Tellus (2007), 59A, 50–70
Printed in Singapore. All rights reserved

TELLUS

Evaluation of the FSU synthetic superensemble performance for seasonal forecasts over the Euro-Mediterranean region

By SEVINC SIRDAS[1,2], ROBERT S. ROSS[1]*, T. N. KRISHNAMURTI[1] and A. CHAKRABORTY[1], [1]*Department of Meteorology, The Florida State University, Tallahassee, FL 32306-4520, USA;* [2]*Department of Meteorology, Istanbul Technical University, Maslak, Istanbul 34469, Turkey*

(Manuscript received 23 June 2005; in final form 17 July 2006)

ABSTRACT

This paper deals with seasonal climate forecasting using as many as 13 coupled ocean–atmosphere models. An analysis of the individual model, multimodel ensemble, and FSU synthetic superensemble (FSUSSE) climate forecasts was performed for monthly and seasonal forecasts of precipitation, sea surface temperature (SST) and surface air temperature over the Euro-Mediterranean region including the land areas of Europe, North Africa and the Near East, during the period 1989–2001. In the FSUSSE methodology, forecasts are obtained by a weighted combination of the individual coupled ocean–atmosphere model forecasts based on a training period. The set of 13 individual climate forecast models utilized in this research is comprised of the seven models in the European suite of DEMETER models, a suite of four Florida State University models, the Australian POAMA model and the NCAR CCM3 model.

The FSUSSE forecasts of seasonal precipitation anomalies were found to have the lowest root mean square (RMS) errors in comparison to the models in the multimodel ensemble, and their ensemble mean. However, the anomaly correlation (AC) coefficient results for seasonal precipitation anomaly forecasts by the FSUSSE were less impressive. The equitable threat scores for the FSUSSE forecasts of seasonal precipitation were found to be better than the various models in the multimodel ensemble, but those scores for the forecasts of positive seasonal anomalies were found to be worse than most of the models in the multimodel ensemble.

The FSUSSE seasonal forecasts for SST and surface air temperature for the season considered (winter) were found to be excellent for both AC coefficients and RMS errors in the forecast anomalies.

1. Introduction

In recent years a number of papers have addressed seasonal climate prediction using global coupled ocean–atmosphere models. Seasonal forecasts are clearly of value to a wide cross section of society, including personal, commercial and humanitarian interests (e.g., Thomson et al., 2000; Hartmann et al., 2002; Pielke and Carbone, 2002). The Euro-Mediterranean area is a climate-sensitive region, which is climatically stressed by limited water resources and extremes of heat, which can cause or exacerbate existing socio-political tensions (Mann, 2001). In particular, high frequency variations (monthly, seasonal, annual and interannual), as well as low-frequency variations (inter-decadal) of precipitation play a crucial role in the management of regional agriculture, ecosystems, environment, socio-economics, and water resources (Xoplaki et al., 2000).

Seasonal climate predictions are now made routinely at a number of operational forecast centers around the world, using coupled models of the atmosphere, oceans and land surface (e.g., Mason et al., 1999; Alves et al., 2002; Kanamitsu et al., 2002). These predictions have been made possible by our understanding of the coupled ocean–atmosphere system, which has come about in the second half of the 20th century (Neelin et al., 1998). Such predictions have also been fostered by the development and deployment of specialized buoys to observe and measure the evolution of near-surface waters in the tropical Pacific (McPhaden et al., 1998), and by better coupled model predictions of El Niño (Zebiak and Cane, 1987). These developments were also aided by the very successful international Tropical Ocean Global Atmosphere (TOGA) Program (World Climate Research Programme, 1985).

In the beginning, the notion of ensemble forecasting was evident in the studies of Lorenz (1963), where he examined the

*Corresponding author.
e-mail: bross@coven.met.fsu.edu
DOI: 10.1111/j.1600-0870.2006.00209.x

initial state uncertainties in a simple nonlinear system. Much progress has been made in ensemble forecasting for the conventional weather prediction problem, using the singular vector based perturbations (Molteni et al., 1996) and the breeding modes (Toth and Kalnay, 1997). Several other formulations have appeared in the current literature, including the simpler Monte Carlo methods (Mullen and Baumhefner, 1994). In seasonal climate forecasts, the ensemble forecasts can be constructed by using initial perturbations from adjacent start dates (LaRow and Krishnamurti, 1998). Most deterministic and probabilistic ensemble forecasts are produced with a single dynamical model, although sometimes a set of multimodels is used (Yun et al., 2005).

Schemes for objective combination of predictions from different models have been applied to seasonal climate forecasts for several years (e.g., Krishnamurti et al., 1999; Krishnamurti et al., 2000a; Krishnamurti et al., 2000b; Krishnamurti et al., 2002; Pavan and Doblas-Reyes, 2000). In these schemes a systematic improvement in root mean square (RMS) errors is observed for a multimodel forecast over that of the individual model forecasts.

DEMETER models have been used for seasonal predictions of precipitation, sea surface (SST) temperature and surface air temperature, and the DEMETER models comprise 7 of the 13 models used in the multimodel ensemble employed in the present study. The Development of a European Multimodel Ensemble System for Seasonal to Interannual Prediction project (DEMETER) was conceived and funded under the European Union Vth Framework Environment Programme. Palmer et al. (2004) stated that the principal aim of DEMETER was to advance the concept of the multimodel ensemble as a method for better representing forecast uncertainties and providing more reliable seasonal forecasts.

This paper deals with the evaluation of the climate forecasts of 13 ocean–atmosphere models over the Euro-Mediterranean and North African regions, as well as the climate forecasts of the Florida State University synthetic superensemble (FSUSSE). In general, the forecasts from the superensemble technique are found to be better than those from the individual models in the multimodel ensemble and their ensemble mean. This is because the technique uses weights for combining the individual model forecasts that are based on past model performance. These weights are determined for each model and for each variable, grid point, and vertical level. It is important to note that several possibilities exist for combining the forecasts from multimodels and that it is not a trivial problem to determine robust weights, due to overfitting of the data, as was described by Kharin and Zweirs (2002).

The large-scale multimodel forecasts are obtained for seasonal forecasts of precipitation, sea surface and surface air temperature, as well as the anomalies of these forecast parameters. The forecast skills of the 13 models, their ensemble mean, and the FSUSSE are compared by using RMS error, anomaly correlation (AC) coefficient, equitable threat and bias scores. In Section 2, the different climates of the large Euro-Mediterranean region are described to indicate the complexity of the region and to underscore the inherent difficulties in attempting to predict such a diverse climate system. The models and data utilized are explained in Section 3. Section 4 addresses the quality of the seasonal climate simulations for the Euro-Mediterranean region by the Florida State University Coupled Ocean–atmosphere General Circulation Model (FSUCGCM), as well as other models in the multimodel ensemble. The methodologies of the conventional FSU superensemble and synthetic superensemble models are presented in Section 5. The precipitation forecast skills are discussed in detail in Section 6. The probabilistic measure of precipitation forecast skill is presented in Section 7 using equitable threat and bias scores. The RMS errors and the AC coefficients for the predictions of SST and surface air temperature are presented in Section 8. Concluding remarks are made in Section 9.

