The Early Twentieth-Century Warming in the Arctic—A Possible Mechanism

LENNART BENGTSSON
Max Planck Institute for Meteorology, Hamburg, Germany, and Environmental Systems Science Centre, University of Reading, Reading, United Kingdom

VLADIMIR A. SEMENOV
Max Planck Institute for Meteorology, Hamburg, Germany, and Obukhov Institute of Atmospheric Physics, Moscow, Russia

OLA M. JOHANNESEN
Nansen Environmental and Remote Sensing Center/Geophysical Institute, University of Bergen, Bergen, Norway

(Manuscript received 7 April 2003, in final form 23 February 2004)

ABSTRACT

The huge warming of the Arctic that started in the early 1920s and lasted for almost two decades is one of the most spectacular climate events of the twentieth century. During the peak period 1930–40, the annually averaged temperature anomaly for the area 60°–90°N amounted to some 1.7°C. Whether this event is an example of an internal climate mode or is externally forced, such as by enhanced solar effects, is presently under debate. This study suggests that natural variability is a likely cause, with reduced sea ice cover being crucial for the warming. A robust sea ice–air temperature relationship was demonstrated by a set of four simulations with the atmospheric ECHAM model forced with observed SST and sea ice concentrations. An analysis of the spatial characteristics of the observed early twentieth-century surface air temperature anomaly revealed that it was associated with similar sea ice variations. Further investigation of the variability of Arctic surface temperature and sea ice cover was performed by analyzing data from a coupled ocean–atmosphere model. By analyzing climate anomalies in the model that are similar to those that occurred in the early twentieth century, it was found that the simulated temperature increase in the Arctic was related to enhanced wind-driven oceanic inflow into the Barents Sea with an associated sea ice retreat. The magnitude of the inflow is linked to the strength of westerlies into the Barents Sea. This study proposes a mechanism sustaining the enhanced westerly winds by a cyclonic atmospheric circulation in the Barents Sea region created by a strong surface heat flux over the ice-free areas. Observational data suggest a similar series of events during the early twentieth-century Arctic warming, including increasing westerly winds between Spitsbergen and Norway, reduced sea ice, and enhanced cyclonic circulation over the Barents Sea. At the same time, the North Atlantic Oscillation was weakening.

1. Introduction

The warming event in the first part of the twentieth century, considered at the time by some as the first sign of climate warming caused by increasing CO₂ (Callendar 1938), had its largest amplitude in the higher latitudes of the Northern Hemisphere. The largest warming occurred in the Arctic (60°–90°N) (Johannessen et al. 2004) averaged for the 1940s with some 1.7°C (2.2°C for the winter half of the year) relative to the 1910s. As can be seen from Fig. 1, it was a long-lasting event commencing in the early 1920s and reaching its maximum some 20 years later. The decades after were much colder, although not as cold as in the early years of the last century. It is interesting to note that the ongoing present warming has just reached the peak value of the 1940s, and this has underpinned some views that even the present Arctic warming is dominated by factors other than increasing greenhouse gases (Polyakov and Johannessen 2000; Polyakov et al. 2002). However, other authors (e.g., Johannessen et al. 2004) concluded that the present warming in the Arctic is dominated by anthropogenic greenhouse gas forcing.

Four possible mechanisms, individually or in combination, could have contributed to the early twentieth-century warming: anthropogenic effects, increased solar irradiation, reduced volcanic activity, and internal variability of the climate system.

It seems unlikely that anthropogenic forcing on its own could have caused the warming, since the change in greenhouse gas forcing in the early decades of the twentieth century was only some 20% of the present (Roeckner et al. 1999). Second, it remains to explain
the marked cooling trend between 1940 and 1960, a period with a similar or faster increase of the greenhouse gases than between 1920 and 1940 (Joos and Bruno 1998). The direct and indirect effect of sulfate aerosols though had a rapidly increasing trend (Roeckner et al. 1999), which partly compensated the increase of greenhouse gas forcing during the 1940–60 period. Nevertheless, the total anthropogenic forcing was larger in the 1940–60 period, when cooling occurred, than in the 1920–40 warming period, thereby rejecting the idea that anthropogenic forcing caused the 1920–40 warming.

Changes in solar forcing have been suggested (e.g., Lean and Rind 1998; Beer et al. 2000 and references therein) as the cause of the warming. It has attracted considerable interest because of the apparent similarity between the assumed solar variability and the temperature trend of the Northern Hemisphere extratropics (Reid 1991; Friis-Christensen and Lassen 1991; Hoyt and Schatten 1993). Available reconstructions of solar variability (Hoyt and Schatten 1993; Lean et al. 1995) have been used in modeling studies (Cubasch et al. 1997; Cubasch and Voss 2000; Stott et al. 2001; Tett et al. 1999) providing results broadly consistent with long-term estimated or observed temperature trends. This is not surprising since multidecadal and longer trends in solar forcing generate a response in global average modeled temperatures on similar time scales as the anticipated variation in solar forcing (Cubasch et al. 1997).

