see their Figure 8). For April we found a correlation of 0.73, for June of 0.60, and for August of 0.55 (Figure 19). These correlations are only slightly lower than values given by *Divine and Dick* for the correlation between their data and the ice-extent data of Vinje (0.8 and 0.85 for the Greenland Sea and the Barents Sea for April, respectively and 0.65 for the Barents Sea in August). Thus the uncertainties of the modeling results are comparable to the uncertainties due to the different techniques employed by *Divine and Dick* and *Vinje*.

[46] The annual ice extent data of Zakharov for the Northern Icy Ocean are reconstructed by means of a strong correlation with the Nordic Sea's ice extent established in the period 1959 to 1988. The model results allow us to check if this strong correlation is stationary, i.e., also holds for earlier periods. For April, the running-correlation in the model is indeed close to the one over the whole period, essentially because the only areas with ice cover less than 15% resides in the Nordic Seas (Figure 20). For August no significant correlation is found while for the annual mean ice extent correlations greater than 8 are obtained for almost the whole period. Only the World War II period shows a slightly smaller correlation. The correlation coefficients are

lower than the one obtained by Zakharov for 1958–1997 (r = 0.94) which we attribute to the fact that Zakharov's data do not include the eastern Chukchi and the Beaufort Seas. However, we found that the correlation is rather stationary, supporting Zakharov's reconstruction.
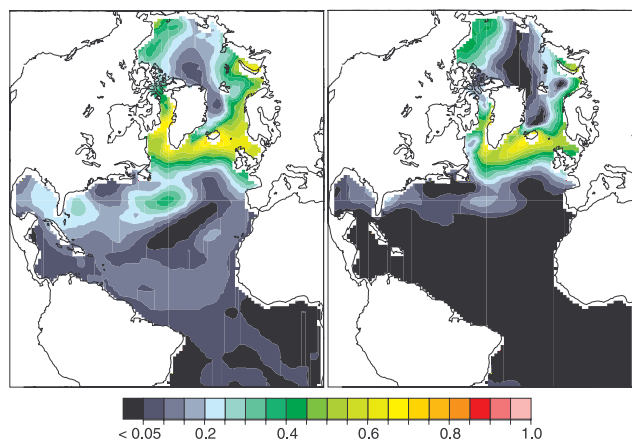

Figure 15. As in Figure 11 but for the SAT reconstruction based on the AICSEX/IARC station data.
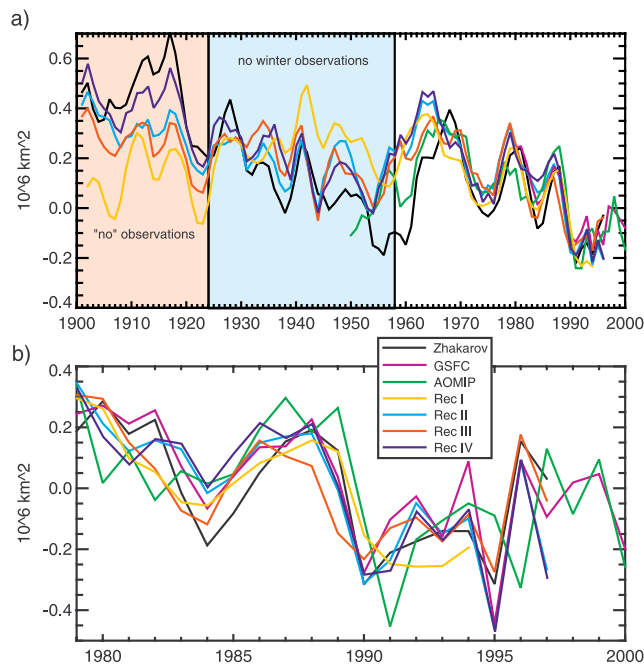
Figure 16. (a) The anomalous monthly ice extent (10^6 km^2) of the Northern Icy Ocean filtered by a 36-month running mean of the simulations run with Rec I to Rec IV, an AOMIP simulation, the Zakharov data, and the data from the GSFC. (b) Annual anomalous ice extent for 1979 to 2000. The anomalies are calculated relative to the mean over the period 1979 to 1997.