2. Climate

In this section we describe the complexity of the climate regimes found in the vast region considered in this paper in order to convey the inherent difficulties involved in attempting to predict such a diverse climate system. The Euro-Mediterranean and North Africa regions, as defined in this paper, are shown in Fig. 1. These regions have three different types of climate classification and regional geomorphology. The area includes regions that are characterized by surface temperature and precipitation anomalies of opposite sign in seasons that have strong North Atlantic Oscillation (NAO) anomalies (Rodo et al., 1997; Quadrelli et al.,

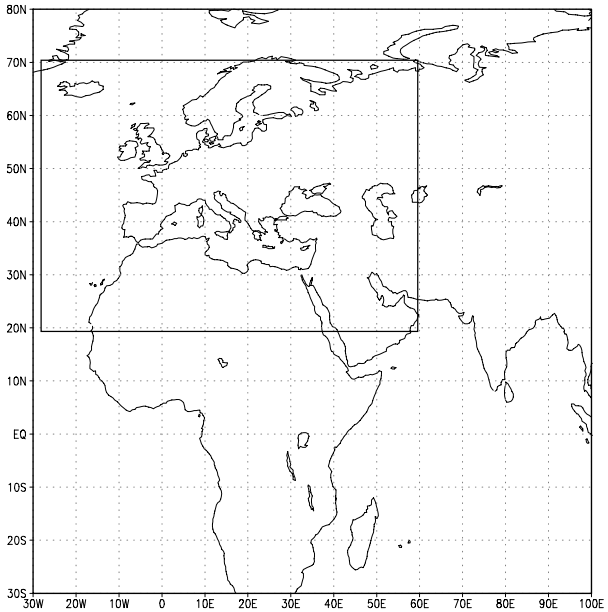

Fig. 1. Areas considered in this study: Euro-Mediterranean region and North Africa (20°N–70°N and 30°W–60°E).

2001). The Mediterranean climate is a special type of climate characterized by a regime of hot summer drought and winter rain in the mid-latitudes, located to the north of the subtropical climate zone. Frei and Schar (1998), in describing the climatology of the Alpine region, state that over the Northern Alps the main wet season is in summer due to convection. Thus, the seasonality for part of the Mediterranean region is bimodal. North Africa has a dry climate that is defined by 0–30 mm average annual rainfall due to the presence of the large Saharan Desert. Over Northern Europe most of the rain falls in winter when the storm track is stronger.

In boreal summer (during May to August), the high-pressure belts of the subtropics drift northward in the Northern Hemisphere. They are coincident with substantially higher temperatures and little rainfall over the Euro-Mediterranean region. During the boreal winter, the high-pressure belts drift back towards the equator, and the Euro-Mediterranean weather becomes more dominated by the rain-bearing low-pressure depressions. While usually mild, such areas can experience cold snaps when exposed to the icy winds of the large continental interiors, where temperatures can drop to $-40\ ^{\circ}$C in the extreme continental climates.

The Euro-Mediterranean region lies in an area of great climatic interest. It is influenced by some of the most relevant mechanisms influencing the global climate system; it marks a transitional zone between the deserts of North Africa, which are situated within the arid zone of the subtropical high, and central and northern Europe, which is influenced by the westerly flow during the whole year. In addition, the Euro-Mediterranean climate is exposed to the South Asian Euro-Mediterranean region (SAM), the Siberian High Pressure System, the Southern Oscillation (SO) and the NAO. There is recent evidence that the teleconnection of the El-Niño-Southern Oscillation (ENSO) has extended its reach into parts of the Euro-Mediterranean in recent decades (Moron et al., 1998). The main physical and physico-geographical factors controlling the spatial distribution of the climatic conditions over the Euro-Mediterranean region are the atmospheric circulation, the latitude, the altitude, the topography, the Atlantic and Euro-Mediterranean SST distribution, the land-sea interactions (distance from the sea) and smaller-scale processes (Xoplaki et al., 2000).

3. Data sets and models

3.1. Data sets

Xie and Arkin (1997) developed a new technique, which they called the outgoing long wave radiation (OLR)-based Precipitation Index (OPI), to estimate global monthly precipitation from the satellite-observed OLR data that are available from June 1974 to the present, except for a 10-month period of missing data from March to December 1978. Estimates based on the new technique are able to provide nearly complete global coverage of large-scale precipitation with high quality for all seasons. By combining the gauge based analysis of Xie et al. (1996) over land with the oceanic estimates of Spencer (1993), based on observations from the Microwave Sounding Unit (MSU), the OPI estimates provide a useful means to quantitatively describe the global distributions of large-scale precipitation continuously for an extended period from 1979 to the present. The full precipitation data set of Xie and Arkin (1997), which is referred to as CMAP, is used in this research. The observed climatology for SST is from Reynolds et al. (2002), while the observed climatology of surface air temperature is from the ECMWF reanalysis.

3.2. Models

3.2.1. FSU suite of coupled models. The FSU suite of coupled models is comprised of four identical models that share the same initial coupled data assimilation, dynamics, and most of the physical parameterizations except for the cumulus parameterization and the radiative transfer algorithms. Two versions of cumulus parameterization and two versions of radiative transfer are used following Krishnamurti et al. (2002) in the design of these FSU coupled models. The relevant references for these algorithms for convection and radiation are Krishnamurti and Bedi (1988), Grell (1993), and Lacis and Hansen (1974). Details of the four versions of the FSU coupled model are provided in Table 1.

The FSU coupled model combines the FSU global spectral model (Krishnamurti et al., 1998) and the Hamburg Ocean model (Latif, 1987). The model's performance is described in Krishnamurti et al. (2002). Model forecasts are initialized using data assimilation in a two-step process:

(1) An ocean spin up component: Using monthly mean surface winds from the ECMWF reanalysis and the Reynold's SST fields, we first spin up the ocean with the ocean model alone. Here Newtonian relaxation of the observed SST and imposed wind stresses drive the ocean for an 11-yr period. Details are presented in LaRow and Krishnamurti (1998) and Krishnamurti et al. (2000b).

(2) Coupled assimilation: A coupled assimilation phase utilizes the FSU coupled model, where a physical initialization following Krishnamurti et al. (1991) and a Newtonian relaxation assimilate the atmospheric and oceanic data sets. Here the observed rain rates used in the physical initialization are derived from the Defense Meteorological Satellite Program/Special Sensor Microwave Imager (DMSP/SSMI) based on microwave scattering temperatures and the NASA Tropical Rainfall Measuring Mission (TRMM) satellites. This assimilation also includes a Newtonian relaxation for the observed SSTs. This is all done with daily rainfall and weekly Reynold's SST data sets. In this process, the model based daily rainfall totals are brought to a close agreement with observed (analyzed) precipitation at the same resolution. This assimilation tries to retain the temperature of the air and the rotational part of the wind in the ECMWF

