SweClim
SWEDISH REGIONAL CLIMATE
MODELLING PROGRAMME

SMHI

a) CCC b) CERFACS c) CSIRO
d) GFDL e) GISS f) LMD/IPSL
g) MPI h) MRI i) NCAR−CSM
j) NCAR−WM k) NRL l) UKMO

CO_2-induced climate change in northern Europe: comparison of 12 CMIP2 experiments

Jouni Räisänen
Rossby Centre, SMHI

CO$_2$-induced climate change in northern Europe: comparison of 12 CMIP2 experiments

Jouni Räisänen
Rossby Centre, SMHI

Report Summary / Rapportsammanfattning

Issuing Agency/Utgivare	Report number/Publikation
Swedish Meteorological and Hydrological Institute S-601 76 NORRKÖPING Sweden	RMK No. 87
	Report date/Utgivningsdatum
	January 2000

Author (s)/Författare

Jouni Räisänen

Title (and Subtitle/Titel

CO$_2$-induced climate change in northern Europe: comparison of 12 CMIP2 experiments

Abstract/Sammandrag

The results of 12 coupled atmosphere-ocean general circulation model experiments participating in the second phase of the Coupled Model Intercomparison Project (CMIP2) are studied with focus on the area of northern Europe. The variables considered are surface air temperature, precipitation and sea level pressure. The 80-year control simulations are first compared with observational estimates of the present climate. Several aspects of the simulated CO$_2$-induced climate changes, defined by subtracting the control run seasonal or annual means from 20-year perturbation run means around the transient doubling of CO$_2$, are then studied. The common features and individual variations in the simulated climate change are documented. Particular attention is put on expressing the interexperiment agreement in quantitative terms and on estimating the relative contribution of model-simulated internal variability to the interexperiment variance. For that purpose, a new statistical framework is developed. Finally, an attempt is made to statistically relate the interexperiment differences in the simulated climate change in northern Europe to aspects of the control climates, global climate change and some of the basic model characteristics.

A summary of the main findings is given in the last section of the report.

Key words/sök-, nyckelord

Climate change, climate model, CMIP2, intercomparison, internal variability

Supplementary notes/Tillägg	Number of pages/Antal sidor	Language/Språk
This work is a part of the SWECLIM programme and of the Coupled Model Intercomparison Project	59	English

ISSN and title/ISSN och titel

0347-2116 SMHI Reports Meteorology Climatology

Report available from/Rapporten kan köpas från:

SMHI
S-601 76 NORRKÖPING
Sweden

Contents

1 Introduction

The possibility of substantial climate change during the next century, primarily as a result of anthropogenic increases in CO_2 and other so-called greenhouse gases, is of great international concern (e.g., Houghton et al. 1996). Estimating the magnitude and, in several cases, even the direction of the forthcoming changes is unfortunately complicated by several sources of uncertainty. One is the fact that even our most sophisticated tools for studying climate change, coupled general circulation models of atmosphere and ocean (AOGCMs), necessarily contain a large number of approximations. Even when forced by similar changes in external conditions, different models simulate different climate changes. This is the case in particular on the regional scale that is most important for predicting the practical impacts of climate change (Grotch and MacCracken 1991; Räisänen 1997; Kittel et al. 1998 (hereafter KGM98)).

To understand, quantify, and ultimately reduce the model-related uncertainty, it is of vital importance to systematically intercompare climate change experiments made with different models. A recently established effort for such intercomparison is the second phase of the Coupled Model Intercomparison Project CMIP2 (Meehl et al. 1997; 1999). CMIP2 is a "level 2" (Gates 1992) intercomparison in which all the model experiments share the same forcing, a gradual (1% per year compound) increase in atmospheric CO_2.

The present study compares the results of 12 CMIP2 experiments in the area of northern Europe. It can be seen as continuation to an earlier intercomparison made for the same area by Räisänen (1994; hereafter R94), but it uses a wider set of generally more recent experiments. In addition, some effort will be put in expressing the intercomparison results in a quantitative form. How good or bad is the agreement between the different experiments, and to which extent may the interexperiment differences in climate change be explained by internal variability ("noise") in the model simulations? To achieve a solid view on these issues, some new statistics will be introduced which are likely to be useful in other studies as well.

As indicated above, the main focus of this study are the simulated CO_2-induced climate changes. As in R94, however, the ability of the models to simulate the present observed climate in northern Europe will also be investigated. Apart from serving as an update to R94, this gives background information for the analysis of climate changes. The occasionally large interexperiment differences in the simulated climate response call for explanations, and differences in control climate are one candidate.

The design of the CMIP2 experiments, the models participating in this comparison and some technical issues related to the processing of the model data in the present study are described in the following section. In Section 3, a look is taken at the ability of the models to simulate the present observed climate in northern Europe. The following three sections focus on the simulated CO_2-induced climate changes in the 12 experiments used for this study. Section 4 documents the changes in seasonal and annual mean surface air temperature, precipitation and sea level pressure in these

experiments. In Section 5, the interexperiment agreement is quantified in a statistical sense, and the relative roles of model differences and internal variability in explaining the differences in the simulated climate response are estimated. To put the results obtained in the computations for northern Europe in a wider perspective, similar statistics are also calculated in a global domain. In Section 6, connections are searched between the interexperiment differences in the simulated climate changes in northern Europe (with focus on area mean temperature and precipitation) and other factors including the control climates, the simulated global climate changes and basic model characteristics. A summary of the main findings with some further discussion is given in Section 7.

2   The CMIP2 data set

The model experiments that participate in the CMIP2 intercomparison follow a standard setup. Each of them consists of a control run with constant ("present-day") atmospheric CO_2 and a perturbation run in which the CO_2 concentration starts from the control run level and then increases 1% per year compound, doubling in 70 years. All these models belong to the same basic category, coupled general circulation models of atmosphere and ocean (AOGCMs), but there are also several important differences between them. These include the horizontal and vertical resolution of the atmospheric and oceanic components, the basic numerical scheme (grid box or spectral), various parameterization methods, and possible use of flux corrections.

A brief summary of the 12 CMIP2 models used in the present study is given in Table 1. The experiments are abbreviated according to the respective modelling centres; see Table 2 for explanation and key references. At the time of finalising this report, five additional models had just joined the CMIP2 comparison. The results of these new models will be included in a journal article planned to follow this report.

The control and the perturbation runs in the CMIP2 experiments both last for 80 simulated years, with two exceptions. In the NCAR-WM experiment both runs only last 75 years, and the NRL control run is only three years long. The present study is based on the time series of surface air temperature, precipitation and sea level pressure which are available at monthly resolution but were stratified to 5-year seasonal and annual means in order to reduce the computations. For quantitative grid box level intercomparison, the results of all other models were interpolated horizontally to the UKMO model grid (2.5° lat x 3.75° lon). This choice was motivated by the use of another UKMO experiment in the Swedish Regional Climate Modelling Programme SWECLIM.

The control climates (Section 3) are in general calculated as seasonal or annual means over the whole (usually 80-year) duration of the control runs. The NRL control run, which is only 3 years long, is complemented by the first 17 years of the perturbation run to reduce the effects of interannual variability. What is called control climate is therefore in this case slightly affected by the initial increase in CO_2 during the early part of the perturbation run. The climate changes discussed in Sections 4-6 are calculated by

subtracting the control run means from the means over the years 61-80 in the perturbations runs (20 years around the doubling of CO_2). For NCAR-WM, with only 75 years of data, the 20-year perturbation run mean is formed by assuming that the missing five years (76-80) would have been identical with the previous five-year period.

Table 1. Characteristics of the CMIP2 models.

Model	Resolution (atmosphere)	Resolution (ocean)	Sea ice	Land surface	Flux corrections
CCC	T32 (3.8° × 3.8°) L10	1.9° × 1.9° L29	T	BBR	HW
CERFACS	T31 (3.8° × 3.8°) L19	1.5° × 2.0° L31	T	PC	None
CSIRO	R21 (3.2° × 5.6°) L9	3.2° × 5.6° L21	TR	PC	HWM
GFDL	R15 (4.5° × 7.5°) L9	4.5° × 3.7° L12	TF	BR	HW
GISS	4.0° × 5.0° L9	4.0° × 5.0° L13	T	PR	None
LMD/IPSL	3.6°s × 5.6° L15	2.0° × 2.0° L31	T	PC	None
MPI	T21 (5.6° × 5.6°) L19	4.0° × 4.0° L11	T	BRC	HWM
MRI	4.0° × 5.0° L15	2.0° × 2.5° L21	TF	PR	HW
NCAR-CSM	T42 (2.8° × 2.8°) L18	2.0° × 2.4° L45	TR	PC	None
NCAR-WM	R15 (4.5° × 7.5°) L9	1.0° × 1.0° L20	TR	B	None
NRL	T47 (2.5° × 2.5°) L18	1.0° × 2.0° L25	T?	BB	HW
UKMO	2.5° × 3.75° L19	2.5° × 3.75° L20	TF	BBRC	HW

Resolution (example): the CSIRO atmospheric model is a spectral model with rhomboidal truncation to maximum wave number 21 (roughly equivalent to a grid box size of 3.2° in latitude and 5.6° in longitude) and 9 vertical levels. **3.6°s** (LMD/IPSL) indicates a meridional discretization of 50 grid points evenly spaced in sine of latitude (i.e., coarser meridional spacing at high latitudes than near the equator).

Sea ice: **T** denotes simulated ice thermodynamics, **R** simulated ice dynamics with rheology and **F** advective ("free-drift") ice dynamics.

Land surface: **B** denotes the use of a standard bucket scheme in predicting soil moisture (single soil layer with constant moisture capacity), **BB** a modified bucket scheme (e.g., a single layer with spatially varying moisture capacity) and **P** some more physically based approach. **R** indicates inclusion of a routing model for treatment of runoff as as supply of freshwater to the ocean model, and **C** the inclusion of an explicit vegetation canopy for moisture interception and transpiration.

Flux corrections: **H** = heat, **W** = water, **M** = momentum. The corrections are temporally constant in NRL and vary seasonally in the other models.

Some of the results related to the control run means and simulated changes in temperature and precipitation will be represented as area means over a Northern Europe Land Area (NELA). This is defined as the 112 UKMO grid land grid boxes within 10°W-60E, 50°-75°N. At a few occasions, a "Nordic Subregion" will also be used. This subdomain of NELA only includes the 25 land grid boxes at least in part within the borders of Finland, Sweden, Norway and Denmark. NELA is the northern Europe domain used in R94 (and it only differs from the domain used by KGM98 by excluding

3

Iceland) whereas the Nordic Subregion is the main area of interest in the Swedish SWECLIM project. See Fig. 1 for the definition of these areas.

Table 2. *The CMIP2 modelling centres and (where appropriate) model versions, together with key references to the models.*

CCC	Canadian Centre for Climate, CGCM1 (Fyfe and Flato 1999)
CERFACS	Centre Européen de Recherche et de Formation Avancée en Calcul Scientifique (France), ARPEGE-OASIS-OPAICE (Barthelet et al. 1998)
CSIRO	Commonwealth Scientific and Industrial Research Organization (Australia), CSIRO Mk2 (Gordon and O'Farrell 1997; Hirst et al. 1999)
GFDL	Geophysical Fluid Dynamics Laboratory (U.S.A.) (Manabe et al. 1991)
GISS	Goddard Institute for Space Studies (U.S.A.), "Russel model" (Russell et al. 1995; Russell and Rind 1999)
LMD/IPSL	Laboratoire de Météorologie Dynamique / Institut Pierre Simon Laplace (France), LMD-OPA (Braconnot et al. 1997)
MPI	Max Planck Institute for Meteorology (Germany), ECHAM3-LSG (Voss et al. 1998)
MRI	Meteorological Research Institute (Japan) (Tokioka et al. 1995)
NCAR-CSM	National Center for Atmospheric Research (U.S.A.), Climate System Model (Boville and Gent 1998)
NCAR-WM	National Center for Atmospheric Research (U.S.A.), "Washington and Meehl model" (Meehl and Washington 1995; Washington and Meehl 1996)
NRL	Naval Research Laboratory (U.S.A.) (Li and Hogan 1999)
UKMO	United Kingdom Meteorological Office / Hadley Centre, HadCM2 (Johns 1996; Johns et al. 1997; Mitchell and Johns 1997)
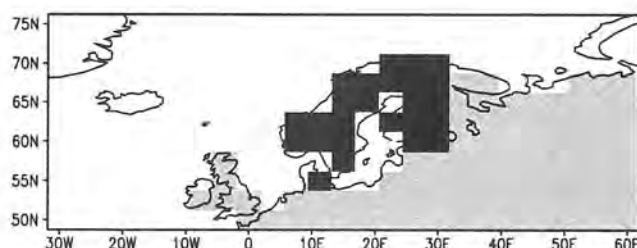
Figure 1. *The Northern Europe Land Area (all shaded grid boxes) and a subdomain of this, the Nordic Subregion (dark shading).*

3 Control climates

In this section the CMIP2 control simulations in northern Europe are discussed. The seasonal and annual means of surface air temperature, precipitation and sea level pressure are compared between the 12 models and observational estimates of the current climate. The distributions of present-day surface air temperature and precipitation are inferred from the CRU (Climatic Research Unit, University of East Anglia) land-area climatology for 1961-1990 (Hulme et al. 1995; New et al. 1999). The simulated sea level pressure is compared with ECMWF reanalyses for the period 1979-1993 (Gibson et al. 1997). In doing this, the simulated pressure values are corrected for biases in the global annual mean. This correction is −6.8 hPa for GFDL, 3.0 hPa for UKMO, 2.2 hPa for NRL, and within ±1.2 hPa for the other nine models. As absolute pressure values are of less meteorological interest than horizontal pressure differences, the corrected values give probably a more meaningful indication of where the simulated pressure should be considered too low or too high.

3.1 12-model mean control climates, intermodel variation and systematic differences from observational data

We first consider the group of the 12 CMIP2 control simulations as a whole and then take a closer look at the results of the individual models. Figs. 2, 3 and 4 give the 12-model means of surface air temperature, precipitation and sea level pressure, the intermodel standard deviation, and the difference of the 12-model mean from the observational estimate. Areas where the absolute difference between the 12-model mean and the observational estimate exceeds one intermodel standard deviation are also indicated.

The 12-model average annual mean temperatures, shown in the bottom row of Fig. 2, are mostly within 2°C of the observed values. The models have, however, a tendency to overestimate the meridional temperature gradient in the eastern part of the map domain in all seasons. The average simulated annual temperatures are slightly below those observed in Finland and northwestern Russia but slightly above the observed values further south in the former Soviet Union. The simulated temperatures are also somewhat too high in northern and western Scandinavia, especially in winter. This is likely to be associated with the relatively coarse resolution of the models, which allows the strongest influence of the warm Atlantic Ocean to extend further inland in the simulations than in reality. In addition, coarse resolution tends to smooth out and lower the Scandinavian mountains, although the magnitude of this effect depends on the choices made in presenting the orography.
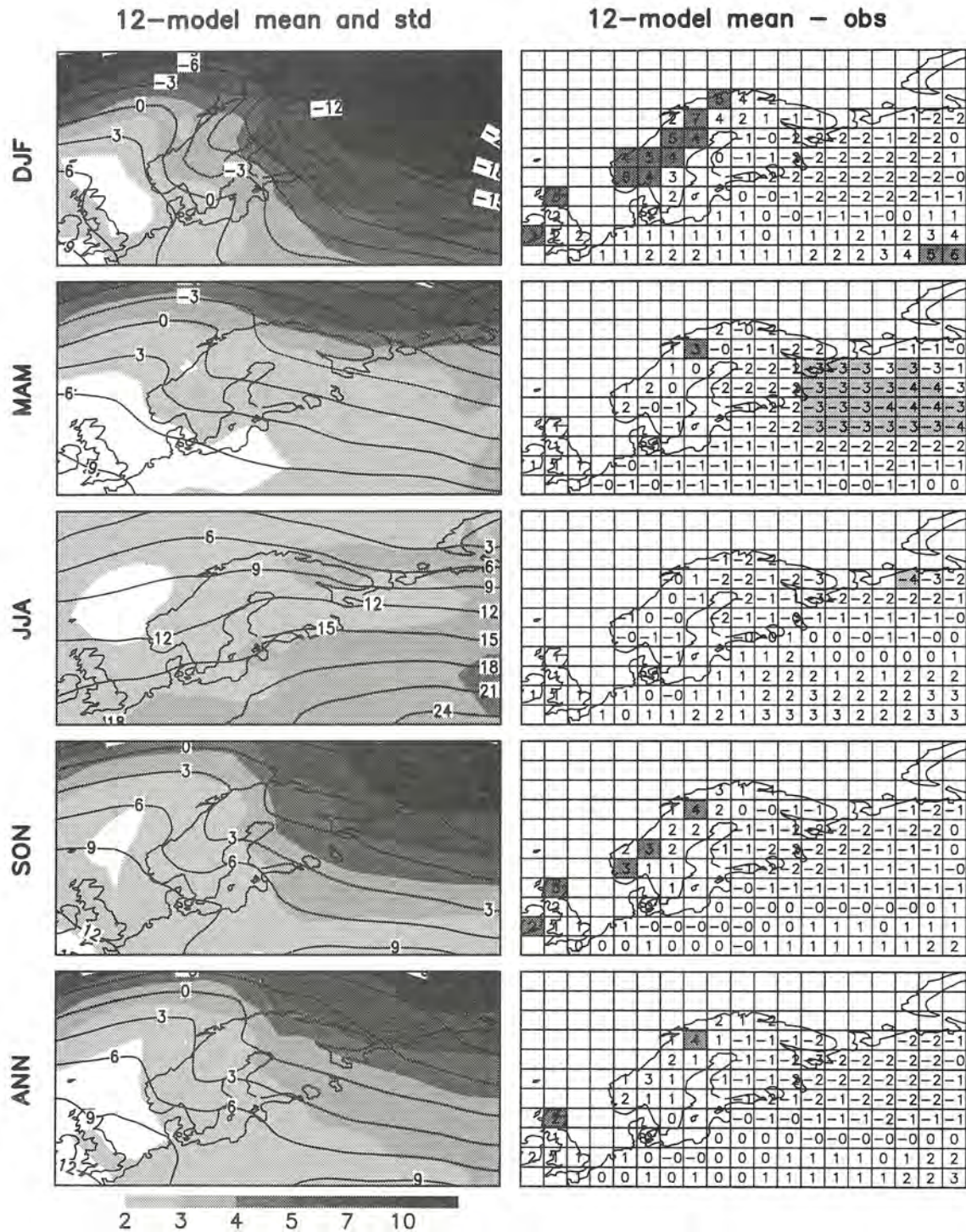
Figure 2. *12-model statistics for seasonal (DJF, MAM, JJA and SON) and annual (ANN) mean surface air temperatures in the control runs. The left column shows the 12-model mean (contours every 3°C) and the standard deviation between the models (shaded; scale below the bottom panel). The numbers in the right column give the difference between the 12-model mean and the CRU climatology for 1961-1990. Grid boxes in which the 12-model mean is at least one intermodel standard deviation above (below) the climatological estimate are shaded in dark (light).*

The 12-model annual NELA mean temperature bias is, incidentally, 0.0°C (Table 3 in the next subsection). A cold bias of -1.3°C in spring (MAM) is balanced by negligible or

6

slightly positive biases in the other seasons. The 12-model mean simulated temperatures in spring are below the observed values in almost the whole map domain, excluding Norway and the British Isles. In northwestern Russia, the cold bias is 3-4°C and exceeds the intermodel standard deviation in a wide continuous area. In the other seasons, this condition of systematic difference is met in only single grid boxes, with the exception of Scandinavia in winter. However, the scatter between the individual models is substantial. The standard deviation is below 2°C locally over the Atlantic Ocean in all seasons and in western Europe in spring, but it varies more commonly between 2 and 4°C. In winter and to some extent in autumn, considerably higher values occur in northwestern Russia and, in particular, over the Arctic Ocean.

The 12-model mean precipitation is generally larger than the CRU observational estimate (Fig. 3). The difference is relatively small and unsystematic in summer and autumn but more pronounced in the other seasons. It exceeds the intermodel standard deviation in wide areas in winter and, in particular, in spring. In spring, the all-model mean is locally by a factor of three larger than the observational estimate, although this only happens in the southeastern corner of the map domain where observed precipitation is small. The NELA mean difference between the 12-model mean and the climatology is 48% of the climatological mean in winter, 72% in spring, 9% in summer, 15% in autumn, and 31% in the annual mean (Table 3).

One most likely contributor to the large winter- and springtime discrepancy between the models and the observations are errors in the simulated atmospheric circulation. As will be further discussed below, the 12-model mean simulated sea level pressure in northern Europe in spring is too low, with largest negative bias to the south of 60°N. This indicates that the North Atlantic storm track takes a too zonal course in the models and vigorous cyclone activity thus extends too east and too south in continental northern Europe. The fact that the average model-simulated winter precipitation exceeds the observational estimate by a much larger factor in the southern part of northern Europe than in northwestern Russia is likewise consistent with a south-north gradient in the pressure bias.