4.3. Mean Ice Extent and Trends During the Satellite Observation Period

[47] While the historic observations are only available as time series of ice extent for some ocean basins or shelf seas, the model simulations allow us to look into the regional details of sea ice changes. Before discussing the ice concentration changes in the 1920s to mid 1950s we need to compare the simulated ice concentration changes in the time period where satellite data are available. Figure 21 depicts the mean ice concentration and trend for March and September for 1979 to 1997. The satellite data used are available from the National Snow and Ice Data Center (NSIDC) and are derived from multichannel passive microwave sensors SMM/R and SSM/I with the help of the NASATEAM algorithm [*Cavalieri et al.*, 1996]. NSIDC gives uncertainties for the ice concentration of 5% in winter and of 15% in summer (in summer melt ponds on the ice

Table 6. Correlation Coefficients and Explained Variances $(1 - \sigma_{(sim-obs)}/\sigma_{obs})$ for the 20th Century and for 1979 to 1997 Between the Simulations and Observations

r/σ_{exp}	Zakharov rm3 1901–1996	GSFC ann 1979–1997
Zakharov	–/–	0.914/83.5%
AOMIP	–/–	**0.825/57.3%**
Rec I	**0.012/<0%**	807/72.0%
Rec II	0.715/50.2%	0.864/69.2%
Rec III	0.628/38.9%	0.782/57.5%
Rec IV	0.817/64.3%	0.858/65.6%

Bold numbers refer to the simulation with the lowest skill compared to observations.
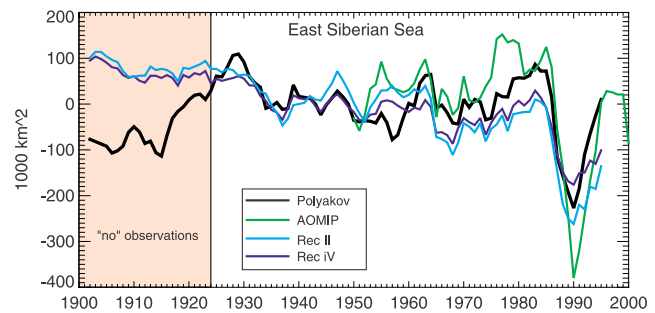

Figure 17. Time series of August ice extent anomalies for the East Siberian Sea. The solid lines are the data of *Polyakov et al.* [2003], the light and dark blue lines are the data of Rec II and Rec IV, and the green line is taken from the AOMIP hindcast. The data are 5-year running-mean filtered.

surface increase the uncertainty). The corresponding means and trends in ice concentration in Rec IV are shown in Figure 22. Taking the relatively coarse horizontal resolution of the model into account, the March mean ice margin in the simulation agrees well with the corresponding satellite data. However, the September means differ considerably. Except for the East Siberian Sea the ice margin resides too far to the north compared to the satellite data. Especially in the Laptev and Kara Sea, between Svalbard and Franz-Josef Land and in the Greenland Sea the model underestimates the ice cover. This bias of the model is known from comparisons in the AOMIP project [*Johnson et al.*, 2007] where the model was forced with NCAR/NCEP reanalysis data. The bias can thus not be regarded as fault of the reconstructed forcing data although the bias is somewhat larger in Rec IV than in the AOMIP results.

[48] The trend in March shows a reduction of the ice cover in the Barents and Greenland Sea and an increase of the ice cover in the Labrador Sea in both the satellite and the model results. The observed area of reduced sea ice cover reaches further south in the Barents Sea but is of similar
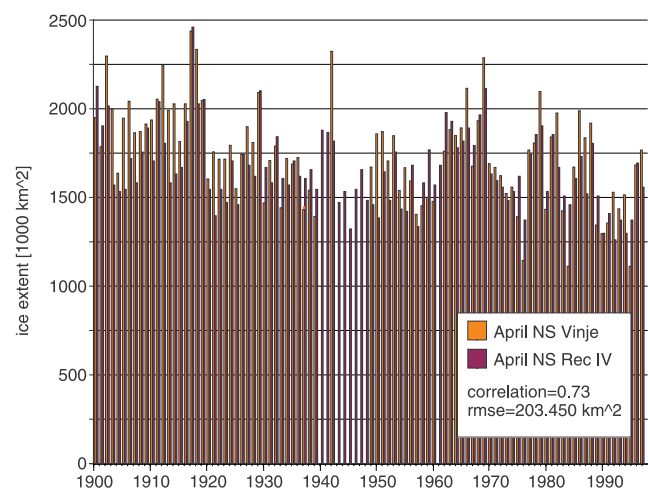

