Tellus SERIES A DYNAMIC METEOROLOGY AND OCEANOGRAPHY
PUBLISHED BY THE INTERNATIONAL METEOROLOGICAL INSTITUTE IN STOCKHOLM

Tellus (2011), 63A, 324–337

TELLUS

Sensitivity of summer 2-m temperature to sea ice conditions

By R. E. BENESTAD[1]*, R. SENAN[1], M. BALMASEDA[2], L. FERRANTI[2], Y. ORSOLINI[3] and A. MELSOM[1], [1]*The Norwegian Meteorological Institute, PO Box 43, 0313 Oslo, Norway;* [2]*The European Centre for Medium-range Weather Forecasts, Reading, United Kingdom;* [3]*Norwegian Institute for Air Research, Kjeller, Norway*

(Manuscript received 11 January 2010; in final form 20 September 2010)

ABSTRACT

Current seasonal forecast models involve simple schemes for representing sea ice, such as imposing climatological values. The spread of ensemble forecasts may in principle be biased due to common boundary conditions prescribed in the high latitudes. The degree of sensitivity in the 2-metre temperature, associated with seasonal time scales and the state of the June–August sea ice, is examined through a set of experiments with a state-of-the-art coupled ocean-atmosphere model. Here we present a suite of numerical experiments examining the effect of different sea ice configurations on the final ensemble distribution. We also compare the sensitivity of the 2-metre temperature to sea ice boundary conditions and sea surface temperature perturbation in the initial conditions. One objective of this work was to test a simple scheme for a more realistic representation of sea ice variations that allows for a spread in the Polar surface boundary conditions, captures the recent trends and doesn't smudge the sea ice edges. We find that the use of one common set of boundary conditions in the polar regions has little effect on the subsequent seasonal temperatures in the low latitudes, but nevertheless a profound influence on the local temperatures in the mid-to-high latitudes.

1. Introduction

According to the fourth assessment report (AR4) of the Intergovernmental Panel on Climate Change, a changing climate is likely to bring more severe and frequent extreme weather and climate conditions in the future, such as heat waves, floods and droughts (Meehl et al., 2007, p. 750). Moreover, AR4 explicitly states that 'A warmer climate, with its increased climate variability, will increase the risk of both floods and droughts' (Kundzewicz et al., 2007, p. 186) and an analysis of changes in the upper tail of the monthly precipitation simulated by global climate models (GCMs) suggest that record-wet months will become more frequent in regions already characterized by a wet climate (Benestad, 2006). Cold waves and frost days are also expected to become less frequent in a warmer world. Thus, it will be of increasing value to provide skillful seasonal forecasts in order to be prepared for adverse or beneficial effects from seasonal anomalies. Furthermore, in Europe, there is great media interest in seasonal forecasts for the summer holiday season, coinciding with large reductions in northern hemisphere sea ice. Sivakumar (2007) argues that national climate policies are needed for making the most out of the seasonal forecasts, which also could include exploring ways to improve the seasonal forecasts. In this respect, the purpose of the present work was to see if such forecasts could potentially improve by using a more up-to-date sea ice representation.

1.1. *Sea ice*

There has been a downward trend (−11% per decade[1]) in the September Arctic sea ice extent[2] (Stroeve et al., 2007, 2008), which also may have an effect on seasonal forecasts. Such trends are not reflected in the mean seasonal sea ice climatology. According to a model study by Holland et al. (2006), the Arctic September sea ice may be almost gone within a time horizon of 30–50 years and future transitions in the sea ice may take place as rapid spurts.

*Corresponding author.
e-mail: rasmus.benestad@met.no
DOI: 10.1111/j.1600-0870.2010.00488.x

[1] The observed trends in sea ice extent according to the US National Snow and Ice Data center is: −3.2, −2.9, −2.6, −2.6, −2.5, −3.3 −6.1 −8.7 −11.2 −5.9 −4.5 and −3.3 %/decade for January–December with greatest decrease in September; source http://nsidc.org/cgi-bin/bist/bist.pl?config=seaice_extent_trends

[2] It is important to note that there is a distinction between the sea ice *extent* (area enclosed within the line of 15% ice concentration) and the sea ice *area* (the surface integral of the sea ice concentration) (Kauker et al., 2009)

Kauker et al. (2009) used an adjoint ocean-ice model to study the causes for the September 2007 Arctic sea ice minimum. They argued that most of the reduction between 2005 and 2007 could be attributed to the March sea ice thickness, May and June wind conditions and September surface temperature. However, they did not take into account feedbacks related to air–sea coupling and thus their results represent only a first-order estimate. Since the summer sea ice conditions were found to be sensitive to the March sea ice thickness, Kauker et al. suggested that there may be precursory signal in preceding sea ice conditions.

Rowell (1998) argued, based on simulations with the Hadley Centre Atmospheric Model, that the ice extent sometimes influences the local weather but rarely influences the large-scale flow. He analysed the potential sensitivity based on 'analysis of variance' (ANOVA): $\sigma^2_{SST}/\sigma^2_{tot}$ (the variance of an atmospheric property associated with sea surface temperatures (SSTs) and sea ice, divided by the property's total variance). However, such diagnostics capture only a predictable response in terms of the model, but does not reveal the total (predictable + unpredictable in terms of the model) influence of sea ice.

Empirical evidence supports the notion of the sea ice having a strong local effect on the climate, as Benestad et al. (2002) showed that the year-to-year variations in the local temperature near the ice edge were to a large extent driven by changes in the geographical sea ice distribution. Serreze et al. (2009) concluded that the Arctic amplification largely was driven by the loss of sea ice cover and the associated increase in heat transfer from ocean to atmosphere.

Singarayer et al. (2006) carried out a numerical study with an atmospheric general circulation model (AGCM), in which they explored the response to different sea ice and SST conditions. When prescribing climatological SSTs, they found that a decrease in the sea ice extent only had a significant impact on wintertime surface air temperatures. However, with corresponding increases in SSTs, changes were seen in both winter and summer. They also concluded that reductions in sea ice do induce more widespread changes in atmospheric circulation, however, one limitation of the Singarayer et al. (2006) study was that their work was based on an atmosphere only model, not taking into account the coupled ocean–atmosphere effect from changes in the sea ice.

A significant part of the variance in temperature, pressure and precipitation over the North Atlantic region can be associated with the North Atlantic Oscillation (NAO) and Pacific-North-American (PNA) patterns. Johansson (2007) argued the NAO and PNA are inherently internal to the atmosphere and that oceans are not essential for their existence and behaviour. The conclusion from Rowell (1998) also questions whether the difference between sea ice and non-sea ice affect the larger scales and according to these results, the question of whether the sea ice plays a role for the wider weather on seasonal time scales is still not adequately answered.

However, several publications have suggested that there may be a link between sea ice and circulation modes such as the NAO (Balmaseda et al., 2009, 2010; Seierstad and Bader, 2008; Sorteberg and Kvingedal, 2006; Francis et al., 2009; Overland and Wang, 2010). Based on numerical experiments with an AGCM and imposed SST or sea ice conditions, Magnusdottir et al. (2004) managed to reproduce in their model a response with a spatial character similar to the NAO. But they had to exaggerate the magnitude of the imposed surface conditions to achieve statistically significant results. Deser et al. (2004) analysed the results from the same set of experiments further and tried to resolve the direct and indirect response from changes in the SST and sea ice. The total response and the indirect response to changes in sea ice was found to be equivalent barotropic, whereas the direct response was baroclinic between 1000 and 850 hPa.

Balmaseda et al. (2009, 2010) also carried out simulations for the summers of 2007 and 2008 with observed and climatological sea ice conditions and found a substantial response in the 850 hPa geopotential height field from prescribed combined SST and sea ice, but a weak response from a coupled model with prescribed sea ice extent. They concluded that the response of the circulation on sea ice extent depends critically on the background state, in particular SST.

Sorteberg and Kvingedal (2006) furthermore suggested there was a connection between the east Siberian cyclone statistics and the sea ice in the Barents Sea and Seierstad and Bader (2008) studied the atmospheric sensitivity to sea ice conditions by setting the Arctic sea ice in an AGCM to present annual cycle and a projected sea ice state for the future. They found a significant reduction in the storminess during December and January over both the mid-latitudes and the Arctic (high latitudes), as well as a large reduction over the extra-tropics in March. The reduction was accompanied by a negative phase of the NAO.