The gap between the simulated and observed values of precipitation is, however, also affected by the well-known tendency of the observations to underestimate the actual precipitation. At Swedish stations, for example, the average annual mean precipitation in 1961-1990 was according to Raab and Vedin (1995) 14% larger than the measured precipitation. The corrected value includes measurement losses due to evaporation, adhesion and aerodynamic factors, but it is still more likely an under- than overestimate of the true precipitation (Hans Alexandersson, personal communication). Because the aerodynamic loss is largest for solid precipitation, the correction is much larger in midwinter (on the average 25-30%) than in summer (around 10%). Christensen et al. (1998) estimate the annual precipitation in several regions of Scandinavia as a residual in the water budget, combining measurements of river runoff with a model-based estimate of evaporation. According to their Table 8, the actual precipitation exceeded the CRU climatology by 20% in Denmark and by as much as 47% in northern Norway.
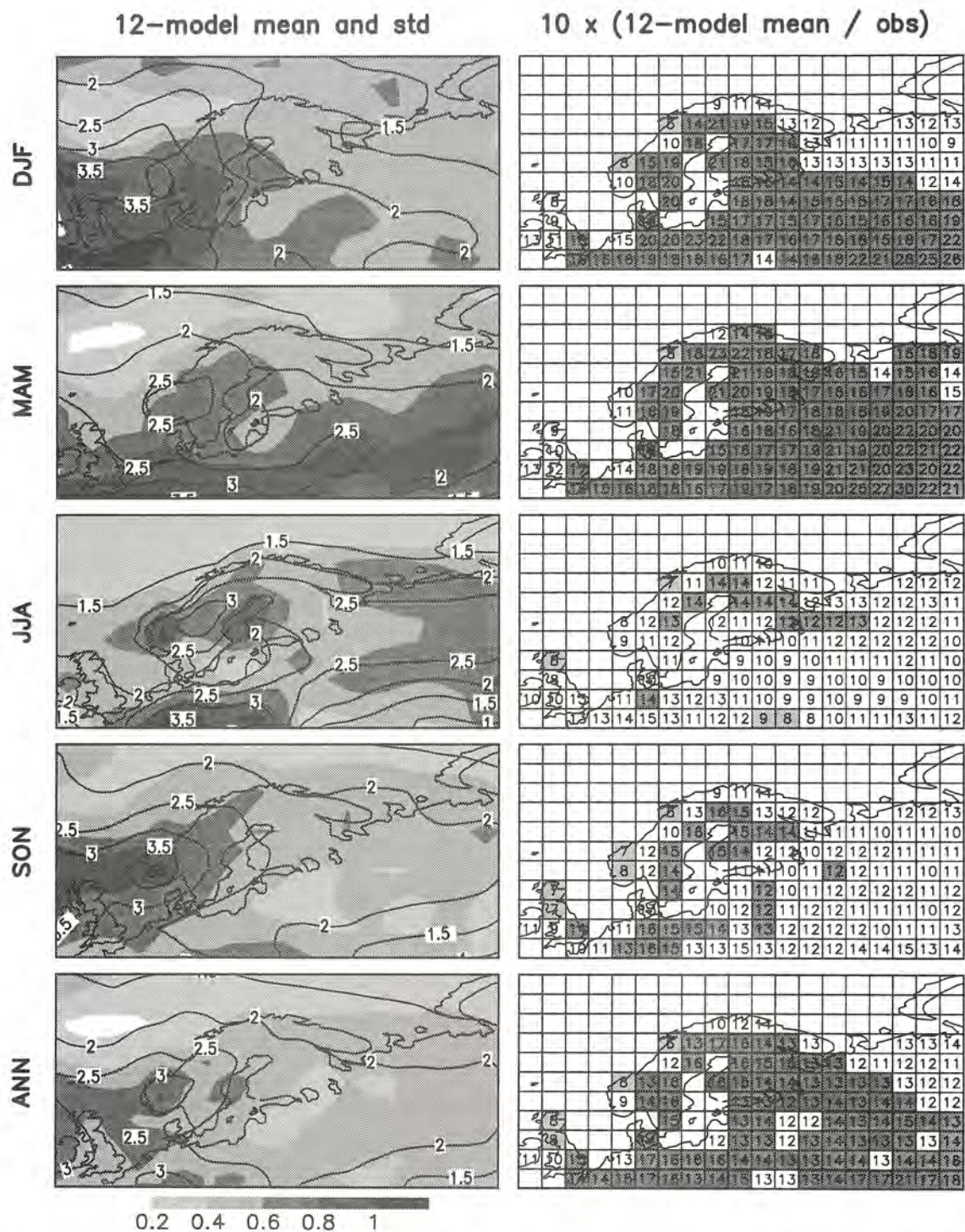
Figure 3. *12-model statistics for seasonal and annual mean precipitation in the control runs. The left column shows the 12-model mean (contours every 0.5 mm d⁻¹) and the standard deviation between the models (shaded; scale below the bottom panel). The numbers in the right column give 10 times the ratio between the 12-model mean and the CRU precipitation climatology for 1961-1990. Grid boxes in which the 12-model mean is at least one intermodel standard deviation above (below) the climatological estimate are shaded in dark (light).*

For a presumably less biased comparison with the model simulations, a corrected precipitation climatology was formed in an *ad hoc* manner. The corrected values were obtained by multiplying the CRU precipitation by the factor

$$f = \begin{cases} 1.1, & T > 5°C \\ 1.1 + 0.2 \times (5°C - T)/10°C, & -5°C < T < 5°C \\ 1.3, & T < -5°C \end{cases} \qquad (1)$$

where T is the monthly mean temperature from CRU. The lower limit of the correction (10%) follows the mentioned Swedish estimate for summer, whereas the upper limit of 30% is a (probably) conservative estimate of measurement errors for solid precipitation. The implicit assumption that the proportion of liquid precipitation changes linearly from 0 to 1 in the mean temperature interval of 5°C to -5°C is of course simplistic. Several factors that affect the actual measurement error (e.g., that different rain gauges have been used in different countries) are neglected. While roughly mimicking the corrections used in Raab and Vedin (1995) for Swedish stations, this subjectively chosen correction is smaller than the corrections implied by the results of Christensen et al. (1998). On the other hand, the routinely reported precipitation data in the former Soviet Union have since 1966 been corrected for the adhesion loss (Groisman et al. 1991), and this correction (typically 5-15%) is thus present even in the CRU climatology for this area (New et al. 1999).

The resulting correction in the NELA mean precipitation is 16% in the annual mean, ranging between 10% in summer and 27% in winter. If these values are realistic (which is not known), the modest positive area mean differences between the 12-model mean and the CRU climatology in summer and autumn are not indicative of model errors. In winter as well, (1) attributes over a half of the difference between the 12-model mean and the original climatology to measurement losses, even though the upper limit of the correction (30%) may well be too modest. However, the positive bias in spring is far too large to be explained by measurement losses alone.

As expected, the models do not properly capture sharp local maxima of precipitation associated with mountain ranges. Hence, in contrast with the general positive bias relative to the uncorrected climatology, the 12-model mean precipitation at the west coast of Norway and in Scotland is somewhat below the observed values. The lack of orographic precipitation in these mountain areas likely increases the positive precipitation bias immediately to the east of them.

The 12-model mean distribution of sea level pressure (Fig. 4) correctly reproduces the observed seasonal cycle. The near-surface circulation in winter is dominated by an intense Icelandic low and an Eurasian high, whereas the average pressure pattern in summer is much flatter. Nevertheless, there are quantitative differences between the 12-model mean and the climatological pressure field. In spite of a substantial variation between the models, some of these differences are relatively systematic, especially in winter and in spring.

9
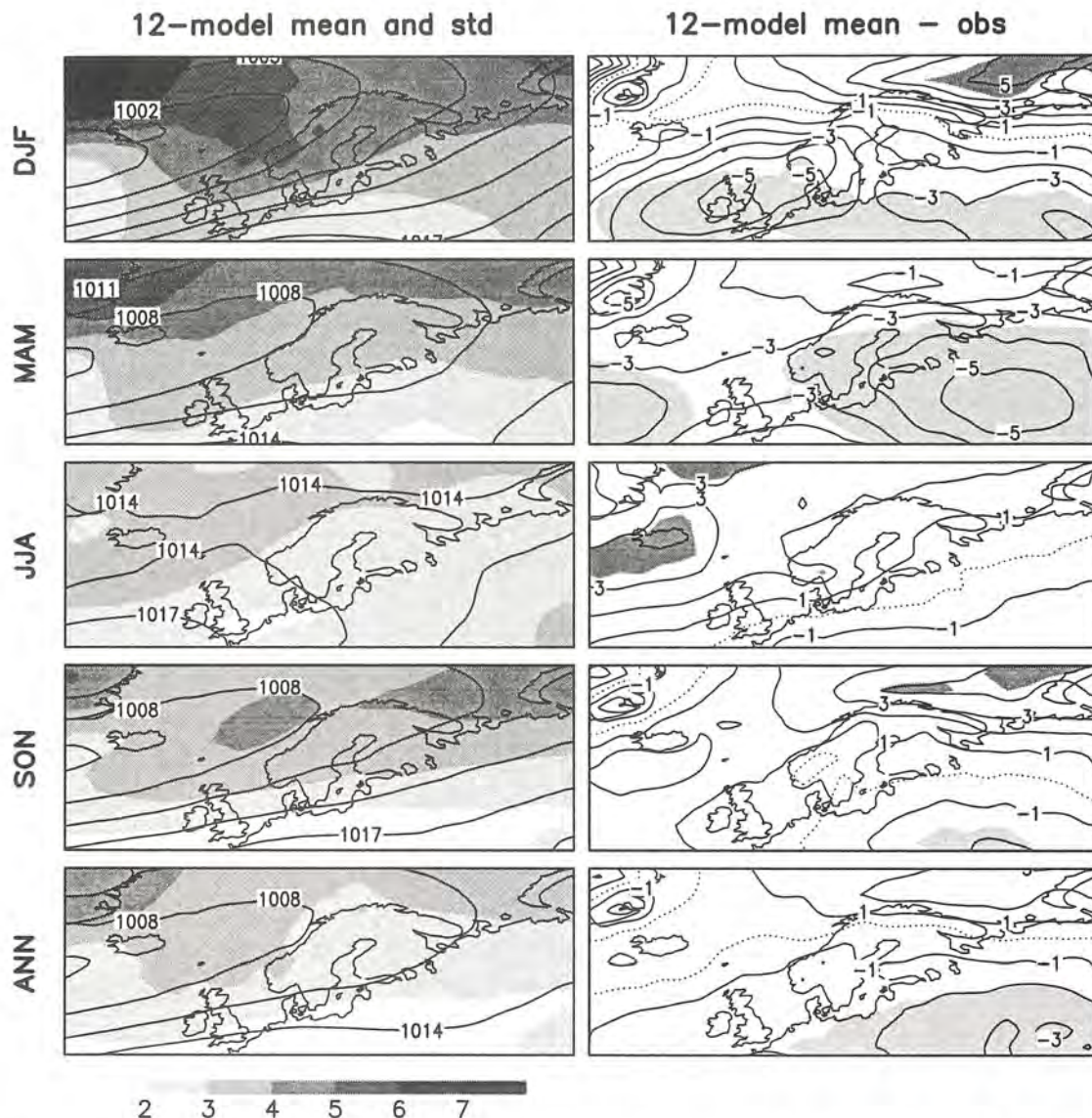
Figure 4. *12-model statistics for seasonal and annual mean sea level pressure in the control runs. The left column shows the 12-model mean (contours every 3 hPa) and the standard deviation between the models (shaded; scale below the bottom panel). In the right column, the difference between the 12-model mean and the ERA reanalysis for 1979-1993 is contoured at every 1 hPa. Areas where the 12-model mean is at least one intermodel standard deviation above (below) the reanalysis are shaded in dark (light). The simulated pressure values are corrected for biases in the global annual mean (see text).*

In winter, the 12-model mean meridional pressure gradient across northern Europe is too small, with too low (high) pressure in the southern part of the map domain (over the Arctic Ocean). The extension of the simulated Icelandic low towards northern Europe lacks, on the average, a sufficient northeastward tilt. In spring, the average simulated pressure is below the climatological values in the whole map domain. As in winter, however, the largest negative bias occurs in the southern part of northern Europe, where its magnitude locally exceeds the intermodel standard deviation by a factor of two.

10

The differences between the models and the climatology in summer and autumn are less systematic than in winter and spring, but the 12-model mean bias shows qualitatively the same meridional pattern even in these seasons. In both summer and autumn, the simulated pressure on the average exceeds the climatological values over the Arctic Ocean and the northern North Atlantic but is slightly too low in the southeastern part of northern Europe.

3.2 Control climates in individual models

Some results concerning the control climates in the 12 individual models are given in Table 3 (biases of the NELA mean temperature and precipitation relative to the uncorrected CRU climatology) and in Figs. 5 and 6 (geographical distributions of temperature, precipitation and sea level pressure in winter and in summer). The warmest model in the annual area mean is NCAR-WM (bias +5.3°C) and the coldest LMD/IPSL (-5.9°C). These two models also have the largest and smallest area mean precipitation (biases relative to the uncorrected climatology 90% and 6%, respectively). Overall, warm models tend to be wet models and vice versa. The cross correlation between the annual area means of temperature and precipitation is 0.84, which is significant at the 99% level. The correlation in the individual seasons varies between 0.31 in JJA and 0.82 in MAM.

Table 3. *Differences between the simulated values of NELA mean temperature and precipitation and the CRU climatology for 1961-1990. Temperature differences are given in °C and differences in precipitation in per cent of the climatological area mean. The climatological area means are shown in the second last row (precipitation is uncorrected and given in units of mm day⁻¹). The correction in climatological precipitation suggested by (1) is given in per cent in the last row.*

	Temperature					Precipitation				
	DJF	MAM	JJA	SON	ANN	DJF	MAM	JJA	SON	ANN
CCC	3.2	-0.2	1.0	2.7	1.6	54	78	34	26	45
CERFACS	6.4	1.4	-0.6	1.1	2.0	110	101	12	35	57
CSIRO	0.0	-2.7	0.7	0.1	-0.5	19	35	20	5	18
GFDL	-0.9	-3.2	0.1	-1.6	-1.4	39	56	-15	10	17
GISS	0.6	-1.8	-1.9	0.2	-0.8	38	85	14	4	30
LMD/IPSL	-10.8	-4.7	-0.3	-7.6	-5.9	4	26	15	-16	6
MPI	-1.6	-3.1	-2.2	-1.2	-2.0	50	29	-16	11	14
MRI	3.9	1.6	3.4	1.9	2.7	46	93	8	17	35
NCAR-CSM	-2.2	-3.1	-3.2	-2.2	-2.7	30	41	9	16	21
NCAR-WM	5.4	2.1	8.1	5.6	5.3	130	206	15	63	90
NRL	4.1	-0.7	1.3	2.6	1.8	25	74	25	-7	25
UKMO	0.1	-1.6	-1.2	-0.8	-0.9	30	37	-8	6	13
Mean	0.7	-1.3	0.4	0.0	0.0	48	72	9	14	31
St. dev.	4.4	2.1	2.9	3.1	2.8	35	48	15	20	22
CRU	-8.4	3.7	15.8	4.4	3.9	1.47	1.35	2.19	1.93	1.74
Corr. in P (%)						27	16	10	15	16

Among the warmest models in the annual area mean are, in addition to NCAR-WM, MRI, CERFACS, NRL and CCC; excluding NRL these all are within the four wettest models as well. The other seven models have annual mean cold biases of varying magnitude; the second largest underestimate in temperature after LMD/IPSL occurs in NCAR-CSM (-2.7°C). CSIRO, GISS and UKMO are within 1°C of the climatology. The simulated annual mean precipitation exceeds the uncorrected CRU climatology in all models, but this is not the case in comparison with our corrected values. Seven of the 12 models (CSIRO, GFDL, LMD/IPSL, NCAR-CSM, MPI, NRL and UKMO) are within 10% of the latter estimate.

Taking a look at the individual seasons, the intermodel variation of the NELA mean temperature is largest in winter. In particular, the negative area mean bias in LMD/IPSL exceeds 10°C, which mainly reflects a very large cold bias (up to over 20°C) in northern Russia; in the western part of northern Europe the bias is relatively small (see Fig. 5). The warmest model in this season is CERFACS. In spring, there is less variation between the individual models but a slight overall tendency towards too cold simulated temperatures. In summer, the simulated area mean temperatures are within 3.5°C of the CRU climatology in all models except for NCAR-WM, which has a warm bias of over 8°C. This warm bias is substantial in the whole of northern Europe, but it is largest (over 10°C) in the southern part of the area (Fig. 6). In contrast with the other seasons, the area mean temperature bias in LMD/IPSL in summer is close to zero, although this results from a compensation between warm and cold biases in the southern and northern parts of the domain. The area mean temperature biases in autumn are relatively small, except in LMD/IPSL and NCAR-WM.

The area mean precipitation in winter is largest (more than twice the uncorrected CRU climatology) in the two warmest models, NCAR-WM and CERFACS. The nominal area mean bias is marginally positive even in LMD/IPSL, but as expected from the very low simulated temperatures, there is much too little winter precipitation in this model in northern Russia. In spring, the area mean difference from the uncorrected climatology exceeds 25% in all models and 50% in seven of them, but it is by far the largest, over 200%, in NCAR-WM. In summer, the intermodel variation is relatively small, and, also in contrast with winter and spring, the simulated area mean is slightly below the climatological value in three models (GFDL, MPI and UKMO). The simulated area means are also below the climatology in two models (LMD/IPSL and NRL) in autumn. In one of these (LMD/IPSL), the negative precipitation bias appears to be linked to a large cold bias in temperature and to a substantially too high time mean sea level pressure over the northern part of northern Europe and the northern North Atlantic (not shown).

Some of the observations that one can make from the geographical distributions of winter- and summertime temperature, precipitation and sea level pressure in the individual models (Figs. 5 and 6) are the following:

12
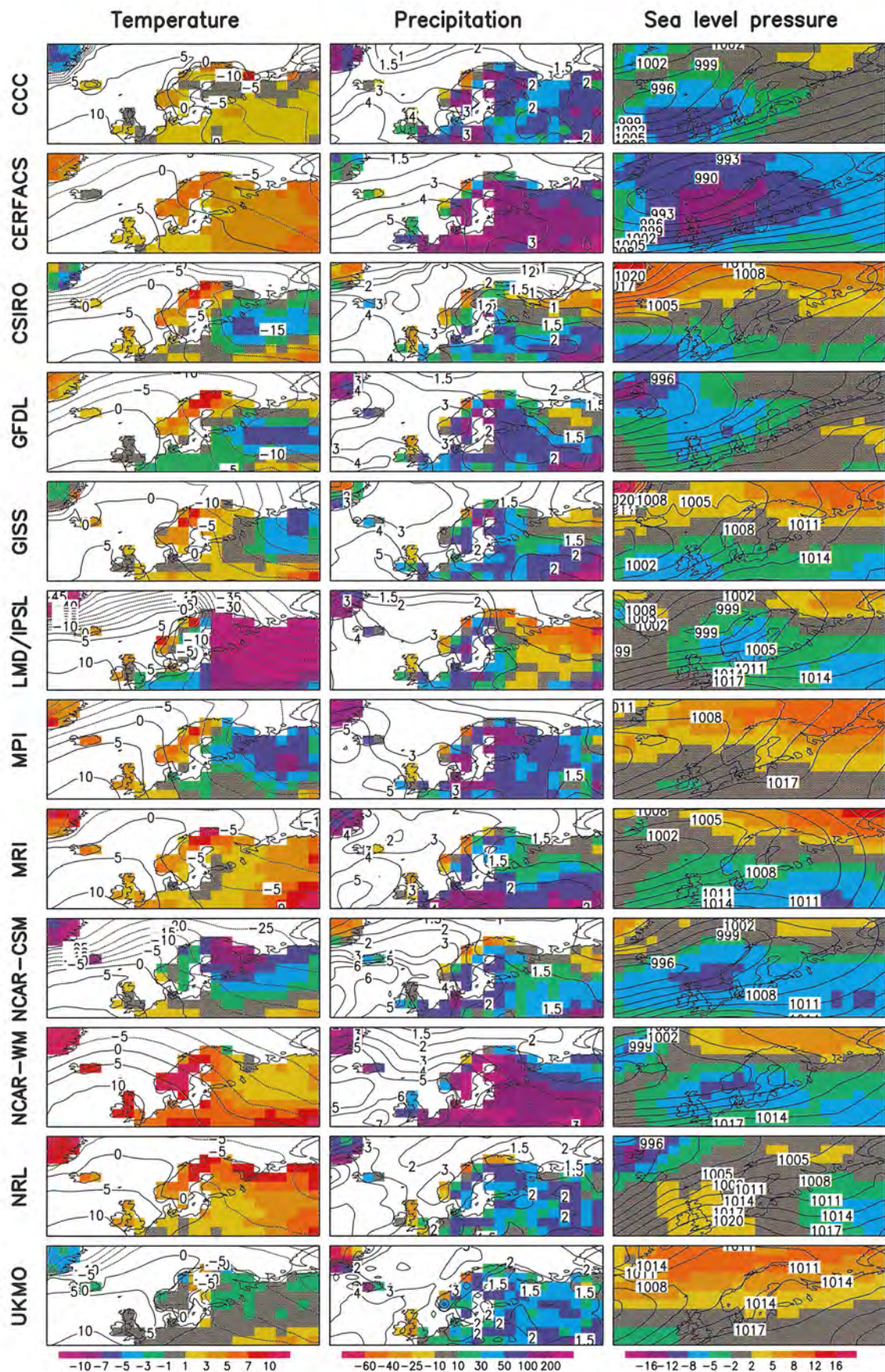
Figure 5. Control run December-February mean temperature (left), precipitation (middle) and sea level pressure (right) in the 12 CMIP2 simulations, and the differences of the simulated values from observational data. The simulated temperature is contoured at every 5°C, precipitation at 0.5, 1, 1.5, 2, 3, 4, 5, 6 and 7 mm/day, and sea level pressure at every 3 hPa. Sea level pressure is adjusted for global mean biases (see text). The differences from the observational estimates (temperature and precipitation: CRU; sea level pressure: ECMWF reanalysis) are indicated with colours (scale below the bottom panels; precipitation bias in %).
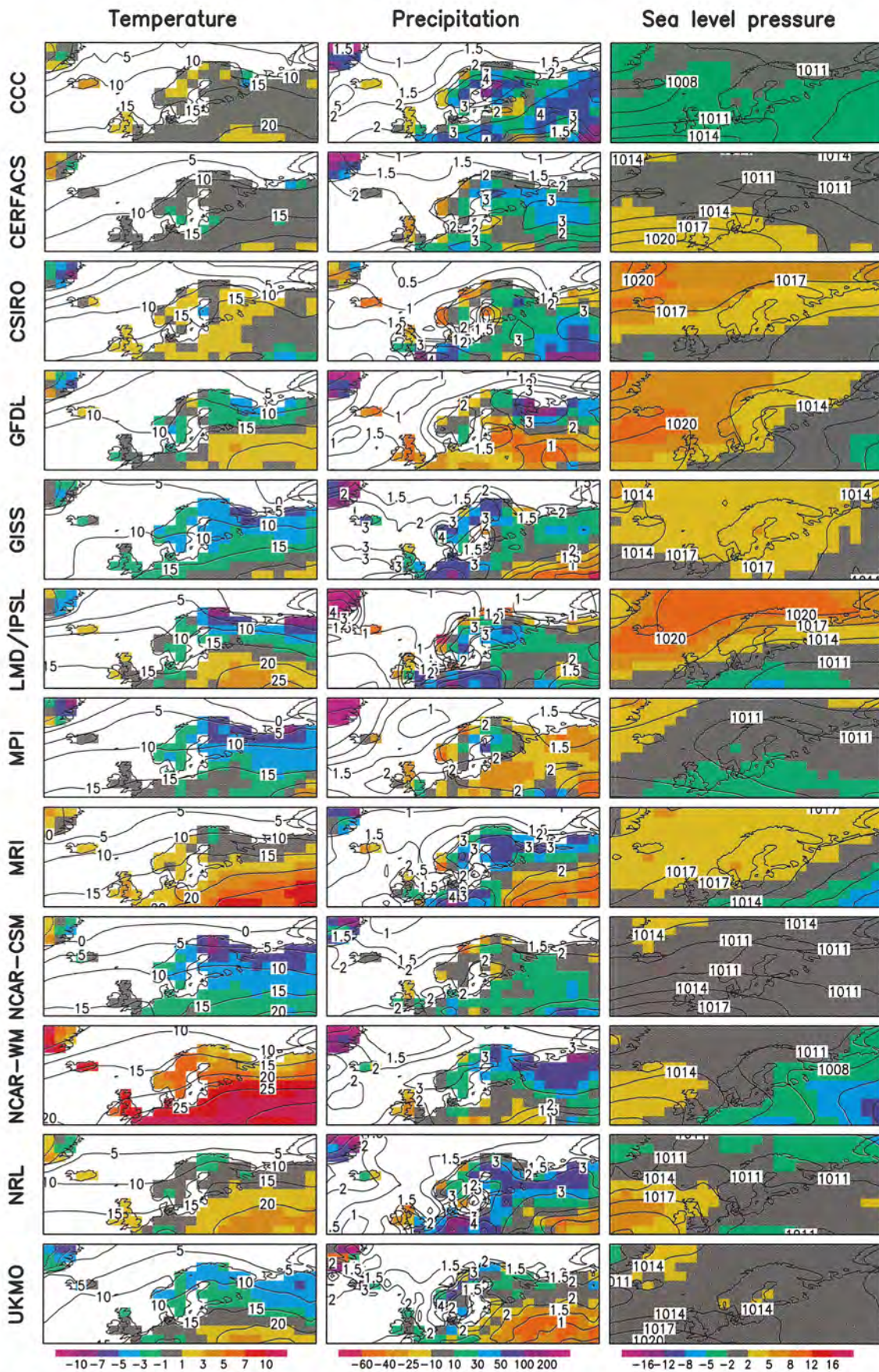
Figure 6. *As Fig. 5, but for June-August.*

- The biases in temperature and precipitation are not geographically constant across the whole of northern Europe but vary substantially in magnitude and, in several cases, even in sign between the different parts of the area.
- Some of the biases in temperature and precipitation are at least qualitatively consistent with the deficiencies in the lower tropospheric circulation indicated by

14

biases in sea level pressure. For example, the substantial warm bias and very abundant precipitation in CERFACS in winter are accompanied by a too deep and far too eastern Icelandic low, which indicates too strong advection of warm and moist air from the Atlantic Ocean to northern Europe.