Figure 18. Vinje's observed and the modeled (Rec IV) April sea-ice extent (1000 km^2) in the Nordic Seas area (bounded by $30°W - 70°E$ and $80°N$).
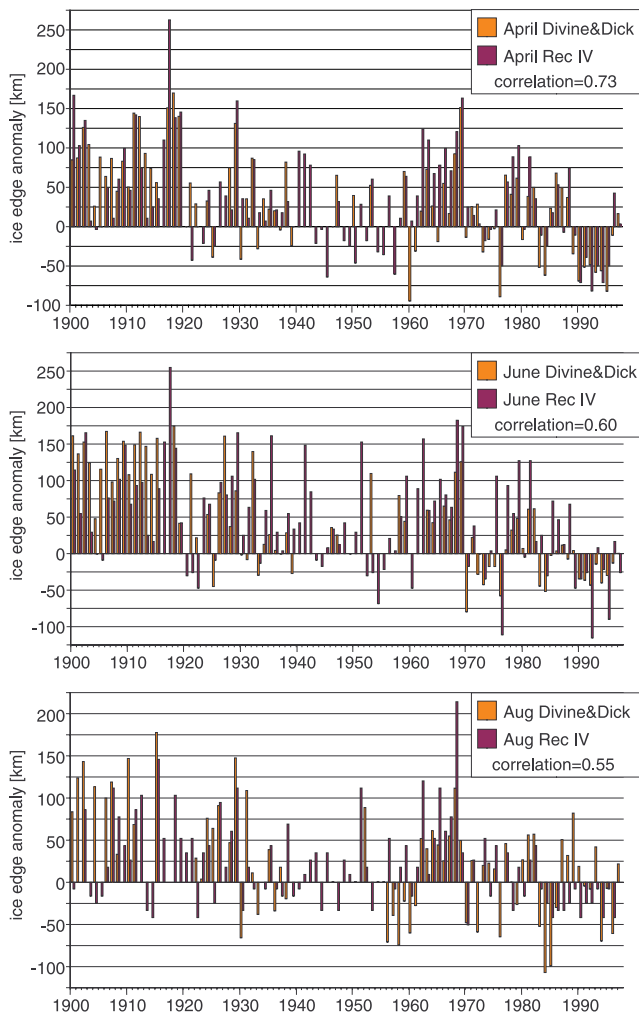
Figure 19. The time series of April, June, and August ice edge anomalies from *Divine and Dick* [2006, Figure 8] of the whole study area and the corresponding ice extent time series from Rec IV. The modeled ice extent time series are rescaled to have the same mean and standard deviation as the data of Divine and Dick.

4.4. The Early 20th Century Warming

[50] The relatively good agreement between the simulated and satellite-derived trend of the sea-ice concentration gives confidence to extend the analysis to earlier periods. We are especially interested in sea ice cover changes during 1930s Arctic warming. There is a relatively homogeneous trend in total Arctic ice extent between 1915 and 1955 (Figure 16). A corresponding map of local March and September ice concentration trends is given in Figure 23. In the Barents and Greenland seas, the trend for March reveals a decrease similar to the most recent warming. In the Labrador Sea, both periods show increasing sea ice concentration trends. Quantitatively, the earlier warming is accompanied by a much smaller increase. The pronounced dipole between Barents and Greenland seas on the one hand and the Labrador Sea on the other hand in recent trends is the response of the ice cover to the increasing trend in the strength of the NAO from the mid-1960s to the mid-1990s [*Kauker et al.*, 2003]. The different spatial pattern in the earlier trends indicates that the earlier warming was not associated with a corresponding change in the atmospheric circulation. This is consistent with findings of *Bengtsson et al.* [2004] based on ensemble integration with the atmospheric GCM ECHAM4. They note low correlations between the annual mean Arctic SAT and the NAO index but high correlations between the annual mean Arctic SAT and the winter time SLP difference between Svalbard and the northernmost Norwegian coast [see *Bengtsson et al.*, 2004, Figure 8]. According to Bengtsson et al., the early 20th century warming is not associated with large-scale atmospheric anomalies but rather caused by anomalies in the Nordic Seas itself. The higher than normal SLP difference between Svalbard and Norway causes higher oceanic heat transport into the Barents Sea, leading to the northward retreat of the sea ice and much enhanced ocean-atmosphere heat flux and higher air temperatures. It is this higher temperature that is captured in the reconstruction and causes the sea ice retreat in the ocean-sea ice model.

