

The influence of single model ensemble on the simulated extratropical interannual variability

Chia-Chi Wang^{1,*}, Huang-Hsiung Hsu², Yu-Luen Chen², Jun-Kai Yang², and Meng-Pai Hung¹

¹Department of Atmospheric Sciences, Chinese Culture University, Taipei City, Taiwan

²Research Center for Environmental Changes, Academia Sinica, Taipei City, Taiwan

Article history:

Received 3 October 2017

Revised 2 July 2018

Accepted 18 July 2018

Keywords:

Single model ensemble, Interannual variability, Climate simulation

Citation:

Wang, C.-C., H.-H. Hsu, Y.-L. Chen, J.-K. Yang, and M.-P. Hung, 2018: The influence of single model ensemble on the simulated extratropical interannual variability. *Terr. Atmos. Ocean. Sci.*, 29, 679–694, doi: 10.3319/TAO.2018.07.18.01

ABSTRACT

This study compares the interannual variance of boreal winter near-surface temperature (DJF T2m) with and without performing single model ensemble (SME) in seasonal hindcasts (DEMETER, ENSEMBLES, and NCEP CFSv2) and historical climate simulations (CMIP5). The results demonstrate that the extratropical temperature variability is significantly reduced after performing SME even though the signal in the tropical Pacific remains strong. Cancellation between positive and negative perturbations simulated by individual model members, of both tropical and extratropical origins, leads to the under-simulation. The atmospheric circulation induced by tropical Pacific sea surface temperature is not well represented in global climate models and the simulation is further deteriorated by SME, leading to an unrealistically weak interannual variance of simulated winter temperature in North America. Similar effect was also found in North Eurasia where winter temperature is strongly influenced by atmospheric internal variability and its interaction with land and ice/snow in the middle-high latitudes. The SME procedure should be avoided when evaluating the model performance in simulating the higher-order long-term statistics (such as variance). Variance of individual models should be calculated first and then averaged among members. Models used in seasonal forecast and long-term climate simulation already have good capability in simulating the long-term statistics of stochastic processes in the extratropics, although the capability in accurately simulating the temporal variation is still poor.

1. INTRODUCTION

The weather prediction and climate simulation have been made possible through the continuous development and improvement of general circulation models (GCMs) in the past few decades. Seasonal forecasts and climate simulations are carried out by different strategies using basically the same tools, the GCMs. The main difference between these two types of simulation is the initiation stage. Seasonal forecasts are highly dependent on initial conditions while climate simulations would be spun up until the system reaches a steady state to avoid the influence from initial conditions. Both types of simulations make use of the ensemble method to estimate the uncertainties that may come from imperfections of observations or model formulations (such as errors from numerical computation and parameterization) and the

chaotic nature of the climate system. For seasonal forecast, when a large number of ensemble members are conducted, the ensemble result can produce a probability distribution that may provide information for decision making and risk assessment. For climate simulations, the ensemble result is more like an uncertainty range in which a steady state of the climate system may reach (Slingo and Palmer 2011).

The performance of seasonal forecasts and climate simulations are evaluated by two different approaches. For seasonal forecast, the single model ensemble (SME) is first calculated, some may involve multi-model ensemble (MME) as well. Then the ensemble means are evaluated based on their standardized verification system, which usually includes global and zonal means of hindcast anomalies, time series of climate indices, and various skill scores, such as anomaly correlation coefficients (ACC) (Miyakoda et al. 1972). Long-term climate simulations are evaluated by

* Corresponding author
E-mail: wang1794@gate.sinica.edu.tw

comparing the long-term mean of climatology and the dominant modes of variability with reanalysis or observational data. The model performance is often shown as a Taylor diagram (Taylor 2001) to gather a quick overall comparison between different climate models. In general, the seasonal forecasts are evaluated based on how close the simulated seasonal mean is to the observed mean at a given season. The long-term climate simulation is evaluated by the statistical characteristics of the Earth system, rather than how close the model emulates the exact condition in the reality.

The interannual variability is one of the climate variations that strongly influence human activities and can be simulated much better by GCMs than other climate variations, such as the Madden-Julian Oscillation (Madden and Julian 1971). Therefore, interannual variability has been extensively studied for both seasonal forecasts and climate simulations purpose. The seasonal forecasts have been shown to have good and stable predictability, especially for the winter season during which the most important influencing factor is the El Niño and Southern Oscillation (ENSO) (e.g., Weisheimer et al. 2009). However, extreme climatic events of temperature on seasonal timescale, especially in the extratropics, have been observed to increase in recent years (e.g., the record-breaking cold events in East Asia in late January to early February, 2016; an unusually warm winter in Alaska and the Bering Sea and coldness in Northeastern America during 2013 - 2014 and 2015 - 2016; Palmer 2014; Lee et al. 2015). The ability in simulating and forecasting the extratropical interannual variability that may arise due to factors other than ENSO is relatively unexplored.

Several studies have suggested that the cold winter extremes in the northern continents may be associated with Arctic sea-ice loss (e.g., Tang et al. 2013; Cohen et al. 2014; Mori et al. 2014; Lee et al. 2015; and reference therein), but large uncertainties remain due to our limited knowledge about the climate system. It is therefore important to understand the ability of the current state-of-the-art GCMs in simulating and forecasting the interannual variability in the extratropics. The long-term climate simulations may provide us information to find the possible factors or mechanisms that influence the climate in the extratropics. In this study, we evaluate the performance of climate models in this aspect using boreal winter (December, January, February) near-surface (2-m height) temperature (DJF T2m). The result shows that SME significantly decreases the interannual variability in both seasonal forecasts and climate simulations, especially in the extratropics. Next, we show the damping impact of SME on physical processes that may then weaken the variation of T2m over different Northern Hemisphere (NH) continents.

The data and methods for analysis are described in section 2. The interannual variability computed via different procedures and the simulation spreads of different sets of data are compared in section 3. The corresponding physical

processes that influence extratropical winter temperature and how these processes are influenced by SME are discussed in section 4. Section 5 is the discussion and summary.

2. DATA AND METHODS

In the past, the European Union had completed two consecutive projects that evaluated the capability of state-of-the-art coupled atmosphere-ocean models in seasonal forecast for the past (or hindcasts). The first project is called the Development of a European Multimodel Ensemble System for Seasonal to Interannual Prediction (DEMETER) project (Palmer et al. 2004). The second project is a newer generation called EMSEMBLES (Weisheimer et al. 2009). Both projects had contributed valuable results to improve seasonal forecasts. The DEMETER project includes 7 coupled atmosphere-ocean models. Each model produces a set of 6-month simulations with 9 different initial perturbations (by perturbing wind stress and sea surface temperature, SST) from 4 starting dates (see Table 1 and Fig. 1 in Palmer et al. 2004 for the details of the models and initial conditions). The hindcast data range from 1980 - 2001 (some back to 1958). The ENSEMBLES project is a 46-year (1960 - 2005) hindcast dataset (the Stream 2 simulations) produced from 6 coupled models (without flux adjustments) that consider radiative forcings from major greenhouse gases, aerosols, and clouds. Each model was initialized using European Centre for Medium-Range Weather Forecast (ECMWF) operational analyses or reanalysis (ERA-40) with 9 different initial perturbations as in DEMETER (see <https://www.ecmwf.int/en/research/projects/ensembles> and Table 1 in Weisheimer et al. 2009 for the details). For both datasets, the DJF forecasts initialized at November 1 are used in this study.

A more recent dataset, the National Centers for Environmental Prediction (NCEP) climate forecast system version 2 (CFSv2), is used to provide an updated vision of the seasonal forecast. CFSv2 is a real-time forecast system that operates on a regular basis using a fully coupled model, including the atmosphere, oceans, land, and sea ice (Saha et al. 2014). The 9-month hindcast is initiated every 5 days and run 4 cycles (00Z, 06Z, 12Z, 18Z) of that day (Fig. A1 in Saha et al. 2014). We use the monthly mean time series initialized at October 29 and November 2 that provides 8 ensemble members for DJF forecasts, starting from 1982.

For the long-term climate simulations, the World Climate Research Programme's (WCRP's) Coupled model intercomparison project phase 5 (CMIP5) (Taylor et al. 2012) historical simulations are carried out from 1850 - 2005 forced by historical emissions. The ensemble members of each model consist of simulations initialized at different points in the pre-industrial control. Models that provide only one ensemble member are excluded and therefore there are only 20 models in this study (listed in Table 1). For comparison, the ERA40 DJF T2m is used to evaluate the performance

of model simulations. ERA40 monthly means are available from 1957 - 2002 in 2.5° spatial resolution and after the analysis processes only the data during 1980 - 1998 remain in this study.

The analysis processes are shown in Fig. 1. Because the spatial resolutions are different between these datasets, all data are re-gridded to 64×128 global resolution (approximately 2.8°). The SME is the information disseminated to the weather and climate agencies for making regional and seasonal forecast. Therefore, we first calculate the SME of DJF seasonal mean for each model following this concept. Because we are focusing on interannual timescale, the 9-year running mean is subtracted to remove variations

with periods longer than this timescale. After the removal of 9-year running mean, the remaining range of year is 1980 - 1998 for DEMETER, 1985 - 1997 for ENSEMBLES, and 1987 - 2000 for CFSv2. The CMIP5 dataset is long-term simulations which have monthly data from 1980 - 2000. The temporal variance of the remaining time series is then calculated to obtain the interannual variability. The multi-model mean variance of each project is calculated at the end, except CFSv2 because the system consists of only one model. We use the mean variance of multi models to indicate the overall performance of the interannual variability in each project. The ensemble average calculated through these steps is indicated as SME+MME. In order to show the

Table 1. The list of seasonal hindcasts and climate simulations analyzed in this study. Numbers in parenthesis indicate the number of models used in analysis.