- In other cases, deficiencies in circulation may act to reduce biases associated with other factors. For example, the wintertime mean sea level pressure over the Arctic Ocean and the northernmost north Atlantic Ocean in the UKMO simulation is too high, which indicates too weak time-mean westerly flow from the Atlantic Ocean to northern Europe. The simulated temperatures are close to those observed but, if the pressure distribution were correctly simulated, they would probably be somewhat too high.

- Although the intermodel correlation between area mean temperature and precipitation biases is positive even in summer, the geographical distributions of temperature and precipitation biases in many individual models are anticorrelated in this season. In GFDL, MRI, NRL and UKMO, the simulated temperatures in the southeastern part of northern Europe exceed the observed values and coincide with a local negative bias in precipitation, whereas biases of the opposite sign occur further north. The warm bias in NCAR-WM covers the whole area but is largest in the southern part where precipitation is close or below the observed values, in contrast with too abundant precipitation in the north. Where too warm temperatures and too small precipitation coincide, they may be connected by a common factor such as underestimated cloudiness or overestimated drying out of soil.

- The relation of temperature and precipitation biases in summer to biases in sea level pressure appears vague. However, one may speculate that the cold bias in LMD/IPSL in the northernmost northern Europe is worsened by excessive northeasterly winds from the Arctic Ocean, where the simulated pressure is too high.

- In NCAR-WM and MRI, negative JJA pressure biases in the southeastern part of the map domain coincide with substantially too high temperatures and may be caused by the warm bias via the thermal low mechanism. However, both these models also overestimate the strength of the continental scale Asian monsoon low, and the local pressure biases in northern Europe are not necessarily independent from this.

Evident from the maps and tables presented here is that the 12 CMIP2 models simulate the current surface climate in northern Europe with varying degrees of skill. The relative performance of the models depends on the variable and season considered, but some of the intermodel differences are still quite systematic. An objective verification statistics indicates that, taking into account all seasons and all three variables, the surface climate in northern Europe is best reproduced in the UKMO control simulation (see Section 6.2). Other models that do well in this comparison include GISS, CSIRO and MPI. The largest control climate biases in northern Europe occur, in general, in the NCAR-WM and LMD/IPSL models. These two models are among those five that do not use flux corrections to improve the agreement with the observed climate, and they both also have a relatively low horizontal resolution (Table 1; note, in particular, that the meridional spacing of grid boxes in LMD/IPSL is coarsest at high latitudes). Excluding these two models, no clear relationship between control run quality and resolution or the use of flux corrections is apparent.

15

The success in the simulation of the present climate is frequently taken as an indication of a model's reliability in simulating climate changes. Screening out of models with bad control climate is therefore sometimes used as a first step before analysing the response of different models to, for example, increased CO_2 (e.g., Whetton et al. 1994). A difficulty with this approach is, however, that the threshold between acceptable and unacceptable control climate is ill-defined. In addition, the relative ranking of models may depend on the aspects of climate studied and the geographical area considered (the verification statistics used in Section 6.2 in fact yields some differences in ranking between northern Europe and a global domain). The analysis of climate changes in the next two sections therefore includes all 12 CMIP2 experiments. The sensitivity of the results to the set of models used will, however, be studied afterwards, in Section 6.2.

3.3 Comparison with Räisänen (1994)

In a similar regional intercomparison by R94, the control climates and CO_2-induced climate changes in five AOGCMs and two atmospheric GCMs with mixed-layer ocean were analysed. Of these models, GFDL (present in R94 both as an AOGCM and as a mixed-layer ocean version) also participates in CMIP2. The remaining models in R94 were earlier versions of the MPI (Cubasch et al. 1992; Lunkeit et al. 1996), NCAR (Washington and Meehl 1989) and UKMO (Mitchell et al. 1989; Murphy 1995) climate models. Regarding the control climates, two major differences between R94 and the present study are evident:

- All the models studied in R94 exhibited a cold bias in northern Europe winter mean temperature. The average area mean bias was about –6°C, in strong contrast with the marginally positive average bias found in the present study. In the cases in which different model versions from the same modelling centres participate in the two intercomparisons, the earlier versions studied in R94 invariably simulate a colder winter climate. In particular, the two UKMO models in R94 both showed a cold bias of -9° – -12°C (in contrast with the negligible bias in HadCM2), and the NCAR model version used in that study was likewise substantially colder (bias almost -9°C) than either of NCAR-CSM or NCAR-WM. In addition, the CMIP2 models from those modelling centres that were not presented in R94 simulate, with the exception of LMD/IPSL, winter temperatures well above (CCC, CERFCACS, MRI, NRL) or close (CSIRO, GISS) to those observed.
- Unlike in any of the 12 CMIP2 models, the northern Europe annual area mean precipitation in all three models with precipitation data available in R94 (the two MPI models and NCAR) was slightly below uncorrected observations. In particular, summer precipitation was underestimated by about 40% in one of these models and by as much as two thirds in the other two.

Winter temperatures and summer precipitation are generally better simulated in the CMIP2 control runs than in the experiments analysed in R94. In many other aspects of the simulated climate (for example, summer temperatures and winter precipitation), however, the distinction between the two sets of control runs is much less clear.

16

4 Climate changes induced by transient doubling of CO_2

In this section, the simulated CO_2-induced changes in surface air temperature, precipitation and sea level pressure in northern Europe are described. Climate change is defined, with the exceptions specified in Section 2, as the seasonal or annual mean difference between the years 61-80 in the perturbation run (20 years around the doubling of CO_2) and the whole 80-year control run.

A potential disadvantage of using the whole 80-year means from the control runs is the existence of climate drift (spurious trend-like behaviour) in some of these (see Appendix B). As far as the transient increase in CO_2 can be regarded as a small perturbation, similar trends are expected to be superimposed on the true CO_2-induced climate change in the perturbation runs. From this point of view, it might be a better choice to define the control climate using the same period (years 61-80) that was used in the perturbation runs (climate change definition 2 of Cubasch et al. 1992). We prefer to use the whole 80-year control runs because (I) symptoms of climate drift in northern Europe are generally weak, and (II) the noise caused by internal variability decreases with increasing averaging period. The latter is an advantage, in particular, in defining regional details of simulated climate change, especially regarding such highly variable parameters as precipitation and sea level pressure. On the area mean scale (whole northern Europe), the choice of period in defining the control climate has some impact on the results of individual models but little impact on the general conclusions.

The impact of internal variability on the inferred climate changes could be further reduced by using a longer than 20-year average even in defining the perturbed climate. However, because the CMIP2 simulations are only 80 years long, this could only be achieved by moving the beginning of the averaging period backwards in time. The perturbation run averaging period would then be no longer centred at the doubling of CO_2, which would complicate comparison with other studies.

4.1 Surface air temperature

The simulated annual mean temperature changes in the 12 experiments[1] are shown in Fig. 7. There are variations in both the magnitude and the patterns of the simulated temperature change, but excluding GISS, all models indicate some annual mean warming in the whole NELA. In most cases, the warming is between 1° and 4°C at least in most of this domain. There are three exceptions:

- In NCAR-WM, the warming is generally larger than 4°C. This model also exhibits the largest global mean warming, almost a factor of two larger than any of the other models (see Table 4).
- In NRL, the warming is below 1°C. This model is also the one with by far the smallest global mean warming (see Table 4). The fact that our definition of the NRL control climate includes the first 17 years of the perturbation run (see Section 2) only explains a minor part of the contrast between this experiment and the others.

---

[1] The word "experiment" is preferred over "model" since the climate change simulated by any single model in response to the same forcing depends, in its details, in a chaotic manner on the initial conditions.

- In GISS, the eastern and southern parts of NELA warm by 1-2°C. By contrast, marginal cooling occurs in northern and western Scandinavia and over the northern British Isles, reflecting a more substantial decrease in temperature over the surrounding sea areas.

The cooling of air over the northern North Atlantic in the GISS model is caused by a cooling of the surface water, and the maximum in this over the Arctic Ocean to the north of Finland (5°C in the annual mean and 8°C in winter) is due to a resulting local expansion of sea ice. Russell and Rind (1999) attribute the decrease in sea surface temperatures to a 30% weakening of the North Atlantic thermohaline circulation. However, there is no definite explanation to why such cooling takes place in the GISS model but not in other AOGCMs, although a similar CO_2-induced weakening of the North Atlantic thermohaline circulation has been reported for many other models as well (e.g., Kattenberg et al. 1996, p. 304). Russell and Rind (1999) note, though, that the GISS AOGCM does not incorporate explicit horizontal diffusion in the ocean, unlike the present AOGCMs in general. This and the use of a sophisticated advection scheme allow for sharper gradients of temperature and salinity and may, therefore, help to create a more localized but larger reduction in the oceanic heat flux convergence. However, a more recent version of the GISS model with improvements in the treatment of sea ice and oceanic vertical mixing (Russell et al. 1999) gives in this respect more conventional results. Even in this model version, some greenhouse gas induced cooling is simulated over a relatively large part of the northern North Atlantic, but this is weaker than the cooling in the CMIP2 model. The cooling no longer extends to land areas of northern Europe, although warming is still weak in western Scandinavia and the British Isles.

The seasonal variation of the simulated temperature change in the 12 models is characterized in Table 4 by area means calculated over NELA and the Nordic Subregion shown in Fig. 1. The 12-model average annual area mean warming in NELA is 2.5°C, with slightly stronger warming in winter (3.0°C) than in summer (2.2°C). The average annual mean temperature increase in the Nordic Subregion is marginally smaller but the seasonal contrast between winter and summer is similar. However, the seasonal cycle of the temperature change varies somewhat between the individual experiments. In the whole NELA, winter is the season with largest warming in only six models and summer the season with smallest warming in seven models, and both of these are true only for GFDL, MPI and NCAR-CSM. In both this area and the Nordic Subregion, the largest warming in CSIRO and UKMO occurs in spring, and that in CERFACS and MRI in autumn. In NCAR-WM, the warming averaged over NELA actually maximizes in summer, and its large magnitude (5.5°C) is in marked contrast with the results of the other 11 models for this season. In NRL, the simulated warming is weak throughout the year. In GISS, the annual mean temperature change in the Nordic Subregion is virtually zero, with small positive and negative changes in the individual seasons.
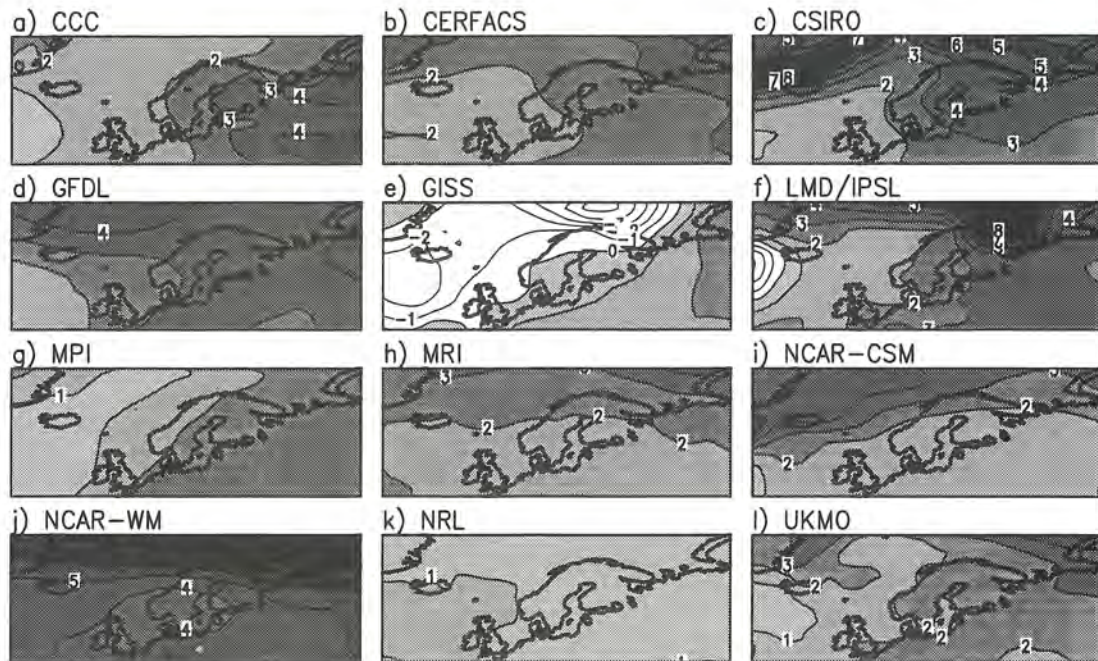
Figure 7. *Changes in annual mean temperature in the 12 experiments, defined as the difference between the perturbation run mean in years 61-80 and the control run mean in years 1-80 (see Section 2 for exceptions). Contours every 1°C. Positive values are shaded with darkest shading for largest warming.*

Table 4. *Area mean temperature changes (°C) in the 12 experiments. Shown are the global annual mean change and the seasonal and annual mean changes averaged over the Northern Europe Land Area and the Nordic Subregion depicted in Fig. 1.*

	Global	Northern Europe Land Area					Nordic Subregion				
	Annual	DJF	MAM	JJA	SON	Annual	DJF	MAM	JJA	SON	Annual
CCC	2.0	4.6	3.6	2.5	2.2	3.2	3.9	2.6	2.1	1.8	2.6
CERFACS	1.9	2.6	2.5	2.2	2.7	2.5	2.4	2.4	1.8	2.5	2.3
CSIRO	2.0	3.2	3.4	2.4	2.8	3.0	3.2	4.5	2.7	2.6	3.3
GFDL	2.2	3.6	3.2	3.1	3.4	3.3	3.5	3.2	2.9	3.5	3.3
GISS	1.7	1.4	0.9	1.0	1.3	1.1	-0.2	-0.2	0.3	0.2	0.0
LMD/IPSL	2.0	5.0	1.6	2.8	3.4	3.2	5.2	1.2	2.7	2.6	2.9
MPI	1.6	3.5	1.9	1.5	2.4	2.3	3.2	1.7	1.2	2.3	2.1
MRI	1.6	2.0	2.0	1.2	2.2	1.8	1.9	1.9	1.6	2.3	1.9
NCAR-CSM	1.4	1.9	1.4	1.3	1.7	1.6	2.4	1.6	1.2	1.7	1.7
NCAR-WM	4.0	5.1	3.8	5.5	4.3	4.7	5.2	4.3	5.0	3.5	4.5
NRL	0.8	0.6	0.6	0.7	0.7	0.6	0.4	0.7	0.7	0.5	0.6
UKMO	1.8	2.2	2.7	1.8	2.6	2.3	2.6	2.8	1.7	2.2	2.4
Mean	1.91	2.98	2.29	2.18	2.47	2.48	2.84	2.24	2.00	2.14	2.30
St. dev.	0.73	1.39	1.01	1.25	0.95	1.06	1.55	1.32	1.18	0.95	1.14

For comparison, Table 4 also includes the global annual mean warming. In both NELA and the Nordic Subregion, the simulated warming is larger than its global mean in all experiments apart from GISS and NRL. The 12-experiment mean warming in NELA

19

(the Nordic Subregion) exceeds the global mean by 30% (20%). Using the years 61-80 instead of the whole control runs for defining the control climate would have increased this difference to 34% (26%). The average global mean warming would have decreased from 1.91°C to 1.85°C, because many of the control runs exhibit a slight warming trend in the global mean temperature (see Appendix B). By contrast, the 12-experiment mean warming in NELA would have remained unchanged at 2.48°C and that in the Nordic Subregion would have increased marginally from 2.30°C to 2.34°C.

The horizontal distribution of the 12-experiment mean temperature change in different seasons and in the annual mean is shown in the left column of Fig. 8. The geographical variations are largest in winter, with a southwest-northeast gradient between smaller warming over the northern North Atlantic and larger warming in northern Russia and the Arctic Ocean. A similar but weaker gradient also occurs in spring and autumn and in the annual mean. In summer, the 12-model mean warming is close to or slightly above 2°C in all of NELA. Slightly weaker average warming occurs in this season in the surrounding sea areas, not only over the northern North Atlantic but also over the Arctic Ocean.
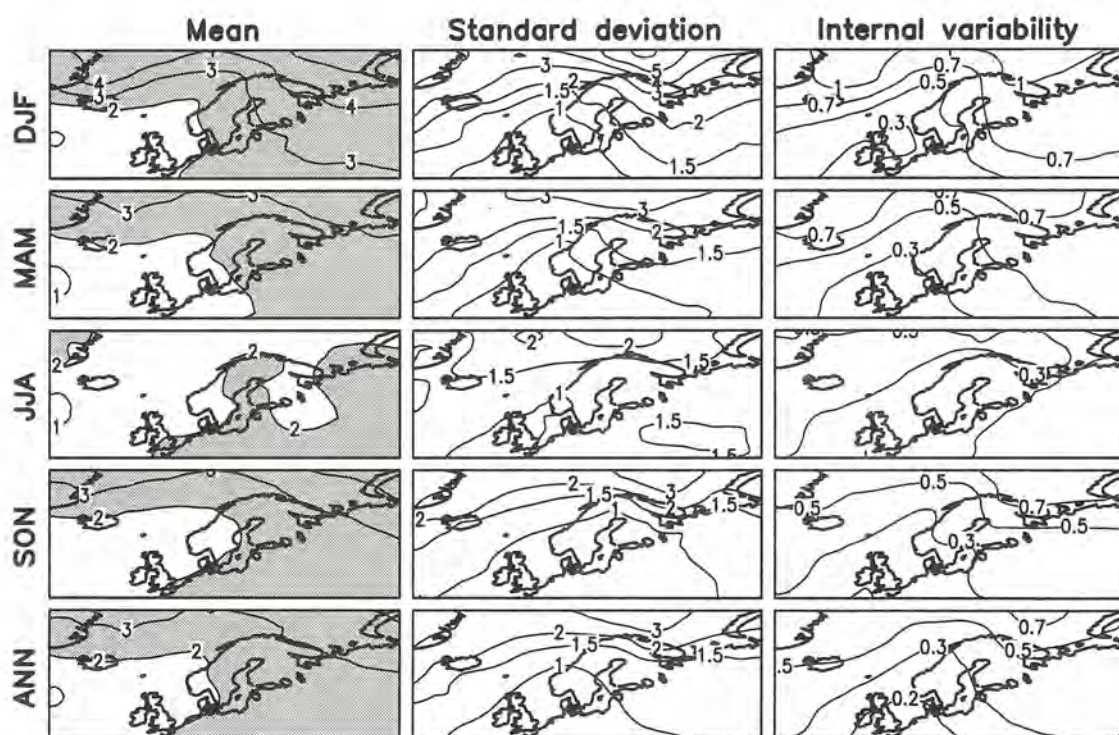
Figure 8. *12-experiment statistics of temperature change in individual seasons (DJF, MAM, JJA and SON) and in the annual mean (ANN). The left column shows the 12-experiment mean change (contours every 1°C; values of over 2°C shaded) and the midcolumn the interexperiment standard deviation (contours at 0.5, 1, 1.5, 2, 3, 4, 5 and 6°C). The right column gives an estimate for the interexperiment standard deviation explained by internal variability alone (contours at 0.2, 0.3, 0.5, 0.7 and 1°C).*

The interexperiment standard deviation (midcolumn of Fig. 8) in the simulated temperature change shows an even stronger gradient toward the Arctic Ocean than the 12-model mean warming, especially in winter. In several models (e.g., CSIRO, LMD/IPSL and NCAR-WM), decrease of sea ice in the Arctic Ocean leads to very large warming, but in GISS a local increase of ice to the north of Finland causes cooling (see Fig. 7). Further away from the Arctic Ocean, a typical value of the standard deviation is 1-1.5°C, which is roughly a half of the 12-experiment mean warming. As a whole, the interexperiment variation is slightly smaller in autumn than in the other seasons.

The right column of Fig. 8 gives an estimate of how large the interexperiment standard deviation would have been *in the absence of model differences*, that is, if differences in the calculated climate response were solely caused by internal variability in the simulations. This is calculated as

$$\sqrt{\frac{11}{12}N^2},$$

where N^2 is interpreted and estimated as described in Sections 5.1-5.2. This estimated standard deviation is substantially smaller than the actual interexperiment standard deviation, which implies that most of the differences in the simulated temperature change are caused by model differences. The estimated contribution of internal variability is larger in winter than in summer and annual mean, and in all seasons except for summer it increases from the southern part of the map domain toward the Arctic Ocean.

4.2 Precipitation

The relative changes in annual mean precipitation in the individual models are shown in Fig. 9. Again, both the pattern and the magnitude of the simulated change vary. Overall, however, increases of precipitation dominate in land areas of northern Europe. Local decreases occur in some models (mainly CCC, CERFACS, GISS and MPI) mostly in the southern part of the domain. The model results over the surrounding oceans are somewhat more variable. In particular, one may note that the cooling in GISS over the northern North Atlantic and in the Arctic Ocean to the north of Finland is accompanied with reduced precipitation – in the latter area the decrease reaches 40%. On the other hand, the simulated increase in precipitation in the southern and central parts of the Nordic Subregion in GISS is substantial, despite the minimal warming. This pattern, which is strongest in winter but evident in all seasons, suggests that the cooling of the North Atlantic surface water and the expansion of sea ice force a southward shift in the eastern part of the North Atlantic storm track.