[51] The September trend from 1916 to 1955 exhibits the strongest decrease of about 60% north of the eastern part of the East Siberian Sea. In all other areas the decrease is much
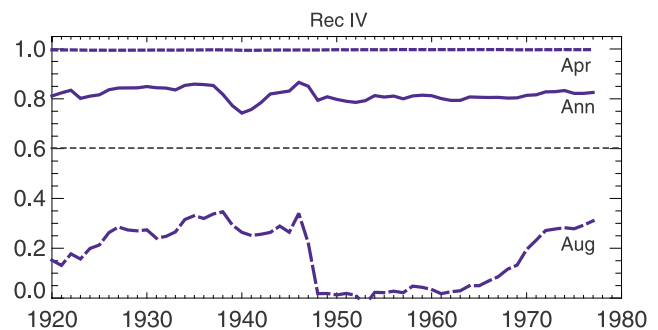
overall magnitude as in the model. In the Labrador Sea, the simulation shows an area with more ice cover than the satellite data. In the period 1979 to 1997 the NAO shows a considerable trend and the sea ice concentration trends are very similar to the pattern of the sea ice concentrations associated with an increase of the NAO [see, e.g., *Kauker et al.*, 2003, Figure 2].

[49] The trend in September is largest in the East Siberian Sea where sea ice concentration decreases by about 60% over the period 1979 to 1997 in both the satellite and simulated ice cover. Model and satellite data agree in a 20 to 30% decrease in the Chukchi and western Beaufort seas. The decrease north of the Laptev Sea is exaggerated in the simulation and is shifted north compared to the satellite data. The simulation shows large increases of ice concentration north of Svalbard where the satellite data shows only small changes.

Figure 20. The 41-year running correlation of the sea-ice extent between the Nordic Seas and the Northern Icy Ocean for April (short dashed line), for August (long dashed line), and for the annual mean (straight line). The solid short dashed line gives the 99% significance level estimated by a Monte Carlo test in which AR(1)-random variables were fitted to the data.
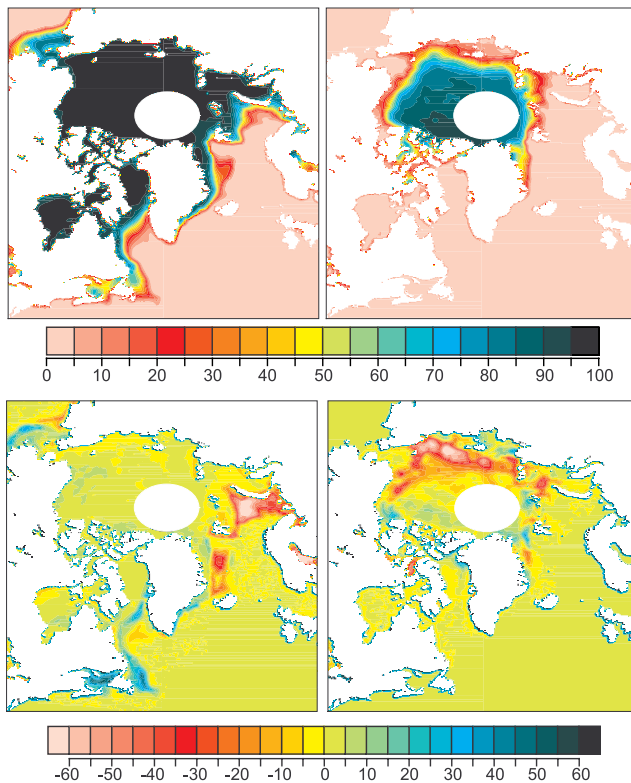
Figure 21. Mean ice concentration (%) (top) and the trend (%) from 1979 to 1997 (bottom) for March (left) and September (right) derived from SSMI/R data. The trend is given in % change for the whole period of 1979 to 1997.

time series of gridded data and individual meteorological stations with 1958–1987 reanalysis data through a redundancy analysis. With the available long time series we found that only reconstructed surface air temperature and SLP (or 10m winds) had sufficient correlation with the reanalysis modes and could explain a substantial part of the variability during two validation periods (1948–1957 and 1988–1997).

[54] For further validation, the reconstructed fields were applied in hindcast simulations with the coarse resolution version of NAOSIM. The results were compared with independent sea-ice extent data. Unfortunately, these data are not always based on direct observations but are partly based on simple statistical models themselves. Deviations thus cannot unambiguously be attributed to deficiencies in the reconstruction. Overall, we judge the reconstruction based on SLP station data from a combined data set of IARC and SMHI and the gridded surface air temperature atlas of the AARI (Rec IV) as superior.