1.2. Forecasting

Doblas-Reyes et al. (2007) give an account of the skill of up-to-date dynamical seasonal forecasting models and their use for short-term planning. So far, seasonal forecasts based on dynamical models exhibit some skill in the Tropics, but their performance is poor in high latitude regions such as northern Europe. These models take precursory signals mainly from the oceanic state in the Tropics, but often neglect to take the actual polar conditions (sea ice) properly into account.

Derome et al. (2001) performed a study on seasonal predictability with two dynamic models for which SST anomalies were persisted throughout the integrations, but where the sea ice extent in one model was specified to be that observed during the previous month and then relaxed to climatology over a period of 15 days. In the other model, the sea ice was taken to be the climatology. No attempt was made to relate the predictability specifically to sea ice, but Derome et al. (2001) argued that a substantial part of the winter-time skill in surface temperatures were associated with the forcing from tropical SSTs [e.g. El Niño Southern Oscillation (ENSO)], while the skill in the other seasons could not be related to the tropics.

Seasonal forecasts based on dynamical models tend to involve *ensembles*, which are many parallel integrations with one or several models, but with each run either starting with slightly different initial conditions or using different model configurations (Dix and Hunt, 1995; Stockdale et al., 1998). Since the models describe a 'chaotic' system (Lorenz, 1967), the individual members tend to follow trajectories that will diverge after a certain time, but where all the model trajectories will be part of the same strange attractor in phase space. Each integration should in principle be free to evolve completely independent of the other runs, except for being subject to the same underlying attractor and having similar, but not identical, initial conditions. A common set of sea ice may, however, bias the solutions when the sea ice itself is part of the variables that evolve over time.

The idea of ensemble forecasts is that the spread in the solutions will provide an indication of the range of possible solutions and that highly probable outcomes will lead to clusters in phase space (Wilks, 1995), which again will shape the statistical distribution of the prognostic variables. The basis for seasonal forecasts is that different boundary conditions may either affect the geometry of the attractor, or that different initial conditions coupled with slow dynamics (e.g. heat stored in the oceans) have a significant impact on the subsequent evolution. Hence, ensemble forecasts may in principle provide a probabilistic prognosis through proper post-processing.

Seasonal forecasts based on ensembles of GCMs have often used climatological values for sea ice (Balmaseda et al., 2009, 2010), both as initial and boundary conditions, since most of the predictability has been assumed to have its source in the Tropics (Derome et al., 2001). At the European Centre for Medium-range Weather Forecasts (ECMWF), however, the model making seasonal forecasts (Anderson et al., 2006) uses the observed sea ice as initial conditions, but employs a gradual interpolation towards climatological values over the first month and then relaxes towards climatological values for the remaining part of the integration.

All these strategies imply that the different runs with seasonal forecasting models are artificially constrained to a common set of surface boundary conditions in regions with sea ice. Furthermore, the climatology provides a sea ice description where the edges have been 'smudged out' through an averaging process and may therefore contain unrealistic features. Hence, the final distribution of the ensemble forecasts may be affected by the common constraints in the polar regions in terms of the surface fluxes, rather than following separate and independent trajectories for the regional state. In other words, the individual members are not entirely mutually independent and the distribution yielded by such ensemble forecasts may be biased and should not automatically be interpreted as a probability distribution function (PDF) or to represent a true probabilistic forecast.

Here, we follow up on the work of Balmaseda et al. (2009, 2010), who compared the ECMWF operational forecast, for which the sea ice is relaxed to climatology after one model month of integration, with corresponding runs where the sea ice climatology had been replaced by the actual sea ice observations. Our experiments are similar but not identical to theirs and the present study looks at other diagnostics of the model results. We also wanted to evaluate a simple scheme where an ensemble consists of different sea ice conditions, using a resample taken from the 8 yr during 2000–2007, to see if this has an impact on the distribution. Thus, our experiments mimic a simple sea ice scheme with a more realistic description of the cryospheric state, facilitating a way to describe a range of different sea ice configurations (hence avoid one common regional solution), capturing the recent downward trend and avoiding the 'smudging' of sea ice edges in taking the mean over several years.

We also wanted to test the sensitivity of the solutions for the 2-metre temperature to the sea ice and compare the response with the effect of SST perturbations, as Balmaseda et al. (2009, 2010) found that the background SST played an important role in the atmospheric response to sea ice anomalies. This paper will mainly focus on northern Europe, which is a region with low seasonal sensitivity skill in coupled dynamical models (anomalous correlation coefficient $\in [0.2, 0.5]$)[3]

2. Methods

2.1. Design of experiments

We aimed to explore the sensitivity of the seasonal forecasts to different sea ice conditions and evolutions thereof, rather than to predict this evolution during particular years. To this end, we examined whether the different sea ice descriptions had an effect on the simulated results. We did not involve actual observations in the analysis of the model results.

Another objective was to assess a new simple statistical scheme for representing sea ice more realistically than the present climatological values. The new scheme is based on eight 5-member ensembles with prescribed sea ice evolutions taken from the eight previous years, respectively.

The model set-up was the same as in Balmaseda et al. (2009, 2010) where we used the IFS/HOPE system 3 coupled ocean-atmospheric model (Anderson et al., 2006) to run a set of 9 ensemble integrations in total: One 40-member control ensemble and eight 5-member experimental ensembles. The former will

[3] http://iri.columbia.edu/cgitest-bin/skill_indvl_req.pl
http://www.ecmwf.int/products/forecasts/d/charts/seasonal/verification/previous/scores_determ/acc!scop!2mTemperature!Europe!February!2-4 months!latest!

Table 1. List of experiments

Subset	ECMWF experiment ID	Sea ice year
Subset 1	'b0ne'	2000
Subset 2	'b0nf'	2001
Subset 3	'b0ng'	2002
Subset 4	'b0nm'	2003
Subset 5	'b0nl'	2004
Subset 6	'b0n0'	2005
Subset 7	'b0na'	2006
Subset 8	'b0mq'	2007
Control	'ezuz'	climatology

henceforth be referred to as the 'control', while each of the latter ensembles will be referred to as a 'subset'. Although the integrations were performed for May–September, the analysis herein will mainly be restricted to the June–August mean 2-metre temperature [hereafter T(2 m)].

The model set-up with T159L62 ($\sim 1^\circ \times 1^\circ$) represents significantly higher spatial resolution in the atmosphere—both horizontally and vertically—than in earlier works, such as Singarayer et al. (2006) (2.5° latitude $\times$ 3.75° longitude $\times$ 19 vertical levels), Seierstad and Bader (2008) (T42L19), Magnusdottir et al. (2004) (T42L18), Deser et al. (2004) (T42L18) and Derome et al. (2001) (T32L10 & T63L23).

Since our simulations involved a coupled ocean-atmosphere model with prognostic SST, the solutions should respond to the imposed sea ice in a more consistent way than for atmosphere only simulations (Seierstad and Bader, 2008; Singarayer et al., 2006; Magnusdottir et al., 2004; Deser et al., 2004).

The results are available from the ECMWF archive system (MARS) under special projects (Table 1 provides an overview of the experiment names used by the MARS archive).

The observed sea ice was prescribed in both hemispheres throughout the model integration in 8 subsets, taking it to be the observed sea ice cover from May–September for each respective year in the period 2000–2007. All the simulations started with a consistent set of initial conditions taken from 1st May 2007. Hence, the evaluation of the experiments was based on the following summer season, June–August 2007 and involved comparisons between ensembles of model simulations of one episode rather than examining performance over time. The 'skill' of these experiments is therefore not affected by long-term trends.

We used a similar model set-up as in the operational forecasts, which did not archive the prescribed sea ice. The initial sea ice cover was derived from the ECMWF operational ocean analysis (Balmaseda et al., 2007), taking regions with SST values that were below the freezing point (-1.7°C) to be covered with sea ice. In this analysis, SST values were nudged strongly to Reynolds OIv2 SST product (Reynolds et al., 2002), using stronger relaxation (2–3 h time scale) when the value of the interpolated Reynolds SST fell below a threshold defining the freezing point.

The SSTs were perturbed in the initial conditions for the ensemble integrations (see Section 5.2 of Persson and Grazzini, 2005), leading to different temporal evolutions in the ensemble members of the same subset (and the control). Corresponding members in the various subsets were subject to the same SST-perturbation. Since the control consisted of 40 members, it involved a larger set of SST perturbations than in the subsets.