However, reliable observational data of solar irradiation only exist for the last two decades (Fröhlich and Lean 1998), which is insufficient to show any longer-term trends. Therefore, solar forcing can at the present time only be considered as a hypothesis of climate change and will presumably need another couple of decades of accurate irradiation measurements before it can be accepted or dismissed.

From 1912 onward, volcanic activity entered a more quiescent phase and thus eliminated this climate cooling factor. Following the major eruptions at the beginning of the last century (Santa Maria in 1902, Ksudach in 1907, and Katmai in 1912) no substantial volcanic eruption occurred until Mount Agung in 1963 (Robock 2000). Of these eruptions Katmai was the most intense at least with respect to the anticipated effect on climate (stratospheric aerosols). Nevertheless, the eruption by Mount Pinatubo in 1991 was more powerful by about a factor of 2 when compared with Katmai (Robock 2000). We may therefore assume that the climate effect from Katmai was less than that from Pinatubo and we have no reason to believe that the effect should have lasted longer than the effect of Pinatubo. According to several modeling studies as well as to observational assessments, the cooling effect of Pinatubo had disappeared after a period shorter than 3 yr (Bengtsson et al. 1999). Consequently the cooling effect of Katmai could hardly have lasted longer than 1915.

Coupled model experiments have demonstrated marked natural fluctuations on decadal time scales in the Arctic (Johannessen et al. 2004). There are strong indications that such fluctuations also exist in nature although the mechanisms causing them are open to debate (e.g., Ikeda 1990; Ikeda et al. 2001; Mysak et al. 1990, 2001; Mysak and Venegas 1998; Johnson et al. 1999; Delworth and Mann 2000; Goosse et al., 2002). However, there are strong indications that they are chaotic and unpredictable, at least on a time scale longer than the fluctuation itself. The lack of long-term predictability is also indicated by the results of ensemble integrations with a coupled atmosphere–ocean model (Delworth and Knutson 2000).

To attribute a particular forcing mechanism with an observed pattern of climate change is hardly feasible, since the pattern forcing and the pattern of response are essentially uncorrelated (Hansen et al. 1997; Bengtsson 2001). The forcing by CO₂, for example, is largest in the Tropics but the largest surface warming occurs at higher latitudes. The same is true for solar forcing. Characteristic for all of the models used in a Coupled Model
Intercomparison Project (CMIP) intercomparison study (Räisänen 2002) was a maximum warming in the Arctic, a modest warming in the Tropics, and a minimum warming at the higher latitudes of the Southern Hemisphere. In fact, in the study reported by Bengtsson (2001, his Fig 7) the actual forcing was negative over parts of the Northern Hemisphere (greenhouse gases and sulfate aerosols, 1950–90) but the actual warming was among the largest in these areas. Similarly, the forcing was positive for the Southern Hemisphere (practically only greenhouse gases), yet here the warming was the smallest. The approach generally applied to study natural and anthropogenic variability is one of pattern recognition between forced coupled models and control experiments without forcing (Hasselmann 1997; Hegerl et al. 1997). This is probably a possible approach in cases of more substantial forcing, and when global patterns are considered. However, it requires that the climate models are capable of reproducing characteristic internal patterns of the climate system. If models underestimate such fluctuations, observed patterns outside the range of such models could then incorrectly be ascribed to external forcing. Alternatively, if models overestimate internal low-frequency fluctuations the opposite will hold. It seems that for present models neither of these alternatives can be excluded.

In the case of regional anomalies an additional level of complexity arises. Here the role of atmospheric and ocean circulation anomalies must be considered. As demonstrated by Hurrell (1995), for example, the positive surface temperature anomaly of the Northern Hemisphere extratropics during 1981–95 strongly depended on the mode of circulation of the North Atlantic Oscillation (NAO) and Southern Oscillation (SO). The NAO in particular has a marked stochastic variability, although there are indications from numerical experiments that SST anomalies, mainly in the tropical oceans, may influence the probability distribution function of the NAO (Hoerling et al. 2001; Schneider et al. 2003) or that the NAO can be forced through changes in the stratosphere (e.g., Shindell et al. 2001).

However, the atmospheric response to forcing from ocean SST anomalies is not particularly robust, as has recently been demonstrated by Schneider et al. (2003). The North Atlantic–European–Arctic region is strongly exposed to stirring by transient synoptic atmospheric eddies, and has a less robust response to forcing from SST anomalies than for example the North Pacific–North American region (Schneider et al. 2003). The large chaotic atmospheric variability in the North Europe–Arctic region is also indicated by the CMIP intercomparison (Räisänen 2002), and by the large differences between the five members of the ensemble simulation study reported by Delworth and Knutson (2000). The differences are particularly large in the Arctic region where they extend over several decades. An issue to be addressed in this paper is a possible climate feedback mechanism that can contribute toward long-lasting climate anomalies in the Arctic.