[55] Although differences between different reconstructions are noticeable, three of them agree in a long-term trend of declining sea ice extent in the Northern Icy Ocean (the Arctic Ocean proper, the Barents Sea, and the Greenland Sea). This century long trend is consistent with the results of many coupled climate models under natural and anthropogenic forcing [*Gerdes and Köberle*, 2007]. However, superimposed on the century long trend, we see pronounced multi-decadal variability. This low-frequency variability is manifest in a strong decline of sea ice extent between around 1920 and 1960 that leads into a minimum of sea ice at the end of the 1950s. The decrease in the simulation is
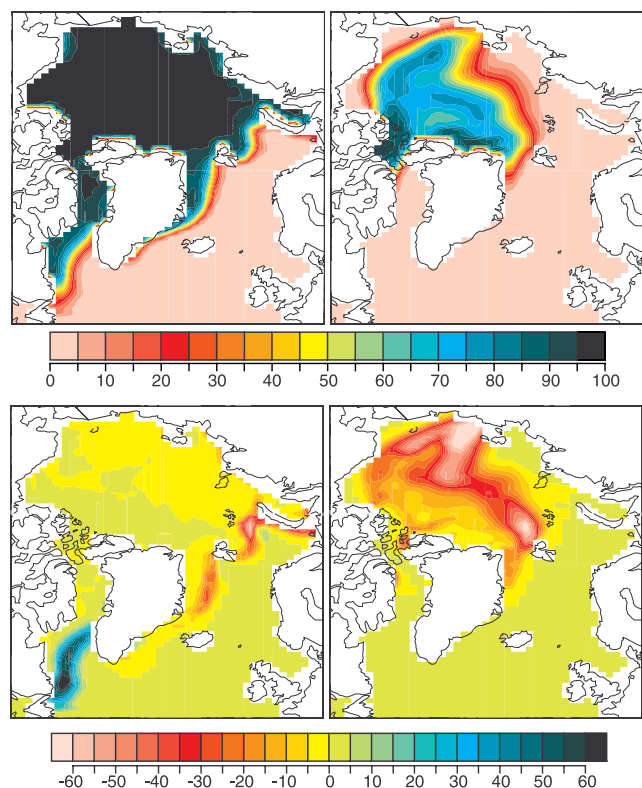
lower with values of the order of 10 to 30%. In contrast to the most recent warming the increase is confined to the Eurasian part of the Arctic. Only very localized anomalies are found in the eastern Chukchi and Beaufort Sea which counterbalance each other. This finding is consistent with the argument of *Johannessen et al.* [2004] that the mid-1950s minimum in the Zakharov data (compare Figure 16) is almost as pronounced as the mid-1990s minimum because the eastern Chukchi and Beaufort seas are not taken into account in these data. Note that all reconstructions show a decrease of the ice extent between the mid-1950s and the mid-1990s of about 200,000 km^2.

[52] Rec IV suggests that the effect of the early 20th century warming on sea ice is restricted to the Greenland, Barents, and the Siberian part of the Arctic. Thus the situation appears different from the recent warming where basically the whole Arctic Ocean is affected. From the results of Rec IV we estimate a decrease in the Northern Icy Ocean sea ice extent for the mid-1910s to the mid-1950s of about 500,000 km, a loss of about 6%. For the mid-1960s to the mid-1990s the loss is about 600,000 km^2 or 7% of the total sea ice extent.

5. Summary and Conclusions

[53] We have introduced an atlas of atmospheric surface fields for the 20th century that is suitable to force an ocean-sea ice model. The data were constructed by linking long

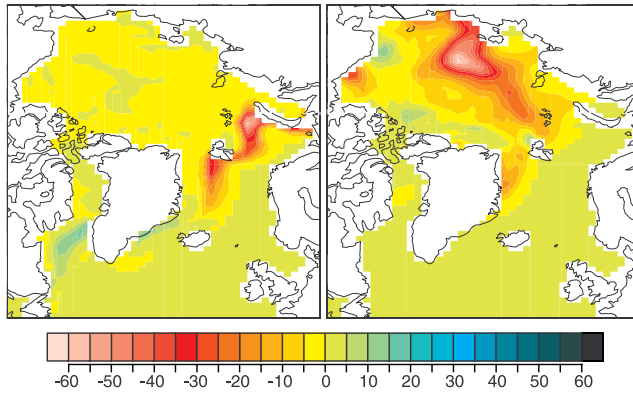**Figure 22.** As in Figure 21 but for the simulation Rec IV.

Figure 23. The trend of the sea ice concentration (%) from 1916 to 1955 for March (left) and September (right) for Rec IV. The trend is given in % change for the whole period of 1916 to 1955.

almost as large as the decrease from the mid 1990s to end of the 20th century, however the minimum of the end of the 20th century is lower by about 200,000 km^2 than the minimum at the end of the 1950s. A brief but intense build-up in the first half of the 1960s is followed by the second long-term decline toward the end of the century. We suppose that the century-long trend is anthropogenically forced but we warn that even time series of five to six decades length might be strongly influenced by natural variability of the climate system and trends estimated from such time series might not faithfully reflect the effect of increasing concentration of greenhouse gases in the atmosphere.