Table 1. Characteristics of the 13 models

Name (Source)	Atmospheric component			Oceanic component		
	Model	Resolution	Initial condition	Model	Resolution	Initial condition
ANR (FSU)	FSUGSM with Arakawa-Schubert convection and new radiation (band model)	T63L14	ECMWF with Physical Initialization	HOPE global	5° longitude, 0.5–5° latitude, 17 levels	Coupled assimilation relaxed to observed SST
AOR (FSU)	FSUGSM with Arakawa-Schubert convection and old radiation (emissivity/absorptivity based)	T63L14	ECMWF with Physical Initialization	HOPE global	5° longitude, 0.5–5° latitude, 17 levels	Coupled assimilation relaxed to observed SST
KNR (FSU)	FSUGSM with Kuo convection and new radiation (band model)	T63L14	ECMWF with Physical Initialization	HOPE global	5° longitude, 0.5–5° latitude, 17 levels	Coupled assimilation relaxed to observed SST
KOR (FSU)	FSUGSM with Kuo convection and old radiation (emissivity/absorptivity based)	T63L14	ECMWF with Physical Initialization	HOPE global	5° longitude, 0.5–5° latitude, 17 levels	Coupled assimilation relaxed to observed SST
CERFACS (France)	ARPEGE	T63L31	ERA40	OPA 8.2	2° × 2°, 31 Levels	Forced by ERA40
ECMWF (Europe)	IFS	T95L40	ERA40	HOPE-E	1.4° × 0.3°-1.4°, 29 levels	Forced by ERA40
INGV (Italy)	ECHAM-4	T42L19	Coupled AMIP type	OPA 8.1	2° × 0.5°-1.5°, 31 levels	Forced by ERA40
LODYC (France)	IFS	T95L40	ERA40	OPA 8.2	2° × 2°, 31 levels	Forced by ERA40
MPI (Germany)	ECHAM-5	T42L19	Coupled run relaxed to observed SST	MPI-OMI	2.5° × 0.5°-2.5°, 23 levels	Coupled run relaxed to observed SST
MetFr (France)	ARPEGE	T63L31	ERA40	OPA 8.0	182 × 152 GP, 31 levels	Forced by ERA40
UKMO (England)	ARPEGE	2.5 × 3.75, 19 levels	ERA40	GloSea OGCM HadCM3 based	1.25° × 0.3°-1.25°, 40 levels	Forced by ERA40
CCM3 (NCAR)	CCM3 atmospheric model	T63L18	ECMWF	SOM	2.4 ° × 1.2 ° -2.4°	Coupled assimilation
POAMA1 (Australia)	BMRC Atmospheric Model (BAM3)	R47L17	From latest atmosphere and ocean conditions from GASP	ACOM2	2° × 0.5°-1.5°, 25 levels	From ocean assimilation which was based on optimum interpolation (OI) technique.

daily analysis using a hard nudging. The choices of relaxation coefficients for the nudging are presented in Krishnamurti et al. (2002).

Using the aforementioned four versions of the FSU coupled model, a total of 624 experiments (4 models × 13 yr × 12 months), one experiment per month, were carried out over the period 1989–2001. The two-step process of data assimilation used to initialize the model forecasts is carried out for an initial 2-yr period first, and the initial state for the coupled model is thus prepared. Then this coupled assimilation continues through all 13 yr, so that seasonal forecasts can be carried out thereafter for any start time. Data sets of the various forecast parameters were archived for all forecasts at intervals of 15 d.

3.2.2. DEMETER models. The DEMETER models are described in Palmer et al. (2004). The following is a list of European coupled seasonal climate models that fall in this category:

1. The coupled models from France (CERFACS, LODYC and METEO FRANCE)
2. The ECMWF coupled model
3. The UK Met office coupled model
4. The Italian coupled model INGV
5. The German coupled model MPI

Each seasonal forecast had a 1-month lead time. Data from 4-month forecasts was selected for each initial condition, covering the years 1989–2001. The seven coupled models each carried an ensemble of nine forecasts for each start date. These ensembles were constructed using wind and SST perturbations of the respective initial conditions. In all, 3276 seasonal forecasts were available from the DEMETER database (7 models × 9 ensemble members per model × 13 yr × 4 months). In this research the nine ensemble members for each model were averaged to yield one forecast for each of the seven models. Table 1 describes some features of these coupled models.

3.2.3. POAMA and CCM3 Models. The Predictive Ocean Atmosphere Model for Australia (POAMA) is a state-of-the-art ocean–atmosphere coupled model for seasonal to inter-annual prediction. Real-time oceanic and atmospheric initial states are used to initialize the coupled model. These are provided by an ocean data assimilation system that is run in real time as part of the POAMA system, and by operational weather analyses from BMRC. The atmospheric component of the coupled model used in POAMA is the BMRC unified atmospheric model (BAM) with specifications of T47L17. The ocean model component is the Australian Community Ocean Model version 2 (ACOM2). The model is described in Zhong et al. (2001), and several features of the model are presented in Table 1.

The National Center for Atmospheric Research (NCAR) CCM3 model solves the dynamical equations in spectral space. The configuration used is T63L26. Kiehl et al. (1998) provide a complete description of the physical and numerical methods used in CCM3. Several features of the model are presented in Table 1.

Combining the DEMETER, POAMA, CCM3 and the FSU components, a total of 13 global coupled ocean–atmosphere models provided data sets for 4500 seasonal forecast experiments (624 FSU + 3276 DEMETER + 600 POAMA/CCM3).

4. Model simulation of mean Euro-Mediterranean climate

In the validation of any climate model, it is important that the current climates from the model and observation/analysis be carefully compared. In this section, we focus on the quality of the seasonal climate simulations for the Euro-Mediterranean region by the Florida State University Coupled Ocean–atmosphere General Circulation Model (FSUCGCM). The predictability and the interannual variability of rainfall in a model are dependent on how well that model simulates the seasonal climatology (Sperber and Palmer, 1996). Models with better rainfall climatology generally have less systematic errors and somewhat higher ability to simulate interannual variations. Climate drift is almost inevitable in coupled ocean–atmosphere models and frequently it can be larger than the interannual signal.

The mean features of seasonal climate and its variability produced by the FSUCGCM seasonal forecasts were investigated for the period 1989–2001. Fig. 2 shows the 13-yr time-series of precipitation from this model, from which dry and wet periods may be determined. In this section, we will focus on the

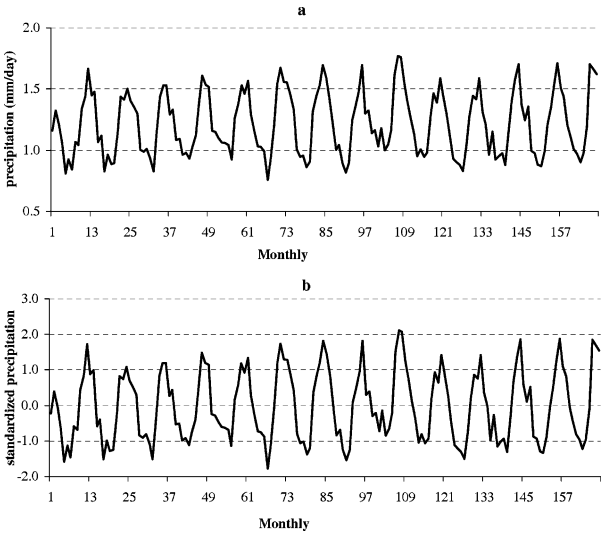

Fig. 2. (a) Time-series of area average monthly precipitation. (b) Time-series of area average standardized precipitation from January 1989 to December 2002. The standardization method is based on a monthly precipitation time-series, $X_1, X_2, .., X_n$. The standardized monthly precipitation series in (b) is defined as, $x_i = \frac{X_i - \bar{X}}{S_x}$, where $\bar{X}$ is the arithmetic mean and S_x is the standard deviation of the series.