Here we will provide in a semiquantitative way using observations and model simulations a possible explanation of the high-latitude warming in 1930–40. During the first and second decade of the last century there was a high proportion of years with stronger-than-normal westerly circulation over the North Atlantic (NAO in a positive phase; Hurrell 1995). As has been shown by Curry and McCartney (2001), the long-term domination of such an atmospheric flow pattern drives the ocean circulation and results in the advection of warm water into the northeastern North Atlantic. The fact that the NAO was in a positive phase for several years could very likely be a consequence of chance along the lines pointed out by Wunsch (1999), where it was shown that longer episodes of either positive or negative NAO anomalies could occur because of aggregation of random atmospheric events.

Despite recent attempts to make the NAO the key driver for explaining Arctic climate variations (e.g., Mortitz et al. 2002), the NAO cannot explain why the Arctic rapidly started to warm up from 1920 onward, since in fact the NAO at the same time had a reverse trend (and should instead have supported an Arctic cooling!) and remained close to the average climate state for several decades (without any longer periods of either high or low values; see also Fig. 8b). Instead, and much more indisputable, we propose here that the warming was caused by the steadily increasing transport of warm water into the Barents Sea driven by increasing south-westerly to westerly winds between Spitsbergen and the northernmost Norwegian coast. Between 1920 and 1940 the observed pressure gradient increased by some 8 hPa corresponding to an average geostrophic wind anomaly of 6 m s\(^{-1}\). This led to increased transport of warm water into the Barents Sea, with a major reduction of sea ice in this region, where the largest atmospheric temperature anomalies also occur. As we will further demonstrate using model simulations, the reduced sea ice coverage particularly in the Barents Sea is the main reason for the increased Arctic temperature. This is in line with results of Ikeda (1990), who found Barents Sea ice cover changes due to varying oceanic inflow to be an important modulator of the decadal climate oscillations in the Arctic. A close link between observed sea ice and temperature variability has also been established by a long-term sea ice analysis (Mysak et al. 1990; Zakharov 1997; Johannessen et al. 2004), supporting the model simulations.

We will describe the early twentieth-century Arctic warming pattern in section 2, and in section 3 and 4 discuss analyses of numerical experiments with atmospheric and coupled ocean–atmosphere models, respectively. In section 5, we outline a positive feedback mechanism. In section 6, we undertake a comprehensive discussion of the results and possible consequences for climate change in the Arctic.
2. Arctic warming pattern

Temperature observations in the Arctic and adjacent regions include a number of records from land-based weather stations, several of them extending back to the nineteenth century (Przybylak 2000, for review). Since the 1950s data for the interior Arctic have been collected primarily by manned drifting polar stations, buoys, and dropsondes (Kahl et al. 1993; Martin et al. 1997; Rigor et al. 2000). Because of different temporal–spatial coverage and measurement techniques, the analyses of these data usually embrace only recent decades and sometimes lead to contradictory conclusions about magnitude and direction of the Arctic temperature trends (Serreze et al. 2000; Kahl et al. 1993; Polyakov et al. 2002).

Global gridded surface air temperatures (SAT; Jones et al. 1999; Hansen et al. 1999), which have been widely used for climate change and variability studies, have major gaps over the highest northern latitudes, in particular over the ice-covered ocean areas. This complicates an adequate analysis of the SAT spatial–temporal variability in the Arctic during the twentieth century, especially for the first half of the century.

Here, we use a century-long gridded Russian SAT dataset, which is more complete than the Jones data in the high latitudes (Alekseev and Svyaschennikov 1991; Johannessen et al. 2004). The dataset was compiled in the Arctic and Antarctic Research Institute, St. Petersburg. It consists of monthly mean values on $5^\circ \times 10^\circ$ latitude–longitude resolution and comprises land and drifting station meteorological observations. The land stations used to compile the gridded data were the same as used by Polyakov et al. (2003a,b). The gridded dataset covers the extratropical part of the Northern Hemisphere for 1892–1999.