Appendix A: Redundancy Analysis

[56] The redundancy analysis was developed in the early 1970s by *Tyler* [1982]. See *Kauker and Meier* [2003] for a detailed mathematical description of the method.

[57] Here, we will only briefly describe the method. Technically, the method can be reduced to solving two eigenequations:

$$\Sigma_{YX}\Sigma_{XX}^{-1}\Sigma_{XY}\vec{a}_j = \lambda_j\vec{a}_j \tag{A1}$$

$$\Sigma_{XX}^{-1}\Sigma_{XY}\Sigma_{YX}\vec{b}_j = \lambda_j\vec{b}_j, \tag{A2}$$

with Σ_{YX} being the covariance matrix between the predictor field X and the predictand field Y and Σ_{XX}^{-1} being the inverse of the autocovariance matrix of X (assuring that the eigenvectors are independent of the predictor pattern's variance). Note that the eigenvalues of both eigenequations are equal. $\vec{X}$ can be expanded in the usual manner

$$\vec{X} = \sum_{j=1}^{m_X}\left(\vec{X}^T\vec{b}_j\right)\vec{p}_j, \tag{A3}$$

where the adjoint patterns $P = (\vec{p}_1|\ldots|\vec{p}_{mX})$ are given by $P^T = B^{-1}$ (the columns of B are formed by the eigenstates $\vec{b}_j$).

The part $\widehat{\vec{Y}}$ of $\vec{Y}$ that can be represented by $\vec{X}$ can be expanded as

$$\widehat{\vec{Y}} = \sum_{j=1}^{k}\left(\widehat{\vec{Y}}^T\vec{a}_j\right)\vec{a}_j \tag{A4}$$

(Note that A is self-adjoint because A is orthogonal). The expansion coefficient for $\widehat{\vec{Y}}$ can be rewritten

$$\widehat{\vec{Y}}^T\vec{a}_j = \sqrt{\lambda_j}\vec{X}^T\vec{b}_j. \tag{A5}$$

$\vec{p}_1$ is the pattern of the predictor $\vec{X}$ which provides maximal variance of the predictand $\vec{Y}$ (the pattern $\vec{a}_1$). $\vec{p}_2$ is the pattern which provides the second most variance and so forth.

[58] The predictand is reconstructed with the help of equation (A5). The reconstruction reads

$$\vec{Y}_{rec} = \sum_{j=1}^{k}\sqrt{\lambda_j}\left(\vec{X}^T\vec{b}_j\right)\vec{a}_j. \tag{A6}$$

For example, if $\vec{X}$ is given for 100 years, equation (A6) allows us to reconstruct $\vec{Y}$ for 100 years.

[59] To avoid collinearity problems of the predictor and predictands an Empirical Orthogonal Function (EOF) analysis has been performed on the data prior to the redundancy analysis. Then, Σ_{XX} and Σ_{YY} are identity matrices and the computational effort is reduced considerably. A disadvantage is that the predictor variance is reintroduced implicitly. To detect overfitting we increased the number of EOFs successively by taking into account modes of greater than 5%, 2%, 1%, 0.5%, 0.2%, and 0.1% described variances. First, we examined the reconstructed time series of each mode by comparing its variance during the fitting period (1958 to 1987) with the variance in the period 1900 to 1957 and 1988 to 1997. For reconstructions using EOF modes with described variances greater than 0.5%, 0.2% and 0.1% we found increasingly higher variances of the modal time series outside of the fitting period compared to within. This is an indication for overfitting. Second, we compared the locally explained variances (Brier skill score) of the reconstructed fields in the two validation periods 1948 to 1957 and 1988 to 1997 and the fitting period. Although the explained variances in the fitting period increases with the number of EOF modes used in the validation periods, for the reconstruction using PCA modes with less than 1% variance the locally described variances decrease for some areas. This is again an indication for overfitting.

[60] The results depend on the reconstructed variable. We found weaker tendency for overfitting for the SAT reconstructions and stronger dependency for the SLP and wind stress reconstruction. We concluded that limiting the number of EOF modes to modes of more than 1% described variance prevents overfitting for all variables and used this criteria for all reconstructions shown.