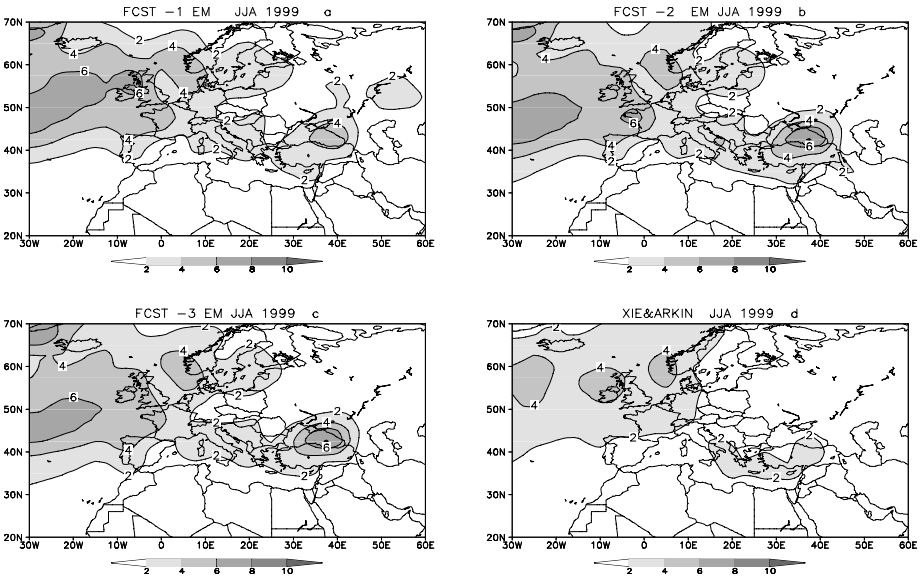

Fig. 3. Precipitation for JJA 1999 in units of mm/d. Panels (a), (b) and (c) show the precipitation climatology for the ensemble mean of forecasts from the four FSU models based on strings of month-1, month-2 and month-3 forecasts, respectively. The observed rain in panel (d) is obtained from Xie and Arkin (1997) data sets.

climate phenomena of dry summers (JJA) and wet winters (DJF). We will examine JJA 1999 and DJF 2000/2001 as representative examples of the dry and wet seasons, respectively, over the Euro-Mediterranean region. JJA 1999 represents one of the driest summers (see months 126–128 in Fig. 2) while DFJ represents one of the wettest winters (see months 144–146 in Fig. 2).

Lower level wind flow patterns and rainfall associated with the boreal summer and winter seasons in the Euro-Mediterranean region were examined. The observation based mean fields were compared with fields simulated by the coupled model for the December, January and February (DJF) and June, July and August (JJA) periods. The seasonal rainfall forecasts from the FSU coupled model in mm/d are presented in Figs. 3–4 and they show almost similar spatial distribution. Fig. 3 compares the mean observed (Xie and Arkin, 1997) and model boreal summer (dry period) Euro-Mediterranean region rainfall patterns for the JJA 1999 period. The mean observed (Xie and Arkin, 1997) and model boreal winter (wet period) Euro-Mediterranean region rainfall patterns for the DJF 2001/2002 period are shown in Fig. 4. In both figures the observed mean (panel d) shows a rain area of 2 mm/d near the east coast of the Euro-Mediterranean region. Over the northwestern region of the map in both figures a center of maximum rainfall of 4 mm/d is noted. The model forecasts (month-1 through month-3 in panels (a), (b) and (c), respectively) valid for DJF 2001/2002 and JJA 1999 captured realistic Euro-Mediterranean region rainfall patterns when compared to these observations. The coupled model captured both the rainfall maximum over the east coast of the Euro-Mediterranean and over the northwestern map region. However, the rainfall amounts show rather large errors. For example, in Fig. 4 the observed precipitation mean is 2 mm/d in the eastern Euro-Mediterranean, while the coupled model forecasts show values of 6–8 mm/d for the same region. In Fig. 3 the model forecasts show a similar pattern to the observed precipitation, although once again there are errors in the precipitation amounts in the model. In both Figs. 3 and 4 the Euro-Mediterranean rainfall amounts are seen to increase from month-1 forecasts to month-3 forecasts in the coupled model, representing a drift in the coupled model's rainfall forecasts. Results from the month-3 forecasts are in the least agreement with respect to the observed precipitation. Therefore, the FSU coupled model fails to reasonably simulate observed wet and dry season rainfall climatology in the longer range in terms of rainfall amounts. The individual member model rainfall forecasts from the coupled suite were not very impressive. The models had a tendency to over predict by 30–40 % the Euro-Mediterranean region rainfall from month-1 through month-3 when compared to the observed climatology. The spatial distributions of precipitation over Europe and the North Africa continent, the Near East, and adjacent oceanic regions in the coupled model were comparable to the respective monthly mean and seasonal mean analyses. In this sense, the coupled model was able to capture the large-scale features of the Euro-Mediterranean region circulation and the associated rainfall reasonably well. Even though the rainfall patterns were reasonably well simulated by the coupled models in comparison to observation, the actual rainfall amounts often showed rather large errors.

Next, we shall illustrate the climatology of seasonal rainfall forecasts from the DEMETER, CCM3, POAMA and FSU coupled models. In Table 1, we have provided a brief summary of these models. The models vary considerably from each other in

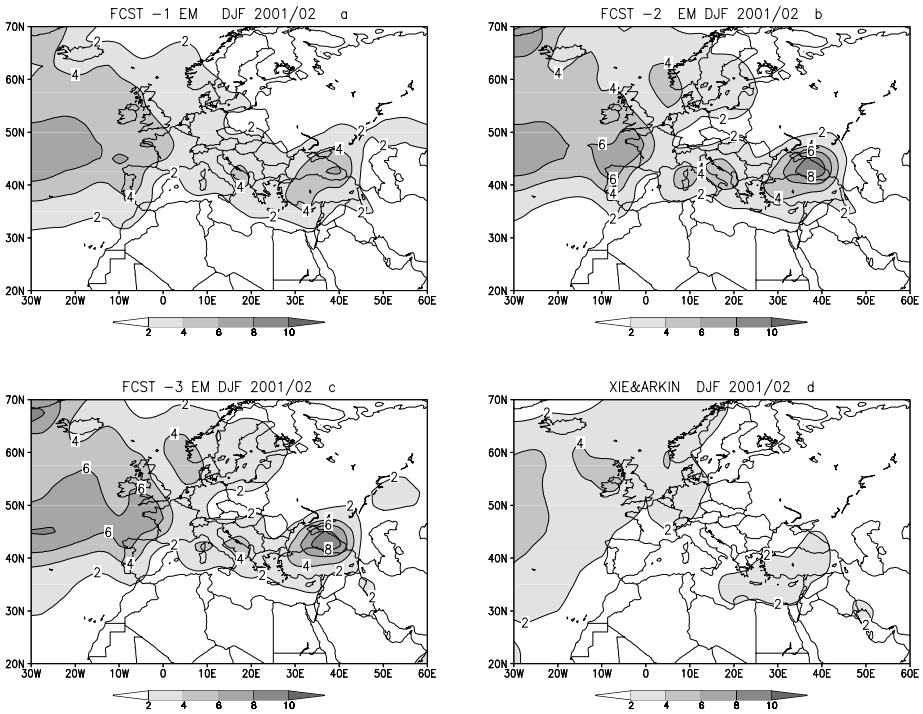

Fig. 4. Precipitation for DJF 2001/2002 in units of mm/d. Panels (a), (b) and (c) show the precipitation climatology for the ensemble mean of forecasts from the four FSU models based on strings of month-1, month-2 and month-3 forecasts, respectively. The observed rain in panel (d) is obtained from Xie and Arkin (1997) data sets.

terms of resolution, physics and the ocean model algorithms. The resolutions vary from T42 (roughly 2.5° lat/lon) to T95 (almost double that resolution). The vertical resolution varies from 19 to 40 levels. Table 1 provides references that outline the atmosphere and the ocean model components for each of these models. Thirteen model runs were carried out for each experiment of the DEMETER, CCM3, and POAMA models by perturbing the initial states. Perturbing the wind stresses generated ensembles of 13 forecasts of SSTs at the interface of the atmosphere-ocean coupled system. Details of the procedures used in producing the DEMETER model data sets may be found in Palmer et al. (2004).