The subsequent evolution of sea ice in the integrations was controlled by the same SST-based criteria as the initial conditions. In the control run a relaxation towards climatology was adopted, taking interpolated values between initial conditions and the monthly climatology for the first month and between different months of the climatology thereafter. For the subsets, however, different sets of sea ice were imposed after the first time step throughout the integration. The Reynolds OIv2 SST product consists of weekly mean values, causing step changes every 7 days (Fig. 1) in the sea ice extend since August 2002. There was also a change in ocean analysis affecting the sea ice in the end of July 2002 (Senan and Benestad, 2009).

We also immediately note that there were inconsistencies as compared with observations in standard data sets, where the September 2007 Arctic sea ice was at a record low (Stroeve et al., 2008). This problem is due to changes in the sea ice analysis over time, as is discussed in Senan and Benestad (2009) and the abrupt changes also occur in the sea ice product from the operational analysis. These issues, however, are not believed to be crucial for the present analysis, which merely explored the *simulated* sensitivity to the model description of the sea ice conditions.

2.2. *Statistical methods*

The control experiment involved a larger set of SST-perturbations than the experiments carried out here and ideally, each experiment ought to involve a 40-member ensemble to be directly comparable with the control. However, resource constraints did not permit this and since part of the analysis addressing the sensitivity to sea ice involved inter-comparisons between each of the subsets only (all being identical in terms of SST-perturbations), this part was not affected by the different number of SST perturbations in the control and the subsets.

Rowell (1998) reviewed three different approaches for estimating the potential sensitivity and we used a variant of his ANOVA approach. Rather than presenting the ratio of predicted and total variances from an ordinary linear regression, we examined the coefficients and the p-values associated with a *factorial* regression (Hill and Lewicki, 2005; Wilkinson and Rogers, 1973). The factorial regression involved two variables representing combinations of five (SST perturbations) and eight (sea ice conditions) categorical predictors, respectively. It was used to compare the sensitivity of T(2 m) to sea ice and SST perturbations, respectively, with 7 degrees of freedom for sea ice conditions and 4

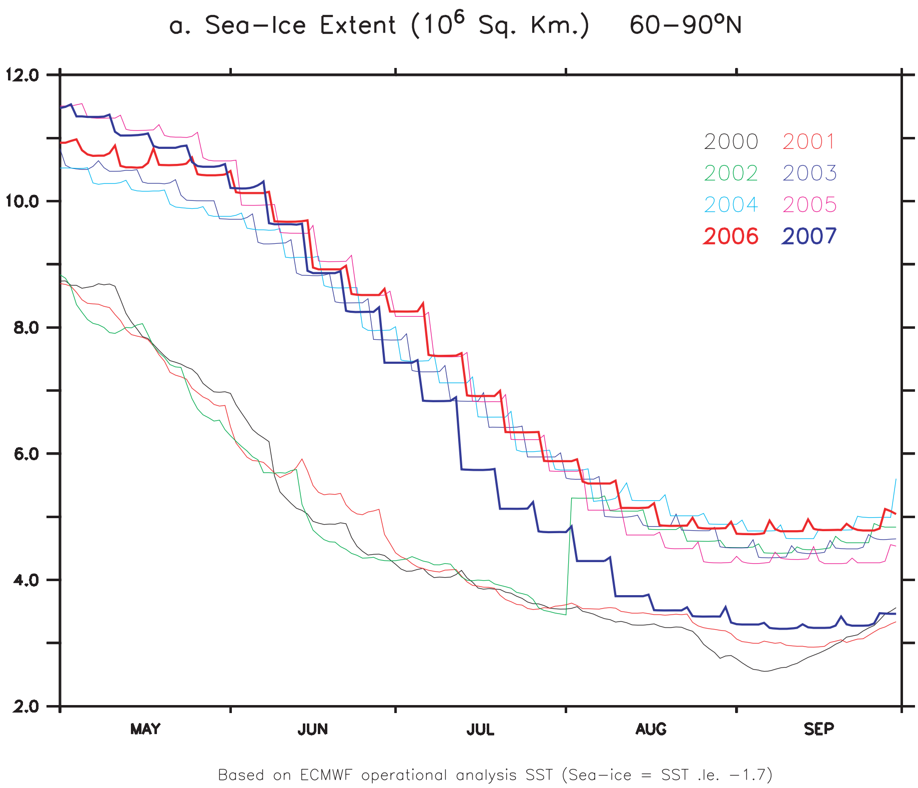

Fig. 1. Time evolution of the sea ice area from ECMWF operational analysis that was prescribed in the experiments.

for SST perturbations. Both ordinary linear models (LM) and generalized linear models (GLM)(Yan et al., 2006) were used.

The factorial regression design provided information about the relationships between categorical predictor variables and responses on the dependent variables beyond a corresponding one-way or main effect analysis. Unlike traditional regression, it did not try to determine the degree to which the outcome varies with the degree of sea ice or SST change, but instead examined whether a different set-up (sea ice or SST-perturbation category) had a predictable effect on the results. The factorial regression was also less clouded by non-linear or complex effects that otherwise may make a linear analysis obscure. The dynamics was non-linear and the different trajectories (especially for the SST perturbation initial conditions) may be folded or stretched in different ways depending on for example, the sea ice conditions (Balmaseda et al., 2009, 2010). The factorial regression analysis was only applied to the subsets and thus excluded the control. All the analyses have been implemented in the R-environment (R Development Core Team, 2004)[4].

A Kolmogorov–Smirnov test (henceforth referred to as 'KS-test') (Wilks, 1995) was used to determine whether the distributions from the subsets and the control were different. In addition, we supplemented the KS-test with Student's t-test and a χ^2-test (the latter is normally applied to contingency tables. Here, a contingency table was generated from the frequency distributions). The results from the Student's t-test and the χ^2-test were in good agreement with the KS-test and thus the results seem to be robust. The term 'p-value' will, unless stated otherwise, refer to the probability that the null-hypothesis H_0 is true, for example, that the distribution of the two samples compared are similar.

The Walker test (Wilks, 2006), expressed as $p_{\mathrm{Walker}} = 1 - (1 - \alpha_{\mathrm{global}})^{1/K}$, was used to evaluate the global significance of all the local hypothesis tests combined. It was also used to test the field significance along a latitude circle. The Walker test involved a comparison between the global minimum p-value, $p_{(1)}$ of all grid cells and the critical threshold p_{Walker}. When $p_{(1)} \leq p_{\mathrm{Walker}}$, the global null-hypothesis (i.e. that there is no effect from the prescribed sea ice) may be rejected at a given confidence level. In this case, the local p-values were derived through KS-tests, Student's t-tests and factorial regressions. Here $\alpha_{\mathrm{global}} = 0.01$ is the level of global significance, and $p_{\mathrm{Walker}} = 1.54 \times 10^{-7}$ is a critical threshold value when $K = 65160$ is the total number of local tests for the whole planet [using the same notation as in Wilks (2006)]. The parameter K refers to the number of different tests rather than the actual 'degree of freedom' reflecting the 'field significance' (Wilks, 1995, pp. 151–157).

[4] The R-software is a multi-platform open source environment, freely available from http://cran.r-project.org. The implementation in R was '`lm(y ~ exp.id + sst.pert)`', `exp.id` and `sst.pert` being variables of factors.

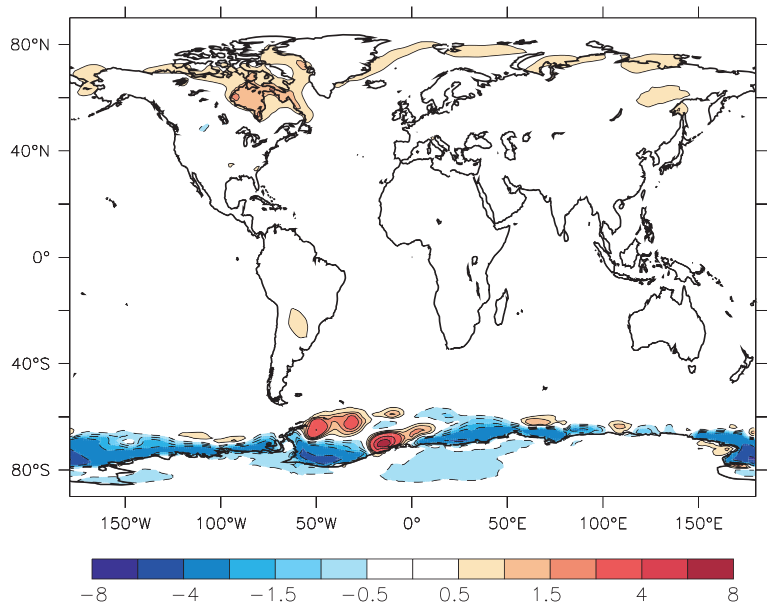

Fig. 2. The ensemble mean difference between T(2 m) from the subsets combined to that of the control.