[61] To assess if the statistical model is able to reproduce the longest resolved time scales we tested the residual trends (predictand time series minus predictor time series) against the variability for 1948 to 1997. We used a Monte-Carlo test by fitting AR(1) random time series to the residual. Then we

calculated the trends of the random time series. The statistics of these trends allow us to test the null hypothesis of zero trend for the residual time series. This yields a fair test because the serial correlation of the residual time series is taken into account. Even for a relatively low p-value of 90% the null hypothesis of zero trend could not be rejected for all modes and for all reconstructions, i.e., the residual time series are trendless.

References

Alekseev, G. V., Ye.I. Aleksandrov, R. V. Bekriayev, P. N. Svyashchennikov, and N. Ya (1999), Harlanienkova, Surface air temperature from meteorological data, in *Detection and Modelling of Greenhouse Warming in the Arctic and sub-Arctic, INTAS Grant 97-1277 Technical Report on Task 1*, Arctic and Antarctic Research Institute, St. Petersburg, Russia.

Alexandersson, H., H. Tuomenvirta, and K. Iden (2000), Trends of storms in NW Europe derived from an updated pressure data set, *Clim. Res.*, *14*, 71–73.

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

Cavalieri, D. J., C. L. Parkinson, and K. Y. Vinnikov (2003), 30-year satellite record reveals contrasting Arctic and Antarctic decadal sea ice variability, *Geophys. Res. Lett.*, *30*(8), 1970, doi:10.1029/2003GL018031.

Cavalieri, D. J., C. L. Parkinson, P. Gloersen, and H. J. Zwally (1996), *Sea Ice Concentration from Nimbus-7 SMMR and DMSP SSM/I Passive Microwave Data* [Digital media, 1996, updated], National Snow and Sea Ice Center, Boulder, Colo.

Chapman, W. K., and J. E. Walsh (1993), Recent variations of the sea ice and air temperature in high latitudes, *Bull. Am. Meteorol. Soc.*, *74*, 33–47.

Divine, V. D., and C. Dick (2006), Historical variability of sea ice edge position in the Nordic Seas, *J. Geophys. Res.*, *111*, C01001, doi:10.1029/2004JC002851.

Gerdes, R., and C. Köberle (2007), Comparison of Arctic sea ice thickness variability in IPCC Climate of the 20th Century experiments and in ocean-sea ice hindcasts, *J. Geophys. Res.*, *112*, C04S01, doi:10.1029/2006JC003616.

Hanssen-Bauer, I., C. Achberger, R. E. Benestad, and D. Forland (2005), Statistical downscaling of climate scenarios over Scandinavia: A review, *Clim. Res.*, *29*, 255–268.

Hatun, H., A. B. Sando, H. Drange, B. Hansen, and H. Valdimarsson (2005), Influence of the Atlantic Subpolar Gyre on the Thermohaline circulation, *Science*, *309*, 1841–1844, doi:10.1126/science.1114777.

Hibler, W. D. (1979), A dynamic thermodynamic sea ice model, *Phys. Oceanogr.*, *9*, 815–846.

Hibler, W. D., III, and K. Bryan (1987), A diagnostic ice-ocean model, *J. Phys. Oceanogr*, *17*, 987–1015.

Hilmer, M. (2001), A model study of Arctic sea ice variability, in *Berichte aus dem Institut für Meereskunde*, vol. 320, Institut für Meereskunde Kiel, Düsternbrooker Weg 20, 24105, Kiel, Germany.

Hotelling, H. (1936), Relations between two sets of variates, *Biometrika*, *28*, 321–377.

Hurrel, J. W. (1995), Decadal trends in the North Atlantic oscillation: Regional temperatures and precipitation, *Science*, *269*, 676–679.

Johannessen, O. M., et al. (2004), Arctic climate change: Observed and modelled temperature and sea-ice variability, *Tellus*, *56A*, 328–341.

Johnson, M., S. Gaffigan, E. Hunke, and R. Gerdes (2007), A comparison of Arctic Ocean sea ice concentration among the coordinated AOMIP model experiments, *J. Geophys. Res.*, *112*, C04S11, doi:10.1029/2006JC003690.

Jones, P. D. (1987), The early twentieth century Arctic high—fact or fiction?, *Clim. Dyn.*, *1*, 63–75.

Kalnay, E., et al. (1996), The NCEP/NCAR 40-year reanalysis project, *Bull. Am. Meteorol. Soc.*, *77*(3), 437–495.

Kauker, F., and H. E. M. Meier (2003), Modeling decadal variability of the Baltic Sea: 1. Reconstructing atmospheric surface data for the period 1902–1998, *J. Geophys. Res.*, *108*(C8), 3267, doi:10.1029/2003JC001797.