Projects	Models			
CMIP5 (20)	BCC-CSM1.1	CCSM4	CESM1-FASTCHEM	CESM1-WACCM
	CNRM-CM5	CanESM2	CSIRO-Mk3-6-0	FGOALS-g2
	FGOALS-s2	GFDL-ESM2G	GFDL-CM3	GISS-E2-H
	GISS-E2-R	HadGEM2-ES	HadGEM2-CC	MPI-ESM-LR
	MIROC5	MIROC4	MRI-CGCM3	NorESM1-M
DEMETER (7)	CERFACS	ECMWF	INGV	LODYC
	MeteoFrance	MPI	UK	
ENSEMBLES (6)	ARPEGEClimate4.6	DePreSys_HadCM3	ECHAM5_OPA8.2	ECHAM5_T63L31
	HadGEM2	IFS31R1_HOPE-E		

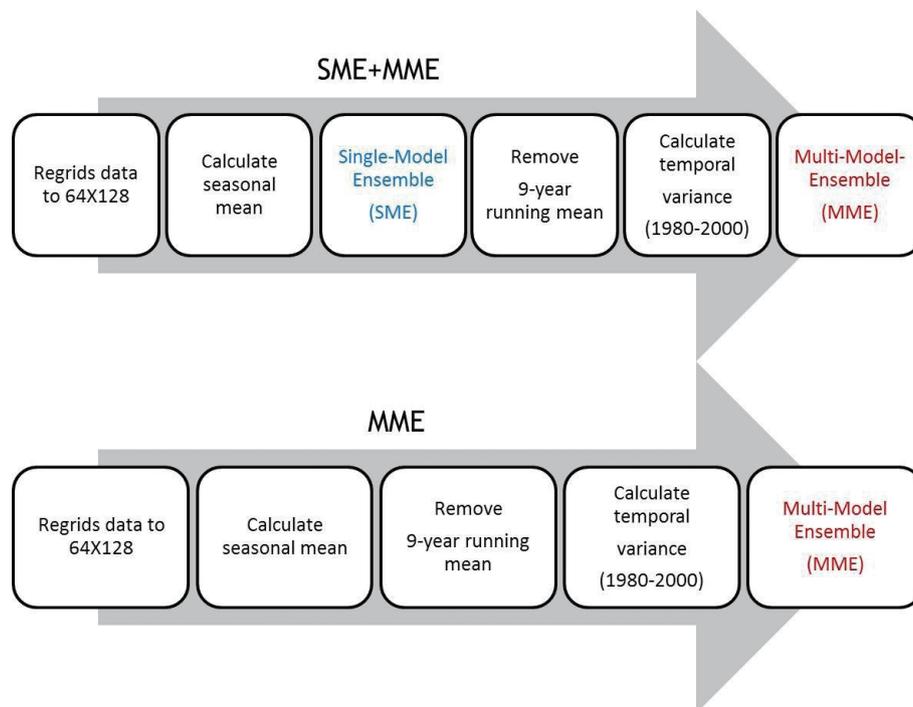


Fig. 1. The schematic for analysis processes used in this study. See the main text for the description of analysis methods.

impact of SME on interannual variability, we calculate ensemble average without SME, meaning all ensemble members are treated as individuals, and the ensemble mean of all individual members is indicated as MME (lower panel in Fig. 1). Both types of ensemble are computed by straight averaging. Note that both SME and MME can decrease the amplitude of interannual variability if the ensemble step is placed before the computation of variance. The reason that we emphasize on SME is because it is often placed at the first step in data processing.

The long-term-statistics based evaluation is different from the year-by-year performance evaluation, which is routinely conducted in the seasonal hindcast projects and forecast centers by computing the skill score of the forecast seasonal mean fields. The mostly used method is the ACC, which only measures the similarity of spatial pattern between forecast and observed deviations from climatology (Miyakoda et al. 1972), but not the actual variance. Therefore, to compare the contrast between the information revealed by two different approaches, the skill score of DJF T2m, using the definition of Hirota and Takayabu (2013), is calculated to evaluate the similarity between simulations and the observation in terms of the distribution and amplitude of the spatial pattern. The skill score used in this study is defined as the following

$$S \equiv \frac{(1 + R)^4}{4\left(\text{SDR} + \frac{1}{\text{SDR}}\right)^2} \quad (1)$$

where R is the pattern correlation between the models and the observational data and SDR is the ratio of the spatial standard deviation of the models against to that of the observational data. The score will be 1 for a simulation that is exactly the same as the observation.

3. INTERANNUAL VARIABILITY OF DJF T2M

The spatial pattern of DJF T2m interannual variances is shown in Fig. 2. In ERA-40 (Fig. 2a), the variance is large over the NH continents and tropical eastern Pacific, owing to the influences from extratropical large-scale circulation systems and the ENSO, respectively. After the standard ensemble procedure (SME+MME), the simulated variance over the NH continents is significantly weaker than observed in all hindcast datasets (Figs. 2b - e). In addition, the variance in the Southern Hemisphere (SH) is all missing. The variance of ENSEMBLES (Fig. 2c) and CFSv2 (Fig. 2d) is particularly weak, even though both systems are more advanced than DEMETER (Fig. 2b). The tropical variance associated with ENSO is overestimated in DEMETER, but it is close to ERA40 in ENSEMBLES and CFSv2. The interannual variance of ENSO is well captured because the ENSO simulation is highly constrained by the

slow evolving SST assimilated into seasonal forecasts. By contrast, in the CMIP5 ensemble (Fig. 2e), the interannual variance in the tropical eastern Pacific nearly disappears after the SME procedure. It can be easily understood that the occurring year and amplitude of ENSO events in the long-term climate simulations do not necessarily synchronize between members due to the internal variability of the coupled atmosphere-ocean system and the signals will be partially canceled during the SME procedure. As shown in Fig. 3 (taking MeteoFrance from DEMETER, ECHAM5 from ENSEMBLE, and CFSv2 as examples), the DJF T2m hindcasts over the tropical Pacific region (150°E - 80°W, 15°S - 15°N) show highly consistent behavior between ensemble members because it is influenced by the initial SST. However, the simulations using CanCM4 from CMIP5 varies markedly between members (green lines in Fig. 3d), and therefore leading to an ensemble with much smaller variance compared to seasonal forecasts (black lines in Fig. 3). The CMIP5 tropical Pacific signal may be further canceled during the MME procedure and results in small variances.

If the ensemble is made without the SME procedure (i.e., MME, Fig. 2 right panel), the extratropical variance in all projects and the tropical variance in CMIP5 are well retained. The difference between the two methods reveals that similar cancelation effect as for the tropical Pacific signal in CMIP5 may be involved for extratropical T2m during the SME procedure because the extratropical temperature fluctuation is less regulated by the tropical oceans. The impact of SME on the variance can be more than half of the amplitude. By contrast, the variance in the tropical Pacific shows very small difference with and without the SME procedure in seasonal forecasts, again reflecting the continuous influence of initial condition.

In order to show the influence from SME on each member, Fig. 4 compares the area-averaged variances of each simulation with ERA40. The box and whisker indicate the median and spread of variance of all model members (see figure caption for the details). For the extratropical region (Fig. 4a), most of the models simulated variances comparable to that in ERA40 (the horizontal dash line), except the much larger value from CFSv2. However, the variance of T2m with SME (red open circles) drops significantly in all projects. As for tropical Pacific variance (Fig. 4b), models in both ENSEMBLE and CFSv2 produced reasonable simulation, while those in DEMETER tended to simulate much larger variance comparing to ERA40. In contrast to the extratropical variance, the SME procedure does not affect the performance in the tropical Pacific (red open circles in Fig. 4b) because of the strong influence of initial condition. The contrasted performance in the tropics and extratropics suggests that good ENSO simulation (forecast) does not necessarily yield reliable simulation in the extratropics. The median of tropical variance simulated by CMIP5 models is close to the observed but with a large spread. The influence