The field significance is complicated by spatial auto-correlation in general (Wilks, 2006) and the degrees of freedom of gridded data in zonal bands vary with the latitude (Benestad, 2005), but according to Wilks (2006) the Walker test is not very sensitive to spatial auto-correlation and it becomes more conservative with mutual dependencies.

3. Results

3.1. Properties of probability distributions

The statistical distributions of T(2 m) in the experiments may differ from that of the control ensemble in their mean value, range, or by having different shapes. The ensemble mean of the pooled subsets was similar to the mean of the control over most of the planet (Fig. 2), but there were clear differences in the regions in vicinity of sea ice. The differences between ensemble mean T(2 m) from the combined subsets and the control were also much larger near the ice edge in the Southern Hemisphere (SH) than in the Northern Hemisphere (NH). There is no large contrast between the SST and the temperature of the sea ice surface during the summer season and during winter the sea ice surface is of the order of 10 K colder than the freezing point (Melsom, 2009). Our study focused on the NH summer, for which the impact on the ensemble mean T(2 m) was modest.

Figure 3 shows a comparison of the spread of the pooled subsets versus the control. Along the ice edges in both hemispheres, the distributions of T(2 m) from the combined subsets showed a substantially greater range, mainly at high latitudes. Elsewhere the T(2 m) distributions had similar ranges and hence, the ensemble range appeared to be less affected outside the mid-to-high latitude regions.

Figure 4 shows maps for Europe of the June–August ensemble mean T(2 m) for the control as well as the ensemble mean difference between each subset and the control. Despite the differences between control and experiments, the anomalies with respect to climatology exhibit some features that were robust in the sense that their sign appeared to be independent of the prescribed sea ice boundary conditions (not shown). These features included the anomalies over Turkey, along the northern coast of the Mediterranean, parts of Spain and the Baltic Sea. There were nevertheless fairly pronounced differences between each subset that could be attributed to the different sea ice conditions representing the only difference between these simulations: the 2002 (subset 3) and 2006 (subset 7) sea ice conditions were associated with a warmer June–August season over Fennoscandia than in the other subsets.

Although the ensemble means from the different subsets in Fig. 4 seemed to vary, it does not mean that they responded to the different sea ice conditions in a systematic and predictable fashion. The spread between the subsets can be compared with each individual subset spread and there was no systematic pattern whereby a specific SST perturbation gave lower or higher temperature than others (not shown). Depending on the sea ice conditions, the different integrations followed different trajectories and the sea ice affected the way the different SST-perturbations lead to different solutions in a more convoluted way, in agreement with the findings by Balmaseda et al. (2009, 2010).

3.2. Hypothesis testing

The question whether the differences between the subsets and the control were statistically significant can be addressed through

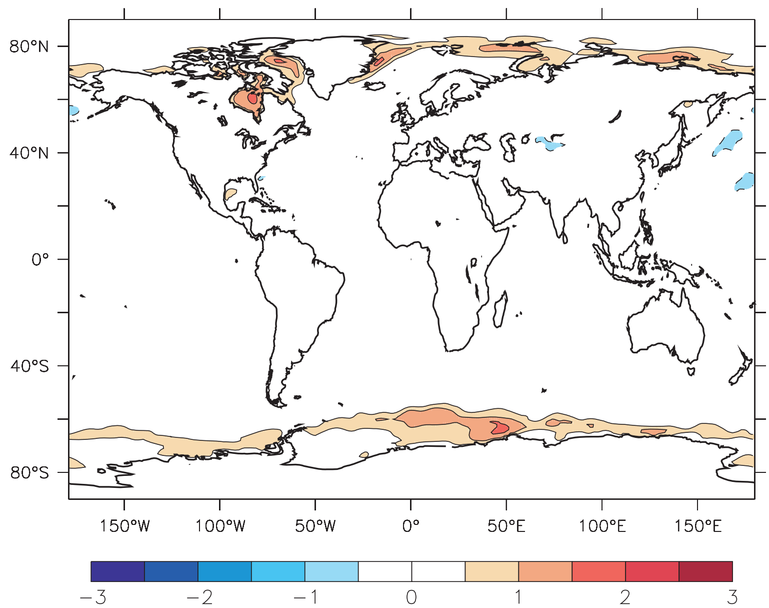

Fig. 3. The logarithm of the ratio of the T(2 m) ensemble standard deviation from the subsets combined to that of the control.

more formal tests, such as KS tests, Student's *t*-tests, χ^2-tests and factorial regression. In other words, did the sea ice scheme consisting of a re-sampling of the eight last years give different statistical distributions, hence implying a bias in the statistical distribution?

Figure 5 shows a global map of *p*-values derived from the KS-test. As expected, the lowest *p*-values (e.g. greatest probability that the T(2 m) really responds to the sea ice cover) were found in the polar regions near the ice edge. The results from a student's *t*-test and a χ^2-test (not shown) were similar, suggesting that the estimates were robust.

There were some regions at low latitudes where the tests found differences that appeared to be significant at the 1% level too, notably near the 'cold tongue' in the tropical east Pacific and over the South China Sea. These low *p*-values were due to the use of a more extensive set of SST perturbations in the control (40 perturbations) compared to the experiments (5 perturbations). Figure 6 shows a similar analysis as Fig. 5, but now applied to the two subsets 7 and 8 (prescribed sea ice data from 2006 and 2007 respectively) instead of the pooled subsets versus the control. These two subsets represent high and low sea ice extent respectively (Fig. 1) and the analyses suggested that the differences were statistically significant at the 10%-level over similar regions as between the control and the subsets combined, except that there was no signal in the tropics.

The Walker test was applied to the *p*-values of the local KS and Student's *t*-tests in order to assess the 'field significance' of the local tests for the whole world as well as a range of zonal bands. The global value for p_{Walker} was $1.5 \times 10^{-7} (K = 65160)$, whereas $p_{(1)}$ was 2.9×10^{-18} for the KS-test and 3.1×10^{-14} for the Student's *t*-test. Thus, the results were highly statistically significant.

Figure 7 shows how the minimum p-value, $p_{(1)}$, varied with latitude ($K = 360$), and the pink hatched regions indicate the latitudes where the Walker test suggested that the p-value associated with sea ice from the KS-test was statistically significant at the 1%-level. The blue curve shows the same analysis for the Student's t-test. The minimum p-value was only lower than p_{Walker} in the mid-latitudes and in a very narrow zonal band in the Tropics due to the differences in the SST perturbations.

3.3. *Factorial regression*

While the distribution based on all subsets combined was indistinguishable to that of the control, the differences between each subset (Fig. 4) had magnitudes that were about 50% of the anomalies associated with the initial conditions (not shown). One objective of the factorial regression analysis was therefore to identify predictable responses in terms of a linear regression analysis. It is possible that the effect of each individual sea ice condition was 'blurred out' when combining the results from the 8 subsets into one and comparing them with the control. Hence, we wanted to test whether an improvement in the sea ice description could affect the prediction skill for mid-latitude locations. To this end we carried out a factorial regression for T(2 m) based on SST perturbation and sea ice categories. In this context, the *p*-values refer to the H_0 that T(2 m) was insensitive to the set-up.

Figure 8 shows the coefficients from the factorial regression for Europe (the results from the GLM are not shown, but the results are similar) along with their error estimates (bottom row). The regression coefficients for sea ice were associated with greater magnitudes than for both the SST-perturbations and the standard errors.