Kauker, F., R. Gerdes, M. Karcher, C. Köberle, and J. L. Lieser (2003), Variability of Arctic and North Atlantic sea ice: A combined analysis of model results and observations from 1978 to 2001, *J. Geophys. Res.*, *108*(C6), 3182, doi:10.1029/2002JC001573.

Kistler, R., et al. (2001), The NCEP-NCAR 50-year reanalysis: Monthly means CD-ROM and documentation, *Bull. Am. Meteorol. Soc.*, *82*, 247–267.

Köberle, C., and R. Gerdes (2003), Mechanisms determining the variability of arctic sea ice conditions and export, *J. Clim.*, *16*, 2843–2858.

Köberle, C., and R. Gerdes (2007), Simulated variability of the Arctic Ocean fresh water balance 1948–2001, *J. Phys. Oceanogr*, *37*, 1628–1644, doi:10.1175/JPO3063.1.

Levitus, S., R. Burgett, and T. Boyer (1994), *World Ocean Atlas 1994, vol. 3, Salinity, and vol. 4, Temperature*, NOAA Atlas NESDIS 3 and 4, U. S. Department of Commerce, Washington D.C.

Meier, H. E. M., and F. Kauker (2003), Modeling decadal variability of the Baltic Sea: 2. Role of freshwater inflow and large-scale atmospheric circulation for salinity, *J. Geophys. Res.*, *108*(C11), 3368, doi:10.1029/2003JC001799.

Parkinson, C. L., and W. M. Washington (1979), A large-scale numerical model of sea ice, *J. Geophys. Res.*, *84*(C1), 311–337.

Polyakow, I. V., G. V. Alekseev, R. V. Bekryaev, U. S. Bhatt, R. Colony, M. A. Johnson, V. P. Karklin, D. Walsh, and A. V. Yulin (2003), Long-term ice variability in Arctic marginal seas, *J. Clim.*, *16*, 2078–2085.

Polyakov, I., et al. (2005), One more step toward a warmer Arctic, *Geophys. Res. Lett.*, *32*, L17605, doi:10.1029/2005GL023740.

Röske, F. (2006), A global heat and fresh water forcing dataset for ocean models, *Ocean Model.*, *11*(3–4), 235–297.

Stevens, D. P. (1991), On open boundary condition in the United Kingdom Fine-Resolution Antarctic Model, *J. Phys. Oceanogr.*, *21*, 1494–1499.

Trenberth, K., and E. Paolino (1980), The Northern Hemisphere sea level pressure data set: Trends, errors, and discontinuities, *Mon. Weather Rev.*, *108*, 855–872.

Trenberth, K., and E. Paolino (1981), Characteristic patterns of variability of sea level pressure in the Northern Hemisphere, *Mon. Weather Rev.*, *108*, 1169–1189.

Tyler, D. E. (1982), On the optimality of the simultaneous redundancy transformations, *Psychometrica*, *47*, 77–86.

van Loon, H., and J. C. Rodgers (1978), The seasaw in winter + temperatures between Greenland and Northern Europe. part I: General description, *Mon. Weather Rev.*, *106*, 296–310.

von Storch, H., and F. Zwiers (1998), *Statistical Analysis in Climate Research*, Cambridge Univ. Press, New York.

von Storch, H., E. Zorita, J. Jones, Y. Dimitriev, F. Gonzalez-Rouco, and S. Tett (2004), Reconstructing past climate from noisy data, *Science*, doi:10.1126/science.1096109.

Vinje, T. (2001), Anomalies and trends of sea-ice extent and atmospheric circulation in the Nordic Seas during the period 1864–1998, *J. Clim.*, *14*(3), 255–267.

Wallace, J. M., C. Smith, and C. S. Bretherton (1992), Singular value decomposition of wintertime sea surface temperature and 500-mb height anomalies, *J. Clim.*, *5*, 561–576.

Zakharov, V. F. (1997), *Sea Ice in the Climate System, World Climate Research Programme/Arctic Climate System Study, WMO/TD 782*, 80 pp., World Meteorological Organization, Geneva.

Zorita, E., F. Gonzalez-Rouco, and S. and (2003), Testing the Mann et al. approach to paleoclimate reconstructions in the context of a 1000-yr control simulation with the ECHO-G coupled climate model, *J. Clim.*, 161,378–161,390.

———————

F. Kauker, C. Köberle, and R. Gerdes, Climate System, Alfred Wegener Institute for Polar and Marine Research, Bussestr 24, P.O. Box 120161, Bremerhaven 25175, Germany. (fkauker@awi-bremerhaven.de)

M. Karcher, O.A.Sys, Ocean Atmosphere System, Hamburg, Germany.