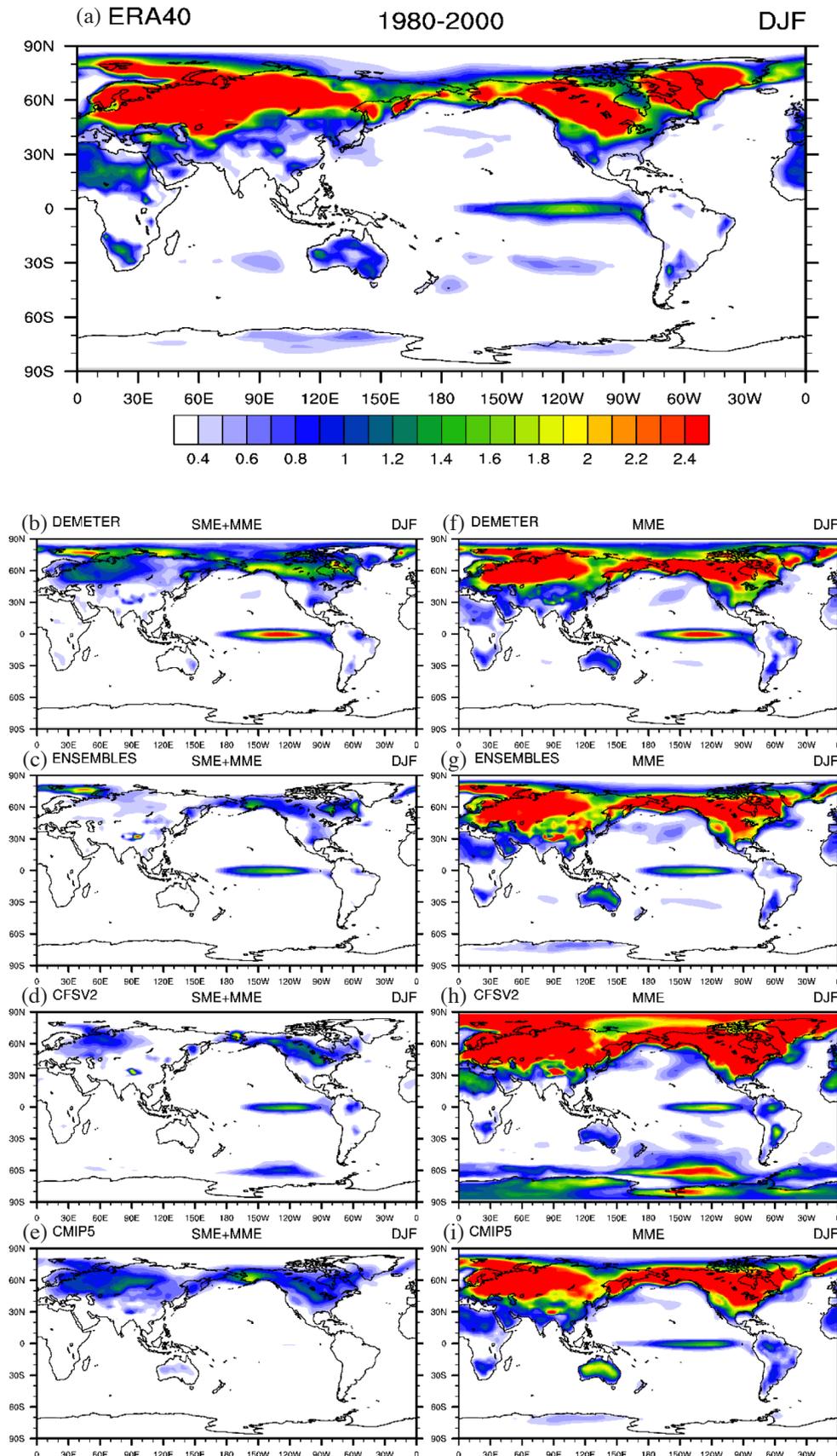


Fig. 2. The interannual variability of boreal winter near-surface temperature (DJF T2m). The top row shows ERA40. Results from each project are shown below with the project name indicated on the upper-left corners. The left panel shows ensemble variances using SME+MME method and the right panel shows results using MME.

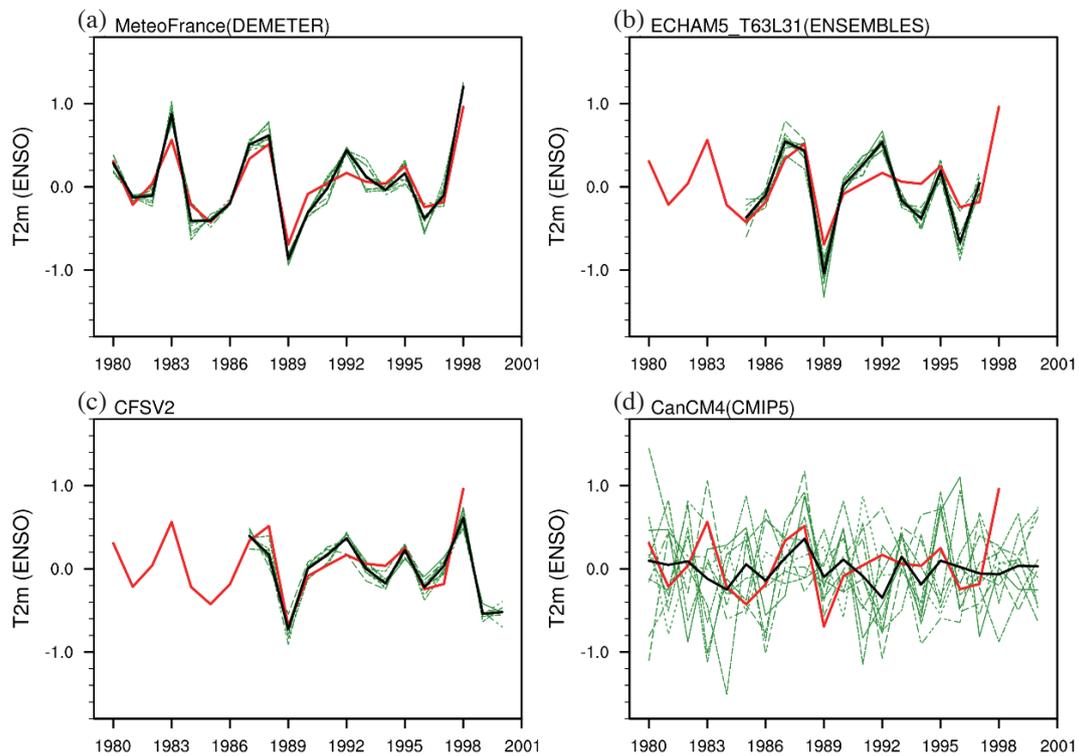


Fig. 3. The time series of averaged DJF T2m anomaly over tropical Pacific (150°E - 80°W, 15°S - 15°N) for (a) MeteoFrance (DEMETER), (b) ECHAM5 (ENSEMBLES), (c) CFSv2, and (d) CanCM4 (CMIP5). Green lines indicate ensemble members and black lines indicate single-model ensemble means. The red line indicates ERA40.

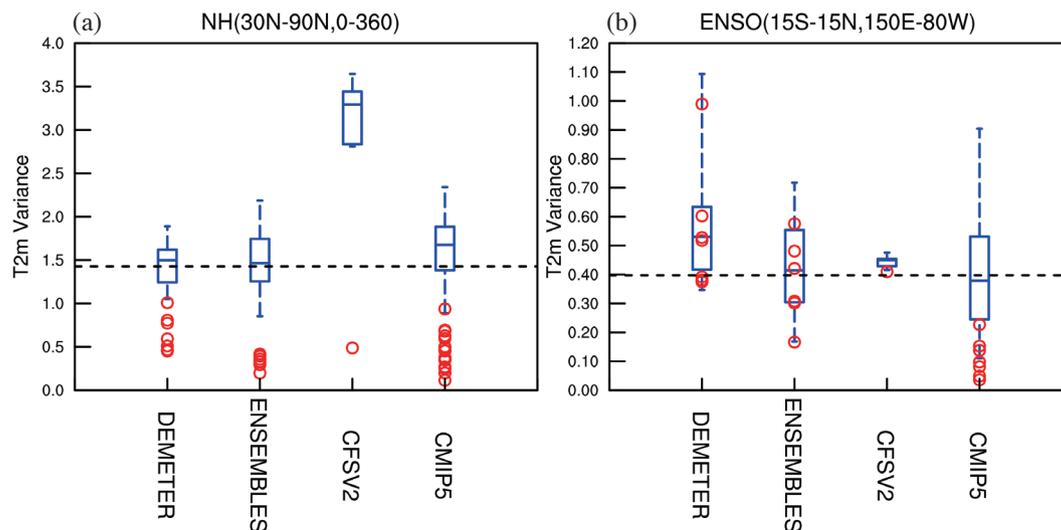


Fig. 4. The spread of area-averaged variances of all simulations (indicated by the box-and-whisker representation), comparing to SME (red open circles) and ERA40 (the dash line), for (a) the NH extratropics (30° - 90°N, 0° - 360°) and (b) the tropical Pacific (15°S - 15°N, 150°E - 80°W). The whisker lower and upper ends indicate the minimum and maximum of the ensemble spread, the line in the box marks the ensemble median, and the box bottom and top are the first and third quartiles.

of SME for CMIP5 tropical Pacific is seen clearly again as the values of SME (red open circles) are much smaller than the variance of individual members (the box and whisker) and ERA40 (the horizontal dash line). The results presented in this section indicate that if the interannual variance of individual member is calculated individually (i.e., the MME calculation in this study), the extratropical variance can be retained much better in seasonal forecast. For climate simulation, both extratropical and tropical variances are kept better as shown in Fig. 2.

The skill score of DJF T2m is calculated for the NH extratropical and the tropical Pacific areas (Fig. 5). The score of each winter season is computed and then averaged over the available period. For the NH extratropics (Fig. 5a), the scores of both individual member (indicated by the box and whisker) and the SME mean (red open circle) are very low in all projects, indicating poor capability of current AGCMs in hindcasting the spatial pattern of extratropical winter temperature. For the tropical Pacific (Fig. 5b), the scores are much higher and the SME (red open circles) increases the score slightly for seasonal forecasts but has little effect on CMIP5 as expected. Our calculation also indicates that most of the scores for the tropical Pacific are contributed from ENSO years (the scores can be more than 0.7 for strong ENSO years) whereas skill scores of non-ENSO years are often below 0.5 (figures not shown). Since the skill score for the NH extratropics shows no clear association with ENSO, it is suggested that ENSO's contribution to model performance is mainly confined in the tropics during strong ENSO events and is quite minimum in the extratropics.