The spatial pattern of the wintertime [November–April (NDJFMA)] warming is presented in Fig. 2 as the anomaly (relative to the long-term mean, 1892–1998) of the SAT averaged over the period 1935–44 (around the maximum warming time). As can be seen, the strongest warming (more than 2°C) occurred in the Kara and Barents Seas and (to a smaller extent) in Baffin Bay. Significant warming (exceeding 1°C) covered the interior Arctic, eastern Siberia, and the northern part of the Greenland Sea. This pattern is rather robust, which is confirmed by the analysis of Arctic SAT trends (Johannessen et al. 2004) and EOF decomposition of the wintertime SAT variability in the Arctic (Semenov and Bengtsson 2003). A very similar pattern was identified as the first EOF of the annual SAT variability in the Arctic (60°–85°N) for 1881–1980 (Kelly et al. 1982). The EOF decomposition of the SAT dataset (1892–1998) (not shown) also reveals a variability mode with maximum in the Kara and Barents Seas, which is found to be strongly related to the averaged Arctic temperature variability (Semenov and Bengtsson 2003).

The pattern shown in Fig. 2 is different from the temperature changes associated with large-scale atmospheric circulation variability patterns, such as the NAO and SO (Hurrell 1996). At the same time, the regions of the largest warming are located in (or close to) the areas of the highest variability of the wintertime sea ice concentration (on interannual to interdecadal time scales) in the Barents and Greenland Seas (Venegas and Mysak 2000; Deser et al. 2000). A 40–50-yr variability signal consistent with the SAT variations was identified by Venegas and Mysak (2000) in the Arctic sea ice concentrations and sea level pressure during the twentieth century. Thus, it is reasonable to assume that the warming pattern in the Barents–Greenland Seas reflects SAT variability associated with the wintertime sea ice changes.

There are several indications that the Arctic sea ice was reduced during the period of the early twentieth-century warming. Reconstructed annual mean values of the sea ice cover in the Barents Sea based on the May–August observations (Zakharov 1997) and August sea ice extent observations in the eastern Arctic (Polyakov et al. 2003a) support this. Furthermore a correlation between the observed (detrended) wintertime SAT over the Barents Sea and March Barents Sea ice extent is $r = 0.62$ for the 1953–98 period when the reliable sea ice data became available (Walsh and Johnson 1979; Chapman and Walsh 1993). Correlations between Arctic sea ice area ($S_{\text{ice}}$) and SAT and sensitivity of the SAT to ice area changes ($T/S_{\text{ice}}$) are shown in Table 1. Additional support for a negative correlation between sea ice cover and temperature was found in the numerical experiments.

3. Atmospheric modeling experiments

In order to analyze the sensitivity of the Arctic SAT to sea ice changes, we used ensemble simulations with
A correlation with Zakharov’s data is not significant at the 90% level because of high autocorrelation of the sea ice time series.

Changes and simulated (ensemble mean) SAT response and low ones (1954–83). Corresponding sea ice extent

3a. Two 30-yr periods have been chosen for the com-

sensitivity study because we found that the spatial
distributions of ice extent changes and their typical sea-

lonal variations features before and after 1949 were in-

distribution of ice extent changes and their typical sea-

2 m for every ice-covered grid. Sea ice albedo was

house gases concentrations were included. The sea ice
temperature representative for the upper 10-cm layer is

obtained from the total heat flux at the ice surface and

the heat conduction through the ice slab. The ocean

temperature below the slab is fixed at the freezing point

for the seawater, and the ice slab thickness is taken as

2 m for every ice-covered grid. Sea ice albedo was

specified as a linear function of surface temperature. An

ensemble of four global simulations using the Global

Sea Ice and Sea Surface Temperature dataset (GISST2.2) SST/sea ice extent analysis for 1903–94

(Rayner et al. 1996) as boundary conditions was carried

out. The experiments started from slightly different ini-
tial atmospheric states, but all had identical surface

boundary conditions. The large-scale Northern Hemi-

sphere circulation varied considerably among the four

experiments. There is a discontinuity in the GISST2.2

sea ice data in 1949 (prior to 1949 only climatological

sea ice data were used) due to introduction of a new

procedure for sea ice concentration derivation (Rayner

et al. 1996). This resulted in a sudden sea ice decrease

of some 2 Mkm$^2$ (Fig. 3a). This is, of course, incorrect

but has here been used as suitable boundary conditions

for a sensitivity study because we found that the spatial
distribution of ice extent changes and their typical sea-

onal variations features before and after 1949 were in-
general similar to the observed change of 1978–99 (Par-
kinson et al. 1999), in particular with the largest changes

in the Barents Sea.

Time series of the wintertime (NDJFMA) Arctic SAT

for ensemble mean and individual ensemble members

and the corresponding sea ice area are shown in Fig.