As averaged over all 12 experiments and the whole of NELA, the simulated increase in annual mean precipitation is 6.6% (see Table 5). This is more than twice the corresponding global mean increase. The average increase is largest (11%) in winter but negligible (less than 1%) in summer. The 12-experiment mean statistics for the Nordic Subregion are similar, although the minimum of increase in summer is slightly less pronounced.
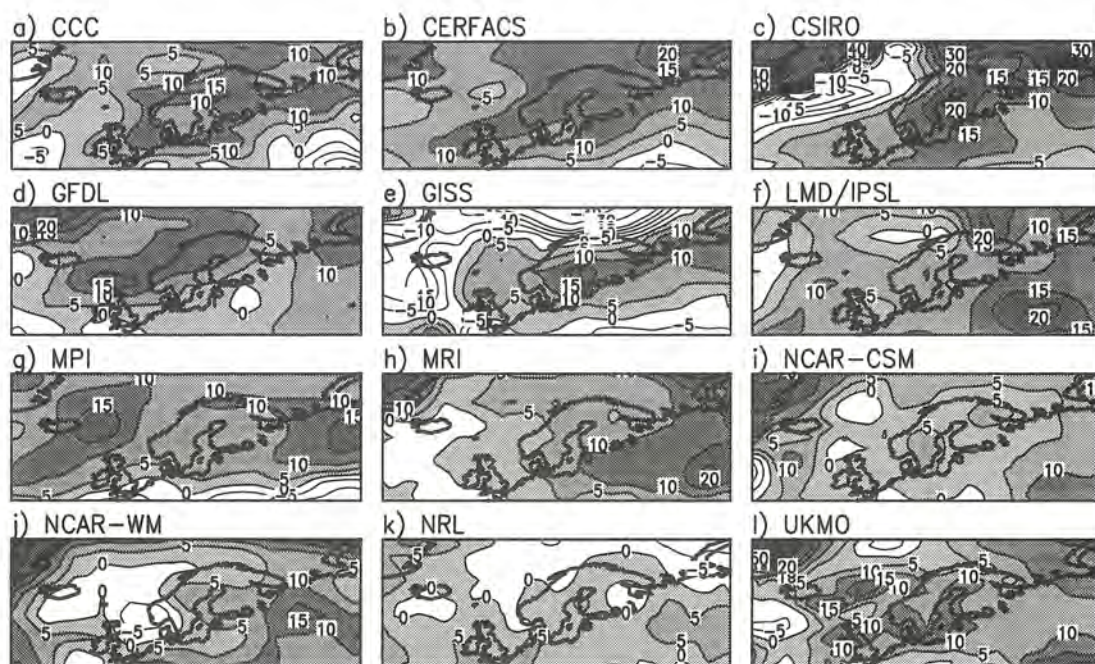
Figure 9. *As Fig. 7, but relative changes in annual mean precipitation (contours at 0, ±5, ±10, ±15, ±20, ±30 and ±40%). Positive values are shaded with darkest shading for largest increase in precipitation.*

Table 5. *As Table 4, but changes in area mean precipitation (per cent of the area mean in the control run).*

	Global	Northern Europe Land Area					Nordic Subregion				
	Annual	DJF	MAM	JJA	SON	Annual	DJF	MAM	JJA	SON	Annual
CCC	1.4	12	5	-2	6	5.0	18	10	6	10	10.7
CERFACS	4.3	11	10	-1	8	7.2	11	10	7	14	11.0
CSIRO	3.5	17	11	4	13	10.5	14	22	8	15	14.0
GFDL	4.5	8	9	-1	5	5.3	11	12	-1	12	8.8
GISS	2.6	11	2	-3	8	3.8	12	8	1	12	7.7
LMD/IPSL	2.7	18	18	-3	18	10.7	17	10	0	8	8.1
MPI	3.0	8	8	-1	5	5.2	10	7	2	11	7.6
MRI	4.0	6	9	13	6	8.7	5	7	9	4	6.2
NCAR-CSM	2.0	8	4	2	6	4.7	2	3	7	4	4.3
NCAR-WM	3.1	14	1	-11	21	7.0	9	1	-10	9	3.0
NRL	0.9	5	-2	7	1	2.9	6	-1	3	-1	1.6
UKMO	3.0	11	10	4	7	8.1	11	11	3	12	9.1
Mean	2.92	10.7	7.2	0.6	8.8	6.6	10.6	8.3	2.8	9.1	7.7
St. Dev.	1.07	3.8	5.1	5.7	5.6	2.4	4.5	5.7	5.1	4.4	3.4

22
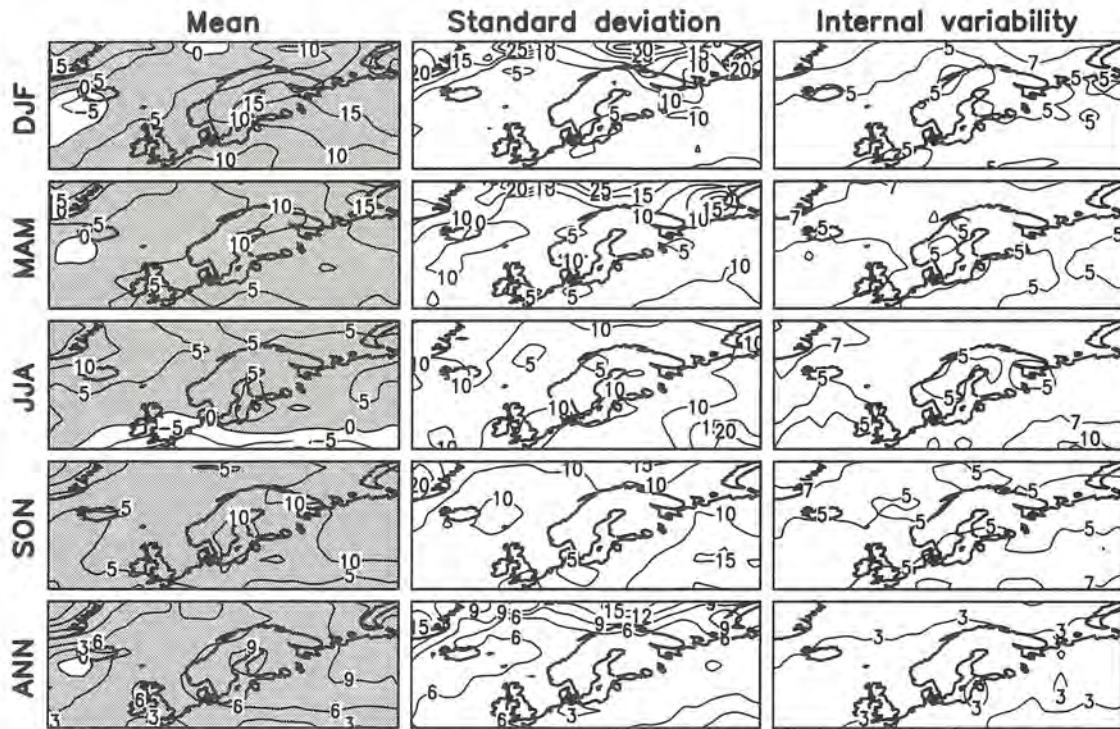
Figure 10. *12-experiment statistics of precipitation change. The 12-experiment mean changes in the left column and the interexperiment standard deviation in the midcolumn are contoured at every 5% (seasonal means) or at every 3% (annual mean). Positive 12-experiment mean changes are shaded. The estimated standard deviation due to internal variability (right column) is contoured at 3, 5, 7 and 10%.*

The 12-experiment mean geographical patterns of change are shown in the left column of Fig. 10. The average annual mean increase ranges, in land areas of northern Europe, from 3% in the south to about 10% in northern Russia. The south-north gradient is largest in summer, when the all-model mean in fact indicates a small decrease of precipitation to the south of about 53°N. However, as is obvious from Fig. 9 and will be further quantified in Section 5.4, the interexperiment agreement on the distribution of precipitation changes within northern Europe is actually very limited.

The local interexperiment standard deviation in the annual mean precipitation change is in most of NELA only 3-6%, but the scatter in the seasonal means is larger (midcolumn of Fig. 10). A similar contrast between the annual mean and the seasonal means is also evident in the standard deviation estimated to be explained by internal variability (the right column). Although it is obvious that internal variability is not able to explain all the interexperiment differences in precipitation change either, its relative importance for these differences appears distinctly larger than was the case with temperature change.

Returning to Table 5, the change in the annual area mean precipitation varies from 3% (NRL) to almost 11% (CSIRO and LMD/IPSL) in the whole NELA, and from less than 2% (NRL) to almost 14% (CSIRO) in the Nordic Subregion. All 12 experiments indicate an increase in the NELA mean precipitation in winter and autumn, and with the exception of a slight decrease in NRL this is the case even in spring. The area mean changes in summer are of varying sign and in most models relatively small, but MRI

simulates an increase of 13% and NCAR-WM a decrease of 11%. Although the seasonal cycle of the simulated area mean change varies between the experiments, the minimum in increase (or maximum in decrease) in summer is absent from only MRI and NRL. In the Nordic Subregion only two models (NCAR-WM and GFDL) simulate a decrease in area mean summer precipitation, but on the whole the differences between this subregion and the whole NELA are rather unsystematic.

4.3 Sea level pressure

As an example of the individual CMIP2 model results on the change in sea level pressure, the simulated changes in winter (December-February) are shown in Fig. 11. The 12 experiments show little agreement with each other; instead there is a remarkable variation in both the magnitude and the sign of the simulated pressure changes. For example, the CCC experiment indicates a change towards more cyclonic circulation in northernmost Europe, with a pressure decrease of 4-5 hPa in central and northern Scandinavia. In NCAR-WM, by contrast, a belt of 2-5 hPa pressure increase extends all the way from Iceland to Finland and northern Russia. Note that the correction for global mean biases used in the comparison of the control runs is omitted in calculating the changes. It is not necessary since the global biases in the perturbation runs are very similar with those in the control runs, apart from a small general decrease (typically −0.2 hPa) expected to accompany simulated global warming (see Räisänen 1997).
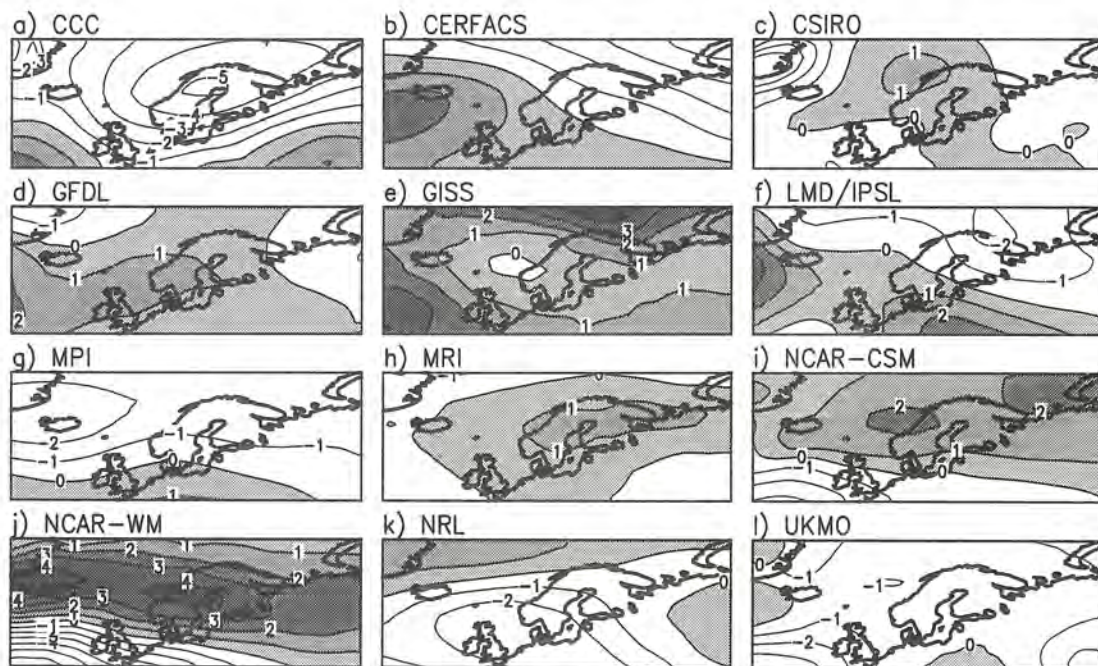
Figure 11. As Fig. 7, but changes in December-February mean sea level pressure (contours at every 1 hPa). Positive values are shaded with darkest shading for largest increase in pressure.

The simulated changes in wintertime sea level pressure are of particular interest because of the recent observed trends in winter climate. In particular from winter 1988-89, the surface circulation in northern Europe has been characterized by stronger-than-average surface westerlies, forced by a dipole pressure anomaly pattern with negative (positive)

anomalies over the Arctic Ocean and northernmost North Atlantic (Central Europe and mid-latitude North Atlantic) (e.g., Watanabe and Nitta 1999). This pattern has been interpreted either as a positive phase of the North Atlantic Oscillation (Hurrell and van Loon 1997) or as part of a hemispheric-scale Arctic Oscillation (Thomson and Wallace 1998). It is not clear, however, if these anomalies belong to natural climate variability or if they are caused by increases in greenhouse gases or by other sources of anthropogenic climate forcing (e.g., Corti et al. 1999).

The variability of results shown in Fig. 11 gives little evidence for the idea that the recent circulation anomalies would be greenhouse gas induced. However, the possibility of common model deficiencies remains. Recent experiments with the GISS climate model suggest that a realistic representation of the stratosphere may be a crucial factor for the simulation of greenhouse gas induced pressure changes (Shindell et al. 1999). In a transient greenhouse gas run with the model top located at a conventional 30 km level, the model simulated no trends in the Arctic Oscillation. When the top was raised to 85 km, however, a trend was obtained that compares well with the recent observed trends.
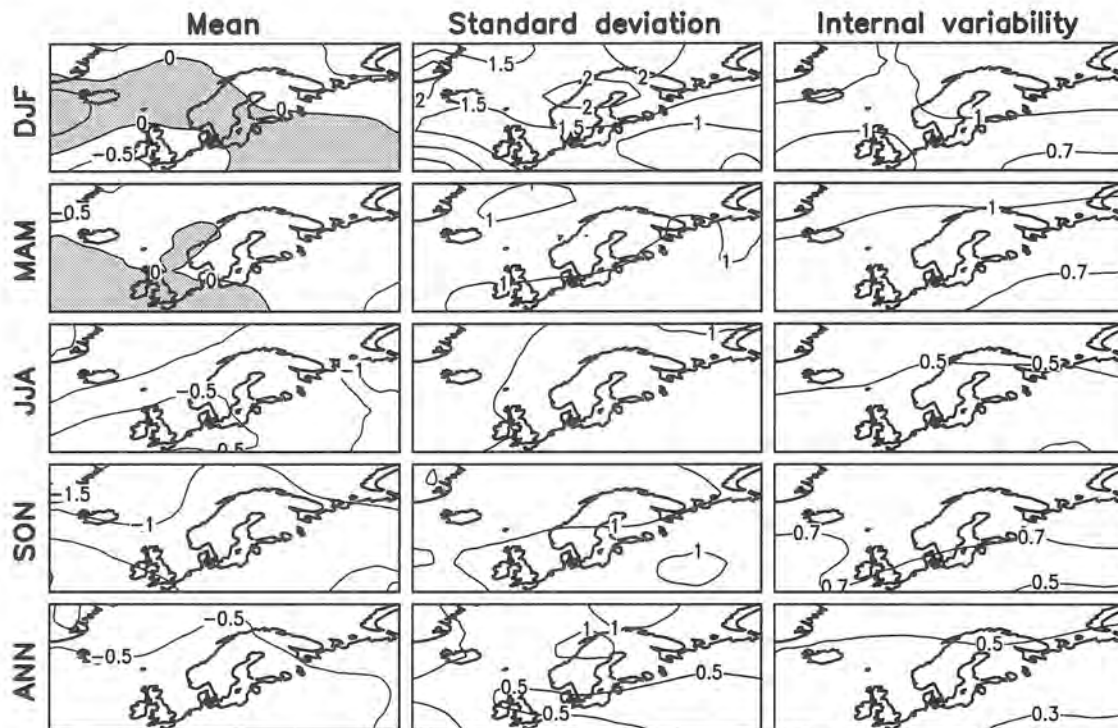
Figure 12. 12-experiment statistics of changes in mean sea level pressure. The 12-experiment mean changes in the left column and the interexperiment standard deviation in the midcolumn are contoured at every 0.5 hPa. Positive 12-experiment mean changes are shaded. The estimated standard deviation due to internal variability (right column) is contoured at 0.3, 0.5, 0.7 and 1 hPa.

In addition to winter, the interexperiment agreement on changes in sea level pressure in northern Europe and in the surrounding sea areas is very poor even in spring. The 12-experiment mean changes in these seasons are generally much smaller than the interexperiment standard deviation (see Fig. 12). In summer and in autumn, the interexperiment agreement is slightly better. In most experiments, the simulated

pressure changes in and around northern Europe are in these seasons predominantly negative, but the more detailed patterns of change vary widely.

The poor agreement between different experiments on CO_2-induced changes in sea level pressure is consistent with the results obtained by Räisänen (1997) for a smaller set of models. From physical arguments, the response of sea level pressure to increased CO_2 may be inherently more difficult to predict than the response of the other two variables considered here. The warming of surface air temperature is qualitatively a direct response to the effect of increased CO_2 on the transfer of long-wave radiation, although its magnitude is affected substantially by several feedbacks. The increase in precipitation in northern Europe in winter and in the annual mean is, in turn, an expected consequence of the fact that warmer air has a larger capacity to transport moisture from lower latitudes and from the Atlantic Ocean. By contrast, it is hard to deduce from simple physical reasoning how the pressure distribution should change. Changes in sea level pressure are affected by both thermodynamic (e.g., horizontal distribution of lower tropospheric warming; Murphy and Mitchell 1995) and dynamic factors (such as changes in transient activity in middle and upper troposphere; Räisänen 1998). They are therefore expected to be sensitive to various other aspects of the simulated climate change.

Another factor that is important in interpreting the simulated pressure changes is the large internally generated variability of sea level pressure in northern Europe especially in the cold season. A comparison of the estimated standard deviation due to internal variability (right column of Fig. 12) with the actual interexperiment standard deviation (midcolumn) suggests that internal variability is, at least, of similar importance for the differences in pressure change than it is for the differences in precipitation change. The issue will be discussed in more quantitative terms in section 5.3.

4.4 Comparison with earlier studies

Of the seven climate change experiments studied by R94, four were quantitatively comparable with those discussed here, that is, transient AOGCM experiments with results available around the doubling of CO_2. One of these (the GFDL experiment) also participates in the CMIP2 intercomparison, and the remaining three were made with two earlier versions of the MPI model and one earlier version of the UKMO model.

Among the results shown by R94 were the winter- (DJF) and summertime (JJA) changes in surface air temperature averaged over the mentioned four experiments. The four-model mean temperature changes (Figs. 8f and 9f of R94) bear in both seasons a close resemblance to the 12-model means shown here, both in the magnitude and the large-scale geographical distribution of the simulated warming. Some of the details naturally differ; for example, the old four-model mean wintertime warming over Scandinavia is about 0.5°C stronger than the new 12-model mean. The overall tendency toward lower summertime sea level pressure was also present in the experiments analyzed by R94; in the four-model mean shown in that study (Fig. 13f) the decrease is slightly larger than in the present 12-model mean. The weak dipole pattern with increasing (decreasing) DJF sea level pressure over the northern North Atlantic to the

south (north) of 60°N indicated by the four-model mean of R94 (Fig. 12f) is, however, not reproduced in the present results.

KGM98 compared the changes in DJF and JJA mean temperature and precipitation in nine transient CO_2 experiments in several subcontinental areas, including northern Europe. Of the seven experiments in which the averaging period for the perturbation run is centred within 10 years of the doubling of CO_2, five are from models (CSIRO, GFDL, MRI, NCAR-WM, UKMO) also present in the CMIP2 intercomparison and the remaining two from earlier versions of the MPI and UKMO models. In these seven experiments, the area mean warming in northern Europe varies, according to Fig. 3 of KGM98, from about 2.4°C to 4.6°C in winter and from 1.4°C to 4.5°C in summer. The lower ends of these ranges are much higher than those reported here, in particular in DJF, because the two CMIP2 models with weakest warming (NRL and GISS) are not present in KGM98. Even in the case of the common CMIP2 models, some differences exist between the changes reported here and the values of KGM98. This is most probably related to differences in the control run and perturbation run averaging periods and in the exact definition of the areas.

The approximate seven-experiment ranges for change in northern Europe area mean precipitation in KGM98 are from 5% to 25% in winter and from -9% to 14% in summer – the seasonal contrast is thus similar to that found here. The largest change in winter (25%) in KGM98 occurs in the older UKMO model (Murphy 1995) that is not present in the CMIP2 intercomparison.

5 Statistical comparison

As illustrated in the previous section, the 12 CMIP2 experiments show both common features and differences in their CO_2-induced climate change in northern Europe. Furthermore, the differences are expected to be caused partly by model differences, partly by internal variability in the simulations. In this section, we attempt to put these qualitative remarks in a more quantitative form.

5.1 Theoretical framework

For a theoretical framework, consider the 12 CMIP2 experiments as a sample of an infinitely large population of similar experiments. The climate change in each member of this hypothetical population can be written as

$$x_i = M + \varepsilon_i \tag{2}$$

where M is the mean value for the whole population and ε_i is the deviation for experiment i. Here, x_i may denote either a local (grid box scale) climate change or a mean over a larger area. Alternatively, one can write

$$x_i = M + \delta_i + \eta_i \tag{3}$$

where ε_i has been divided into a model-related deviation (δ_i) and a deviation associated with unforced internal variability in experiment i (η_i). The effects of internal variability average out in the mean value for the whole infinite population of experiments, so that

the population means of both δ_i and η_i are zero. However, in the usual case that only one climate change experiment has been made with each model, the values of δ_i and η_i in the individual experiments are unknown. To determine these, one should repeat the experiment with the same model a large (in principle, infinite) number of times, using the same forcing but different initial conditions. The climate change averaged over all members of this ensemble would be M + δ_i and the values of η_i for the individual ensemble members would be obtained accordingly.

Substituting from (2) and (3), one obtains analogous expressions for the population mean of the *squared* climate change A^2:

$$A^2 \equiv \left\{ x_i^2 \right\} = M^2 + E^2 \tag{4}$$

where $E^2 \equiv \{\varepsilon_i^2\}$ is the total interexperiment variance[2], or

$$A^2 = M^2 + D^2 + N^2 \tag{5}$$

where $D^2 \equiv \{\delta_i^2\}$ and $N^2 \equiv \{\eta_i^2\}$ denote the variances due to model differences and internal variability, respectively. These two are additive, because $\{\delta_i\eta_i\} = 0$ for the infinite population.

The division (4) provides an elegant quantification of the relative interexperiment agreement in the whole hypothetical population, and (5) further divides the disagreement into contributions of model differences and internal variability. In this section we try to estimate the components of these divisions using the CMIP2 results. This is straightforward with (4) but more complicated with (5).

Analogous to (4), the mean squared climate change in the sample of $k=12$ CMIP2 experiments can be written as

$$a^2 \equiv \left[x_i^2 \right] = m^2 + e^2 \tag{6}$$

where m is the sample mean climate change and e the within-sample standard deviation. Substituting from (2), the expected values of a^2, m^2, and e^2 as a function of A^2, M^2 and E^2 are

$$a_{EXP}^2 = A^2, \tag{7}$$

$$m_{EXP}^2 = M^2 + \frac{1}{k}E^2 \quad \text{and} \tag{8}$$

$$e_{EXP}^2 = \frac{k-1}{k}E^2. \tag{9}$$

Conversely, the best estimates for the population statistics A^2, E^2 and M^2 (which are in reality unknown) are obtained by assuming that the actual sample statistics a^2, e^2 and m^2 equal their expected values:

$$A_{EST}^2 = a^2, \tag{10}$$

---

[2] See Appendix A for the notations used for different averages.

$$M_{EST}^2 = m^2 - \frac{1}{k-1} e^2 \quad \text{and} \tag{11}$$

$$E_{EST}^2 = \frac{k}{k-1} e^2. \tag{12}$$

Similarly, beginning from (3), one obtains an estimate of how the interexperiment variance E^2 is divided between the contributions of model differences (D^2) and internal variability (N^2):

$$D_{EST}^2 = \frac{k}{k-1} e^2 - n^2 \quad \text{and} \tag{13}$$

$$N_{EST}^2 = n^2, \tag{14}$$

where

$$n^2 \equiv \left[\eta_i^2 \right] \tag{15}$$

is the sample mean squared value of the contribution of internal variability to the simulated climate changes. However, because the exact contribution of internal variability in any one experiment is unknown, only the approximate magnitude of n^2 can be estimated. The methods used for this are described in the following subsection.