4. DJF LARGE-SCALE CIRCULATIONS

The temperature variances over NH extratropical continents (North America and North Eurasia) are associated with different processes. The further question is what physical processes are sensitive to the SME procedure. ENSO has been identified as one of the major factors in shaping the interannual variability of temperature, precipitation, and circulation in the extratropics (Rasmusson and Mo 1993; Hurrell 1995; Trenberth et al. 1998; Allan 2000). The atmospheric waves induced by tropical thermal forcing was suggested by Hoskins and Karoly (1981). Recently, several studies suggested that a positive SST anomaly in the tropical western Pacific may induce Rossby waves and cause the extratropical circulation changes (Ding et al. 2014; Wang and Schubert 2014; Lee et al. 2015; Watson et al. 2016). There are also studies debating on the relative contribution from ENSO and the extratropical variability pattern (i.e., Baxter and Nigam 2015).

Other studies suggest connections with the Arctic sea-ice decline (Francis and Vavrus 2012; Tang et al. 2013; Frankignoul et al. 2014; Mori et al. 2014) owing to the anthropogenic forcing. Recent studies identified the marked

effect of the Arctic sea ice on the interannual variability of surface temperature and atmospheric perturbations over the high-latitude Eurasian and North American continents (Petoukhov and Semenov 2010; Kug et al. 2015). Furthermore, the Arctic Oscillation (AO) and the Arctic sea ice were found to be dynamically linked (Yang et al. 2016). It follows that a numerical model should be able to not only simulate the internal variability associated with ENSO and AO/North Atlantic Oscillation (NAO) but also need to realistically capture the effect of the Arctic sea ice and its interaction with the high-latitude circulation. By contrast, Wallace et al. (2014) warns that the connection between Arctic sea-ice loss and mid-latitude extreme winters is not supported by observational studies (Barnes 2013; Barnes et al. 2014).

The connection between regional temperature and other physical variables can be easily shown by the cross-covariance maps (Fig. 6). Figure 6a shows the covariance between T2m averaged over North America (40° - 70° N, 120° - 40° W) and T2m at all other grid points. The figure shows that North American T2m closely varies with tropical Pacific T2m. Figure 6b shows the covariance between North American T2m and 500 hPa geopotential height (Z500) at all grid points. It shows that the North American T2m is not associated with Arctic Z500 but is associated with North Pacific Z500, which is part of the wave-like pattern originating from the tropical Pacific. This wave-like pattern also influences T2m in the North Pacific and produces negative covariance values there. However, it has to be noted that the signal of thermal forcing is still from the tropics. The same calculation is carried out for the North Eurasian continent (40° - 70° N, 1° - 150° E) as well (Figs. 6c and d). In contrast to North America, Fig. 6c shows very low covariance between the North Eurasia and the tropical Pacific, indicating minimum influences from ENSO events. However, the North Eurasian T2m is highly associated with the Arctic pressure variation and anomaly over the North Atlantic (Fig. 6d), suggesting the influences from high-latitude systems. A similar result was shown in Fig. 3a in Higgins et al. (2002). These high-latitude influences have been suggested to be associated with Arctic sea ice loss and AO/NAO. Kug et al. (2015) identified two distinct geopotential height patterns over the Eurasian and North American continents that were associated, respectively, with the temperature variation in the Barents-Kara Sea region and the East Siberian-Chukchi Sea region (their Figs. 3c and d). These two patterns resemble the covariance patterns shown in Figs. 6d and b, respectively.

The performance of each project in turns of simulating the physical processes associated with extratropical T2m is shown in Figs. 7 and 8. Figure 7 shows the covariance between North American T2m and Z500 of each grid point (similar to Fig. 6b, but for models). In the left panel, the variables of ensemble members are averaged first (i.e., SME). Then the covariance is calculated and the multi-model mean covariance is then averaged. In the right panel, the

covariance is calculated first for each ensemble member and then straight averaged. Figure 7 shows that the covariance is clearly damped in the left panel, which means SME smooths out the amplitudes of atmospheric perturbations.

Figure 8 is the same as Fig. 7, but for comparison with Fig. 6d, showing the covariance between North Eur-

asian T2m and Z500 of all grid points. For North Eurasia, the physical processes that influence T2m are suggested to be atmospheric internal modes, such as AO/NAO and feedbacks involved with the Arctic sea ice loss as mentioned previously. All of these processes are not simulated well enough in GCMs. Therefore, the covariances between North

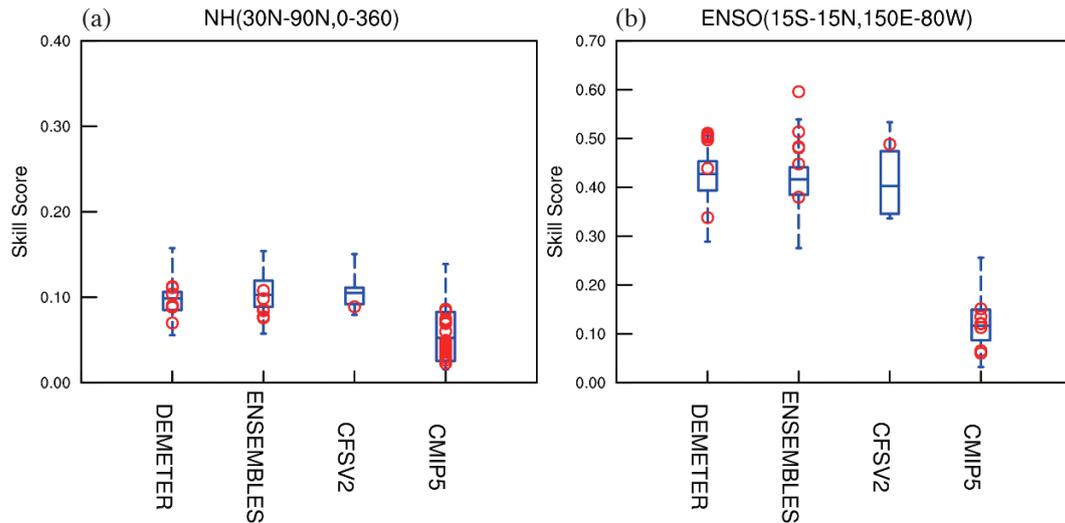


Fig. 5. Skill scores of DJF T2m with respect to ERA40 for (a) the NH extratropics ($30^{\circ} - 90^{\circ}\text{N}$, $0^{\circ} - 360^{\circ}$) and (b) the tropical Pacific ($15^{\circ}\text{S} - 15^{\circ}\text{N}$, $150^{\circ}\text{E} - 80^{\circ}\text{W}$). Scores for ensemble spread are indicated by the box-and-whisker representation, comparing to scores involving SME (red open circles). The whisker lower and upper ends indicate the minimum and maximum of the ensemble spread, the line in the box marks the ensemble median, and the box bottom and top are the first and third quartiles.

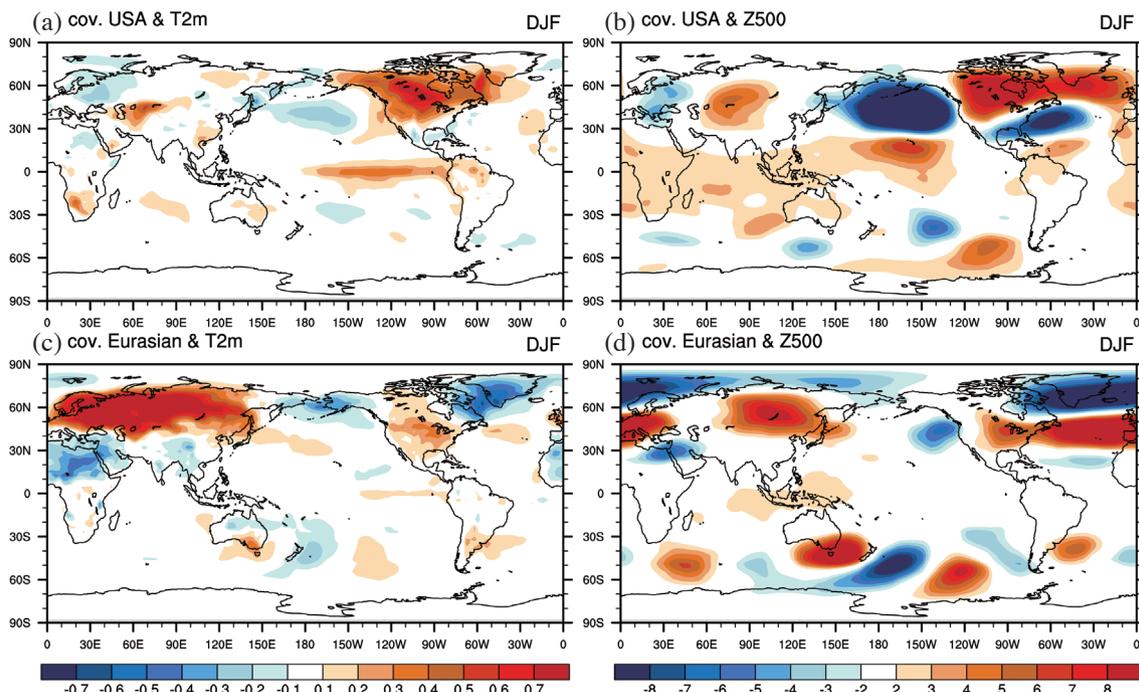


Fig. 6. The covariance between T2m (averaged over selected regions) and selected physical variables at each grid point, using the seasonal mean time series of ERA40 from 1980 - 2000. (a) Covariance between North American T2m ($40^{\circ} - 70^{\circ}\text{N}$, $120^{\circ} - 40^{\circ}\text{W}$) and T2m at each grid point; (b) North American T2m and Z500 at each grid point; (c) North Eurasian T2m ($40^{\circ} - 70^{\circ}\text{N}$, $1^{\circ} - 150^{\circ}\text{E}$) and T2m at each grid point; (d) North Eurasian T2m and Z500 at each grid point. Note the color ranges are different in the left and right columns.