3a. Two 30-yr periods have been chosen for the com-

parison representing high sea ice area values (1910–39)

and low ones (1954–83). Corresponding sea ice extent

changes and simulated (ensemble mean) SAT response

are shown in Figs. 3b and 3c, respectively. The results

of the GISST simulations clearly show a strong response

of the Arctic SAT to the imposed sea ice retreat in 1949

and following variability. The artificial 1949 sea ice
decrease of 1.8 Mkm$^2$ resulted in about 2.0°C wintertime

warming, (Fig. 3a). As can be seen from the ensemble

members’ results, the area-averaged SAT increase is ob-

viously exceeding the differences between the members

of the ensemble. The warming is generally confined to

the Arctic with largest changes (exceeding 6°C) in the

Barents Sea (Fig. 3b). SAT changes north of 60°N are

statistically significant (exceeding two standard intraen-

semble deviations). The warming pattern in general fol-

lows the ice extent changes, including strong warming

in the Greenland Sea similar to the observed warming

pattern (Fig. 2). The summer-half [May–October

(MJJASO)] SAT changes were much lower (0.9°C and

2.3 Mkm$^2$, respectively) and exhibited even some SAT

decrease around the Pole in June. An overall comparison

of the observed warming pattern and simulated SAT

changes (Figs. 2 and 3b) reveals a close similarity. This

suggests corresponding sea ice changes as a reason for

the observed temperature variability. A quasi-stationary

sensitivity of the Arctic temperature change to the sea

ice (difference between two 30-yr-mean climates) re-

sults in $-0.67$ (−1.13)°C Mkm$^{-2}$ for annual (win-
tertime) means. This is similar to the observed sensitivity

(Table 1). The model sensitivity (ensemble average) for

the 1951–94 period is lower (one of the experiments

produced higher sensitivity values of −0.91 for the win-
tertime).

The simulated temperature increase in the areas of

reduced ice extent is related to stronger oceanic heat

flux to the atmosphere. This is illustrated in Figs. 4a,b,

where the ice extent and surface turbulent heat loss (la-
tent and sensible for the ensemble mean) changes

(1954–83 minus 1910–39) are presented (DJF values).

The increased heat flux, exceeded 150 W m$^{-2}$ in

the areas of the strongest ice extent decrease and about 20

W m$^{-2}$ for the Barents Sea average (corresponding to

40% sea ice decrease), generated an atmospheric cir-
culation response. Sea level pressure (SLP) and wind

<table>
<thead>
<tr>
<th>Observations</th>
<th>Correlation</th>
<th>$T/S_{\text{ice}}$ °C (Mkm$^2$)$^{-1}$</th>
</tr>
</thead>
<tbody>
<tr>
<td>Chapman and Walsh (1993), 1953–98</td>
<td>−0.60 (90%)</td>
<td>−0.98</td>
</tr>
<tr>
<td>Zakharov (1997), 1900–93</td>
<td>−0.55</td>
<td>−1.44</td>
</tr>
<tr>
<td><strong>Model results</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>ECHAM4/GISST*</td>
<td></td>
<td></td>
</tr>
<tr>
<td>ECHAM4/GISST 1951–94 (ensemble mean)</td>
<td>−0.48 (90%)</td>
<td>−0.67</td>
</tr>
<tr>
<td>ECHAM4/OPYC3 (200 yr)</td>
<td>−0.68 (99%)</td>
<td>−1.40</td>
</tr>
</tbody>
</table>

* These sensitivities are based on the ensemble mean changes for (1955–83) − (1910–39).
anomalies (ensemble means) are shown in Figs. 4c and 4d, respectively. One can clearly see a cyclonic vortex and a negative pressure anomaly in the northeastern part of the Barents Sea associated with the surface heat source to the west to the Novaya Zemlya. The wind anomalies imply an advection of a relatively warm air from the southern Barents Sea to the Kara Sea and may also result in an increased wind-driven oceanic inflow through a western opening of the Barents Sea (not shown). This suggests a potential positive feedback mechanism, to be discussed in section 5. Such a positive feedback was earlier suggested as an amplifier of decadal oscillations in the Arctic (Ikeda 1990; Mysak and Venegas 1998). However, in our ensemble simulations the effect of the reduced sea ice on the SLP was local, for example, no statistically significant SLP decrease was found in the central Arctic.