It is also important to note that the residual estimate (13) for the model-related variance might be inaccurate even if the magnitude of internal variability were known precisely. The assumption that model-related and variability-related differences are uncorrelated, and therefore additive in terms of variance, is exactly true only for an infinite population of experiments. Within a limited sample of experiments, the deviations due to internal variability and model differences might be either positively of negatively correlated. In the former (latter) case, (13) tends to overestimate (underestimate) the magnitude of the model-related differences. However, even with a limited sample of experiments, the additivity of model-related and variability-related variances becomes a more plausible assumption when variances averaged over a larger area instead of variances in individual grid boxes are considered. Putting this statement in a quantitative form is complicated by the spatial autocorrelation present in all meteorological fields.

Although the statistics A^2, M^2, E^2, D^2 and N^2 can be estimated for each grid box separately, we will only discuss their area means over northern Europe, or, for comparison, over the whole world. For brevity, the notations

$$F = \overline{M_{EST}^2} / \overline{A_{EST}^2} \tag{16}$$

and

$$I = \overline{N_{EST}^2} / \overline{E_{EST}^2} \tag{17}$$

will be used. F will be called the *fraction of common signal*. The value $F=1$ indicates perfect interexperiment agreement and $F=0$ is what one would on the average get when drawing the k experiments from a population in which climate changes are totally

random. One but by no means the only use for this nondimensional measure is in comparing how the interexperiment agreement differs between different variables. I indicates the relative importance of internal variability for the interexperiment differences.

5.2   Estimating the contribution of internal variability

As already noted, the direct calculation of n^2 according to (15) is unfeasible. Rather, one must use the statistical estimate

$$n^2 = [v_i]$$
(18)

where v_i is, in principle, the variance one would obtain by repeating the same climate change experiment with the model i a large number of times with different initial conditions. In the absence of such large ensembles, however, this variance must be estimated in an indirect manner. We use the estimate

$$v_i = \left(\frac{1}{T_1} + \frac{1}{T_2}\right) v5_i$$
(19)

where T_1 and T_2 are the control run and perturbation run averaging periods used in defining the climate change, expressed as multiplies of 5 years. For all experiments excluding NRL, $v5_i$ is calculated as the variance of individual five-year means ($f5_i$) in the control run:

$$v5_i = \frac{T}{T-1}\left(\overline{f5_i^2} - \overline{f5_i}^2\right)$$
(20)

where T is the total length of the control run in multiplies of 5 years (for models with full 80 years of data, $T=16$). However, this estimate cannot be applied to the NRL experiment with a control run length of only 3 years. Rather, the magnitude of internal variability is in this case inferred from the perturbation run after first detrending it:

$$v5_i = \frac{T}{T-2}\overline{g5_i^2}$$
(21)

where $g5_i$ is the deviation from the best-fit linear trend calculated from the 16 5-year means in the perturbation run.

The changes in precipitation are in this study expressed in relative units. The absolute variances obtained from (19) are therefore divided by the square of the control run mean before averaging over the 12 experiments. This procedure is slightly approximate, since the sensitivity of the relative change on variations in the control run mean actually depends on the relative change itself.

The estimates of variance v_i are naturally prone to sampling errors caused by the limited number of 5-year periods in the control run. In addition, they may be systematically biased by at least three inexact assumptions. First, (19)-(21) assume that temporal autocorrelation between subsequent 5-year periods is negligible. Second, they assume that the amplitude of internally generated variability remains the same even in the run with increasing CO_2. Finally, all the variability within the control runs is assumed to

represent true internal variability in the models, which is not necessarily the case. Spin-up effects that reflect a conflict between the initial state of the control simulation and the internal dynamics of the model may create climate drift that leads to an overestimate of the normal internally generated variability.

The importance of the mentioned sources of systematic error was studied by estimating the amplitude of internal variability by some alternative methods (Appendix C). The differences between the different methods were in most cases small. While not completely proving this, this similarity suggests that the uncertainty in estimating internal variability is not critical for the conclusions of the present study.
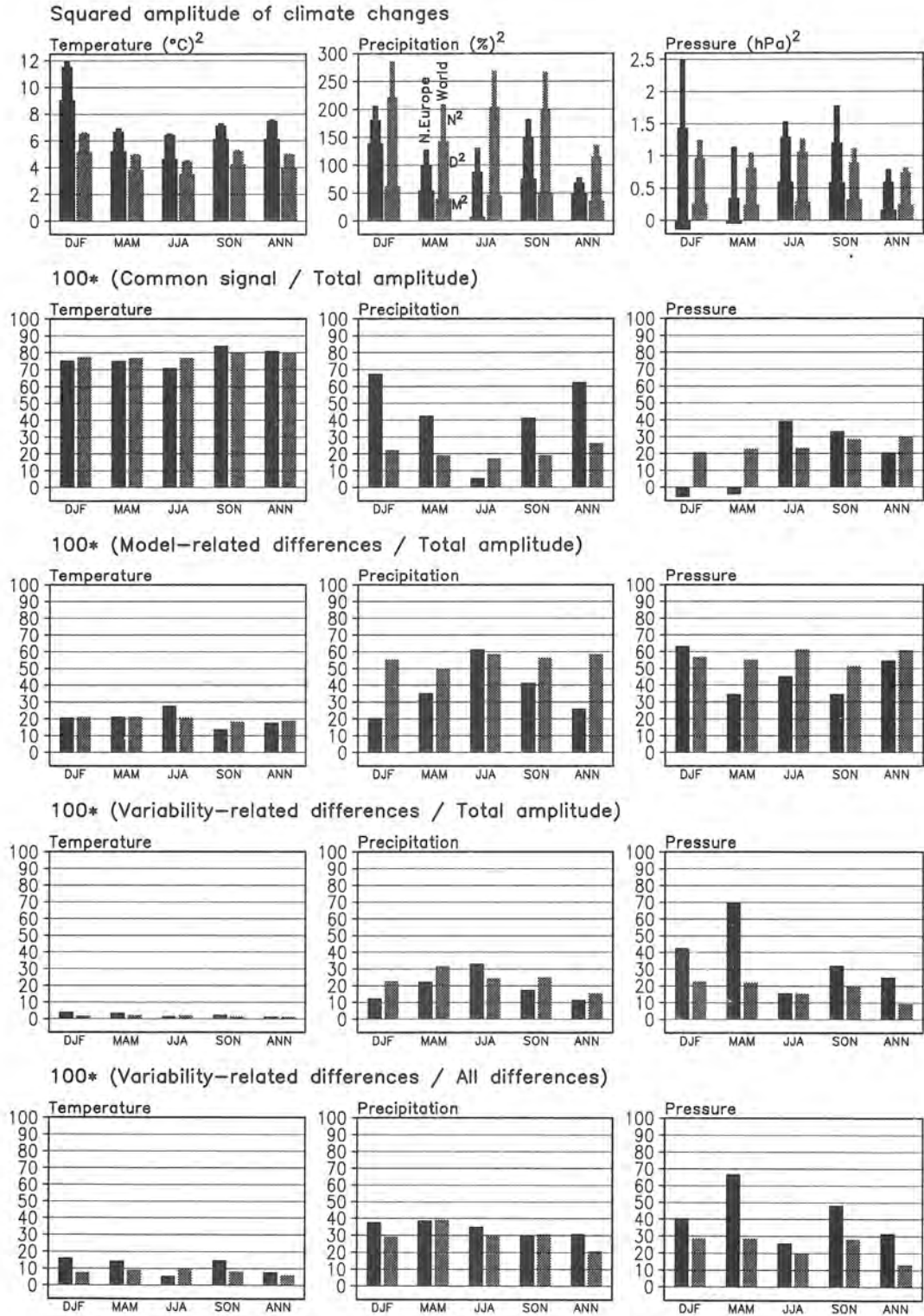
5.3   Statistics for northern Europe and the whole world

A statistical summary of the simulated climate changes in the 12 CMIP2 experiments is given in Fig. 13 using the quantities introduced in Section 5.1. A^2, M^2, E^2, N^2 and D^2 (the subindex EST will not be repeated here) were first calculated for each grid box separately, as described in the two previous subsections. They were then averaged over the Northern Europe Land Area (surface air temperature and precipitation) or over the whole area 30°W-60°E, 50°-75°N shown by Fig. 1 (sea level pressure). The distinction between the variables is motivated by the fact that the surface climate in land areas of northern Europe may be affected by changes in the pressure pattern over the northern North Atlantic, not only by the local pressure changes.

For comparison with northern Europe, Fig. 13 also shows the globally averaged values of the same statistics. To facilitate the global calculation, however, a minor redefinition of precipitation changes is made. These are calculated as

$$\Delta P = 100\% \times \frac{P_{per} - P_{ctrl}}{Max(P_{ctrl}, P_{min})} \qquad (22)$$

where P_{ctrl} and P_{per} are the time means of precipitation in the control run and the perturbation run and $P_{min} = 0.25$ mm/d. In northern Europe this is just the standard relative change, because the 80-year control run means of precipitation exceed 0.25 mm/d in all models in the whole area in all seasons. But, in extremely dry areas, modest absolute changes in precipitation occasionally yield gigantic relative changes, which would completely dominate the global statistics if no lower threshold for the denominator were used. However, these global statistics were found to be reasonably insensitive to the exact choice of P_{min} as far as this was in the range 0.1-1 mm/d.

Squared amplitude of climate changes

Figure 13. Statistics of climate changes in the 12 CMIP2 experiments, averaged over northern Europe (black bars) and over the whole world (grey bars). The first row shows the average squared amplitude (A^2) of the changes, divided to the contributions of the common signal (M^2; thick part of each bar), model-related deviations (D^2; medium-thick part) and variability-related deviations (N^2; thin part). The ratios $M^2/A^2=F$, D^2/A^2 and N^2/A^2 are given in per cent in the second, third and fourth row, respectively. The ratio of variability-related differences to all interexperiment differences $[N^2/(D^2 + N^2)=I]$ is indicated in the last row. See text for further details.

The first row in the figure gives the absolute values of the statistics and the remaining four rows show various ratios associated with these. One can make several conclusions:

- The relative interexperiment agreement on temperature changes in northern Europe is quite good. The F ratio (16) ranges from 0.71 in JJA to 0.84 in SON. These values are similar to the values calculated for the whole world. However, the total squared amplitude of temperature changes in northern Europe exceeds the global average (especially in DJF), because the simulated warming in northern Europe is in most models relatively large. Thus, in spite of the similarity of F in the two areas, the absolute interexperiment variance in northern Europe is somewhat above its global mean as well.

- The estimated relative contribution of internal variability to the total squared amplitude of temperature changes is, at most, 0.04 in northern Europe in DJF. The fraction of interexperiment differences explainable by internal variability, I (17), is likewise quite modest, in northern Europe from 0.06 in JJA to 0.16 in DJF and globally from 0.06 (annual mean) to 0.10 (JJA). Most of the interexperiment variance is therefore inferred to be associated with model differences.

- The fraction of common signal for precipitation changes in northern Europe is relatively high in winter (0.67) and in the annual mean (0.63) and reasonable even in spring (0.43) and autumn (0.41), but very low (0.05) in summer. The low value for summer is in part due to a maximum in the absolute interexperiment variance in this season, but it is more strongly linked to the small squared amplitude of the common signal M^2. The models typically simulate relatively modest changes in summer precipitation to both directions, in contrast with larger positive changes in winter precipitation.

- Excluding JJA, the fraction of common signal for precipitation changes in northern Europe is substantially higher than the global value of the same statistics. The latter varies between 0.17 and 0.22 for seasonal mean changes and is 0.26 for the annual mean change. The magnitude of the common signal M^2 in northern Europe is above its global mean in all seasons but JJA, reflecting the relatively large overall increase in simulated precipitation. By contrast, the interexperiment variance in northern Europe is relatively small. The largest interexperiment differences in relative precipitation change generally occur in the dry subtropics, but even in the wet tropics, the scatter between the model results is larger than it is in northern Europe. The agreement is comparable with northern Europe only in some other mid- to high-latitude areas, such as Siberia, northern North America and the Southern Ocean.

- The potential effect of internal variability on the simulated precipitation changes is larger than is the case with temperature changes. The ratio N^2 / A^2 varies in northern Europe between 0.12 (annual mean) and 0.33 (JJA), and globally between 0.15 (annual mean) and 0.32 (MAM). For the fraction of interexperiment differences explainable by internal variability, the range is from 0.30 in SON to 0.39 in MAM in northern Europe, and from 0.21 in the annual mean to 0.39 in MAM in global averages. Both in northern Europe and in a global perspective, a majority of the interexperiment differences in precipitation change nevertheless appears to be related to model differences.

- The estimated common signal M^2 for changes in sea level pressure in northern Europe and northern North Atlantic in DJF and MAM is slightly negative. This result (which would not actually be possible in a truly infinite population) indicates a worse interexperiment agreement than would on the average be obtained by pure chance. There is some agreement between the experiments in JJA ($F=0.39$), SON ($F=0.33$) and the annual mean ($F=0.20$), which largely reflects the general dominance of simulated pressure decrease. As with precipitation, the global values are also relatively low – between 0.21 in DJF and 0.30 in the annual mean.

- The estimated contribution of internal variability to the interexperiment differences in simulated pressure change in northern Europe varies from 26% in JJA to as much as 67% in MAM. As a whole, the simulated pressure changes in northern Europe are less well discernible from internal variability than the changes in temperature and precipitation, at least excluding JJA. The absolute amplitude of internally generated pressure variability (N^2) in northern Europe is, excluding JJA, markedly higher than its global average, and so is its relative contribution to the simulated pressure changes and interexperiment differences in these.

5.4 Dependence on horizontal scale

The statistics introduced in Section 5.1 allow a convenient look at several issues associated with interexperiment agreement and the impact of internal variability on this. First, it is obvious that the conclusions might depend on the horizontal scale considered – one would not expect GCM predictions to be at their best at the smallest scales nominally discerned by these models. A simple quantification of this is provided in Fig. 14. Here, statistics similar to those represented in Fig. 13 are shown separately for area mean climate changes in northern Europe (temperature and precipitation over land only, sea level pressure over the whole area 30°W-60°E, 50°-75°N) and for the deviations from these area mean changes.

The mean squared amplitudes for the area mean changes and for the deviations from these are shown in the first row in Fig. 14. In the case of surface air temperature, the squared amplitude for the deviations is an order of magnitude smaller than the amplitude for the area mean changes. This reflects the fact that temperature changes are in most models positive throughout the Northern Europe Land Area, with only moderate within-area variations in the magnitude of the change. In the cases of precipitation and sea level pressure, the squared amplitudes for the area mean change and the deviations are much more comparable. For precipitation changes in JJA and pressure changes in DJF and MAM, the deviations in fact have a larger mean squared amplitude than the area mean changes, indicating large within-area variations in the simulated change.

Comparison between the F values for the area mean changes and the deviations (second row of Fig. 14) gives a clear and expected conclusion. Whenever Fig. 13 indicated any substantial interexperiment agreement on the total two-dimensional change, this is largely manifested as good agreement on the area mean change. The fraction of common signal for the deviations is at best about 0.3 (DJF, SON and annual mean) in the case of temperature change, and the values for the other variables are even lower. The interexperiment agreement on the distribution of climate changes within northern

34

Europe is therefore generally very limited. On the other hand, one may note the very good relative agreement on changes in northern Europe area mean precipitation in DJF ($F=0.88$) and annual mean ($F=0.87$) – these values are actually higher than the corresponding values for surface air temperature.

Squared amplitude of climate changes
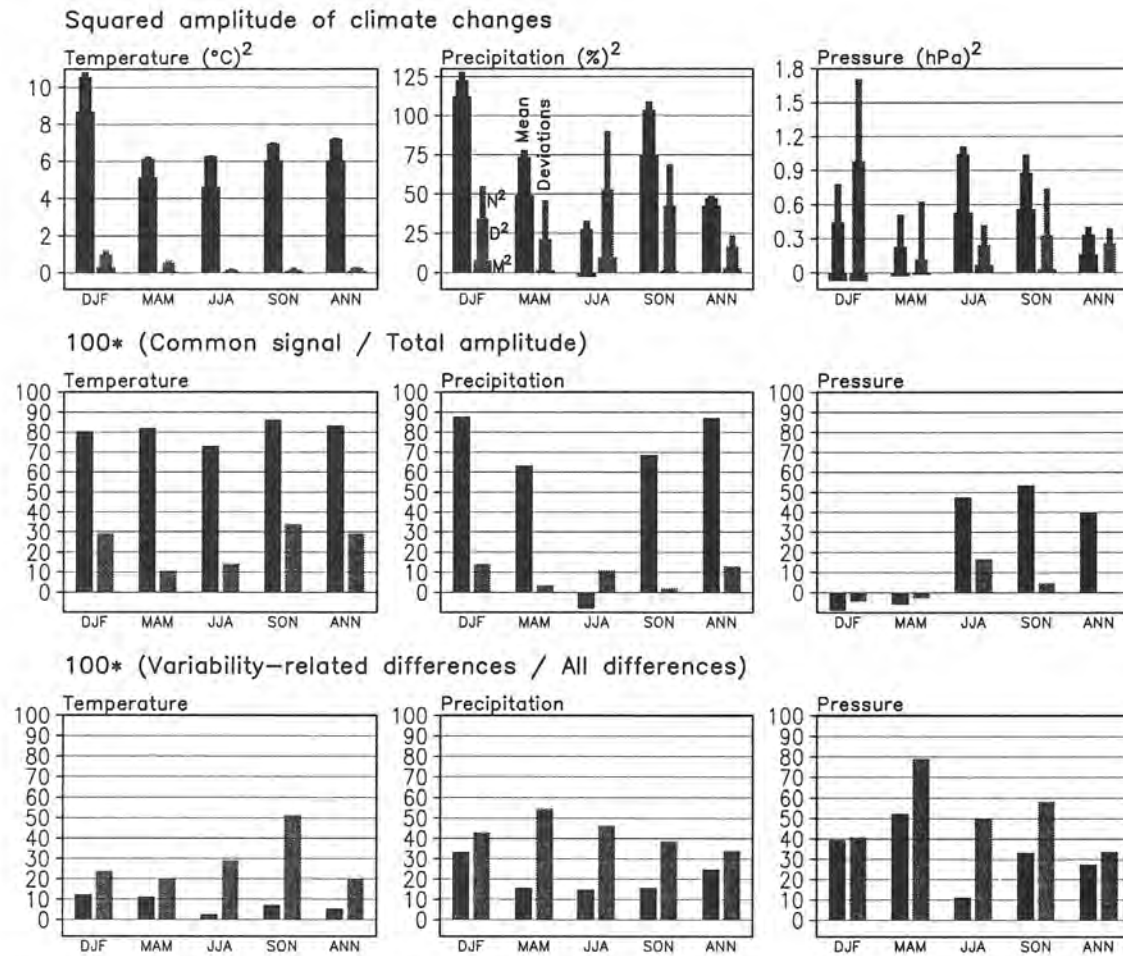
Figure 14. *Statistics for northern Europe area mean climate changes (black bars) and for the deviations from these (grey bars). The first row shows the average squared amplitude (A^2) of the changes, divided to the contributions of the common signal (M^2; thick part of each bar) model-related deviations (D^2; medium-thick part) and variability-related deviations (N^2; thin part). The second row gives the ratio $F=M^2/A^2$, and the third row the ratio $I=N^2/(D^2+N^2)$.*

In general, both the model-related and the variability-related differences take a larger share of the total squared amplitude of the deviations than is the case with the northern Europe mean changes. However, the differences associated with internal variability appear to grow more rapidly with decreasing scale than the model-related differences – the I ratio is in all cases higher for the deviations than for the area mean changes (bottom row of Fig. 14). Thus, although the relative interexperiment differences in the area mean changes are smaller than the differences in the deviations, the former seem to be better discernible against the noise caused by internal variability.

5.5 Dependence on averaging period

The estimates of the impact of internal variability shown above apply to the relatively long averaging periods (generally 80 years for the control run and 20 years for the perturbation run) used. With shorter averaging periods, the effects of internal variability should grow larger, which is expected to lead to a deterioration of the interexperiment agreement. To verify that this actually is the case, the comparison of the 12 CMIP2 experiments was redone using an averaging period of 10 years for both the control and perturbation runs. Such a short averaging period has been frequently used, in lack of more data, in earlier intercomparisons (e.g., R94; Räisänen 1997; KGM98). This exercise was repeated 16 times comparing, in turn, each of the two decades 61-70 and 71-80 in the perturbation runs with each of the 8 decades in the 80-year control runs.
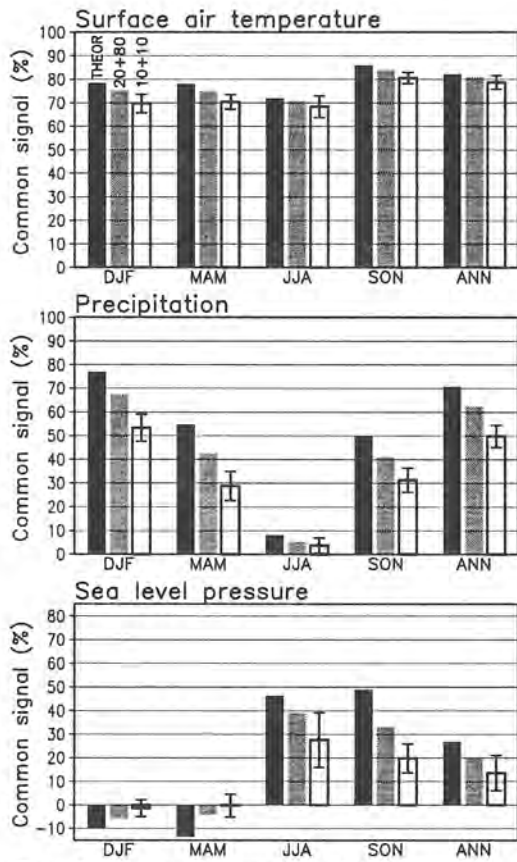
Figure 15. Fraction of common signal (F) for climate changes in northern Europe with different averaging periods. For each variable and season, the first bar shows the estimated theoretical F value in the absence of internal variability (see text) and the second bar the value obtained with the standard (usually 20+80 -year) averaging periods. The third bar gives the arithmetic mean of the 16 F values obtained with different 10+10 -year periods. The standard deviation of these 16 values is also indicated.

The F values obtained in this test are depicted in Fig. 15 together with the values for the standard (usually 20+80 year) averaging periods. Also shown are estimates for the F values that would be obtained in a comparison between these 12 models if the simulated climate changes were not disturbed by internal variability. These are estimated as

$$F_{THEOR} = \frac{M^2}{A^2 - N^2} \tag{23}$$

where M^2, A^2 and N^2 refer to the values obtained using the standard periods.

As indicated by the standard deviations included in Fig. 15, the F values obtained using different 10+10 -year periods were not constant. Depending on the particular decades selected, the detrimental effect of internal variability on the interexperiment agreement

is either larger or smaller than the theoretically expected average; at times internal variability may even act to mask model-related differences. This variation is caused by the limited number of experiments (12) compared and the relatively small area considered. In a global domain, the variation is substantially smaller (not shown).

On the average, however, the F values obtained with a 10+10 -year averaging period were distinctly lower than those in the standard case. The average decrease is in absolute terms largest, 0.14, for precipitation in DJF (from 0.67 to 0.53) and MAM (from 0.43 to 0.29). Other examples of substantial (0.10–0.13) decrease in F are precipitation in SON and in the annual mean, and sea level pressure in JJA and SON. In the case of temperature changes, the decrease is smaller, as expected from the fact that a large majority of the interexperiment differences were inferred to be model-related when using the standard (20+80 -year) averaging periods. The averaging period also has little impact on F when the relative interexperiment agreement is very poor even with the standard periods (precipitation in JJA and sea level pressure in DJF and MAM). As expected, however, the absolute variance between the different experiments increases with decreasing averaging period even in these cases (not shown).