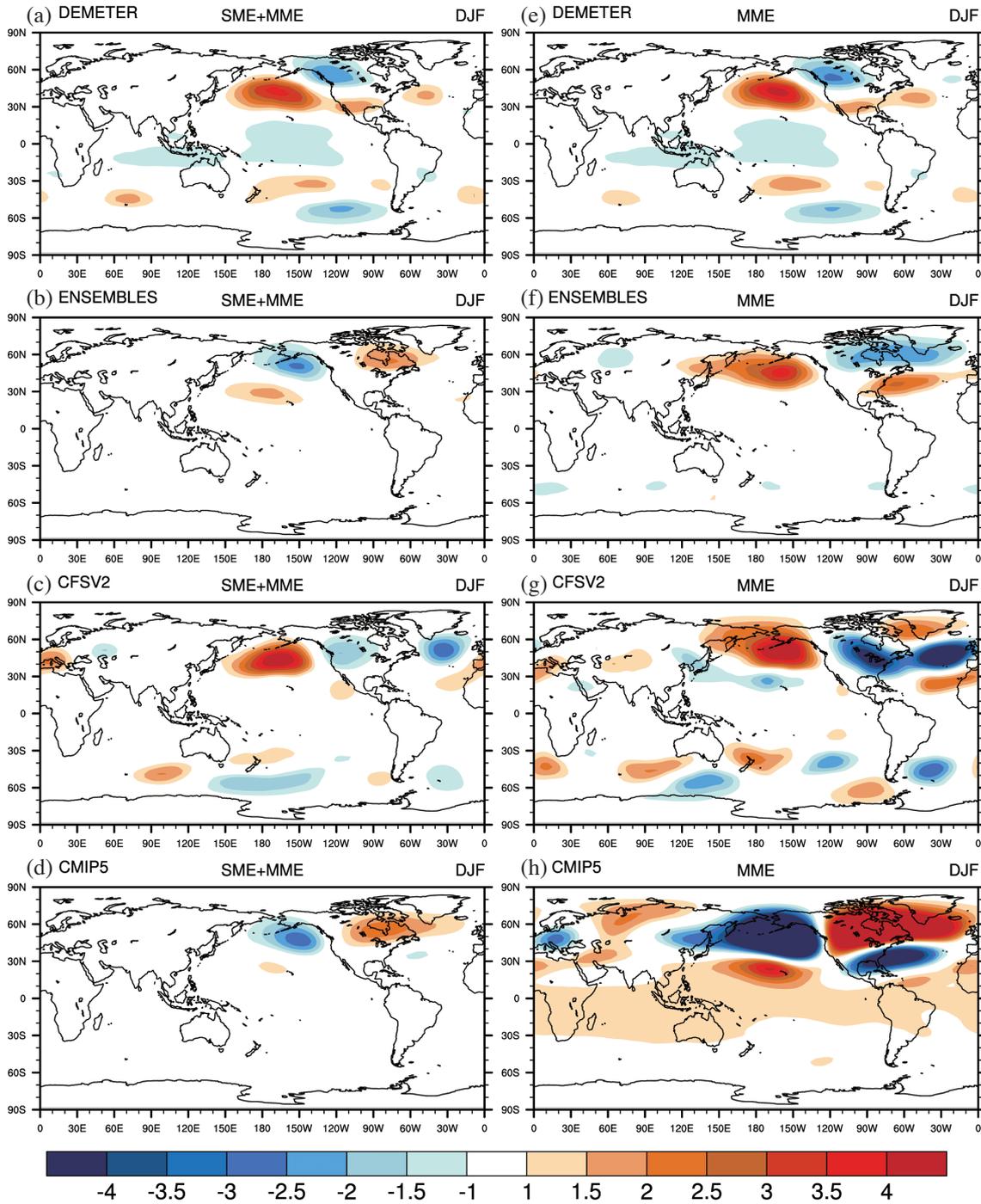


Fig. 7. Covariances between North American T2m and Z500 at each grid point (similar to Fig. 6b) for each set of model simulations (from top to bottom rows) with SME involved (left panel) or not (right panel). Note the color range is half of that in Fig. 6b in order to show details.

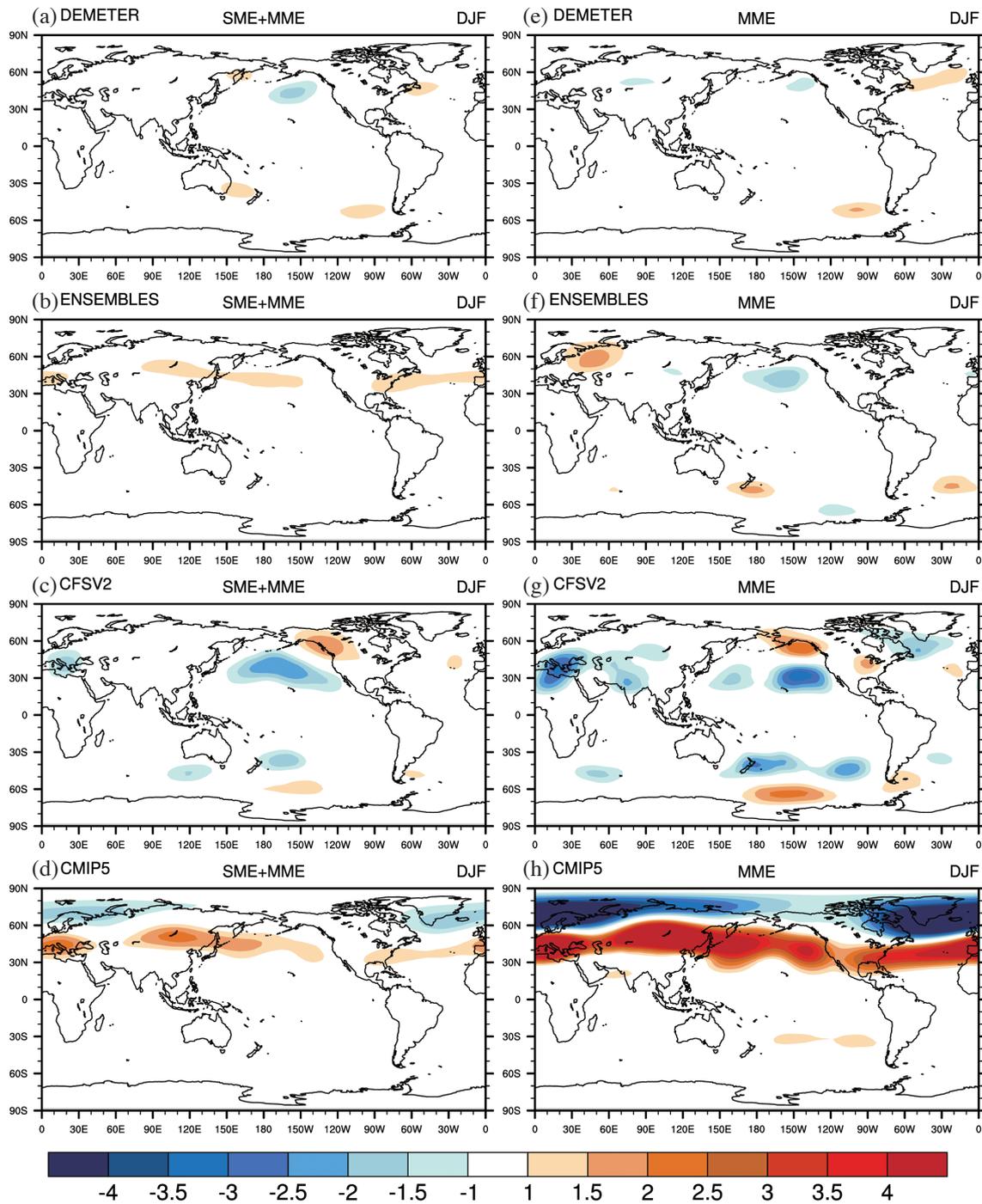


Fig. 8. The same as Fig. 7, but for covariances between North Eurasian T2m and Z500 at each grid point (similar to Fig. 6d).

Eurasian T2m and Z500 are mostly very small, except in CMIP5. This implies that the signals of atmospheric internal modes may be retained better in long-term climate simulations than in seasonal forecast, but the exact reasons need further investigation. Still, the damping effect from SME is also seen in simulations.