4. Coupled atmosphere–ocean experiments

The model experiments described in section 3 only involved the atmospheric component of the climate system. In order to evaluate the role of ocean–sea ice–atmosphere feedbacks, and to search for a similar warming mechanism, we analyzed a control 300-yr simulation with a coupled climate model ECHAM4/OPYC3 (Roeckner et al. 1999). The coupled model uses the same atmospheric model as was used in the sensitivity experiments described in the previous section. The concentrations of the greenhouse gases were set to the observed 1990 values. The model was run for 300 yr with flux adjustment. Details of the coupling procedure and flux adjustment fields can be found in Bacher (1998). Arctic sea ice area drifts for the first 100 yr of the experiment, and so the analysis is restricted to the last 200 yr of the simulation.
The model reproduces rather well the annual mean (9.8 Mkm$^2$) and interannual variability (0.22 Mkm$^2$) of Arctic sea ice area in comparison with the observations (Chapman and Walsh, 1993), which show a steep downward trend during the last 30 yr with the sea ice area of about 10.5 Mkm$^2$ in the 1990s and the standard deviation of 0.24 Mkm$^2$ (after subtracting the trend for the 1953–98 period). The regions of the highest wintertime sea ice variations are located along the sea ice boundary in the Atlantic sector of the Arctic, with maximum variability in the Barents Sea. The model simulates warming events in high latitudes that resemble the observed early twentieth-century warming, having comparable amplitude and spatial distribution, although of a shorter duration (about 25 yr). The simulated Arctic SAT (60$^\circ$–90$^\circ$N) variability and sea ice area values (annual means) are shown in Fig. 5a. The correlation between these time series is $-0.77$ (5-yr running mean) or $-0.70$ (annual mean), and the sensitivity of the annual SAT to the sea ice changes is about $-1.4^\circ$C Mkm$^{-2}$ (see Table 1). As can be seen from the correlations between annual mean Arctic sea ice and SAT (Fig. 5b), the strongest temperature changes associated with sea ice cover variability are located in the Greenland–Barents–Kara Seas with a maximum in the Barents Sea. This pattern is rather similar to the observed warming pattern (Fig. 2) and the SAT changes simulated by the atmospheric model (Fig. 3b).

The sea ice cover variability in the Arctic on interannual to decadal time scales is mainly determined by atmospheric conditions (Deser et al. 2000; Venegas and Mysak 2000), through driving ocean currents that carry relatively warm and saline waters into the Arctic, and through wind-driven sea ice circulation. Particularly sensitive to the advective heat flux is the shallow (about 250 m) Barents Sea, where the highest variability of the wintertime sea ice coverage is found (both observed and simulated). An analysis of the coupled model results suggests that the variation of the Barents Sea ice cover is determined by the oceanic volume inflow from the west. This is demonstrated by Fig. 6, which shows simulated annual mean volume inflow to the Barents Sea (VIB) (in the upper 125 m) through the Spitsbergen–Norwegian cross section (about 20$^\circ$E) and Arctic sea ice area (a correlation between the oceanic volume inflow and heat transport is 0.95 in the model). The figure demonstrates the role of this inflow for the sea ice variability. The correlation is $-0.65$ for the time series presented in the figure (wintertime, 5-yr running mean) and $-0.57$ for annual means. The correlation is the strongest with zero lag. An additional contributing factor could be the forcing of the sea ice in the Barents Sea by the stronger-than-normal winds. This is likely to contribute to enhanced surface heat flux in turn to be balanced by oceanic heat convergence. Several studies based on observational data (Dickson et al. 2000; Furevik 2001) and model results (Loeng et al. 1997) have demonstrated that the VIB variation is linked to the wintertime atmospheric circulation. As was found in the coupled model simulation, the changes of the VIB are...
Fig. 5. (a) Wintertime Arctic SAT anomalies (60°–90°N) (°C) and sea ice area (Mkm²) as simulated by the ECHAM4/OPYC3 coupled GCM in a control experiment (5-yr running mean)—the correlation between the two curves is −0.77; (b) correlation between averaged wintertime Arctic sea ice coverage and local surface air temperatures.

Fig. 6. ECHAM4/OPYC3 control simulation: wintertime Arctic sea ice area (Mkm²) and oceanic volume inflow (upper 125 m) into the Barents Sea (Sv), 5-yr running mean. The correlation between these time series is −0.65.
related to the corresponding variability of the SLP gradient over the western Barents Sea opening. This gradient, represented by the SLP difference between Spitsbergen and the northeastermmost Norwegian coast, is proportional to the strength of geostrophic winds, which drive the surface current. The VIB and SLP gradient are shown in Fig. 7. The strongest correlation between them (found with 1-yr SLP gradient lead) is 0.42 for 5-yr running means and 0.36 for annual data. Several periods can be seen in which decadal variations of the inflow and SLP gradient are particularly well related. A regression pattern of the SLP anomalies associated with VIB (not shown) shows an SLP dipole with maximum in northwestern Russia–Scandinavia and minimum stretched in the Greenland Sea. Some resemblance can be found between this pattern and one of the major atmospheric variability patterns in the high latitudes, the Barents oscillation (Skeie 2000). A weak correlation (0.28) was also found between the VIB and the index of the meridional overturning in the North Atlantic.