The difference between the F values estimated to be obtained in the absence of internal variability and the F values obtained in the standard case is generally smaller than the difference between the latter and the average F with 10+10 -year period. This is not surprising, since the variance associated with internal variability in the standard case is expected to be only about a third of that with 10+10 -year period[3]. The substantially negative F_{THEOR} derived for sea level pressure in DJF (-0.10) and MAM (-0.14) are below the lowest value of F (–0.09) actually allowed by (10), (11) and (16). These spurious values are associated with the uncertainty in "subtracting" internal variability from the model results.

6 Attempts to understand the interexperiment differences – a statistical approach

Although the 12 CMIP2 models show in some respects a relatively large similarity in their response to increased CO_2 in northern Europe, the differences are not negligible either. The results of the previous section attribute a majority of these differences to the models themselves, rather than to internal variability, this being most clearly the case with the area mean changes of temperature and precipitation. Deducing the ultimate origin of these differences is, however, an extremely challenging task, because any two CMIP2 models differ in far more than just one respect.

One can nevertheless try a statistical approach. Would the simulated change in, for example, northern Europe mean temperature have a strong physical connection with some feature of the models themselves (e.g., horizontal resolution) or other aspects of

---

[3] If all experiments had full 20+80 years of data in the standard case, the theoretical ratio would be 3.2, neglecting temporal autocorrelation between different decades. The shorter standard averaging periods in NRL and NCAR-WM make the actual ratio slightly lower.

the model results (say control run temperature), this might be reflected as an interexperiment statistical connection between this feature and the simulated change, provided that this connection is not masked by other differences. A number of surveys for such statistical connections were made and some of them are reported in this section.

The results of this exercise were, as a whole, somewhat discouraging. Only a relatively few statistically significant connections were found, and some of these might have appeared obvious even without making the analysis. This lack of success indicates that the interexperiment differences in simulated climate change indeed depend on several factors. In addition, 12 models is a relatively small sample for a statistical analysis – if a physical connection is to be seen clearly from comparing just 12 models, it must be quite strong. On the other hand, statistical methods also have the risk of suggesting connections where they actually do not exist. In spite of these problems, the statistical approach may give useful guidelines of where more physically based studies of model behaviour should be directed.

6.1 Relating climate change in northern Europe to specific features of control climate and global climate change

We first study the connections of the NELA mean temperature and precipitation changes with other aspects of the model results. The set of "predictors" considered here is somewhat subjectively chosen but it attempts to include the most obvious candidates.

The scatter plots in Fig. 16 illustrate the relationship between the NELA annual mean warming and four temperature-related predictors: annual mean temperature in the control run, the amplitude of internal variability, the annual temperature range, and the global mean warming. Internal variability is characterized by the interpentadal (i.e., inter-5-year) standard deviation of temperature in the control run, excluding NRL for which the standard deviation within the detrended perturbation run is used. Annual temperature range is defined as the temperature difference between the average warmest (generally July in northern Europe) and coldest (generally January or February) months. The first three predictors were first calculated for each grid box separately and then averaged over NELA. The figure suggests that

- The simulated annual mean warming and the mean temperature in the control run are unrelated.
- The simulated warming and the amplitude of interpentadal internal variability are positively correlated. However, the positive correlation suggested by this figure is largely due to two models – NCAR-WM in which both the variability and the temperature change are largest, and NRL, in which these both are smallest.
- The simulated warming also tends to increase with increasing annual temperature range, even though the models with the strongest (LMD/IPSL) and weakest (CERFACS) seasonal cycle are not the two most extreme in terms of the warming.
- Annual mean warming in northern Europe is closely related to the global mean warming.
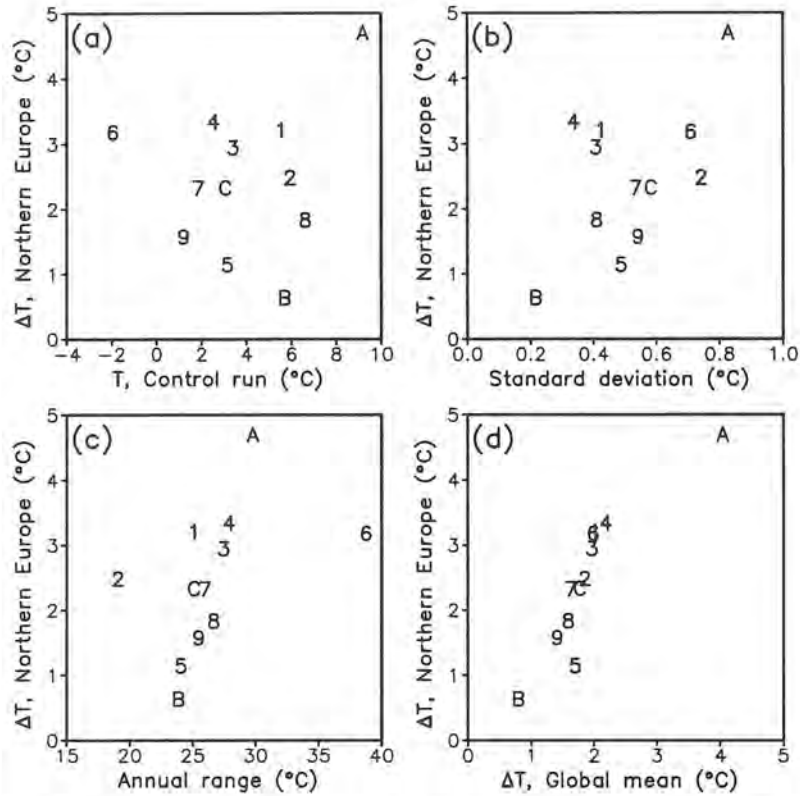
Figure 16. Scatter plots between Northern Europe Land Area annual mean warming (on the vertical axis) and (a) annual area mean temperature in the control run, (b) area mean standard deviation of 5-year mean temperature in the control run, (c) area mean annual temperature range in the control run and (d) global annual mean warming. The 12 experiments are labelled as 1 – CCC; 2 – CERFACS; 3 – CSIRO; 4 – GFDL; 5 – GISS; 6 – LMD/IPSL; 7– MPI; 8 – MRI; 9 – NCAR-CSM; A – NCAR-WM; B – NRL; C – UKMO.

The last point is of fundamental importance. The high correlation between the warming in northern Europe and the global mean warming (0.88 in the annual mean and from 0.72 in MAM to 0.96 in JJA in the individual seasons) indicates that the warming in northern Europe is to a large extent determined by global-scale processes. The interexperiment variations in this can therefore be not understood by considering processes specific to northern Europe alone.

Seasonal and annual correlation coefficients related to the other comparisons are shown in Table 6. In the comparisons involving seasonal mean warming, the control run mean temperature and internal variability are computed from the data for the same season. For the annual mean, the table also shows the correlations obtained when both the predictors and the predictant are averaged globally, and a hybrid case (correlation of warming in northern Europe with global predictors). The two-sided statistical significance of these correlations is tested against the distribution of correlation coefficients obtained from 1000 random permutations of the predictor values among the 12 experiments. Since the usual (Pearson) correlation is inconveniently sensitive to individual outliers (such as NCAR-WM and NRL in Fig. 16b), the (Spearman) rank correlations are also given in

case of the annual mean. These are significant at the 90% (95%, 99%) level when their absolute value exceeds 0.50 (0.59, 0.75). As seen from the table,

- The correlation between the control run temperature and the simulated warming in northern Europe is weak in the other seasons but markedly positive (0.68) in summer. However, although this value is significant at the 95% level, it is largely explained by NCAR-WM, which has a large warm bias in the control run and simulates by far the strongest summertime warming. The rank correlation in this season in only 0.29.
- The correlation between the amplitude of interpentadal variability and the simulated warming in northern Europe is positive in all seasons but statistically insignificant in JJA and significant only at the 90% level in the other seasons. The annual rank correlation is much lower than the Pearson correlation, which is associated with the NRL and NCAR-WM results noted above.
- The correlation between the simulated warming and the annual temperature range is also positive in all seasons but significant at the 90% level or higher only in winter and autumn. In this case, the annual rank correlation is somewhat higher than the Pearson correlation.
- The positive correlation of the simulated warming with both the amplitude of interpentadal variability and the annual range holds even in the global mean and is actually stronger (in both cases 0.75; significant at the 99% level) than in northern Europe.
- Annual mean warming in northern Europe is almost as well correlated with the global average of interpentadal variability as with the average variability in northern Europe, and actually better correlated with the global average annual temperature range than with the average temperature range in northern Europe.

Table 6. *Interexperiment correlation between area mean temperature changes and three predictors related to the control run temperature climate. NELA → NELA indicates that the predictors and the predictant are both averaged over the Northern Europe Land Area, and G→ G refers to global averages of these both. G → NELA shows the correlations between globally averaged predictors and temperature change in northern Europe. For the annual means, the rank correlation is also given (in parentheses). Correlation coefficients shown in italics (bold) are significant at the 90% (95%) level, and those with a star have a significance of at least 99%. See text for further details.*

	NELA → NELA					G→G	G→ NELA
	DJF	MAM	JJA	SON	Ann (rank)	Ann (rank)	Ann (rank)
Mean temperature	-0.30	0.29	**0.68**	-0.07	0.17 (0.08)	0.30 (-0.03)	0.10 (-0.02)
Internal variability	*0.56*	*0.55*	0.36	*0.58*	*0.55* (0.29)	**0.75***(0.35)	0.50 (0.24)
Annual range	**0.62**	0.06	0.41	*0.51*	0.45 (**0.62**)	**0.75***(**0.64**)	**0.76***(**0.68**)

The strong relationship between the warming in northern Europe and the global mean warming is hardly unexpected, and it is certainly physically based. If interpentadal temperature variability and annual temperature range have a real physical connection

with the simulated warming is not equally clear. The correlation coefficients associated with these connections are not statistically significant in all cases, and some significant correlations are expected to be obtained even by pure chance. That these correlations are similar in northern Europe and in a global domain is comforting but not surprising. The warming in northern Europe is strongly correlated with the global mean warming, and intermodel differences in the regional control run temperature climate in northern Europe tend to be related, in terms of both the mean state and the variability, to differences in the global temperature climate.

On the other hand, these correlations might well be physically based. Large unforced variability in a model might reflect the presence of positive feedbacks that act to amplify variations originating from the chaotic dynamics. Some of these feedbacks might be relevant in the simulation of climate change as well (the response of sea ice to temperature perturbations is an obvious example). Similarly, although the radiative forcing creating the seasonal cycle (i.e., the variation in the solar declination) has both a much larger amplitude and shorter timescale than the forcing due to increased CO_2, the feedbacks determining the amplitude of the temperature response are likely to be partly the same.

Attempts to relate the simulated changes in Northern Europe Land Area mean precipitation to the control run precipitation climate were not successful. The interexperiment correlation between the control run area mean precipitation and the simulated change is very weak in the annual mean, and the seasonal correlations are also insignificant even at the 90% level with the exception of spring (Table 7). The variability of the control run precipitation climate, which was measured by the ratio between the areally averaged standard deviation of 5-year means in the control run and the control run area mean, likewise has no statistically significant correlation with the simulated annual and seasonal changes. Global averages of control run mean precipitation and precipitation variability also turned out to be poor predictors of the simulated precipitation change, both in northern Europe and in the global mean. Moreover, in sharp contrast with temperature, the interexperiment cross correlation between the global annual mean precipitation change and the change in the Northern Europe Land Area is only 0.49 (see Fig. 17a)

Table 7. *As Table 6, but for interexperiment correlation between area mean precipitation changes and three predictors (control run mean precipitation and precipitation variability between 5-year periods, and simulated temperature change). See text for further details.*

	NELA → NELA					G→G	G→ NELA
	DJF	MAM	JJA	SON	Ann (rank)	Ann (rank)	Ann (rank)
Mean precipitation	-0.01	*-0.50*	-0.05	0.34	-0.11 (-0.28)	-0.22 (-0.43)	-0.12 (-0.36)
Internal variability	-0.41	0.22	-0.41	-0.17	-0.43 (-0.31)	0.27 (0.26)	0.33 (0.43)
Temperature change	**0.67**	0.21	**-0.71**	**0.74***	0.46 (0.48)	0.33 (0.39)	0.28 (0.39)
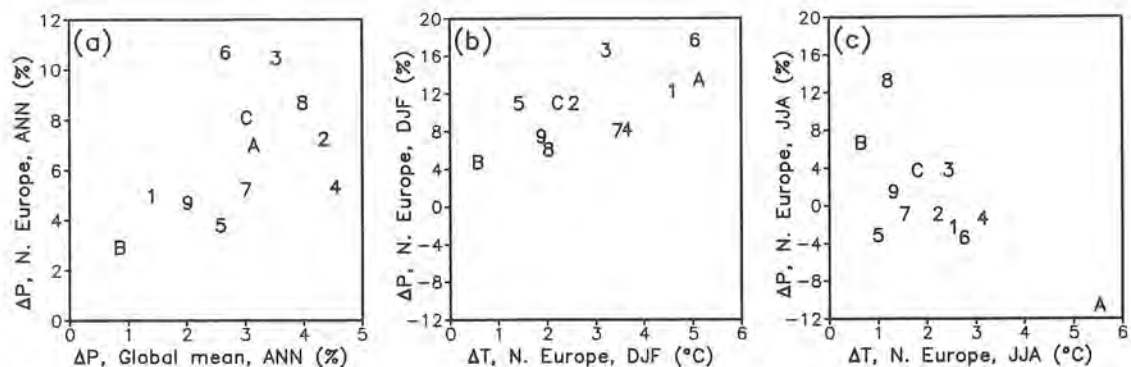
Figure 17. Scatter plots between (a) Global annual mean precipitation change and annual mean precipitation change in northern Europe, (b) temperature change and precipitation change in northern Europe in DJF, and (c) temperature change and precipitation change in northern Europe in JJA. The models are labelled as in Fig. 16. Note that the horizontal and vertical scales in (a) differ and that the vertical scale in (b) and (c) is different from that in (a).

A stronger interexperiment correlation was found between the simulated changes in NELA mean temperature and precipitation. However, the sign of this correlation varies with season (see the last row in Table 7 and Fig. 17b-c). It is strongly positive (about 0.7) and significant at least at the 95% level in winter and autumn, but equally strongly negative and significant in summer. The positive correlation in winter and autumn is as expected from the Clausius-Clayperon relationship. The larger the increase in temperature, the larger the increase in the capacity of air to bring moisture from lower latitudes and from the Atlantic Ocean to northern Europe, where local evaporation in winter is small (e.g., Numaguti 1999). In summer, however, large increases in area mean temperature have a tendency to coincide with decreases or only small increases in precipitation (in particular, NCAR-WM simulates both a very large warming and a substantial decrease in precipitation). This apparent relationship may be explained by feedbacks related to soil moisture and/or cloud cover. For example, when evaporation becomes restricted by a drying out of soil, this tends to induce both a decrease in precipitation and an increase in temperature due to increased sensible heat flux and increased solar radiation allowed by reduced cloudiness (Wetherald and Manabe, 1995; Gregory et al. 1997). This seasonal contrast in the correlation between temperature and precipitation changes was also noted by KGM98.

6.2 Differences between "good" and "bad" models

The correlation coefficients presented above suggest that the simulated time mean control climate is, in general, a poor predictor of the simulated climate changes. However, as correlation analysis only reveals linear connections, there might still be differences between those models that simulate the current climate with high skill and those that simulate it less successfully. Climate changes in those models that are either much above or much below the observed climate in their control simulations might deviate, in principle, to the same direction from the climate changes in those models that have a control climate close to that observed. Even if this were not the case, models with "bad" control climates might have a larger scatter in their simulated climate changes than "good" models.

To study this issue, a subset of six best models was selected out of the 12 and the climate change statistics for these six models were compared with the statistics of the whole 12-model sample. The selection method described below is of course subjectively chosen. There is no unambiguous way to define the "goodness" of control climate, and other equally plausible methods might give somewhat different results.

The selection was based on a total of 30 comparisons between the control simulations and observational estimates of the current climate. For each three variables (temperature, precipitation and sea level pressure), comparisons were made for the four seasonal distributions separately and for the annual mean, and in each case in two domains: northern Europe and the world. These two domains were used since climate changes in northern Europe will be certainly affected by climate changes in the rest of the world (as indicated by the results of the previous subsection) but the local model behaviour in the area of interest is still assumed to deserve some extra attention. The comparisons for sea level pressure included both land and sea areas but those for temperature and precipitation only the Northern Europe Land Area or the domain covered by the CRU data set (New et al. 1999), that is, the global land area excluding Antarctica. As in section 3, biases in global mean sea level pressure were removed. The simple temperature-based correction (1) to the observational estimate of precipitation was also applied (however, the final set of best 6 models was insensitive to excluding this correction).

Table 8. *Ranks averaged over the five time periods (the four seasons and the annual mean) in the comparison between the control simulations and observational estimates of temperature (T), precipitation (P) and sea level pressure (SLP). 'All' refers to ranks averaged over the three variables and 'N+G' to ranks further averaged over the two domains. In each case, the six best models are marked in bold. See text for further details.*

	Northern Europe				Globe				N+G
	T	P	SLP	All	T	P	SLP	All	All
CCC	**5.6**	8.6	7.8	7.3	8.6	7.6	**4.4**	6.9	7.1
CERFACS	**5.0**	9.0	8.6	7.5	6.8	**4.2**	**3.6**	4.9	**6.2**
CSIRO	**3.4**	**5.0**	7.2	**5.2**	3.4	**1.2**	**2.4**	2.3	**3.8**
GFDL	**5.8**	**6.4**	**6.0**	**6.1**	8.8	7.6	9.8	8.7	7.4
GISS	**4.4**	**4.8**	**4.6**	**4.6**	3.8	5.8	9.2	6.3	**5.4**
LMD/IPSL	11.2	7.0	7.6	8.6	9.6	9.6	7.0	8.7	8.7
MPI	7.8	**4.8**	**4.0**	**5.5**	3.6	2.4	6.4	**4.1**	**4.8**
MRI	8.0	9.4	**5.6**	7.7	9.6	10.8	8.0	9.5	8.6
NCAR-CSM	8.2	**2.6**	8.0	6.3	**5.6**	3.8	**3.4**	4.3	**5.3**
NCAR-WM	10.2	10.6	8.4	9.7	11.6	11.8	11.4	11.6	10.7
NRL	6.0	7.0	**5.6**	**6.2**	**5.6**	9.6	10.4	8.5	7.4
UKMO	**2.4**	**2.8**	**4.6**	**3.3**	**1.0**	3.6	**2.0**	2.2	**2.7**

In each comparison, the models were ranked from 1 to 12 according to the area-weighted mean absolute error from the observational estimate. The somewhat more commonly used root-mean-square error yielded in most cases a very similar ranking as the mean absolute error, but it appeared too sensitive to the simulation of strong localized precipitation maxima in the tropics. The ranks were then averaged over all 30 comparisons. The assumption implied by this averaging (equal importance of each variable, each season and the two domains) is naturally impossible to validate.

The six best models picked up by this procedure were UKMO, CSIRO, MPI, NCAR-CSM, GISS and CERFACS. As indicated by Table 8, the same set of models would have been obtained, although in different order, by only using the global domain. If only the local performance in northern Europe had been considered, NCAR-CSM and CERFACS would have been replaced by GFDL and NRL. As expected, the ranking of the models depends on the variable, season and domain considered. Nevertheless, some models performed well (in particular, UKMO) and some poorly (in particular, NCAR-WM) in almost all comparisons.

Table 9. *Comparison between climate change statistics for the six "best" models and the statistics for all 12 models. Mean6 is the NELA mean change averaged over the six models and Δmean the difference between this and the 12-model mean. V6 is the area mean of the local nominally unbiased interexperiment variance for the six experiments and V6/V12 the ratio of this to the 12-model variance. Cases in which the six-model variance is at the 90% (95%) significance level smaller than the 12-model variance are marked in italics (bold). The differences between the 6-model and 12-model area means are in all cases insignificant even at the 90% level.*

	Area mean change				Local interexperiment variance					
	Temperature		Precipitation		Temperature		Precipitation		Pressure	
	Mean6 ($^\circ$C)	Δmean ($^\circ$C)	Mean6 (%)	Δmean (%)	V6 ($^\circ$C^2)	V6/V12	V6 (%^{2})	V6/V12	V6 (hPa^2)	V6/V12
DJF	2.45	-0.53	10.9	0.2	1.12	**0.38**	51	0.76	2.04	0.77
MAM	2.13	-0.16	7.4	0.2	1.30	0.76	50	0.68	0.96	0.81
JJA	1.72	-0.46	0.7	0.1	0.48	**0.25**	96	0.78	0.66	0.84
SON	2.25	-0.22	8.0	-0.8	0.53	*0.46*	44	**0.41**	1.23	1.04
Ann	2.14	-0.34	6.6	0.0	0.64	*0.44*	26	0.89	0.74	1.19

Climate change statistics for the "best" six models are compared with the statistics for all 12 models in Table 9. Two aspects are studied: the northern Europe area mean changes of temperature and precipitation, and area means of the local interexperiment variance (sea level pressure is also included). The significance of the differences between the six-model and 12-model statistics was judged against the distribution obtained by forming all 924 possible six-model combinations. For the difference between the six-model and 12-model area means, a two-sided test for used. In the case of the variance, however, a one-sided significance test was applied (one would assume the variance between the best models to be relatively small and this hypothesis is

supported, at the 90% level, if the six-model interexperiment variance is among the 92 lowest possible values). From this table it is seen that

- The six-model mean warming is slightly smaller than the 12-model mean warming. However, the difference is in no season statistically significant even at the 90% level. A large part of the difference is explained by NCAR-WM being outside the six-model group.
- The six-model area mean changes in precipitation are in all seasons very close to the 12-model mean.
- The six-model variance is smaller than the 12-model variance with the exception of the autumn and annual means of sea level pressure. The decrease in variance is as a whole largest in the case of temperature changes, being at the 95% level significant in winter and summer and at the 90% level significant in autumn and the annual mean. It is also significant at the 95% level in the case of autumn precipitation, but insignificant in the remaining cases.

Thus, this exercise gives little evidence for any systematic differences between the temperature and precipitation changes simulated by those models that perform well in simulating the present climate and those models that perform less well. By contrast, it does suggest a decrease in the interexperiment variance when models with large errors in the control climate are excluded from the comparison. However, the evidence is not convincing because this decrease is statistically significant in only a few cases.

The present method of ranking models is, of course, only one of a large number of possibilities. Some variations of the method were tested, including the use of northern Europe verification statistics alone and the use of only the same variable in the same season. In terms of the interexperiment variance, these alternative methods gave generally worse results (i.e., larger variance) than the original method. This might either indicate that the original method was in some sense optimal or that the promising results given by it were partly fortuitous.

It is naturally possible that models might be ranked more meaningfully by considering some other aspects of their behaviour than just the seasonal and annual means of temperature, precipitation and sea level pressure. For example, the results of the previous subsection suggest that the annual temperature range might be a more useful predictor of the simulated climate changes than the control run mean temperatures as such. The mean absolute error of the annual temperature range yielded, however, the same set of best six models as the original method except that CERFACS was replaced by MRI. This single change had little impact on the six-model statistics.