The tropical-extratropical connection and the influences of SME can be examined by Z500 and U200. Z500 can represent the extratropical variability while U200 represents the tropical-extratropical variability. The interannual variability of Z500 (Fig. 9a) shows a strong signal in the North Pacific and secondary maxima in the North Atlantic and northwestern Eurasia. The latter is likely associated with the AO/NAO. The one in the North Pacific, which is located in the pathway of the wave-like perturbations (as shown in Fig. 6b) emanating from the tropical Pacific and also near one action center of the AO, is more likely to be associated with both ENSO and AO. This interpretation is confirmed by a comparison with the regression maps between Z500 and the indices of ENSO, AO, and NAO (figures can be obtained from the NOAA/ESRL Physical Sciences Division, Boulder Colorado from their Web site at <http://www.esrl.noaa.gov/psd/>). Accordingly, the maximum variance of Z500 in the North Pacific is contributed from both tropical Pacific SST forcing and the AO while the one over Northwestern Eurasia is contributed mostly from the AO/NAO. In addition, the Arctic sea ice variation is likely to influence both maximum variance as discussed above.

The tropical connection of the variability in the North Pacific can be further demonstrated in the observed interannual variability of U200 (Fig. 10a), which shows a tripole pattern over the tropics and subtropical areas in the eastern Pacific and Atlantic Ocean. The large variance of U200 over the Pacific can be attributed to the Rossby wave-like perturbations forced by the variation of tropical heating associated with ENSO (Hoskins and Karoly 1981; Ropelewski and Halpert 1986; Trenberth et al. 1998; Yang and DelSole 2012). The perturbation influences the atmospheric circulations and geopotential heights as propagating northward and influences the winter temperature in North America.

The influence of the SME procedure can be observed in Z500 and U200 as well (compare the left and right panels in Figs. 9 and 10). The interannual variation of Z500 (Fig. 9) shows that the variance over NH becomes much weaker after SME. The wave-like pattern in U200 (Fig. 10), although weaker, remains in all SME+MME seasonal hindcasts over the Pacific in which ENSO is included in the initial conditions, but it is totally missing in the Atlantic Ocean and in CMIP5 ensemble mean (Fig. 10e). This result suggests that as the Rossby-wave like perturbation propagates from the tropics to the extratropics, the wave is dispersed and may interact with local perturbations that are less constrained by the tropical SST. Therefore, the amplitudes of extratropical circulation and T2m perturbations induced by the tropical

forcing through this wave activity are both partially canceled between members during the SME procedure and the interannual variability is significantly reduced. This result suggests that the high latitude signal induced by the tropical perturbation through this wave activity may also be weakened in CMIP5. The remaining DJF T2m interannual variability in NH extratropics are therefore likely originated in the extratropics.

However, as suggested by Cohen et al. (2014), the winter temperature in the northern Eurasian continent is more likely associated with the AO/NAO, which also partially influences the North America temperature (Van Loon and Rogers 1978; Barnston and Livezey 1987; Hurrell 1995; Thompson and Wallace 1998; Visbeck et al. 2001; Higgins et al. 2002). Both AO and NAO are likely part of internal atmospheric variability (Overland and Wang 2015) that are sensitive to the initial conditions and small perturbations. The related mechanisms are complicated and are out of our scope, but we do see that the SME procedure can easily result in the cancellation of forecasted anomalies made by individual members and inaccurately suggest the poor model performance in simulating long-term statistics as revealed here.

5. CONCLUSIONS

The standard ensemble procedure is usually done with SME first and then MME. The evaluation of seasonal forecast considers mostly the mean state either from single model or multi models, and the evaluation of variance is included in ACC which only considers pattern correlation not the actual values. The SME procedure may not influence the evaluation of the mean state significantly but have strong impact on the variance (or interannual variability). The reason is likely due to the cancellation between positive and negative anomalies between ensemble members during the SME procedure. Therefore, a seasonal forecast can be evaluated as good in terms of the mean, but not in the interannual variance. For example, the interannual variability of predicted seasonal means may be realistic for individual members but is significantly reduced in the SME values because of the cancellation effect, which leads to the reduced amplitudes or even ambiguous signs of ensemble-mean predicted fields. Such cancellation effect is more significant in the extratropics where atmospheric internal variability is larger than in the tropics where the ENSO dominates. This is much less a problem in the tropics for the seasonal forecasts that are initiated with well-observed initial condition in both atmosphere and ocean. As it has been known that the well-simulated ENSO and tropical temperature contribute to most of the forecast skill (Weisheimer et al. 2009; Alessandri et al. 2011) in the tropical variability. By contrast, the cancellation effect in the tropics remains significant for CMIP type simulations because of independence of initial conditions.

The extratropical temperature variances are contributed

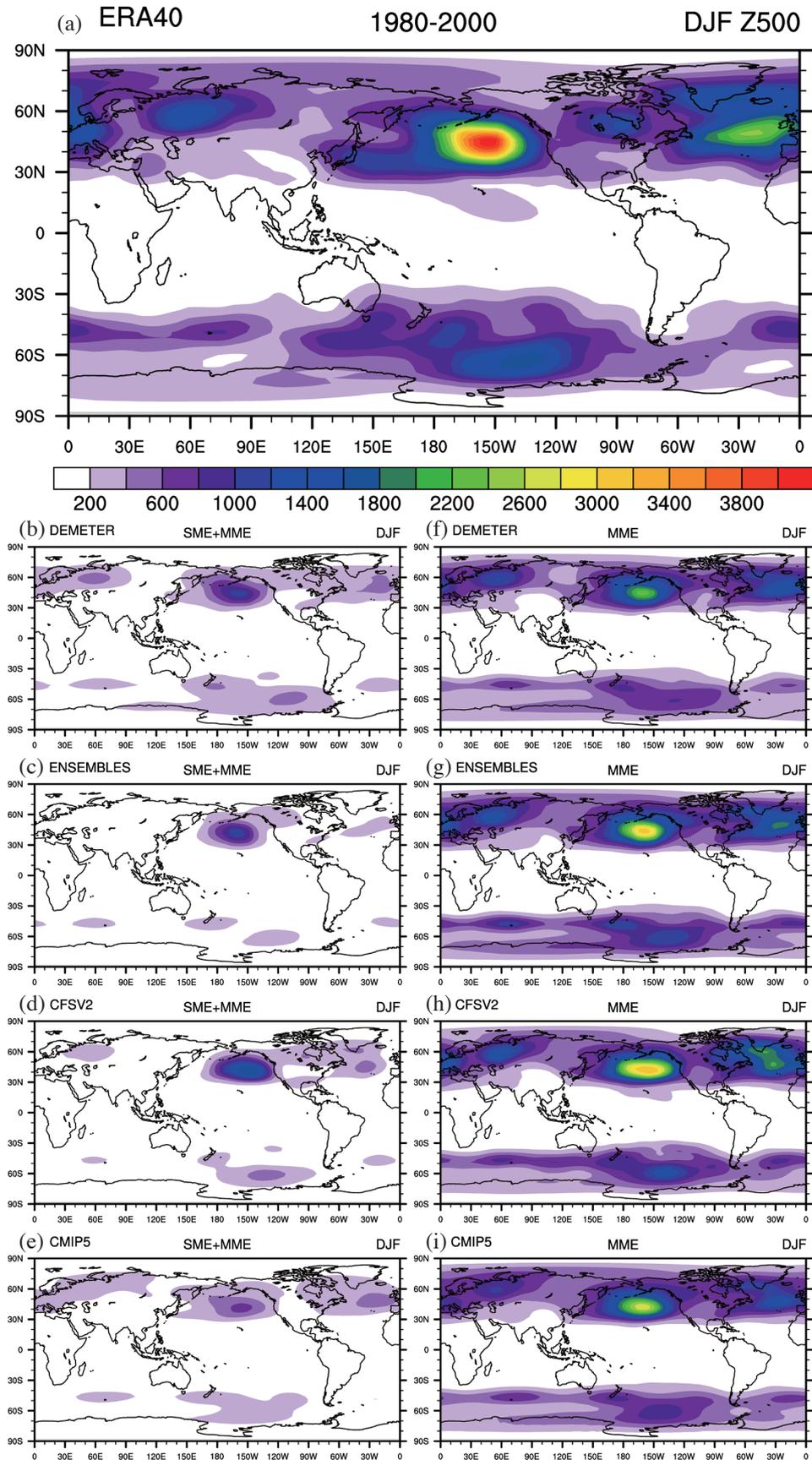


Fig. 9. The interannual variability of the 500 hPa geopotential height (Z500). The top row shows ERA40. Results from each project are shown below with the project name indicated on the upper-left corners. The left panel shows ensemble variances using SME+MME method and the right panel shows results using MME.