5. Mechanism of the early twentieth-century warming

A characteristic feature of the high-latitude winter circulation in the 1930s and 1940s was the strong southwesternerly to westerly flow through the passage between Spitsbergen and northern Norway. As can be seen from Fig. 8 the strength of this flow, as deduced from the SLP difference (Trenberth and Paolino 1980), increased gradually during the 1920s by 6 m s\(^{-1}\) averaged for the winter season. As can further be seen, the Arctic surface temperature is highly related to the intensity of the geostrophic flow essentially for the whole century except for individual years of extreme flows. This suggests that it is the multiyear strength of the atmospheric flow into the Barents Sea that results in a regional warming that provides the main contribution to the averaged Arctic SAT anomalies. It is interesting to note that the NAO in the period 1920–50 is uncorrelated to the Arctic temperature changes. This is not surprising since the NAO, as generally defined, represents the average strength of the eastern Atlantic westerly flow in the region between 65°–40°N.

Observations as well as modeling results show that the Barents Sea region has by far the largest interannual surface temperature variance and is the region that gives a major contribution to the temperature variations of the Arctic as a whole (Fig. 2). It is also a region that is directly influenced by the net atmosphere and ocean heat transport into the Arctic. Observational studies (Dickson et al. 2000) show a clear relation between the wind-driven volume flux and the temperature in the Barents Sea for the last 40 yr. Before that time there is no clear indication of a correlation between the Kola ocean temperature series and the Barents Sea inflow index (R. Dickson 2003, personal communication). However, as has been shown by Toresen and Østvedt (2000), the low-frequency temperature variations in the Kola section agrees well with the Barents Sea inflow index. The temperature variations, in turn, are strongly correlated with herring spawning biomass index (Toresen and Østvedt 2000, their Fig. 3). Atmospheric experiments, discussed in section 3, show a distinct sensitivity of surface air temperature to the extension of sea ice. We therefore hypothesize that the main cause of the Arctic warming was the increase of ocean and atmosphere heat transport into the Barents Sea, leading to reduced sea ice coverage and increased surface temperatures. Reconstructions of the annual mean Arctic sea ice cover based on the May–September observations (Zakharov 1997) and August observations in the marginal Arctic seas (Polyakov et al. 2003a) show a significant reduction of the sea ice extent during the twentieth-century warming episode.

Surface heat fluxes during the cold season in the Barents Sea are large (the area is a pronounced winter heat
FIG. 8. (a) The DJF SLP difference between Spitsbergen and the northernmost Norwegian coast (hPa, vertical bars) and annual mean Arctic SAT anomalies (°C, 5-yr running mean, solid line); (b) the NAO index and SAT anomalies as in (a).

FIG. 9. Correlation between Arctic (60°–90°N area averaged) wintertime SAT anomalies and DJF SLP for the 1920–70 period.

source in the Arctic) with substantial interannual variations (Simonsen and Haugan 1996). Such an enhanced heat source will generate a vorticity source in the lower troposphere with lower surface pressure associated with the vorticity maximum. This effect was considered in a number of studies, for example, Mysak and Venegas (1998) and Deser et al. (2000). The circulation changes will maintain the southwesterly flow into the Barents Sea, and thus the favorable conditions for the heat source (Fig. 4). The correlation pattern between observed sea level pressure (Trenberth and Paolino 1980) and Arctic surface temperature for the period 1920–70 (Fig. 9) is consistent with this hypothesis. The result is underpinned by four atmospheric model experiments described in section 2. The sea ice data, GISST2.2 (Rayner et al. 1996), were later found to be discontinuous, since the dataset only included climatological sea ice data for the period before 1950. However, as has been shown, they were found useful to demonstrate the strong sensitivity of the Arctic temperature to sea ice variation. In all four experiments, a corresponding mechanism is suggested with lower surface pressure and enhanced positive vorticity in the experiments with reduced sea ice. It is further interesting to note that the mechanism
is robust since all four experiments are very similar in this respect.

6. Conclusions

The Arctic 1920–40 warming is one of the most puzzling climate anomalies of the twentieth century. Over a period of some 15 yr the Arctic warmed by 1.7°C and remained warm for more than a decade. This is a warming in the region comparable in magnitude to what is to be expected as a consequence of anthropogenic climate change in the next several decades. A gradual cooling commenced in the late 1940s bringing the temperature back to much lower values although not as cold as before the warming started. Here, we have shown that this warming was associated with and presumably initiated by a major increase in the westerly to southwestierly wind north of Norway leading to enhanced atmospheric and ocean heat transport from the comparatively warm North Atlantic Current through the passage between northern Norway and Spitsbergen into the Barents Sea. It should be stressed that the increased winds were not related to the NAO, which in fact weakened during the 1920s and remained weak for the whole period of the warm Arctic anomaly. We have shown that the process behind the warming was most likely reduced sea ice cover, mainly in the Barents Sea. This is not an unexpected finding because of the climatic effect of sea ice in comparison with that of an open sea but is intriguing since previously available sea ice data (Chapman and Walsh 1993) did not indicate a reduced sea ice cover in the 1930s and 1940s. However, as we have shown here recent sea ice datasets [Johannessen et al. (2004) give a detailed presentation] actually showed a retreat in this period. Experiments with an atmospheric model forced with different sea ice datasets as well as examination of a coupled model integration are in a broad agreement with the observational data, suggesting a 1°C warming for a reduction of the Arctic sea ice of 1 Mkm$^2$.