Finally, the fact that the six best models are in slightly better agreement in their simulated climate changes than all 12 models is unfortunately no proof that these six models actually are close to the reality in their response to increased CO_2.

6.3 Dependence on model resolution and use of flux corrections

The statistical exercises this far discussed in this section have searched to explain the interexperiment differences in given aspects of the simulated climate change in terms of either the control climate or some other aspect of the climate change. A more ambitious goal would be to relate the interexperiment differences in climate change directly to the differences in the models themselves. This is, however, extremely difficult, because the models differ in a large number of ways. Here, only two obvious criteria of classifying the models are considered: the horizontal resolution and the use of flux corrections.

Five of the 12 models (CERFACS, GISS, LMD/IPSL, NCAR-CSM and NCAR-WM) are non-flux-corrected while the other seven use flux corrections (Table 1). In terms of the horizontal resolution, we also divide the models into two groups, with the five models CCC, CERFACS, NCAR-CSM, NRL and UKMO with at least 4000 grid boxes over the world denoted as "high-resolution" models and the other seven as "low-resolution" models. The NELA mean changes in temperature and precipitation were calculated separately, on one hand, for the non-flux-corrected and the flux-corrected models, and on the other hand, for the high- and low-resolution models. The statistical significance of the differences between the two groups in each comparison was then estimated by a permutation procedure similar to that used in the previous subsection.

Table 10. Comparison of Northern Europe Land Area mean temperature and precipitation changes between the five non-flux-corrected (NC) and the seven flux-corrected (FC) models. Cases in which the difference between the two groups is at the 90% (95%) level significant are marked in italics (bold).

	Temperature (°C)			Precipitation (%)		
	NC	FC	NC-FC	NC	FC	NC-FC
DJF	3.21	2.82	0.39	12.2	9.6	2.6
MAM	2.04	2.47	-0.43	7.0	7.3	-0.3
JJA	2.57	1.90	0.67	-3.3	3.3	**-6.6**
SON	2.67	2.33	0.34	12.3	6.3	*6.0*
Ann	2.62	2.38	0.24	6.7	6.5	0.2

The non-corrected models simulate, on the average, slightly stronger annual mean warming in northern Europe than the flux-corrected models (Table 10), and the high-resolution models slightly weaker warming (Table 11) than the low-resolution models. However, these differences are in no season statistically significant even at the 90% level, and the difference between non-corrected and flux-corrected models is actually reversed in spring. The differences in precipitation change between each of the two pairs of groups are also small in most cases. There is a weakly significant (90-95%) difference between the non-corrected and flux-corrected models in summer and autumn, but the sign of the difference is the opposite in these two seasons. Differences in the within-group interexperiment standard deviation between the high- and low-resolution and the non-corrected and flux-corrected models are generally statistically insignificant as well (not shown). At least regarding the aspects studied here, the dependency of the

simulated climate change in northern Europe on model resolution and on the use of flux corrections is therefore quite weak.

Table 11. Comparison of Northern Europe Land Area mean temperature and precipitation changes between five high-resolution (H = CCC, CERFACS, NCAR-CSM, NRL and UKMO) and seven low-resolution (L) models. The statistical significance of the differences is in all cases below 90%.

	Temperature (°C)			Precipitation (%)		
	H	L	H-L	H	L	H-L
DJF	2.36	3.42	-1.06	9.3	11.6	-2.3
MAM	2.16	2.38	-0.22	5.6	8.3	-2.7
JJA	1.71	2.51	-0.80	1.8	-0.3	2.1
SON	1.98	2.83	-0.85	5.8	10.9	-5.1
Ann	2.05	2.79	-0.74	5.6	7.3	-1.7

7 Summary and discussion

The results of 12 coupled atmosphere-ocean general circulation model experiments participating in the second phase of the Coupled Model Intercomparison Project (CMIP2) have been studied with focus on the area of northern Europe and on three variables, surface air temperature, precipitation and sea level pressure. This comparison provides an update to Räisänen (1994; R94) and has included both the control runs and the simulated climate changes due to a transient doubling of CO_2, with more emphasis on the latter. Particular attention has been put on quantifying the interexperiment agreement on the local CO_2-induced climate changes and on estimating the relative contributions of model differences and model-simulated internal variability ("noise") to the interexperiment variance. For that purpose, a new statistical framework was developed. Finally, an attempt was made to statistically relate the interexperiment differences in the simulated climate change in northern Europe to aspects of the control climates, global climate change and some of the basic model characteristics.

The main findings are summarized below and the summary is followed by some further discussion.

7.1 Control climates

The scatter between the individual CMIP2 models extends, in many cases, well to the both sides of observational estimates of the present climate. A few relatively systematic biases were still found:

- The simulated temperatures in spring (March-May) tend to be slightly too cold, in contrast with zero or marginally positive 12-model mean biases in the other seasons.
- The simulated precipitation exceeds, on the average, the CRU climatology in all seasons. This is partly due to measurement errors, but the 12-model mean bias in

spring (72% over the Northern Europe Land Area) is far too large to be explained by this factor alone.

- The wet bias in spring is apparently connected with a negative bias in sea level pressure in particular in the southern half of northern Europe, which suggests that strong cyclone activity in the models extends too far and too south in the continental interior. The models also tend to underestimate the meridional pressure gradient in the other seasons, with more positive biases over the northernmost North Atlantic and the Arctic Ocean than to the south of the 60°N.

Compared with the simulations studied by R94, the CMIP2 control simulations show an overall improvement in winter temperature (systematic cold bias removed) and summer precipitation (severe dry biases no longer present). In many other aspects of the control climate, however, there is no clear distinction between the two sets of simulations.

7.2 CO_2-induced climate changes: general features

The 12-model annual mean warming in response to a transient doubling of CO_2 is, as averaged over the Northern Europe Land Area (NELA), 2.5°C or roughly 30% above the global mean warming. In agreement with R94, the warming is on the average largest in winter (3.0°C) and smallest in summer (2.2°C), but the seasonal cycle varies somewhat between the individual experiments. The average simulated increase in annual area mean precipitation is 7%. The area mean increase is on the average largest in winter (11%) and smallest in summer (1%), when decreases in precipitation actually occur in several models especially in the southern part of the area. The simulations show no consistent pattern of change in sea level pressure in northern Europe and over the northern North Atlantic, except for a vague general decrease in summer and autumn.

In terms of temperature change, three models are substantially outside the range of the others. The largest warming takes place in the NCAR-WM model (in the annual area mean 4.7°C), whereas the average warming in the NRL model is only 0.6°C. These two models are also the two most extreme in terms of the global temperature change. The GISS model simulates a warming of 2°C in the eastern European Russia, but in parts of the Nordic area this experiment indicates a marginal cooling. This reflects a more marked cooling of sea surface temperatures in the northern North Atlantic, which seems to be an unusually strong response to a less unusual simulated weakening of the North Atlantic thermohaline circulation (Russell and Rind 1999). As noted in Section 4.1, the cooling over northwestern Europe is not reproduced in a newer version of the GISS model (Russell et al. 1999), in which even the Atlantic cooling is weaker.

7.3   Quantification of agreement and role of internal variability

Treating the 12 CMIP2 experiments as a sample of an infinite population, a fraction of common signal was calculated as the ratio between the best estimate area means of the square of the population mean climate change and the mean square of climate changes in individual experiments. The seasonal and annual values of this statistics for grid box scale temperature changes in northern Europe varied between 0.71 and 0.84, indicating (despite the few outliers) a good overall agreement. The values for precipitation change were also reasonably high in winter (0.67) and in the annual mean (0.63) when increases

48

in precipitation dominate over land areas of northern Europe, but low (0.05) in summer when the simulated changes are of varying sign in different models. The same statistics indicates some interexperiment agreement on changes in sea level pressure in summer (0.39) and autumn (0.33) but no agreement in winter and spring (slightly negative values). The values of this statistics in northern Europe are similar to the values obtained in a global domain, except that the relative interexperiment agreement on annual and wintertime precipitation changes in northern Europe is well above the global average.

Only a minor part, at most 16% in winter, of the interexperiment variance in temperature change was estimated to be explained by the noise associated with internal variability in the individual experiments. A large majority of the interexperiment differences in temperature change therefore seems to be related to differences in the models themselves. The same fraction for precipitation change was considerably larger, 30-39% depending on the season, but still not dominating. Changes in sea level pressure were as a whole found to be more severely disturbed by internal variability than those in the other two variables, which in part explains the lack of interexperiment agreement on these. It is essential to recall here that climate changes in the present study were defined by using relatively long averaging periods (20 years in perturbation run vs. 80 years in control run). With shorter averaging periods, the effects of internal variability grow larger and the interexperiment agreement is expected to deteriorate. That this deterioration in fact happens was demonstrated in Section 5.5.

As expected, the interexperiment agreement on northern Europe area mean climate changes is generally better than the agreement on the grid box scale changes. In addition, the fraction of common signal was actually found to be slightly higher (0.87 vs. 0.83) for northern Europe annual area mean precipitation than temperature changes. That this is not the case at the grid box scale is explained by the much larger small-scale variability in the simulated precipitation changes and the poor interexperiment agreement on these variations. The 12 CMIP2 experiments in fact have little common in their distribution of climate changes within northern Europe – the fraction of common signal for the deviations from the area mean changes is, at best, about 0.3 for temperature changes in winter, autumn and annual mean.

7.4 Statistical relationships

The interexperiment variations in the northern Europe area mean temperature change are strongly related (correlation in the annual mean 0.88) to the variations in the global mean warming. The temperature change in northern Europe was also found to be positively correlated with the annual temperature range and the amplitude of inter-5-year temperature variability. These correlations are physically plausible but weaker and less statistically significant than the correlation with the global mean warming, and a confirmation of their robustness may still need to wait a study of a larger set of models. The correlation between the simulated warming and control run mean temperature is at the 95% level statistically significant in summer (although this is mainly due to the coincidence of a large warm bias with very strong warming in NCAR-WM) but weak in the other seasons and in the annual mean.

49

The seasonal and annual interexperiment differences in northern Europe area mean precipitation change were in general found to be poorly predicted by characteristics of the control run precipitation climate. The 12 CMIP2 experiments have, however, a stronger but seasonally varying correlation between the area mean temperature and precipitation changes. The positive correlations of about 0.7 in autumn and winter agree with the Clausius-Clayperon relationship, whereas the negative correlation of about -0.7 in summer may reflect feedbacks associated with soil moisture and cloudiness.

The control climates were also used to classify the models as good and bad, on the basis of an objective comparison with observational data. The selected six best models had in almost all respects a somewhat smaller variance in their simulated climate changes than all 12 models (by contrast, there were no statistically significant differences in the average temperature and precipitation changes). A caveat here is that there is no unambiguous way of defining good control climate and the objective selection method itself therefore necessarily includes subjective choices. The promising result might therefore be partly fortuitous.

The seasonal and annual differences between the flux-corrected and non-flux-corrected models in the simulated temperature and precipitation change were generally statistically insignificant. The dependency of the simulated climate changes on horizontal model resolution was likewise found to be weak.

7.5   Discussion

The present study provides a relatively detailed intercomparison of model-simulated CO_2-induced changes in surface climate in northern Europe, and its framework for quantifying the interexperiment agreement and estimating the relative importance of internal variability as a source of disagreement is expected to be useful for other studies as well. It has not, however, provided any ultimate physical explanations to why different models simulate different climate changes to the extent that this is not explained simply by internal variability. Attacking this crucial question will require much more in-depth analysis of model behaviour than is actually possible with time series of a few surface variables. Progress in physical understanding of the interexperiment differences must be based partly on detailed diagnostic studies of atmospheric and oceanic dynamics (or, for example, surface and atmospheric water and energy balance), and partly on additional model experiments. The most direct way to study the link between the simulated climate response and model structure would be to repeat climate change experiments with general circulation models that are identical apart from single well-specified changes, but this is an extremely expensive approach.

At any event, a substantial narrowing of the uncertainties associated with greenhouse gas induced regional climate change is still likely to require (at least) several years of intensive research. Meanwhile, those that apply the model results for estimating the practical impacts of climate change need a sound strategy for handling these uncertainties. The present study indicates that using an ensemble of simulations from a single model is not a satisfactory approach, at least not for time periods around the

doubling of CO_2, when the uncertainty associated with internal climate variability is generally dominated by the effect of model differences. Closer in the future, however, the relative importance of internal variability will be larger. In any case, it is vital to understand that internal variability is not only something that complicates the interpretation of model results – this essentially unpredictable component of climate change will also be present in the real world (Hulme et al. 1999).

Using climate change simulations from several different models implicitly accounts for both the internal variability and the impact of model differences. However, even this approach will underestimate the total uncertainty about future climate change if all the experiments share the same forcing scenario. The standard 1% per year increase in CO_2 in the CMIP2 simulations automatically hides the uncertainty about the real future rate of increase in CO_2 and other well-mixed greenhouse gases (which is in part caused by the lacking knowledge of future emissions). As importantly, other sources of anthropogenic (such as sulphate aerosols) and natural climate forcing (solar variability and volcanic activity) are totally absent in these experiments.

Whether the scatter between the 12 CMIP2 experiments is realistic in a more narrow sense, in indicating the "probability distribution" for the response of the real climate system to a doubling of CO_2, is a more difficult question. Taking a very optimistic view, this scatter might even be too large. This line of thinking is supported by the observation (Section 6.2) that screening out models with "bad" control climates in fact seems to lead to a slight decrease in the scatter. On the other hand, the possible systematic errors that the 12 models share in their response to increased CO_2 are not revealed in an intermodel comparison. The biases in the control simulations appear to be with some exceptions more random than systematic by nature, but this does not prove that the same would be true for the simulated climate changes. In developing and tuning climate models, use can be made of the knowledge about the current observed climate but not of the lacking knowledge about the response of climate to increased CO_2.

Acknowledgements

Appendix A. Notations

Let f be a physical variable (e.g., local or area mean surface air temperature) and x a simulated climate change (difference between the time means of f in a perturbation run and a control run). Averages related to these are indicated as

$$\overline{f} = \frac{1}{T}\sum_{t=1}^{T} f_t = \text{time mean of } f \text{ over a total of } T \text{ samples with constant time interval}$$

$$[x] = \frac{1}{k}\sum_{i=1}^{k} x_i = \text{mean climate change in a sample of } k \text{ experiments}$$

$$\{x\} = \lim_{k\to\infty}[x] = \text{mean climate change in a hypothetical, infinite population of}$$
$$\text{experiments.}$$

Further, letting X be a statistics describing the distribution of climate changes in different experiments, the following additional notations are used:

X_{EXP} = expected value of X for a sample of k experiments drawn
   from a population whose statistical properties are assumed to be known

X_{EST} = value of X for an infinite population of experiments, estimated from a
   sample of k experiments

$\overset{\approx}{X}$ = area mean of the local statistics.

Appendix B. Climate drift in control runs

Some of the CMIP2 control runs show symptoms of climate drift – trend-like behaviour that suggests a conflict between the internal model dynamics and the initial conditions of the control run. Such drift is generally most clearly seen in surface air temperature, but it is in some cases evident in other parameters as well (for example, high-latitude sea areas with a strong warming trend also tend to have an increasing trend in precipitation).

For a simple indicator of climate drift in the different models, the best-fit linear trends in annual mean surface air temperature averaged over the whole globe, the Northern Europe Land Area and the Nordic Subregion are given in the left part of Table B1. NRL with only a 3-year control run is excluded. In most experiments, the global mean trend is positive, and in three models (CCC, CERFACS and GISS) it is statistically significant even at the 99% level (however, the applied significance test neglects autocorrelation between different 5-year periods, which may make the estimated significance levels too high). The largest global mean trend, 0.63°C per 80 years, occurs in the GISS experiment and is mainly associated with strong warming around the Antarctic continent (locally up to 15°C). More generally, the largest local trends occur in high-latitude sea areas, where simulated temperatures are very sensitive to changes in ice cover. Problems are more common in high southern latitudes than in the Northern

Hemisphere, but substantial trends also occur in a few models in areas surrounding Greenland and in the northern North Atlantic. As expected, the trends are generally larger in the five models without flux corrections (mean absolute value of global mean trend 0.37°C) than in the six flux-corrected models (0.07°C)

Although positive trends dominate in the global mean, the 11-model averages of trends in the Northern Europe Land Area and in the Nordic Subregion are very close to zero. The largest positive trends in these areas occur in NCAR-WM and CERFACS, while smaller negative trends exist in most of the other simulations. Finally, although the trends averaged over these areas have in many cases a larger magnitude than the global mean trends, they have generally much lower statistical significance. Excluding the NCAR-WM and CERFACS experiments, none of these regional trends is significant even at the 90% level.

Table B1. *Left: best-fit linear trends in control run annual mean surface air temperature, expressed in °C per the full control run length (NCAR-WM 75 years, other models 80 years). Trends shown in italics (bold) are significant at the 90% (95%) level, and those with a star have a significance of at least 99%. Right: differences in temperature between the last 20 years of the control run and the whole control run. The trends and the differences are both given for the whole globe, the Northern Europe Land Area (NELA) and the Nordic Subregion.*

	Linear trend			Mean 61-80 – mean 1-80		
	Global	NELA	Nordic	Global	NELA	Nordic
CCC	0.16*	-0.27	-0.34	0.09	0.12	0.00
CERFACS	0.44*	1.05	0.82	0.22	0.52	0.61
CSIRO	-0.04	-0.12	-0.16	-0.03	-0.20	-0.39
GFDL	0.06	-0.16	-0.19	0.04	-0.08	-0.06
GISS	0.63*	0.07	-0.49	0.25	-0.01	-0.15
LMD/IPSL	0.26	-0.07	-0.83	0.00	-0.26	-0.54
MPI	0.06	-0.36	-0.33	0.05	-0.11	-0.27
MRI	0.11	0.05	0.47	-0.02	-0.11	0.08
NCAR-CSM	-0.06	-0.57	-0.24	-0.02	-0.14	0.02
NCAR-WM	0.46	1.08	1.33	0.18	0.42	0.53
UKMO	0.01	-0.21	-0.29	0.02	-0.14	-0.27
11 models	0.19	-0.04	-0.02	0.07	0.00	-0.04

As argued in the introductory part of Section 4, the effect of climate drift on the calculated climate changes might have been reduced by using (in analogy with the perturbation runs) the last 20 rather than the full 80 years of the control runs in defining the control climate. As expected, these 20 years are generally warmer (colder) than the 80-year mean in models with a positive (negative) trend (right part of Table B1). The global 11-model mean difference between the 20-year and 80-year means is 0.07°C. The average calculated global mean warming would therefore have been slightly smaller if the 20-year rather than the 80-year period would have been used to define the control climate. In the Northern Europe Land Area and the Nordic Subregion, the differences

between the two periods are less systematic. Judging from the low statistical significance of the trends in these areas, these differences probably reflect more frequently normal internal variability in the models than climate drift.

Appendix C. Estimates of internal variability

The estimates of internal variability used in sections 4 and 5 are based, excluding the NRL experiment, on the variance between the individual 5-year seasonal or annual means in the control runs. In Section 5.2, three sources of systematic error were identified that may bias this estimate. To study the importance of these, the results of this standard method (hereafter M1) were compared with three alternative estimates:

- M2: As M1, but the variance of the 5-year means ($v5_i$) was estimated after first detrending the control runs (i.e., (20) was replaced with (21)). In the presence of climate drift, M2 is expected to give a lower estimate of variance than M1.

- M3: As M2, but five 15-year means instead of 16 5-year means were used as the basis of estimating the variance. (21) was replaced with

$$v5_i = 3 * \frac{5}{5-2} \overline{g15_i^2}$$ (C.1)

where $g15_i$ is the deviation from the best-fit linear trend calculated from the five 15-year means. If there is positive autocorrelation between subsequent 5-year periods, M3 is expected to give a higher estimate of variance than M2.

- M4: As M2, but detrended perturbation runs instead of detrended control runs were used (for NRL with only a 3-year control simulation this estimate was used in all cases). Differences in internal variability between the control run and the perturbation run are expected to be reflected as differences between the estimates M4 and M2.

One should note, though, that the detrending in M2–M4 only removes linear trends. If climate drift in the control runs is not linear with time, some effects of this will be present even in M2 and M3. Similarly, if the "signal" of CO_2-induced climate change is not evolving linearly with time, M4 will overestimate the variability in the perturbation runs (although we expect this to be only a minor problem). In addition, all of M1-M4 are affected by sampling uncertainty, in particular M3 in which the local estimate of variance in any one model has only three degrees of freedom.

We only discuss the results of the four methods at the most general level, that is, how they differ in the resulting estimates of "population mean" internally generated interexperiment variance N^2. Table C1 shows the estimates M2–M4 as averaged over two areas (northern Europe and the whole world) and normalized by M1.

M2 gives slightly lower estimates of variability than M1. The differences are generally small, but the globally averaged variability of surface air temperature is a notable

54

exception. In this case, the detrended variability estimate is substantially lower than the standard estimate, in the annual mean and in JJA by over 30%. The difference is largest in high-latitude sea areas, particularly near the margin of the Antarctic continent. In these areas, large local temperature trends occur in several models, evidently as a result of changes in sea ice (see Appendix B for further discussion).

Table C1. *Area means of the statistics N^2 obtained by methods M2, M3 and M4, divided by the value given by method M1 (see text for the definition of the methods).*

a) N^2 averaged over the Northern Europe Land Area (surface air temperature and precipitation) or over the whole area 30°W-60°E, 50°-75°N (sea level pressure).

	Temperature					Precipitation					Sea level pressure				
	DJF	MAM	JJA	SON	Ann	DJF	MAM	JJA	SON	Ann	DJF	MAM	JJA	SON	Ann
M2	0.99	0.96	0.94	0.96	0.94	1.00	0.99	0.99	0.99	0.97	0.98	0.98	0.95	0.99	0.95
M3	1.09	0.92	1.38	0.94	0.97	0.97	0.91	1.19	1.04	1.16	0.89	1.14	1.15	1.12	1.07
M4	0.84	0.84	1.08	1.07	1.04	1.10	1.07	1.05	1.15	1.08	1.07	0.90	1.06	1.11	1.04

b) N^2 averaged over the whole world.

	Temperature					Precipitation					Sea level pressure				
	DJF	MAM	JJA	SON	Ann	DJF	MAM	JJA	SON	Ann	DJF	MAM	JJA	SON	Ann
M2	0.84	0.77	0.68	0.79	0.68	0.99	0.99	0.98	0.99	0.98	0.98	0.97	0.97	0.97	0.92
M3	1.07	1.01	0.96	1.00	1.01	1.00	1.01	1.00	0.96	1.03	0.98	0.96	1.08	0.96	1.03
M4	0.89	0.78	0.72	0.83	0.75	1.08	1.05	1.09	1.05	1.09	0.99	0.96	1.03	0.93	0.99

M3 gives in most cases a higher estimate of globally averaged N^2 than M2. The differences in estimates of precipitation and pressure variability are relatively small, but the estimates of temperature variability given by M3 are similar to M1 and, therefore, much higher than M2. This implies that autocorrelation between subsequent 5-year periods is not negligible in the control simulations (although, even after the detrending, some part of this autocorrelation might result from climate drift). The absolute differences between M3 and M2 are largest in high-latitude sea areas, but the relative difference is substantial even over many tropical and mid-latitude ocean regions. Differences over land are less systematic. The contrast between land and sea areas is in line with earlier studies of long model simulations (Manabe and Stouffer 1996; Tett et al. 1997) that have found the variability of surface air temperature to have a redder spectrum over the oceans than over the continents.