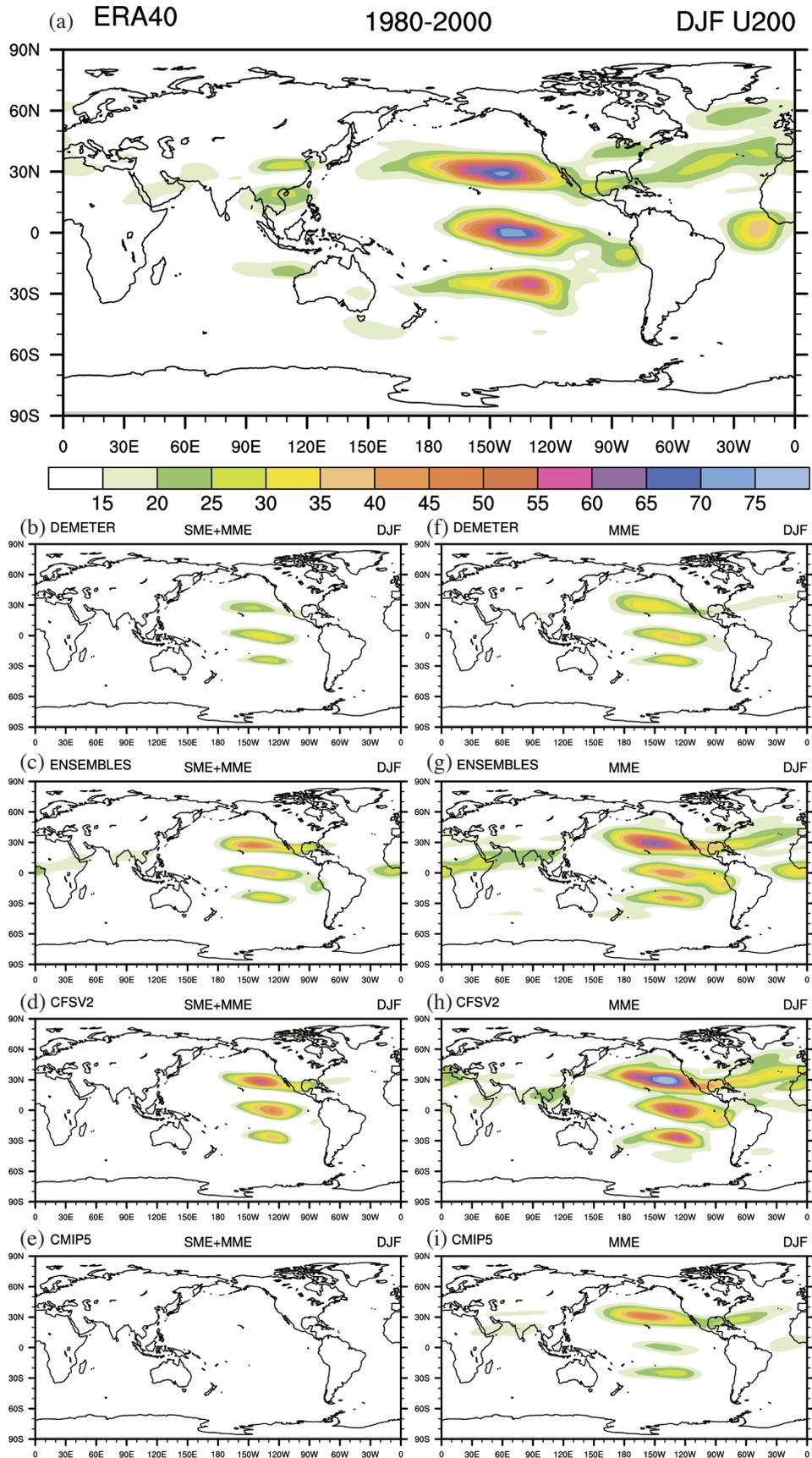


Fig. 10. The same as Fig. 9, but for 200 hPa zonal wind (U200).

from both tropical and extratropical phenomena, i.e., the wave-like perturbations originating from the tropics and the perturbations of extratropical origin such as AO, NAO, and the Arctic sea ice. The SME procedure will smear out the amplitude of forecasted perturbation, even when the tropical SST signals are strong. The atmospheric circulation induced by tropical Pacific SST is not well represented in global climate models and the simulation is further deteriorated by SME, leading to an unrealistically weak interannual variance of simulated winter temperature in North America. Similar effect is also found in North Eurasia where winter temperature is strongly influenced by atmospheric internal variability and its interaction with land and ice/snow in the middle-high latitudes. The same influence of SME procedure occurs in climate simulations (CMIP5) as well. Also inaccurate spatial location of SST anomaly (e.g., simulated warm anomaly in the tropical western Pacific instead of in the central Pacific as it should be) can also easily lead to completely different signs of anomaly in North America because the induced Rossby-wave like perturbation in the extratropics may shift to the west of should-be location. The influences of extratropical factors such as the Arctic sea ice content and land/sea surface temperature are another important influencing factors and should be well captured for reasonable seasonal simulation and forecast.

The SME procedure has little effect on the evaluation of year-by-year seasonal forecast based on the skill score of seasonal mean fields in the extratropics but slightly increase the skill score for the tropics. The skill scores for CMIP5 are very low as expected because of the independence of initial conditions. Interestingly, the skill score with and without SME in the extratropics is only slightly higher in the seasonal forecast than in the CMIP5. This reflects the low predictability of seasonal forecast models for the extratropical fluctuations even though many seasonal forecast models performed quite well in ENSO forecast. The results of this study suggest that the SME procedure should be avoided when evaluating the model performance in simulating the higher-order long-term statistics (such as variance). Variance of individual models should be calculated first and then averaged among members.

Another challenge is the Arctic amplification that has been suggested to be associated with the global warming tendency and may have become a more important factor that influences the extratropical temperature and climate in general. Skillful forecast of ENSO, which has been the major focus during the development of seasonal forecast models in the past few decades, is no longer sufficient for further improvement. Better understanding of the extratropical variability such as the AO/NAO and the influence of Arctic sea ice and the ability to simulate their influences will definitely help improve the extratropical seasonal forecast and long-term climate simulation. One unsolved issue remains: why the seasonal forecast models reasonably simulated

long-term interannual variance in the extratropics while simulated poorly the year-by-year variation. The likely answer may be that the models already have good capability in simulating the long-term statistics of stochastic processes in the extratropics, but the capability in accurately simulating the temporal variation is still poor. This is clearly evident in the CMIP5 long-term climate simulations.

Acknowledgements We thank the anonymous reviewer for his/her helpful comments to improve and clarify this manuscript. The DEMETER and ENSEMBLES datasets were funded by the EU under the Contract EVK2-1999-00024 and 505539, whose support is gratefully acknowledged. The DEMETER data are available at <http://apps.ecmwf.int/datasets/data/demeter-mnth/>. The ENSEMBLES can be obtained at <https://www.ecmwf.int/en/research/projects/ensembles> and NCEP CFSv2 is at <http://cfs.ncep.noaa.gov/>. We acknowledge the World Climate Research Programme's Working Group on Coupled Modelling, which is responsible for CMIP, and we thank the climate modeling groups for producing and making available their model output. The CMIP5 data are available at <https://pcmdi.llnl.gov/search/cmip5/>. The ERA40 data are at <http://apps.ecmwf.int/datasets/data/era40-mnth/>. This research is supported by the Ministry of Science and Technology, Taiwan (Grants: MOST 107-2119-M-034-006- and MOST 107-2119-M-001-010-).

REFERENCES

- Alessandri, A., A. Borrelli, A. Navarra, A. Arribas, M. Déqué, P. Rogel, and A. Weisheimer, 2011: Evaluation of Probabilistic Quality and Value of the ENSEMBLES Multimodel Seasonal Forecasts: Comparison with DEMETER. *Mon. Weather Rev.*, **139**, 581-607, doi: 10.1175/2010MWR3417.1. [[Link](#)]
- Allan, R. J., 2000: ENSO and climatic variability in the past 150 years. In: Diaz, H. F. and V. Markgraf (Eds.), *El Niño and the Southern Oscillation: Multiscale Variability and Global and Regional Impacts*, Cambridge University Press, Cambridge, UK, 3-56, doi: 10.1017/cbo9780511573125.002. [[Link](#)]
- Barnes, E. A., 2013: Revisiting the evidence linking Arctic amplification to extreme weather in midlatitudes. *Geophys. Res. Lett.*, **40**, 4734-4739, doi: 10.1002/grl.50880. [[Link](#)]
- Barnes, E. A., E. Dunn-Sigouin, G. Masato, and T. Woollings, 2014: Exploring recent trends in Northern Hemisphere blocking. *Geophys. Res. Lett.*, **41**, 638-644, doi: 10.1002/2013GL058745. [[Link](#)]
- Barnston, A. G. and R. E. Livezey, 1987: Classification, Seasonality and Persistence of Low-Frequency Atmospheric Circulation Patterns. *Mon. Weather Rev.*, **115**, 1083-1126, doi: 10.1175/1520-0493(1987)115<1083:CSAPOL>2.0.CO;2. [[Link](#)]