An evaluation of the coupled model suggests that a major part of the warming is caused by transport of warm ocean water, in the uppermost 125 m of the ocean model, into the Barents Sea, driven by stronger-than-normal surface winds.

The question as to the cause of the enhanced westerly winds is an open question and we can here only offer possible explanations. It has been shown (Ikeda 1990; Mysak et al. 1990; Mysak and Venegas 1998; Ikeda et al. 2001; Goosse et al. 2002) that decadal-scale internal oscillations are possible in the Arctic with the oscillation period determined by a characteristic time scale of the oceanic or sea ice circulation variability. These mechanisms require a dynamical response of the oceanic/sea ice circulation to the atmospheric circulation (responding relatively fast to changes of the surface conditions), which produces a necessary delay to create a decadal oscillation loop. It is likely that proposed mechanisms may also be responsible for the multidecadal variations such as the early twentieth-century warming. However, from available climate reconstructions of the Arctic SAT (Overpeck et al. 1997), it is hard to identify distinct oscillatory variations with 50–70-yr time scale. We suggest that the warm Arctic event might have occurred through an aggregation of several consecutive winters with a pronounced high-latitude westerly in the Atlantic sector. High-latitude warm events are possible as has been demonstrated by Delworth and Knutson (2000) and have been generated during the course of a long integration with a coupled model. An important factor in setting up a long-lasting event is the indication, both in observations and in the modeling studies, of a dynamical feedback mechanism related to the generation of an atmospheric heat source in the Barents Sea and an associated creation of a cyclonic circulation. Such a circulation will, as we have demonstrated, act to maintain the westerly-to-southwesterly atmospheric flow into the region and the associated heat transport. Alternatively, if sea ice has become more extensive in the Barents Sea any tendency to westerly inflow would be weakened, acting to prolong the conditions for an extensive sea ice cover.

Needless to say, a necessary condition for the Arctic warming event to commence depends on changes in the large-scale atmospheric circulation. A comprehensive discussion of this is outside the scope of this study. As discussed in the introduction, there are many possibilities. Our view is that natural processes in the climate system are the most likely cause; at least, there is hardly any information from observations or from model experiments to the opposite. In fact we have reasons to believe that the natural variability of the ECHAM4/OPYC3 model is underestimated. The latest MPI coupled model experiment (without flux correction) (M. Latif 2003, personal communication) has stronger multidecadal anomalies. We further believe that anthropogenic effects are unlikely because of the modest forcing in the early part of the last century and if it were to be the case the climate feedback would be much stronger than estimated and inconsistent with present observations. Solar forcing cannot be excluded as a possible hypothesis (Lean and Rind 1998) but suffers from the uncomfortable fact that it is not supported by direct observations.

Observational data in the Arctic from the first part of the twentieth century are relatively sparse, especially from the Arctic Ocean. The gridded datasets of surface temperature and sea ice are essentially constructed from a limited number of observations, although the autocorrelation structure used in the analyses are calculated from present data. However, observations from the most sensitive region, Barents Sea, are fairly good and better that previously thought, which means that we can have reasonable confidence in the result.

What consequences may the findings of this study have for understanding the possible evolution of the
Arctic climate? Notwithstanding an expected overall climate warming it is suggested that the Arctic climate would be exposed to considerable internal variations over several years initiated by stochastic variations of the high-latitude atmospheric circulation and subsequently enhanced and maintained by sea ice feedback. The Barents Sea region is identified as a particularly sensitive area in this respect. Realistic simulation of the Arctic climate consequently requires an accurate representation of atmosphere–ocean–sea ice feedback processes.

Acknowledgments. We are thankful to Noel Keenlyside for many useful comments on the text. This study was supported by funding from the European Union within the Arctic Ice Cover Simulation Experiment project (EVK2-CT-2000-00078 AICESEX). One of the authors (O. M. Johannessen) was supported by the Research Council of Norway, projects Marine Climate and Ecosystem in the Seasonal Ice Zone (MACESIZ) and Role of the Arctic Sea Ice–Atmosphere Processes (ROLSARC).

REFERENCES


Toresen, R., and O. J. Óstvedt, 2000: Variation in abundance of Norwegian spring-spawning herring (Clupea harengus, Clupeidae) throughout the 20th century and the influence of climatic fluctuations. Fish Fish., 1, 231–256.