The seasonal climatology forecasts for the years 1989 through 2001 for these models are given in Figs. 5–8 for DJF, MAM, JJA and SON, respectively. The seasonal precipitation forecasts for the boreal winter seasons of DJF 1989–2001 are shown in Fig. 5. This figure represents the relatively wet winter season (defined with respect to the mean values as seen in the Xie-Arkin (1997) data set). Model forecasts depicted include the four FSU coupled models, the seven DEMETER models, the CCM3 model, and the POAMA model, as well as the multimodel ensemble mean of all 13 models, and the FSUSSE, along with the observed precipitation climatology for Xie and Arkin (1997). Fig. 5 shows that the northwestern Euro-Mediterranean region experienced very wet conditions during the boreal winter seasons of DJF 1989–2001. The rest of the Euro-Mediterranean region experienced lesser rains. The individual models vary considerably in their rainfall predictions. Among the seven DEMETER models, the SCNR and SCWF models, that have high resolution ocean components, capture the Euro-Mediterranean features reasonably well in comparison to observation, except for the rainfall all along the Euro-Mediterranean eastern coastline, which is somewhat underestimated. Also, these models fail to capture the wet areas over the northwest Euro-Mediterranean and Northern Atlantic regions mentioned above for the period DJF 1989 through 2001. In addition, these models show a belt of heavy rain stretching from the Atlantic Ocean into the north central Euro-Mediterranean that is somewhat overestimated. The boreal winter (DJF) seasonal forecasts of precipitation in Fig. 5 show that, among the DEMETER models, the UKMO and LODYC have the better forecasts. The CCM3 model has a better forecast than the POAMA model for the Euro-Mediterranean precipitation. The POAMA, CNFC, SMPI and SCWF show excessively heavy rainfall amounts over a very large area. Considering the entire forecast area depicted, the rainfall patterns are best predicted by the FSUSSE (to be described in detail in Section 5.2) in comparison to the observed rainfall, except that the FSUSSE does not capture the 2 mm/d rainfall over the eastern Euro-Mediterranean. For the boreal spring (MAM) forecast in Fig. 6, the SCNR and UKMO models of DEMETER most closely match the observed precipitation. In Fig. 7, where the boreal summer precipitation forecasts are shown, the CNRM, SCWF, and UKMO DEMETER models show the best performance, except that the UKMO

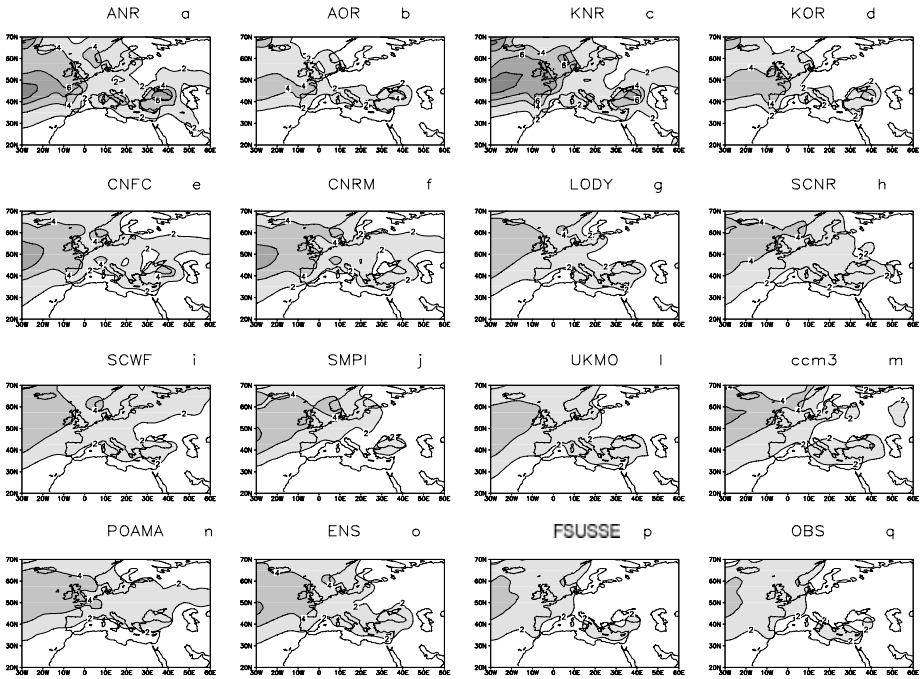

Fig. 5. Climatology of seasonal precipitation (mm/d) over the Euro-Mediterranean region for the boreal winter season of DJF for the period 1989–2001: Climatology from the four FSU models (a–d), the seven DEMETER models (e–l), the CCM3 model (m), the POAMA model (n), the multimodel ensemble mean of all 13 models (o) and the FSUSSE (p) is shown, along with the observed precipitation climatology from Xie and Arkin (1997) (q).

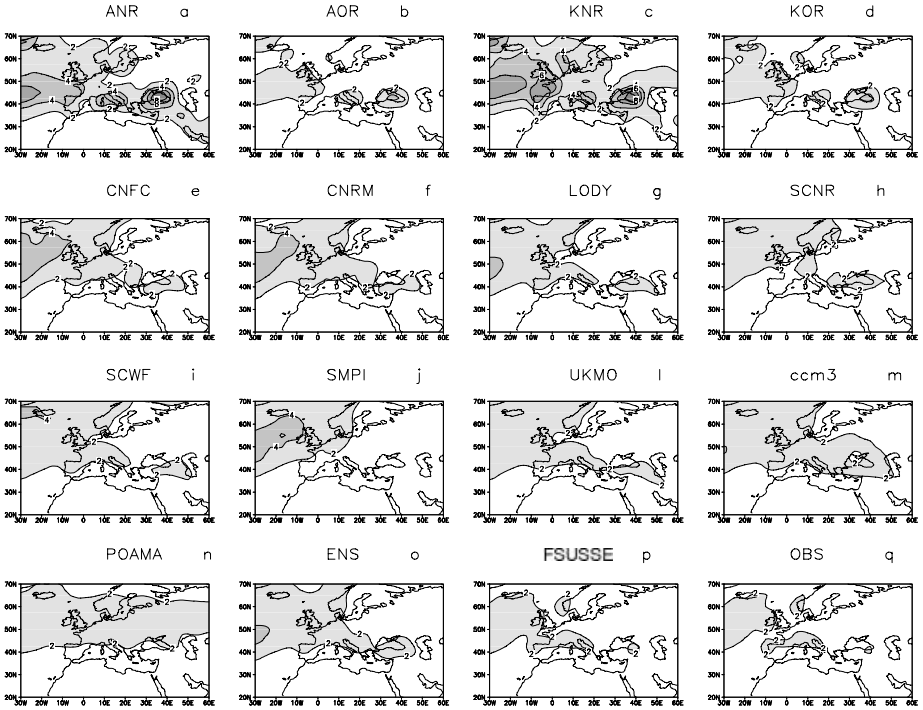

Fig. 6. Same as for Fig. 5 except for the boreal spring season of MAM for the period 1989–2001.