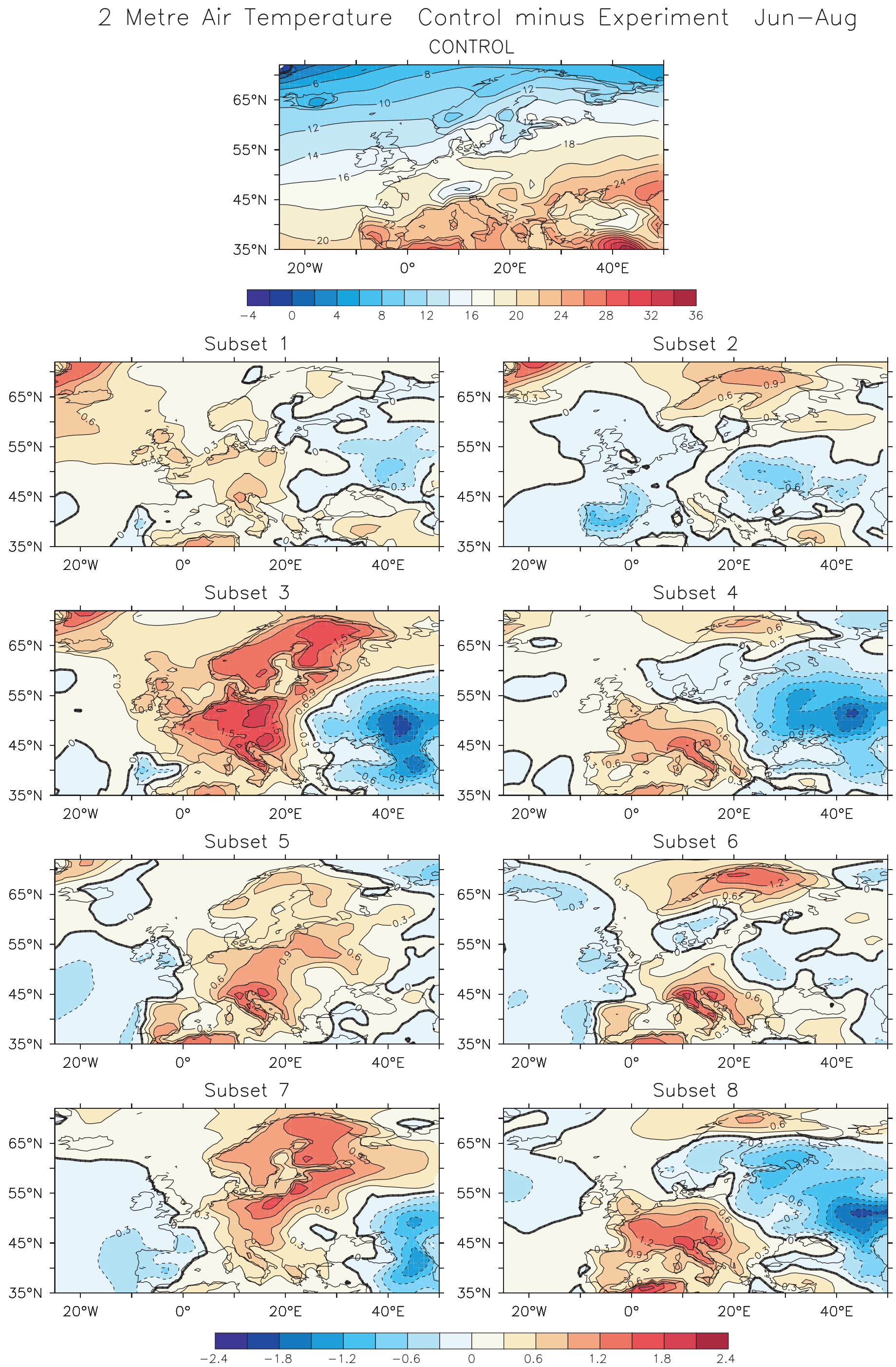

Fig. 4. Differences between the ensemble mean temperature and the control run for the 8 different subsets, with the mean from the control run at the top. The differences in the subset ensemble mean temperatures can be attributed different sea-ice conditions prescribed, as the simulations otherwise involved identical set-up. Note, these are furthermore distinct to anomalies, usually defined as the difference between a particular event and the climatology.

The regression weights for subsets 3 and 7 had similar geographical patterns, albeit with regional details, as do subsets 5 and 6. There were also similar June–August mean sea ice extent in subsets 5 and 6. Different sea ice categories representing similar conditions can result in a failure to detect the response and hence the factorial regression analysis is expected to be conservative with respect to identifying the actual links.

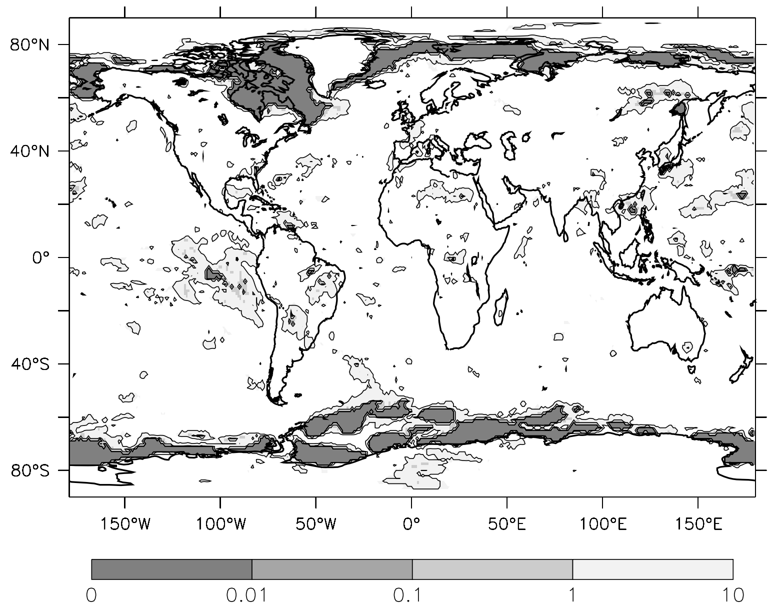

Fig. 5. Kolmogorov–Smirnov tests p-values (%) between T(2 m) values from the experiments combined (40 members) and the control (40 members). Here only p-values $\in [0, 10]$ are shown and white regions represent locations where the results are not statistically significant at the 10%-level.

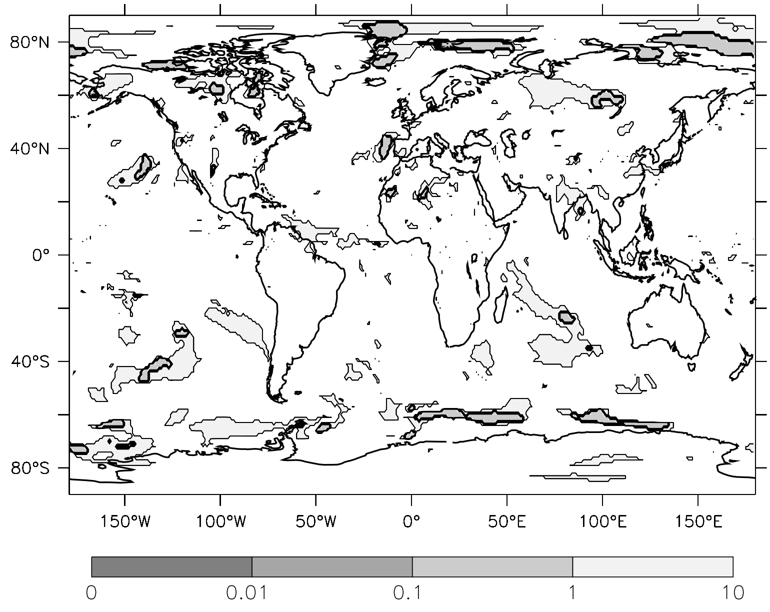

Fig. 6. Kolmogorov–Smirnov tests p-values between T(2 m) values from the subsets with 2006 sea ice ('b0na') and 2007 sea ice ('b0mq'), with five members each. Here only p-values $\in [0, 10]$ are shown and white regions represent locations where the results are not statistically significant at the 10%-level.

The factorial regression was also applied to the whole globe (not shown) and the Walker test was applied to the global results ($K = 65160$ & $\alpha_{\text{global}} = 0.01$). The lowest p-value associated with sea ice was $p_{(1)} = 3.07 \times 10^{-18}$, thus $p_{(1)} \ll p_{\text{Walker}} = 1.54 \times 10^{-7}$ pointing to a virtually certain influence of sea ice on T(2 m) somewhere on the planet (here the effect is strongest near the ice edge). The lowest p-value associated with the SST perturbations was $p_{(1)} = 2.21 \times 10^{-17}$ and again it was established that the SST perturbation are virtually certain to affect the T(2 m) somewhere one the planet.