In northern Europe, the differences between M3 and M2 (or M1) are unsystematic, which might in part reflect the large sampling uncertainty in M3. In JJA, however, M3 gives a distinctly higher estimate of temperature variability in northern Europe than either of M1 and M2 (to a lesser extent, this is also the case with precipitation and sea level pressure in this season). The difference is to the same direction in all models except for MRI, but its physical explanation is not clear. Judging from the fact that the

largest difference occurs in the southeastern part of northern Europe, far from oceans, long time-scale processes in soil models might possibly play a role.

Finally, the results of method M4 for temperature and pressure variability are relatively close to M2, although the globally averaged temperature variability in M4 is slightly larger in all seasons. This difference might well result from small temporal nonlinearities in the genuinely CO_2-induced temperature change. In northern Europe, M4 actually gives a lower estimate of temperature variability in DJF and MAM than any of the other three methods. This might indicate that, in a warmer climate with less snow and ice, positive feedbacks associated with the cryosphere are less efficient in amplifying temperature variability. On the other hand, both in northern Europe and globally, the estimate M4 for precipitation variability is systematically higher than M1 and M2. This is at least partly explained by the general increase in precipitation in a simulated warmer climate. When time mean precipitation increases, the absolute variance will also increase even with no change in relative variability (note that the estimates of precipitation variability are in all cases normalized by the control run mean precipitation).

At any event, the results of M1 are generally reasonably close to the results of the other three methods considered here. The differences are in relative terms largest in the case of surface air temperature, which is, however, the variable whose changes are least strongly affected by internal variability.

References

Barthelet, P., L. Terray, and S. Valcke, 1998: Transient CO_2 experiment using the ARPEGE/OPAICE non flux corrected coupled model. *Geophys. Res. Lett.*, **25**, 2277-2280.

Boville, B. A., and P. R. Gent, 1998: The NCAR Climate System Model, Version One. *J. Climate*, **11**, 1115-1130.

Braconnot, P., O. Marti, and S. Joussaume, 1997: Adjustment and feedbacks in a global coupled ocean-atmosphere model. *Climate Dyn.*, **13**, 507-519.

Christensen, O. B., J. H. Christensen, B. Machenhauer, and M. Botzet, 1998: Very high-resolution regional climate simulations over Scandinavia – present climate. *J. Climate*, **11**, 3204-3229.

Corti, S., F. Molteni, and T. N. Palmer, 1999: Signature of recent climate change in frequencies of natural atmospheric circulation regimes. *Nature*, **398**, 799-802.

Cubasch, U., K. Hasselmann, H. Höck, E. Maier-Reimer, U. Mikolajewics, B. D. Santer, and R. Sausen, 1992: Time-dependent greenhouse warming computations with a coupled ocean-atmosphere model. *Clim. Dyn.*, **8**, 55-69.

Fyfe, J. C., and G. M. Flato, 1999: Enhanced climate change and its detection over the Rocky mountains. *J. Climate*, **12**, 230-243.

Gates, W. L., 1992: AMIP: the Atmospheric Model Intercomparison Project. *Bull. Am. Meteorol. Soc.*, **73**, 1962-1970.

Gibson, J. K., P. Kållberg, S. Uppala, A. Hernandez, A. Nomura, and E. Serrano, 1997: ERA description. ECMWF Reanalysis Project Rep. Series 1, European Centre for Medium-Range Weather Forecasts, Reading, United Kingdom, 66 pp.

Gordon, H. B., and S. P. O'Farrell, 1997: Transient climate change in the CSIRO coupled model with dynamic sea ice. *Mon. Wea. Rev.*, **125**, 875-907.

Gregory, J. M., J. F. B. Mitchell, and A. J. Brady, 1997: Summer drought in northern midlatitudes in a time-dependent CO_2 climate experiment. *J. Climate*, **10**, 662-686.

Groisman, P. Ya., V. V. Koknaeva, T. A. Belokrylova, and T. R. Karl, 1991: Overcoming biases of precipitation measurement: a history of the USSR experience. *Bull. Am. Meteorol. Soc.*, **72**, 1725-1733.

Grotch, S. L., and M. C. MacCracken, 1991: The use of general circulation models to predict regional climate change. *J. Climate*, **4**, 286-303.

Hirst, A., S. P. O'Farrell, and H. B. Gordon, 1999: Comparison of a coupled ocean-atmosphere model with and without oceanic eddy-induced advection. *J. Climate* (in press).

Houghton, J. T., L. G. Meira Filho, B. A. Callander, N. Harris, A. Kattenberg, and K. Maskell, Eds., 1996: *Climate change 1995. The Science of Climate Change*. Cambridge University Press, 572 pp.

Hulme, M., D. Conway, P. D. Jones, T. Jiang, E. M. Barrow, and C. Turney, 1995: Construction of a 1961-1990 European climatology for climate change modelling and impact applications. *Int. J. Climatol.*, **15**, 1333-1363.

Hulme, M., E. M. Barrow, N. W. Arnell, P. A. Harrison, T. C. Johns, and T. E. Downing, 1999: Relative impacts of human-induced climate change and natural climate variability. *Nature*, **397**, 688-691.

Hurrell, J. W., and H. van Loon, 1997: Decadal variations in climate associated with the North Atlantic Oscillation. *Clim. Change*, **36**, 301-326.

Johns, T. C., 1996: A Description of the Second Hadley Centre Coupled Model (HADCM2). Climate Research Technical Note 71, Hadley Centre, United Kingdom Meteorological Office, Bracknell, Berkshire RG12 2SY, United Kingdom, 26 pp.

Johns, T. C, R. E. Carnell, J. F. Crossley, J. M. Gregory, J. F. B. Mitchell, C. A. Senior, S. F. B. Tett, and R. A. Wood, 1997: The second Hadley Centre Coupled ocean-atmosphere GCM: Model description, spinup and validation. *Climate Dyn.*, **13**, 103-134.

Kattenberg, A., F. Giorgi, H. Grassl, G. A. Meehl, J. F. B. Mitchell, R. J. Stouffer, T. Tokioka, A. J. Weaver, and T. M. L. Wigley, 1996: Climate models – projections of future climate. *Climate change 1995. The Science of Climate Change*, J. T. Houghton, L. G. Meira Filho, B. A. Callander, N. Harris, A. Kattenberg, and K. Maskell, Eds., Cambridge University Press, 285-357.

Kittel, T. G. F., F. Giorgi, and G. A. Meehl, 1998: Intercomparison of regional biases and doubled CO_2-sensitivity of coupled atmosphere-ocean general circulation experiments. *Clim. Dyn.*, **14**, 1-15.

Li, T. F., and T. F. Hogan, 1999: The role of the annual mean climate on seasonal and interannual variability of the Tropical Pacific in a coupled GCM. *J. Climate*, **12**, 780-792.

Lunkeit, F., M. Ponater, R. Sausen, M. Sogalla, U. Ulbrich, and M. Windelband, 1996: Cyclonic activity in a warmer climate. *Contrib. Atmos. Phys.*, **69**, 393-407.

Manabe, S., and R. J. Stouffer, 1996: Low-frequency variability of surface air temperature in a 1000-year integration of a coupled atmosphere-ocean-land surface model. *J. Climate*, **9**, 376-393.

Manabe, S., R. J. Stouffer, M. J. Spelman, and K. Bryan, 1991: Transient responses of a coupled ocean-atmosphere model to gradual changes of atmospheric CO_2. Part I: Annual mean response. *J. Climate*, **4**, 785-818.

Meehl, G. A., and W. M. Washington, 1995: Cloud albedo feedback and the super greenhouse effect in a global coupled GCM. *Climate Dyn.*, **11**, 399-411.

Meehl, G. A., G. J. Boer, C. Covey, M. Latif, and R. J. Stouffer, 1997: Intercomparison makes for a better climate model. *Eos*, **78**, 445-446, 451.

Meehl, G. A., G. J. Boer, C. Covey, M. Latif, and R. J. Stouffer, 1999: The Coupled Model Intercomparison Project (CMIP) workshop. *Bull. Am. Meteorol. Soc.* (submitted).

Mitchell, J. F. B., and T. C. Johns, 1997: On modification of global warming by sulphate aerosols. *J. Climate*, **10**, 245-267.

Mitchell, J. F. B., C. A. Senior, and W. J. Ingram, 1989: CO_2 and climate: a missing feedback? *Nature*, **341**, 132-134.

Murphy, J. M., 1995: Transient response of the Hadley Centre coupled ocean-atmosphere model to increasing carbon dioxide. Part I: Control climate and flux adjustment. *J. Climate*, **8**, 36-56.

Murphy, J. M., and J. F. B. Mitchell, 1995: Transient response of the Hadley Centre coupled ocean-atmosphere model to increasing carbon dioxide. Part II: Spatial and temporal structure of response. *J. Climate*, **8**, 57-80.

New, M., M. Hulme, and P. Jones, 1999: Representing twentieth-century space-time climate variability. Part I: Development of a 1961-90 mean monthly terrestrial climatology. *J. Climate*, **12**, 829-856.

Numaguti, A., 1999: Origin and recycling processes of precipitating water over the Eurasian continent: Experiments using an atmospheric general circulation model. *J. Geophys. Res.*, **104**, 1957-1972.

Raab, B., and H. Vedin, Eds., 1995: Climate, Lakes and Rivers. National Atlas of Sweden, vol 14, 176 pp.

Russell, G. L., and D. Rind, 1999: Response to CO_2 transient increase in the GISS coupled model: Regional coolings in a warmer climate. *J. Climate*, **12**, 531-539.

Russell, G. L., J. R. Miller, and D. Rind, 1995: A coupled atmosphere-ocean model for transient climate change studies. *Atmos.-Ocean*, **33**, 683-730.

Russell, G. L., J. R. Miller, D. Rind, R. A. Ruedy, G. A. Schmidt, and S. Sheth, 1999: Comparison of model and observed regional temperature changes during the past 40 years. *J. Geophys. Res.* (submitted).

Räisänen, J., 1994: A comparison of the results of seven GCM experiments in northern Europe. *Geophysica*, **11**, 3-30.

Räisänen, J., 1997: Objective comparison of patterns of CO_2 induced climate change in coupled GCM experiments. *Clim. Dyn.*, **13**, 197-211.

Räisänen, J., 1998: CO_2- and aerosol-induced changes in vertically integrated zonal momentum budget in a GCM experiment. *J. Climate*, **11**, 625-639.

Shindell, D. T., R. L. Miller, G. A. Schmidt, and L. Pandolfo, 1999: Simulation of recent northern winter climate trends by greenhouse-gas forcing. *Nature*, **399**, 452-455.

Tett, S. F. B., T. C. Johns, and J. F. B. Mitchell, 1997: Global and regional variability in a coupled AOGCM. *Clim. Dyn.*, **13**, 303-323.

Thomson, D. W. J., and J. M. Wallace, 1998: The Arctic Oscillation signature in the wintertime geopotential height and temperature fields. *Geophys. Res. Lett.*, **25**, 1297-1300.

Tokioka, T., A. Noda, A. Kitoh, Y. Nikaidou, S. Nakagawa, T. Motoi, S. Yukimoto, and K. Takata, 1995: A transient CO_2 experiment with the MRI CGCM. Quick Report. *J. Meteorol. Soc. Japan*, **73**(4), 817-826.

Voss, R., R. Sausen, and U. Cubasch, 1998: Peridiocally synchronously coupled integrations with the atmosphere-ocean general circulation model ECHAM3/LSG. *Climate Dyn.*, **14**, 249-266.

Washington, W. M., and G. A. Meehl, 1989: Climate sensitivity due to increased CO_2: experiments with a coupled atmosphere and ocean general circulation model. *Climate Dyn.*, **4**, 1-38.

Washington, W. M., and G. A. Meehl, 1996: High-latitude climate change in a global coupled ocean-atmosphere-sea ice model with increased atmospheric CO_2. *J. Geophys. Res.*, **101**(D8), 12795-12801.

Watanabe, M., and T. Nitta, 1999: Decadal changes in the atmospheric circulation and associated surface climate variations in the Northern Hemisphere winter. *J. Climate*, **12**, 494-510.

Wetherald, R. T., and S. Manabe, 1995: The mechanisms of summer dryness induced by greenhouse warming. *J. Climate*, **8**, 3096-3108.

Whetton, P. H., P. J. Rayner, A. B. Pittock, and M. R. Haylock, 1994: An assessment of possible climate change in the Australian region based on an intercomparison of general circulation modeling results. *J. Climate*, **7**, 441-463.

SMHIs publications

SMHI publishes six report series. Three of these, the R-series, are intended for international readers and are in most cases written in English. For the others the Swedish language is used.

Names of the Series	Published since
RMK (Report Meteorology and Climatology)	1974
RH (Report Hydrology)	1990
RO (Report Oceanography)	1986
METEOROLOGI	1985
HYDROLOGI	1985
OCEANOGRAFI	1985

Earlier issues published in serie RMK

1 Thompson, T., Udin, I., and Omstedt, A. (1974)
Sea surface temperatures in waters surrounding Sweden.

2 Bodin, S. (1974)
Development on an unsteady atmospheric boundary layer model.

3 Moen, L. (1975)
A multi-level quasi-geostrophic model for short range weather predictions.

4 Holmström, I. (1976)
Optimization of atmospheric models.

5 Collins, W.G. (1976)
A parameterization model for calculation of vertical fluxes of momentum due to terrain induced gravity waves.

6 Nyberg, A. (1976)
On transport of sulphur over the North Atlantic.

7 Lundqvist, J.-E., and Udin, I. (1977)
Ice accretion on ships with special emphasis on Baltic conditions.

8 Eriksson, B. (1977)
Den dagliga och årliga variationen av temperatur, fuktighet och vindhastighet vid några orter i Sverige.

9 Holmström, I., and Stokes, J. (1978)
Statistical forecasting of sea level changes in the Baltic.

10 Omstedt, A., and Sahlberg, J. (1978)
Some results from a joint Swedish-Finnish sea ice experiment, March, 1977.

11 Haag, T. (1978)
Byggnadsindustrins väderberoende, seminarieuppsats i företagsekonomi, B-nivå.

12 Eriksson, B. (1978)
Vegetationsperioden i Sverige beräknad från temperaturobservationer.

13 Bodin, S. (1979)
En numerisk prognosmodell för det atmosfäriska gränsskiktet, grundad på den turbulenta energiekvationen.

14 Eriksson, B. (1979)
Temperaturfluktuationer under senaste 100 åren.

15 Udin, I., och Mattisson, I. (1979)
Havsis- och snöinformation ur datorbear-
betade satellitdata - en modellstudie.

16 Eriksson, B. (1979)
Statistisk analys av nederbördsdata. Del I.
Arealnederbörd.

17 Eriksson, B. (1980)
Statistisk analys av nederbördsdata. Del II.
Frekvensanalys av månadsnederbörd.

18 Eriksson, B. (1980)
Årsmedelvärden (1931-60) av nederbörd, av-
dunstning och avrinning.

19 Omstedt, A. (1980)
A sensitivity analysis of steady, free floating
ice.

20 Persson, C., och Omstedt, G. (1980)
En modell för beräkning av luftföroreningars
spridning och deposition på mesoskala.

21 Jansson, D. (1980)
Studier av temperaturinversioner och vertikal
vindskjuvning vid Sundsvall-Härnösands
flygplats.

22 Sahlberg, J., and Törnevik, H. (1980)
A study of large scale cooling in the Bay of
Bothnia.

23 Ericson, K., and Hårsmar, P.-O. (1980)
Boundary layer measurements at Klock-rike.
Oct. 1977.

24 Bringfelt, B. (1980)
A comparison of forest evapotranspiration
determined by some independent methods.

25 Bodin, S., and Fredriksson, U. (1980)
Uncertainty in wind forecasting for wind po-
wer networks.

26 Eriksson, B. (1980)
Graddagsstatistik för Sverige.

27 Eriksson, B.(1981)
Statistisk analys av nederbördsdata. Del III.
200-åriga nederbördsserier.

28 Eriksson, B. (1981)
Den "potentiella" evapotranspirationen i
Sverige.

29 Pershagen, H. (1981)
Maximisnödjup i Sverige (perioden
1905-70).

30 Lönnqvist, O. (1981)
Nederbördsstatistik med praktiska tillämp-
ningar.
(Precipitation statistics with practical appli-
cations.)

31 Melgarejo, J.W. (1981)
Similarity theory and resistance laws for the
atmospheric boundary layer.

32 Liljas, E. (1981)
Analys av moln och nederbörd genom
automatisk klassning av AVHRR-data.

33 Ericson, K. (1982)
Atmospheric boundary layer field experiment
in Sweden 1980, GOTEX II, part I.

34 Schoeffler, P. (1982)
Dissipation, dispersion and stability of
numerical schemes for advection and dif-
fusion.

35 Undén, P. (1982)
The Swedish Limited Area Model. Part A.
Formulation.

36 Bringfelt, B. (1982)
A forest evapotranspiration model using sy-
noptic data.

37 Omstedt, G. (1982)
Spridning av luftförorening från skorsten i
konvektiva gränsskikt.

38 Törnevik, H. (1982)
An aerobiological model for operational
forecasts of pollen concentration in the air.

39 Eriksson, B. (1982)
Data rörande Sveriges temperaturklimat.

40 Omstedt, G. (1984)
An operational air pollution model using
routine meteorological data.

41 Persson, C., and Funkquist, L. (1984)
Local scale plume model for nitrogen
oxides. Model description.

42 Gollvik, S. (1984)
Estimation of orographic precipitation by dynamical interpretation of synoptic model data.

43 Lönnqvist, O. (1984)
Congression - A fast regression technique with a great number of functions of all predictors.

44 Laurin, S. (1984)
Population exposure to SO and NO_x from different sources in Stockholm.

45 Svensson, J. (1985)
Remote sensing of atmospheric tempera-ture profiles by TIROS Operational Vertical Sounder.

46 Eriksson, B. (1986)
Nederbörds- och humiditetsklimat i Sverige under vegetationsperioden.

47 Taesler, R. (1986)
Köldperioden av olika längd och förekomst.

48 Wu Zengmao (1986)
Numerical study of lake-land breeze over Lake Vättern, Sweden.

49 Wu Zengmao (1986)
Numerical analysis of initialization procedure in a two-dimensional lake breeze model.

50 Persson, C. (1986)
Local scale plume model for nitrogen oxides. Verification.

51 Melgarejo, J.W. (1986)
An analytical model of the boundary layer above sloping terrain with an application to observations in Antarctica.

52 Bringfelt, B. (1986)
Test of a forest evapotranspiration model.

53 Josefsson, W. (1986)
Solar ultraviolet radiation in Sweden.

54 Dahlström, B. (1986)
Determination of areal precipitation for the Baltic Sea.

55 Persson, C. (SMHI), Rodhe, H. (MISU), De Geer, L.-E. (FOA) (1986)
The Chernobyl accident - A meteorological analysis of how radionucleides reached Sweden.

56 Persson, C., Robertson, L. (SMHI), Grennfelt, P., Kindbom, K., Lövblad, G., och Svanberg, P.-A. (IVL) (1987)
Luftföroreningsepisoden över södra Sverige 2 - 4 februari 1987.

57 Omstedt, G. (1988)
An operational air pollution model.

58 Alexandersson, H., Eriksson, B. (1989)
Climate fluctuations in Sweden 1860 - 1987.

59 Eriksson, B. (1989)
Snödjupsförhållanden i Sverige - Säsongerna 1950/51 - 1979/80.

60 Omstedt, G., Szegö, J. (1990)
Människors exponering för luftföroreningar.

61 Mueller, L., Robertson, L., Andersson, E., Gustafsson, N. (1990)
Meso-γ scale objective analysis of near surface temperature, humidity and wind, and its application in air pollution modelling.

62 Andersson, T., Mattisson, I. (1991)
A field test of thermometer screens.

63 Alexandersson, H., Gollvik, S., Meuller, L. (1991)
An energy balance model for prediction of surface temperatures.

64 Alexandersson, H., Dahlström, B. (1992)
Future climate in the Nordic region - survey and synthesis for the next century.

65 Persson, C., Langner, J., Robertson, L. (1994)
Regional spridningsmodell för Göteborgs och Bohus, Hallands och Älvsborgs län. (A mesoscale air pollution dispersion model for the Swedish west-coast region. In Swedish with captions also in English.)

66 Karlsson, K.-G. (1994)
Satellite-estimated cloudiness from NOAA AVHRR data in the Nordic area during 1993.

67 Karlsson, K-G. (1996)
Cloud classifications with the SCANDIA model.

68 Persson, C., Ullerstig, A. (1996)
Model calculations of dispersion of lindane over Europe. Pilot study with comparisons to measurements around the Baltic Sea and the Kattegat.

69 Langner, J., Persson, C., Robertson, L., and Ullerstig, A. (1996)
Air pollution Assessment Study Using the MATCH Modelling System. Application to sulfur and nitrogen compounds over Sweden 1994.

70 Robertson, L., Langner, J., Engardt, M. (1996)
MATCH - Meso-scale Atmosperic Transport and Chemistry modelling system.

71 Josefsson, W. (1996)
Five years of solar UV-radiation monitoring in Sweden.

72 Persson, C., Ullerstig, A., Robertson, L., Kindbom, K., Sjöberg, K. (1996)
The Swedish Precipitation Chemistry Network. Studies in network design using the MATCH modelling system and statistical methods.

73 Robertson, L. (1996)
Modelling of anthropogenic sulfur deposition to the African and South American continents.

74 Josefsson, W. (1996)
Solar UV-radiation monitoring 1996.

75 Häggmark, L., Ivarsson, K.-I. (SMHI), Olofsson, P.-O. (Militära vädertjänsten). (1997)
MESAN - Mesoskalig analys.

76 Bringfelt, B, Backström, H, Kindell, S, Omstedt, G, Persson, C, Ullerstig, A. (1997)
Calculations of PM-10 concentrations in Swedish cities- Modelling of inhalable particles

77 Gollvik, S. (1997)
The Teleflood project, estimation of precipitation over drainage basins.

78 Persson, C., Ullerstig, A. (1997)
Regional luftmiljöanalys för Västmanlands län baserad på MATCH modell-beräkningar och mätdata - Analys av 1994 års data

79 Josefsson, W., Karlsson, J.-E. (1997)
Measurements of total ozone 1994-1996.

80 Rummukainen, M. (1997)
Methods for statistical downscaling of GCM simulations.

81 Persson, T. (1997)
Solar irradiance modelling using satellite retrieved cloudiness - A pilot study

82 Langner, J., Bergström, R. (SMHI) and Pleijel, K. (IVL) (1998)
European scale modelling of sulfur, oxidized nitrogen and photochemical oxidants. Model development and evaluation for the 1994 growing season.

83 Rummukainen, M., Räisänen, J., Ullerstig, A., Bringfelt, B., Hansson, U., Graham, P., Willén, U. (1998)
RCA - Rossby Centre regional Atmospheric climate model: model description and results from the first multi-year simulation.

84 Räisänen, J., Döscher, R. (1998)
Simulation of present-day climate in Northen Europé in the HadCM2 OAGCM.

85 Räisänen, J., Rummukainen, M., Ullerstig, A., Bringfelt, B., Ulf Hansson, U., Willén, U. (1999)
The First Rossby Centre Regional Climate Scenario - Dynamical Downscaling of CO2-induced Climate Change in the HadCM2 GCM.

86 Rummukainen, Markku. (1999)
On the Climate Change debate

SMHI