- Baxter, S. and S. Nigam, 2015: Key Role of the North Pacific Oscillation–West Pacific Pattern in Generating the Extreme 2013/14 North American Winter. *J. Climate*, **28**, 8109–8117, doi: 10.1175/JCLI-D-14-00726.1. [[Link](#)]
- Cohen, J., J. A. Screen, J. C. Furtado, M. Barlow, D. Whittleston, D. Coumou, J. Francis, K. Dethloff, D. Entekhabi, J. Overland, and J. Jones, 2014: Recent Arctic amplification and extreme mid-latitude weather. *Nat. Geosci.*, **7**, 627–637, doi: 10.1038/ngeo2234. [[Link](#)]
- Ding, Q., J. M. Wallace, D. S. Battisti, E. J. Steig, A. J. E. Gallant, H. J. Kim, and L. Geng, 2014: Tropical forcing of the recent rapid Arctic warming in northeastern Canada and Greenland. *Nature*, **509**, 209–212, doi: 10.1038/nature13260. [[Link](#)]
- Francis, J. A. and S. J. Vavrus, 2012: Evidence linking Arctic amplification to extreme weather in mid-latitudes. *Geophys. Res. Lett.*, **39**, doi: 10.1029/2012GL051000. [[Link](#)]
- Frankignoul, C., N. Sennéchaël, and P. Cauchy, 2014: Observed Atmospheric Response to Cold Season Sea Ice Variability in the Arctic. *J. Climate*, **27**, 1243–1254, doi: 10.1175/JCLI-D-13-00189.1. [[Link](#)]
- Higgins, R. W., A. Leetmaa, and V. E. Kousky, 2002: Relationships between Climate Variability and Winter Temperature Extremes in the United States. *J. Climate*, **15**, 1555–1572, doi: 10.1175/1520-0442(2002)015<1555:RBCVAW>2.0.CO;2. [[Link](#)]
- Hirota, N. and Y. N. Takayabu, 2013: Reproducibility of precipitation distribution over the tropical oceans in CMIP5 multi-climate models compared to CMIP3. *Climate Dyn.*, **41**, 2909–2920, doi: 10.1007/s00382-013-1839-0. [[Link](#)]
- Hoskins, B. J. and D. J. Karoly, 1981: The steady linear response of a spherical atmosphere to thermal and orographic forcing. *J. Atmos. Sci.*, **38**, 1179–1196, doi: 10.1175/1520-0469(1981)038<1179:TSLROA>2.0.CO;2. [[Link](#)]
- Hurrell, J. W., 1995: Decadal Trends in the North Atlantic Oscillation: Regional Temperatures and Precipitation. *Science*, **269**, 676–679, doi: 10.1126/science.269.5224.676. [[Link](#)]
- Kug, J. S., J. H. Jeong, Y. S. Jang, B. M. Kim, C. K. Folland, S. K. Min, and S. W. Son, 2015: Two distinct influences of Arctic warming on cold winters over North America and East Asia. *Nat. Geosci.*, **8**, 759–762, doi: 10.1038/NNGEO2517. [[Link](#)]
- Lee, M. Y., C. C. Hong, and H. H. Hsu, 2015: Compounding effects of warm sea surface temperature and reduced sea ice on the extreme circulation over the extratropical North Pacific and North America during the 2013–2014 boreal winter. *Geophys. Res. Lett.*, **42**, 1612–1618, doi: 10.1002/2014GL062956. [[Link](#)]
- Madden, R. A. and P. R. Julian, 1971: Detection of a 40–50 day oscillation in the zonal wind in the tropical Pacific. *J. Atmos. Sci.*, **28**, 702–708, doi: 10.1175/1520-0469(1971)028<0702:DOADOI>2.0.CO;2. [[Link](#)]
- Miyakoda, K., G. D. Hembree, R. F. Strickler, and I. Shulman, 1972: Cumulative results of extended forecast experiments. Part I: Model performance for winter cases. *Mon. Weather Rev.*, **100**, 836–855, doi: 10.1175/1520-0493(1972)100<0836:CROEFE>2.3.CO;2. [[Link](#)]
- Mori, M., M. Watanabe, H. Shiogama, J. Inoue, and M. Kimoto, 2014: Robust Arctic sea-ice influence on the frequent Eurasian cold winters in past decades. *Nat. Geosci.*, **7**, 869–873, doi: 10.1038/ngeo2277. [[Link](#)]
- Overland, J. E. and M. Wang, 2015: Increased Variability in the Early Winter Subarctic North American Atmospheric Circulation. *J. Climate*, **28**, 7297–7305, doi: 10.1175/JCLI-D-15-0395.1. [[Link](#)]
- Palmer, T., 2014: Record-breaking winters and global climate change. *Science*, **344**, 803–804, doi: 10.1126/science.1255147. [[Link](#)]
- Palmer, T. N., A. Alessandri, U. Andersen, P. Cantelaube, M. Davey, P. Décléuse, M. Déqué, E. Díez, F. J. Doblas-Reyes, H. Feddersen, R. Graham, S. Gualdi, J. F. Guérémy, R. Hagedorn, M. Hoshen, N. Keenlyside, M. Latif, A. Lazar, E. Maisonave, V. Marletto, A. P. Morse, B. Orfila, P. Rogel, J. M. Terres, and M. C. Thomson, 2004: Development of a European multi-model ensemble system for seasonal-to-interannual prediction (DEMETER). *Bull. Amer. Meteorol. Soc.*, **85**, 853–872, doi: 10.1175/BAMS-85-6-853. [[Link](#)]
- Petoukhov, V. and V. A. Semenov, 2010: A link between reduced Barents–Kara sea ice and cold winter extremes over northern continents. *J. Geophys. Res.*, **115**, D21111, doi: 10.1029/2009JD013568. [[Link](#)]
- Rasmusson, E. M. and K. Mo, 1993: Linkages between 200-mb Tropical and Extratropical Circulation Anomalies during the 1986–1989 ENSO Cycle. *J. Climate*, **6**, 595–616, doi: 10.1175/1520-0442(1993)006<0595:LBMTAE>2.0.CO;2. [[Link](#)]
- Ropelewski, C. F. and M. S. Halpert, 1986: North American precipitation and temperature patterns associated with the El Niño/Southern Oscillation (ENSO). *Mon. Weather Rev.*, **114**, 2352–2362, doi: 10.1175/1520-0493(1986)114<2352:NAPATP>2.0.CO;2. [[Link](#)]
- Saha, S., S. Moorthi, X. Wu, J. Wang, S. Nadiga, P. Tripp, D. Behringer, Y.-T. Hou, H. Chuang, M. Iredell, M. Ek, J. Meng, R. Yang, M. P. Mendez, H. van den Dool, Q. Zhang, W. Wang, M. Chen, and E. Becker, 2014: The NCEP Climate Forecast System Version 2. *J. Climate*, **27**, 2185–2208, doi: 10.1175/JCLI-D-12-00823.1. [[Link](#)]
- Slingo, J. and T. Palmer, 2011: Uncertainty in weather and climate prediction. *Phil. Trans. Math. Phys. Eng. Sci.*, **369**, 4751–4767, doi: 10.1098/rsta.2011.0161. [[Link](#)]
- Tang, Q., X. Zhang, X. Yang, and J. A. Francis, 2013: Cold winter extremes in northern continents linked to

- Arctic sea ice loss. *Environ. Res. Lett.*, **8**, 014036, doi: 10.1088/1748-9326/8/1/014036. [[Link](#)]
- Taylor, K. E., 2001: Summarizing multiple aspects of model performance in a single diagram. *J. Geophys. Res.*, **106**, 7183-7192, doi: 10.1029/2000JD900719. [[Link](#)]
- Taylor, K. E., R. J. Stouffer, and G. A. Meehl, 2012: An Overview of CMIP5 and the Experiment Design. *Bull. Amer. Meteorol. Soc.*, **93**, 485-498, doi: 10.1175/BAMS-D-11-00094.1. [[Link](#)]
- Thompson, 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, doi: 10.1029/98GL00950. [[Link](#)]
- Trenberth, K. E., G. W. Branstator, D. Karoly, A. Kumar, N. C. Lau, and C. Ropelewski, 1998: Progress during TOGA in understanding and modeling global teleconnections associated with tropical sea surface temperatures. *J. Geophys. Res.*, **103**, 14291-14324, doi: 10.1029/97JC01444. [[Link](#)]
- Van Loon, H. and J. C. Rogers, 1978: The Seesaw in Winter Temperatures between Greenland and Northern Europe. Part I: General Description. *Mon. Weather Rev.*, **106**, 296-310, doi: 10.1175/1520-0493(1978)106<0296:TSIWTB>2.0.CO;2. [[Link](#)]
- Visbeck, M. H., J. W. Hurrell, L. Polvani, and H. M. Cullen, 2001: The North Atlantic Oscillation: Past, present, and future. *Proc. Natl. Acad. Sci.*, **98**, 12876-12877, doi: 10.1073/pnas.231391598. [[Link](#)]
- Wallace, J. M., I. M. Held, D. W. J. Thompson, K. E. Trenberth, and J. E. Walsh, 2014: Global Warming and Winter Weather. *Science*, **343**, 729-730, doi: 10.1126/science.1251172. [[Link](#)]
- Wang, H. and S. Schubert, 2014: Causes of the extreme dry conditions over California during early 2013. In: Herring, S. C., M. P. Hoerling, T. C. Peterson, and P. A. Stott (Eds.), Explaining Extreme Events of 2013 from a Climate Perspective. *Bull. Amer. Meteorol. Soc.*, **95**, S1-S104, doi: 10.1175/1520-0477-95.9.S1.1. [[Link](#)]
- Watson, P. A. G., A. Weisheimer, J. R. Knight, and T. N. Palmer, 2016: The role of the tropical West Pacific in the extreme Northern Hemisphere winter of 2013/2014. *J. Geophys. Res.*, **121**, 1698-1714, doi: 10.1002/2015JD024048. [[Link](#)]
- Weisheimer, A., F. J. Doblas-Reyes, T. N. Palmer, A. Alessandri, A. Arribas, M. Déqué, N. Keenlyside, M. MacVean, A. Navarra, and P. Rogel, 2009: ENSEMBLES: A new multi-model ensemble for seasonal-to-annual predictions—Skill and progress beyond DEMETER in forecasting tropical Pacific SSTs. *Geophys. Res. Lett.*, **36**, doi: 10.1029/2009GL040896. [[Link](#)]
- Yang, X. and T. DelSole, 2012: Systematic comparison of ENSO teleconnection patterns between models and observations. *J. Climate*, **25**, 425-446, doi: 10.1175/JCLI-D-11-00175.1. [[Link](#)]
- Yang, X. Y., X. Yuan, and M. Ting, 2016: Dynamical link between the Barents-Kara sea ice and the Arctic oscillation. *J. Climate*, **29**, 5103-5122, doi: 10.1175/JCLI-D-15-0669.1. [[Link](#)]