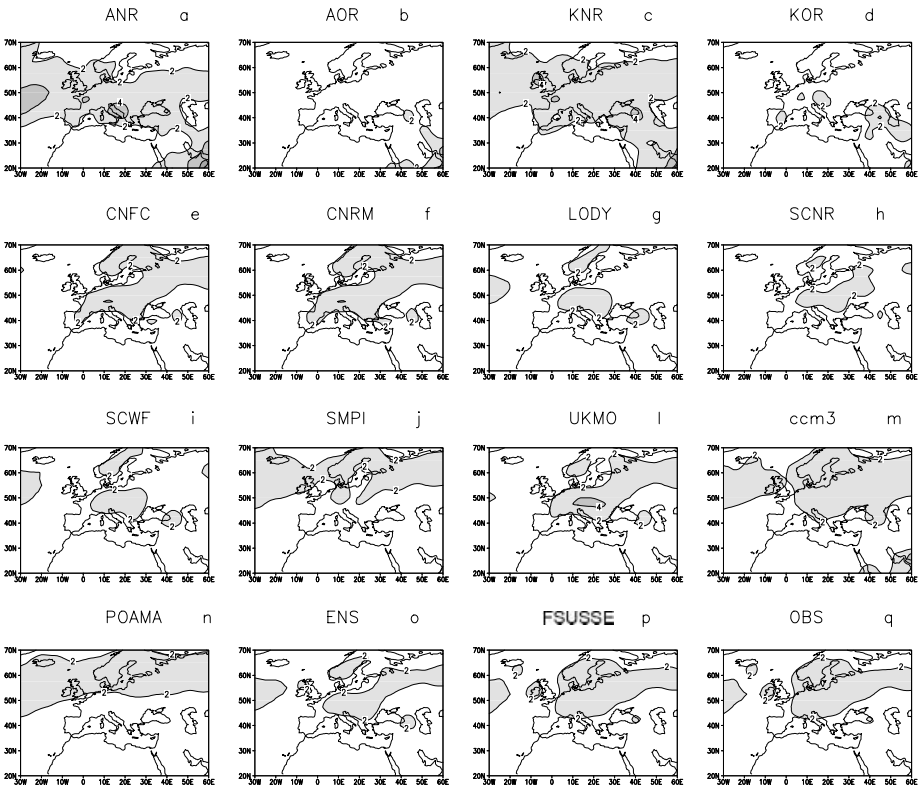

Fig. 7. Same as for Fig. 5 except for the boreal summer season of JJA for the period 1989–2001.

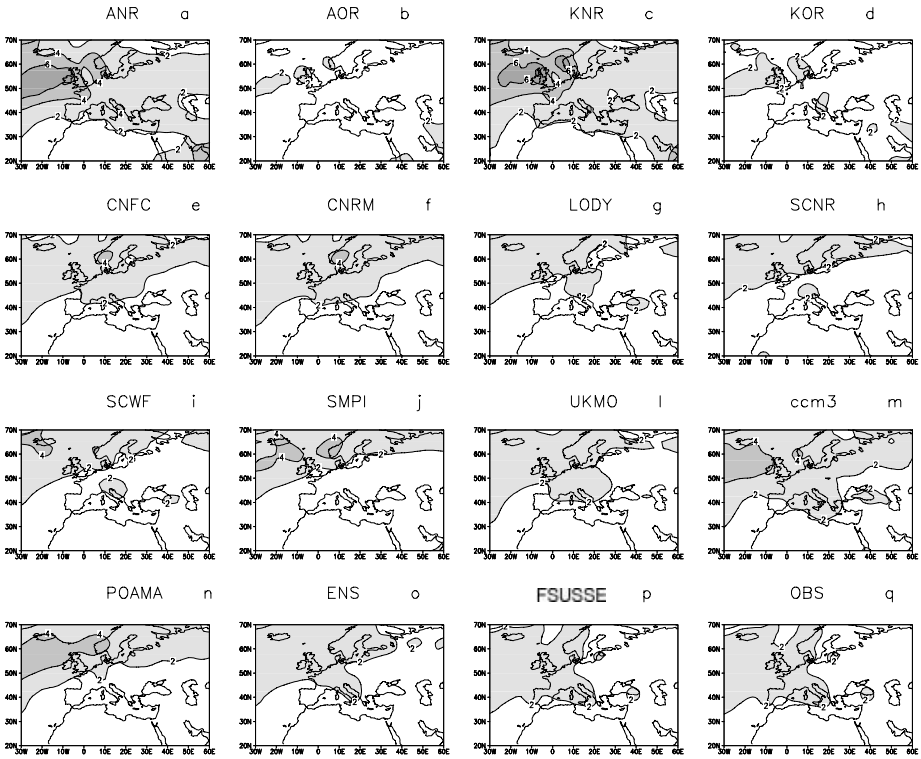

Fig. 8. Same as for Fig. 5 except for the boreal fall season of SON for the period 1989–2001.

model overestimates the seasonal precipitation over central Europe in comparison to observation. For the boreal autumn (SON) seasonal forecasts in Fig. 8, the LODY and UKMO DEMETER models give the best forecasts of precipitation in comparison to observation. For all of these seasons the CCM3 and POAMA models show the least agreement with observation, while the FSUSSE forecasts most closely match the observed precipitation.

Models with poor rainfall climatology show rather large departures in rainfall forecasts for a given season in comparison to observation. The model climatology over the Euro-Mediterranean region is sensitive to many model features that contribute to such errors. The rainfall forecast by the individual models comprising the multimodel ensemble departs considerably from the observed rainfall for all seasons 1989–2001 in comparison to observation. It is difficult to place much credence on these models' anomaly forecasts when they are calculated with respect to their own individual mean fields for a given season. This is one of the reasons why seasonal forecasts have been one of the most difficult forecast problems. Some models, such as the FSU coupled models (ANR, AOR, KNR and KOR), and the SCWF, SMPI, CCM3 and POAMA models, show excessive rainfall in parts of the interior and Euro-Mediterranean coastal areas, as well as over the Atlantic Ocean. The important message here is that with the construction of the FSU synthetic superensemble (FSUSSE) it is possible to remove such biases for seasonal forecasts of rainfall. Such an approach is necessary since the models comprising the ensemble carry large seasonal biases in their rainfall forecasts. These biases are most likely due to inadequate representation of physical processes (physics) in these models. It is important to recognize that the FSUSSE technique circumvents the need to correct the physics in each of these models by developing a forecast equation that accounts for each model's deficiencies through a weighted combination of the various models' forecasts, with the weighting factors being determined through the training procedure, whereby each model's forecasts are compared to observed conditions over an extended period of time. The superensemble technique is described fully in the next section of the paper. In principle, the FSUSSE's success should be evidenced in reduced RMS error scores and in enhanced anomaly correlation coefficients, but this is not always realized, as will be seen for the results of the seasonal precipitation forecasts in Sections 6 and 7. Generally speaking, the RMS error results are better than the anomaly correlation coefficient results for the precipitation forecasts discussed in this paper. However, for SST and surface air temperature forecasts, both of these measures of skill indicate superior performance by the FSUSSE in comparison to the other models, as will be seen in Section 8.

In this section of the paper we have seen that the FSUSSE's simulation of the 13-yr climatology of rainfall for the four selected seasons most closely matches the observed rainfall climatology from Xie and Arkin (1997), when compared to the individual models in the multimodel ensemble and their ensemble mean. Based on the findings of Sperber and Palmer (1996), this success by the FSUSSE in simulating the 13-yr seasonal rainfall climatology would indicate that the model would also have an advantage over the other models in rainfall predictability for individual seasons. Such rainfall predictions by the various models will be examined in Sections 6 and 7 of the paper.