Figure 9 shows the latitude profile of $p_{(1)}$ for sea ice (red curve), suggesting that the sea ice conditions had a systematic effect on the T(2 m) in the northern hemisphere mid-to-high latitudes (from 50°N and up to the ice-edge) as well as the high latitudes in the SH (but not in the Antarctica interior). The sea ice did not appear to have any systematic effect in the low latitudes. The same analysis for the SST perturbations (blue curve), on the other hand, indicated significant p-values in the low latitudes, but not poleward of 60° in either hemisphere.

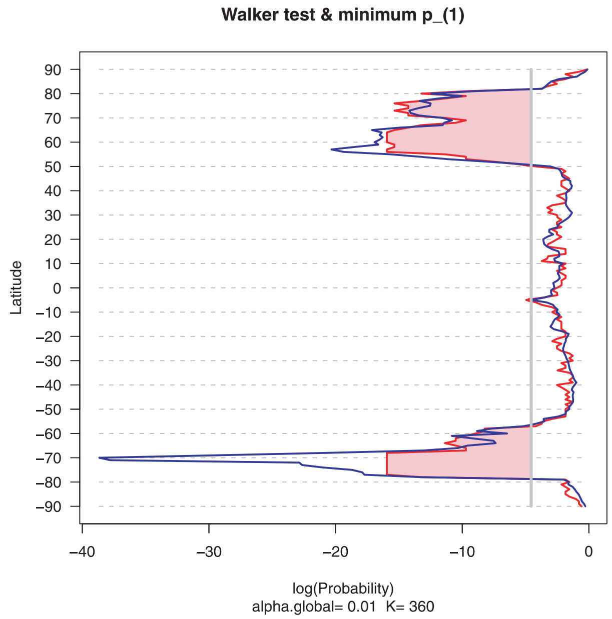

Fig. 7. Latitude profiles of the minimum p-value for each row of (K = 360) grid boxes $p_{(1)}$ from the hypothesis tests applied the experiments combined versus the control in comparison with p_{Walker}. The red curve shows $p_{(1)}$ associated with the KS-test and blue shows corresponding results for the Student's t-test. The pink hatched regions mark the latitudes where $p_{(1)} < p_{\mathrm{Walker}}$ for the KS-test.

Figures 7 and 9 show similar results for $p_{(1)}$ associated with sea ice, both with below-threshold-values in mid-latitude bands even though these were derived through fundamentally different (and independent) approaches. Whereas Fig. 9 shows the p-value from the factorial regression analysis (control excluded) associated with the subset differences, Fig. 7 reflects the differences in the statistical distribution between the combined subsets versus the control. In other words, the sea ice response in the mid-latitudes was a robust feature. The lack of a signal in the tropics and low $p_{(1)}$ associated with the SST-perturbations supports the interpretation that the apparent Tropical 'signal' in Fig. 5 was an artifact from the comparison between the control and experiments due to different SST perturbations.

The most significant and predictable response to different sea ice conditions was seen around the Hudson Bay, the Labrador Sea, the Barents Sea and the sea ice edge. Thus, the latitude profiles of $p_{(1)}$ may not necessarily reflect the sea ice sensitivity over northern Europe (Fig. 8). Therefore, the Walker test was applied also regionally to the $K = 2196$ grid-boxes in the longitude range 10°W–50°E and the latitude range 35°N–70°N in order to explore the regional character of the response to sea ice. This region implied $p_{\mathrm{Walker}} = 4.58 \times 10^{-6}$, but $p_{(1)} = 0.007$ for the KS-test and 0.004 for the Student's t-test. Similarly, the lowest p-value from the factorial regression was 0.0003 over the same region, thus the response to sea ice over northern European land areas was not statistically significant at the 1%-level. The lack of statistical significance does not exclude an effect, but suggests that any influence from sea ice on summer mean temperatures in Europe is too weak to detect in the current experimental set-up. Balmaseda et al. (2009, 2010), on the other hand, detected a statistically significant response to the 500 hPa heights in similar experiments with the same model, but also concluded that the response to sea ice was sensitive to the SST. Hence, the lack of statistical significance may be due to the difficulty of distangling effects from SST and sea ice.

3.4. *Physical processes*

The sea ice cover represents a medium that has a different character (heat conduction) to that of fluid (convection and mixing processes) and is associated with lower temperatures, affecting both the thermal radiation and the latent heat fluxes and has different optical properties than open water, increasing the local albedo.

An inspection of simulated energy fluxes revealed pronounced differences in the short wave (solar) radiation at the top of the atmosphere over the Arctic regions between the subsets and control (not shown), in contrast to other fluxes. There were also pronounced regional differences (∼40 W m^{-2}) in the reflected solar radiation at the surface over the NH ice-edge (not shown). These differences were of local nature, associated with the sea ice albedo.

We found that the surface sensible and latent fluxes, as well as thermal radiation, all vary only with a few W m^{-2} in the Arctic, while their impact was substantially greater in the Antarctic, where contrasting surface conditions during winter lead to substantial gradients from regions with open water to ice covered areas. Thus, the combined effect of latent heat, sensible heat and thermal radiation lead to larger variability during winter than in summer.

The effect of albedo was by far the most dominant type of flux during summer and the effect from changes in the albedo related to sea ice could only affect the weather elsewhere through teleconnections.

Changes in the circulation patterns will have an impact on storm tracks. According to Deser et al. (2004), mid-latitude cyclones can be interpreted as an indirect effect of SST and sea ice anomalies. Storm counts based on the Calculus-based Cyclone Identification (CCI)-method proposed by Benestad and Chen (2006) (not shown) exhibited differences depending on the different prescribed sea ice conditions (the differences between all the subsets were statistically significant at the 1%-level according to a χ^2-test). Exactly what gives rise to these differences is beyond the scope of the present study, but the 2-metre temperature in the mid-latitudes is known to be strongly influenced by low pressure systems through associated cloudiness, precipitation and advection of air.

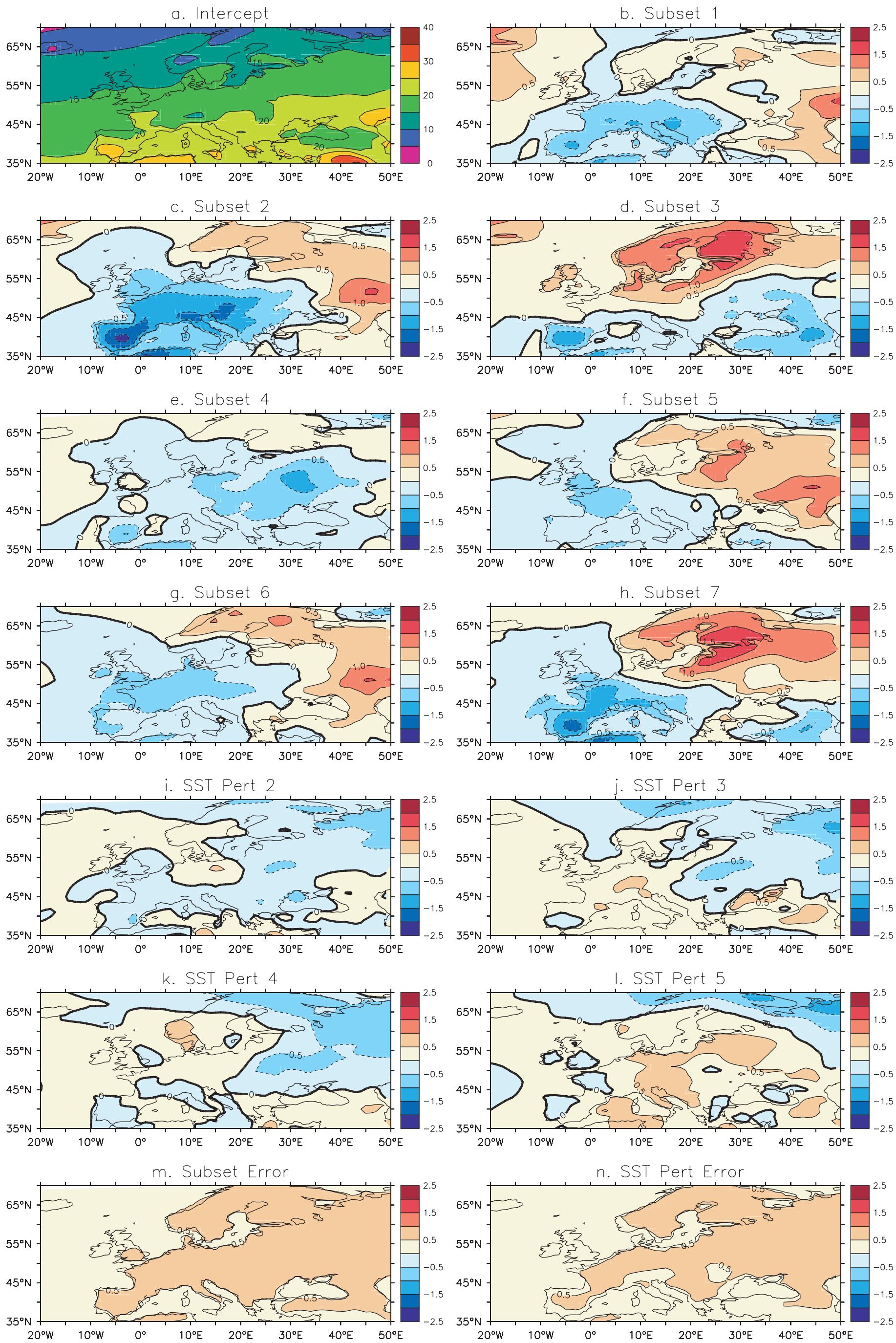

Fig. 8. Regional regression coefficients from factorial regression with an ordinary linear model. 'Intercept', represents the case with sea ice from 2007 and SST-perturbation 1 and the other panels represent deviations from these values. The patterns show the mean difference between set-up using a combination of sea-ice from 2007 (subset 8) and SST-perturbation 1 and the other combinations of sea-ice and SST-perturbation represented by the experiments. The errors refer to standard error estimates from the factorial regression analysis, being sensitive to the ensemble spread.

4. Discussion

One caveat associated with these experiments was that prescribing sea ice for 'wrong' year results in physical inconsistent solutions, that is, taking the sea ice from a different year than the initial ocean conditions. However, the present approach in operational seasonal forecasting and in our control, where climatological ice is used after the first month, have a completely

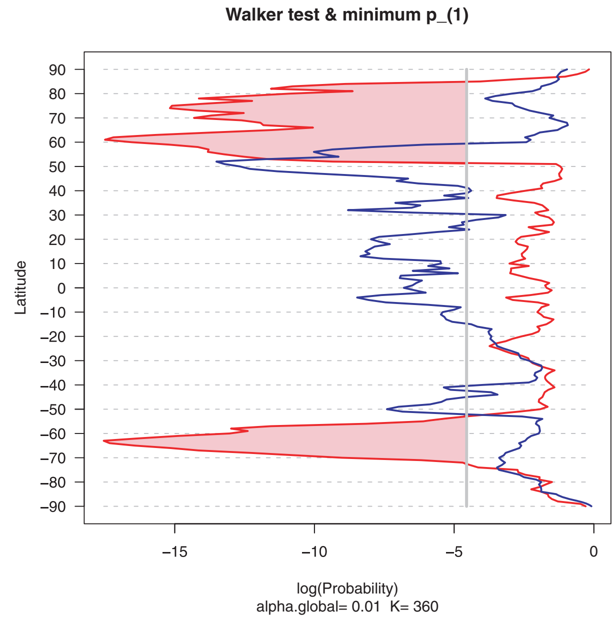

Fig. 9. Latitude profiles of the minimum p-value for each row of ($K = 360$) grid boxes $p_{(1)}$ from the factorial regression for June–August temperature in comparison with p_{Walker}. The red curve shows $p_{(1)}$ associated with the sea ice category (experiment ID) and blue shows corresponding results for the SST-perturbation. The pink hatched regions mark the latitudes where $p_{(1)} < p_{\mathrm{Walker}}$.

analogous discrepancy. Furthermore, the SSTs were prognostic in the coupled ocean–atmosphere model and they would adjust to the imposed sea ice condition. Finally, the set-up was an improvement over previous studies, where sea ice also had been prescribed in conjunction with physically inconsistent SST or vice versa, albeit with atmosphere general circulation models rather than coupled atmosphere-ocean models (Deser et al., 2004; Magnusdottir et al., 2004; Singarayer et al., 2006).

There were some small regions with statistically significant differences in the Tropics, such as near the east Pacific ‘cold tongue’. Similar p-values were not seen in the factorial regression results (not shown) excluding the control and these appeared to be an artefact of the analysis. Whereas all the subsets use identical SST-perturbations for their respective 5 members, the control with 40 members involved 35 additional SST perturbations which can explain some of the differences between the distributions of the control and the pooled experiments. In other words, the distribution from the control may over-estimate the ensemble range when compared to the subsets, thus over-estimating the p-values of the KS and t-tests (making the tests more conservative). It is not clear whether a scheme for operational forecasting, based on a re-sampling the 8 previous years of sea ice, should use only 5 SST perturbations that are identical for each of the subset, or 40 SST perturbations which differ between the subsets but match the control.

Sea ice does to some extent influence the seasonal mean T(2 m) over northern Europe, but the corresponding distributions associated with the two sea ice schemes were not significantly different. Balmaseda et al. (2009, 2010), however, found that the atmospheric response to sea ice anomalies is conditioned by the background SST and hence these results may also be sensitive to the magnitude of the SST perturbations, as the effects from large spread in the initial conditions may in principle dominate over moderate but systematic effect from the sea ice. Here, the factorial regression suggested a higher sensitivity to sea ice representation than SST perturbation in the mid-to-high latitudes, but this is not inconsistent with the conclusion of Balmaseda et al. (2009, 2010).

The present study only looked at the summer season, with initial conditions taken from May 1. It is possible that the sea ice yields a different response in T(2 m) for other seasons, which will be the focus for future work. Furthermore, the current analysis also looked at only one situation—the summer season of 2007—which may represent just one of several regimes. The effect of only a limited range of sea ice conditions on T(2 m) have been investigated, which doesn’t exclude the possibility that more dramatic sea ice changes may give a stronger response. Moreover, the sea ice conditions in some of the subsets were similar, making the attribution by factorial regression more tricky.

It is also important to note that the effect of sea ice is expected to differ on seasonal and decadal-to-centennial time scales since it strongly affects the energy absorbed in the upper ocean layers with a delayed effect on the atmosphere. Thus, the local effect on the seasonal time scales should not be confused with coupled feedbacks on longer times scales, for example, associated with the an Arctic amplification (Serreze et al., 2009).

5. Conclusion

Statistical distributions derived from ensemble forecasts may be biased in some regions if they are artificially constrained by one common description of the sea ice conditions in the polar regions, but a simple re-sampling scheme for representing a range of sea ice conditions may provide more realistic seasonal forecasts for the mid-to-high latitude regions. We compared two sea ice schemes based on (i) climatological values and (ii) a re-sample from eight recent years and their effect on the statistical T(2 m) distribution on seasonal time scales.

The results from these experiments suggest that the distribution of T(2 m) generated by the seasonal forecast ensemble using the current sea ice scheme is not significantly biased for most of the globe, but this does not necessarily mean that the sea ice does not contain precursory signals that may improve the skill. The sea ice has an influence on the June–August T(2 m) response in the mid-latitudes (50°N–80°N & 60°S–80°S), but the number of simulations was too low to derive results that were statistical significant at the 1% level over Europe. The level of statistical significance is a function of the ensemble size and number of ensembles and our analysis does not exclude the possibility for increased skill over northern Europe, given larger ensembles and better representation of the sea ice.

6. Acknowledgments

This work has been supported by the Norwegian Research Council (the SPAR project; project number 178570/S30), the Norwegian Meteorological Institute, and the ECMWF special project SPNOSPAR. We are also grateful for valuable comments from two anonymous reviewers.

References

Anderson, D., Stockdale, T., Balmaseda, M., Ferranti, L., Vitart, F. and co-authors. 2006. Seasonal Forecast System 3. *ECMWF Newsletter*, 19–25.