5. Methodology

5.1. Conventional FSU superensemble methodology

The FSU superensemble technique (Krishnamurti et al., 1999) produces a single forecast derived from a multimodel set of forecasts. Forecasts from this methodology almost universally carry the highest skill compared to the individual models comprising the ensemble, as well as the bias-removed multimodel ensemble mean. In the standard bias removal procedure that is used to produce the bias-removed multimodel ensemble mean, each individual model has its bias removed, and then the mean of the bias removed forecasts is constructed *with each model being given, in effect, a weight of 1*. However, when the FSU superensemble technique is used to produce the bias-removed multimodel ensemble mean, each individual model, in effect, has its bias removed, and then the mean of the bias removed forecasts is formed*, but with the important difference being that each model is given a different weight, based on the training period*. Different models have different predictive skill, and when their forecasts are combined to form the multimodel ensemble mean, the FSU superensemble is equipped to reflect this differential predictive skill in a way that the standard bias removal procedure cannot, through the use of weights that are determined in the training procedure. Thus, the strategy for the multimodel superensemble is based on a partitioning of the forecast time line into two components. The first of these, called the training period, utilizes past multimodel forecasts and observed (analysis) fields to derive model performance statistics. The second phase, called the forecast phase, utilizes current multimodel forecasts and the aforementioned statistics to obtain real-time superensemble forecasts. During training, with the use of benchmark observed (analysis) fields, past forecasts are used to derive statistics on the past behaviour of the models. Given a set of past multimodel forecasts, a multiple regression technique is used, in which the model forecasts are regressed against an observed (analysis) field. This utilizes a least-squares minimization of the difference between anomalies of the model forecasts and the analysis fields in order to determine a distribution of weights. As stated above, these regression coefficients (weights) associated with each individual model can be interpreted as a measure of that model's relative forecast reliability for the given point over the training period. For each model prognostic variable, the purpose of training is to evaluate model biases geographically and vertically. This being done for $\boldsymbol{m}$ multimodels at n gridpoints (along the horizontal and

vertical) for p variables and q time intervals, constitutes as many as $m \times n \times p \times q$ statistical coefficients (order of 10^7 weights). The methodology for this conventional procedure consists of a definition of the superensemble forecast according to the following equation:

$$S = \bar{O} + \sum_{i=1}^{N} a_i (F_i - \bar{F}_i), \tag{5.1.1}$$

where S is the superensemble prediction, $\overline{O}$ is the observed time mean, a_i are the weights for individual models i, F_i is the predicted value from model i, $\bar{F}_i$ is the time mean of predictions by model i for the training period, and N is the number of models. The weights are computed at each of the grid points by minimizing the objective function G for the mean square error of the forecasts as:

$$G = \sum_{t=0}^{t=\text{train}} (S_t - O_t)^2, \tag{5.1.2}$$

where t denotes the length of the training period. The length of the training phase, in which the multiregression coefficients are obtained, varies for each type of forecast addressed in this paper. The statistics, collected during the training phase, are passed on to the forecast phase of the superensemble (eq. 5.1.1). In this forecast phase, we use real-time forecasts from the same member models that were considered in the training procedure. As previously stated, this type of local bias removal is more effective compared to a conventional bias removed ensemble mean. The latter assigns a weight of 1.0 to all models after bias removal. The superensemble includes fractional and even negative weights depending on past behaviours. The negative weights are a statistical fit measure; it is the same as in any application of statistical data that are subjected to multiple regression, where the weights are determined via a least-square minimization. No hard physical interpretation should be expected for these negative weights. One must be bold to use such in meteorological problems. If the weights are positive for some models at some locations and are negative for other models at the same locations, it simply means that for those models with negative weights such a correction provides the best least-squares fit. However, we have established (Chakraborty et al., 2006) that the weights of different models plotted against forecast period during the training phase do very closely resemble the plots of correlations of model output versus the observed counterparts during the same training phase. Thus, negative weights do seem to imply that those models have negative correlation against observations during the training phase. In a probabilistic sense also, the superensemble probability forecasts are somewhat better than the multimodel bias removed ensemble mean forecasts at any threshold level when compared to observations (Stefanova and Krishnamurti, 2002).

Some of our earlier research applications on seasonal climate forecasts utilized the Atmospheric Model Inter-comparison Project (AMIP1 and AMIP2) global multimodel climate data sets involving long integrations over a period of a decade using atmospheric general circulation models (Gates et al., 1999). According to a recent study by Krishnamurti et al. (2000b), 13 models of AMIP1 were selected in order to examine the rainfall forecasts. The selection of these 13 models was based on detailed evaluation of the data sets by Gadgil and Sajani (1998). It was noted that there was a marked improvement in the forecasts of seasonal precipitation from the superensemble in comparison to forecasts from individual models in the multimodel AMIP ensemble (Krishnamurti et al., 2000b). This AMIP exercise did not address how much improvement in rainfall anomalies (above those of climatology) was possible from the conventional superensemble. The issue of rainfall anomaly forecasts in comparison to climatology is addressed in this paper using the FSUSSE and the data from the ensemble of models described in Section 3.2.

5.2. *FSU synthetic superensemble (FSUSSE) methodology*

A variation of the above conventional superensemble formulation was necessary for improved skills for seasonal climate forecasts (Yun et al., 2003). From the member model forecast data sets additional data sets are generated which are called the 'Synthetic Datasets'. This variant of the superensemble contributed towards major improvements of the skills for seasonal climate forecasts. The synthetic data set is created from a combination of the past observations. A consistent spatial pattern is observed among the observations and forecasts. This is simply a linear regression problem in the empirical orthogonal functions (EOF) space. Sets of such synthetic forecasts are then obtained, one for each available forecast, for the creation of superensemble forecasts. The method of creating the synthetic data and the associated statistical procedure is described below.

The time-series of observations can be written as a linear combination of EOFs such as,

$$O(x, t) = \sum_{n} P_n(t) \phi_n(x), \tag{5.2.1}$$

where n is the number of modes selected, which in this study was chosen such that the EOFs explain 99% of the variance in the observed data set. The two variables on the right-hand side of the above equation represent the time (principal component, PC) and space (EOF) decomposition, respectively. PC time-series P(t) represents how EOFs (spatial patterns) evolve in time. PCs and EOFs are defined as below for i member models,

$$F_i(x, t) = \sum_{n} F_{i,n}(t) \varphi_{i,n}(x), \tag{5.2.2}$$

where index i represents a particular member model, from 1 to m.

The interest lies in knowing the spatial patterns of forecast data, which evolve in a consistent way with the EOFs of the observation for the time-series considered. Here, a regression