Balmaseda, M., Freeanti, L., Molteni, F. and Palmer, T. N. 2009. *Impact of 2007 and 2008 Arctic Ice Anomalies on the Atmospheric Circulation: Implications for Long-Range Predictions*. Technical Memorandum 595. ECMWF, http://www.ecmwf.int/publications/library/do/references/list/show?id=89212.

Balmaseda, M. A., Vidard, A. and Anderson, D. L. T. 2007. *The ECMWF System 3 ocean analysis system*. Technical Memorandum 508. ECMWF.

Balmaseda, M. A., Ferranti, L., Molteni, F. and Palmer, T. N. 2010. Impact of 2007 and 2008 Arctic ice anomalies on the atmospheric circulation: implications for long-range predictions. *QJRMS*, doi:10.1002/qj.661.

Benestad, R. E. 2005. On Latitudinal Profiles of Zonal Means. *Geophys. Res. Lett.* **32,** doi:10.1029/2005GL023652.

Benestad, R. E. 2006. Can we expect more extreme precipitation on the monthly time scale? *J. Clim.* **19**, 630–637.

Benestad, R. E. and Chen, D. 2006. The use of a Calculus-based Cyclone Identification method for generating storm statistics. *Tellus*, **58A**, 473–486, doi:10.1111/j.1600-0870.2006.00191.x.

Benestad, R. E., Hanssen-Bauer, I. and Førland, E. J. 2002. Empirically downscaled temperature scenarios for Svalbard. *Atm. Sci. Lett.*, **3**, Issue 2-4, doi.10.1006/asle.2002.0051, 71–93.

Derome, J., Brunet, G., Plante, A., Gagnon, N., Boer, G. J. and co-authors. 2001. Seasonal Predictions Based on Two Dynamical Models. *Atm.-Ocean*, **39**(4), 485–501.

Deser, C., Magnusdottir, G., Saravanan, R. and Phillips, A. 2004. The effects of North Atlantic SST and Sea Ice Anomalies on the Winter Circulation in CCM3. Part II: direct and indirect components of the response. *J. Clim.* **17**, 877–889.

Dix, M. R. and Hunt, B. G. 1995. Chaotic influences and the problem of deterministic seasonal predictions. *Int. J. Climatol.* **15**(7), 729–752.

Doblas-Reyes, F. J., Hagedorn, R. and Palmer, T. N. 2007. Developments in dynamical seasonal forecasting relevant to agricultural management. *Clim. Res.* **33**, 19–26.

Francis, J. A., Chan, W., Leathers, D. J., Miller, J. R. and Veron, D. E. 2009. Winter Northern Hemisphere weather patterns remember summer Arctic sea ice extent. *Geophys. Res. Lett.* **36**, L07503 doi:10.1029/2009GL037274.

Hill, T. and Lewicki, P. 2005. *Statistics: Methods and Applications : A Comprehensive Reference for Science, Industry, and Data Mining*. Tulsa, OK, USA: StatSoft.

Holland, M. M., bitz, C. M. and Tremblay, B. 2006. Future abrupt reductions in the summer Arctic sea ice. *Geophys. Res. Lett.* **33**, L23503, doi:10.1029/2006GL028024.

Johansson, Å. 2007. Prediction skill of the NAO and PNA from Daily to Seasonal Time Scales. *J. Clim.* **20**, 1957–1975.

Kauker, F., Kaminski, T., Karcher, M., Giering, R., Gerdes, R. and co-authors. 2009. Adjoint analysis of the 2007 all time Arctic sea-ice minimum. *Geophys. Res. Lett.* **36**, L03707, doi:10.1029/2008GL036323.

Kundzewicz, Z. W., Mata, L. J., Arnell, N., Döll, P., Kabat, P. and co-authors. 2007. *Climate Change: Impacts, Adaptation and Vulnerabilit. Contribution of Working Group II to the Fourth Assessment Report of the Intergovernmental Panel on Climate Change*. Cambridge University Press, United Kingdom and New York, NY, USA.

Lorenz, E. 1967. *The Nature and Theory of the General Circulation of the Atmosphere*. Publication 218. World Meteorological Organization, Geneva, Switzerland.

Magnusdottir, G., Deser, C. and Saravanan, R. 2004. The Effects of North Atlantic SST and Sea Ice Anomalies on the Winter Circulation in CCM3. Part I:Main Features and Storm Track Characteristics of the Response. *J. Clim.* **17**, 857–875.

Meehl, G. A., Stocker, T. F., Idlingstein, W. D., Gaye, A. T., Gregory, J. M., Kitoh, A., Knutti, R., Murphy, J. M., Noda, A., Raper, S. C. B., Watterson, I. G., Weaver, A. J., Zhao, Z.-C. 2007. *Climate Change: The Physical Science Basis. Chap. Global Climate Projections*. United Kingdom and New York, NY, USA. Cambridge University Press.

Melsom, A. 2009. *Ocean sea ice atmosphere heat fluxes over the Arctic Ocean*. Note 14/2009. met.no, http://met.no/Forskning/Publikasjoner_2009.

Overland, J. E. and Wang, M. 2010. Large-scale atmospheric circulation changes are associated with the recent loss of Arctic sea ice. *Tellus A*. **62**, 1–9.

Persson, A. and Grazzini, F. 2005. *User Guide to ECMWF forecast products*. Meteorological Bulletin M3.2. ECMWF, www.ecmwf.int/products/forecasts/guide/user_guide.pdf.

R Development Core Team. 2004. *R: A language and environment for statistical computing*. R Foundation for Statistical Computing, Vienna, Austria. ISBN 3-900051-07-0.

Reynolds, R. W., Rayner, N. R., Smith, T. M., Stokes, D. C. and Wang, W. 2002. An improved in situ and satellite SST analysis for climate. *J. Clim.* **15**, 1609–1625.

Rowell, D. P. 1998. Assessing Potential Seasonal Predictability with an Ensemble of Multidecadal GCM Simulations. *J. Clim.* **11**, 109–120.

Seierstad, I. A. and Bader, J. 2008. Impact of a projected future Arctic Sea Ice reduction on extratropical storminess and the NAO. *Clim. Dynam*. **33**, 937–943.

Senan, R. and Benestad, R. E. 2009. *Transitional irregularity in sea surface temperature from the ECMWF operational ocean analysis*. Note 22. met.no, www.met.no.

Serreze, M. C., Barrett, A. P., Stroeve, J. C., Kindig, D. N. and Holland, M. M. 2009. The emergence of surface-based Arctic amplification. *The Cryosphere* **3**, 11–19 doi:10.5194/tc-3-11-2009.

Singarayer, J. S., Bamber, J. L. and Valdes, P. J. 2006. Twenty-first-Century Impacts from a Declining Arctic Sea Ice Cover. *J. Clim.* **19**, 1109–1125.

Sivakumar, M. V. K. 2007. climate prediction and agriculture: current status and future challenges. *Clim. Res.* **33**, 3–17.

Sorteberg, A. and Kvingedal, B. 2006. Atmospheric Forcing on the Barents Sea Winter Ice Extent. *J. Clim.* **19**, 4772–4784.

Stockdale, T. N., Anderson, D. L. T., Alves, J. O. S. and Balmaseda, M. A. 1998. Global seasonal rainfall forecasts using a coupled ocean-atmosphere model. *Nature* **392**, 370–373.

Stroeve, J., Holland, M. M., Meier, W., Scambos, T. and Serreze, M. 2007. Artic sea ice decline: Faster than forecast. *Geophys. Res. Lett.* **34**, L09501 doi:10.1029/3007GL029703.

Stroeve, J., Serreze, M., Drobot, S., Gearheard, S., Holland, M. and co-authors. 2008. Arctic Sea Ice Extent Plummets in 2007. *Eos* **89**(2), 13–14.

Wilkinson, G. N. and Rogers, C. E. 1973. Symbolic Description of Factorial Models for Analysis of Variance. *Appl. Stat.* **22**, 392–399.

Wilks, D. S. 2006. On “Field Significance” and the False Discovery Rate. *J. Appl. Meteorol. Climatol.* **45**, 1181–1189 doi:10.1175/JAM2404.1.

Wilks, D. S. 1995. *Statistical Methods in the Atmospheric Sciences*. Orlando, Florida, USA, Academic Press.

Yan, Z., Bate, S., Chandler, R. E., Isham, V. and Wheater, H. 2006. Changes in extreme wind speeds in NW Europe simulated by generalized linear models. *Theoret. Appl. Climatol.* **83**, 121–137.