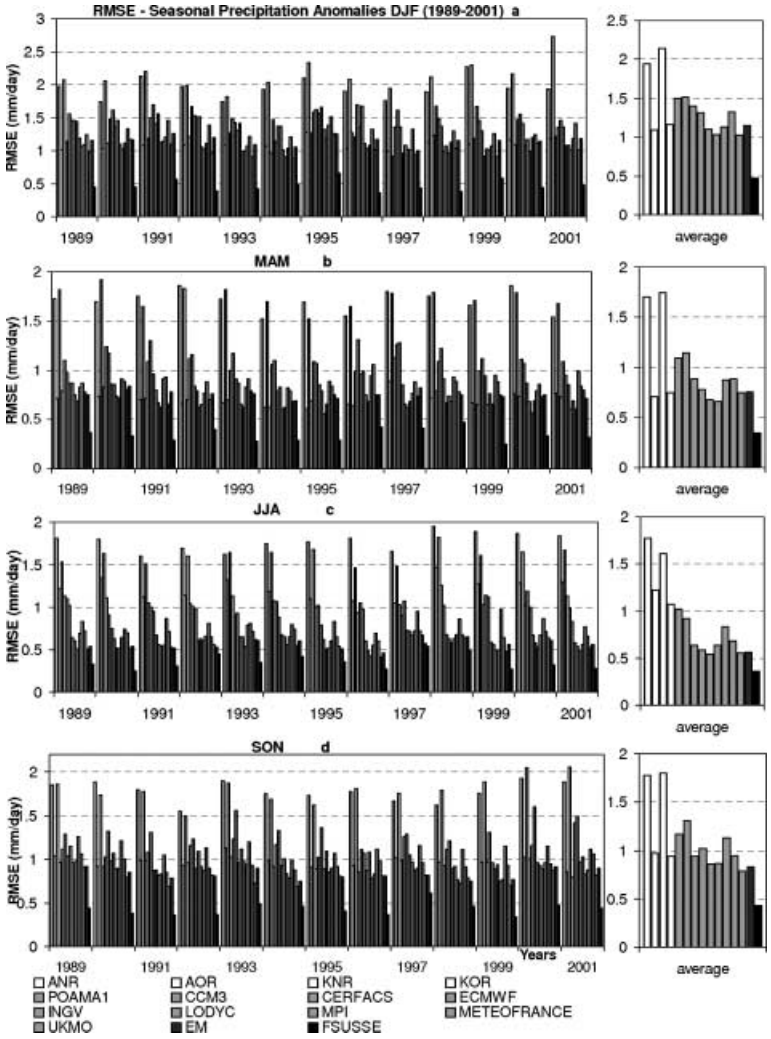

Fig. 9. RMS errors (mm/d) in seasonal precipitation forecast anomalies by the individual models in the ensemble, the multimodel ensemble mean, and the FSUSSE for the period 1989–2001. Inset at the right shows the mean RMS errors for all the years: (a) DJF; (b) MAM; (c) JJA and (d) SON.

relationship is used between the observations PC time-series of forecast data, given as,

$$P_n(t) = \sum_n \alpha_{i,n} F_{i,n}(t) + \varepsilon_{i,n}. \quad (5.2.3)$$

In the above equation, the observation time-series $P(t)$ is expressed in terms of a linear combination of forecast time-series F(t) in EOF space. The regression coefficients, α, are found such that the residual error variance $E(\varepsilon^2)$ is a minimum.

Once the regression coefficients are determined, the PC time-series of synthetic data can be written as:

$$F^{\text{reg}}_{i,n}(t) = \alpha_{i,n} F_{i,n}(t). \quad (5.2.4)$$

Then the synthetic dataset is reconstructed with EOFs and PCs as:

$$F^{\text{syn}}_i(x,t) = \sum_n F^{\text{reg}}_{i,n}(t)\phi_n(x). \quad (5.3.5)$$

The synthetic data (m sets) generated from m member models' forecasts are now subjected to the conventional FSU superensemble technique (Krishnamurti et al., 2000a) described in Section 5.1. Further details on the FSU synthetic superensemble (FSUSSE) are described in Yun et al. (2003 and 2005).

6. Precipitation forecast skills

We shall next address specific improvements in the skill of seasonal climate forecasts for individual seasons. Here we include skills of member models, their ensemble mean and the superensemble. In Figs. 9 and 10 we illustrate the error statistics (RMS errors and anomaly correlations, respectively) for seasonal forecasts of precipitation covering the four seasons of DJF, MAM, JJA and SON (panels a, b, c and d, respectively, in each figure). The four panels of each figure show the results from DEMETER, CCM3, POAMA and FSU coupled models, along with the multimodel ensemble mean and the FSUSSE. It needs to be clarified here that both the RMS errors and the anomaly correlations shown in Figs. 9 and 10, respectively, are based on anomalies for the model forecast fields versus anomalies for the

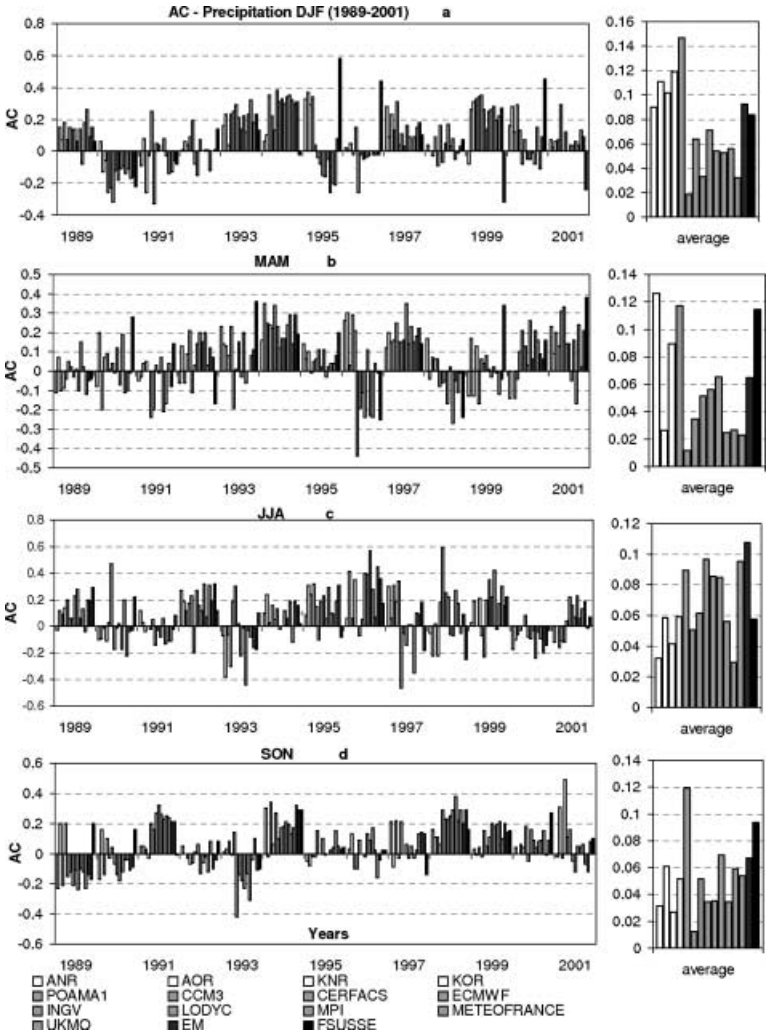

Fig. 10. AC between forecasted anomalies and observed anomalies of precipitation by the individual models in the ensemble, the multimodel ensemble mean, and the FSUSSE for the period 1989–2001. Inset at the right shows the mean AC for all the years: (a) DJF; (b) MAM; (c) JJA and (d) SON.

observed fields. These results pertain to the Euro-Mediterranean precipitation over the region 20N–70N and 30W-60E. The results are based on cross-validation, i.e., all the years of the forecasts, except for the one being addressed, are included in the training database. This was necessary since the data length, i.e., numbers of the forecasts, were still quite small for the optimal development of a training phase (Krishnamurti et al., 2000a). Results presented in these illustrations convey the same essential message, i.e., a stepwise reduction of errors occurs from the individual models to the multimodel ensemble mean and finally to the best product, i.e., the FSUSSE. The RMS errors for the individual models are almost a factor of two larger than for those of the FSUSSE. An inset diagram in the far right, in each panel, shows the 13-yr summary of forecast improvements. Overall, the RMS errors are reduced considerably, since a reduction of RMS error is a built in feature for the design of our superensemble (see eq. 5.3.5). With regard to the anomaly correlations, there were some improvements in the FSUSSE forecasts, but these were not as obvious as for the RMS errors. Looking at the 13-yr average graphs in Fig. 10, the FSUSSE has its best performance in SON when only one member model has a higher AC than the FSUSSE. The worst FSUSSE performance is in JJA when eight member models and the multimodel ensemble mean have higher AC values than the FSUSSE. In principle, it is possible to design a superensemble for the improvement of the anomaly correlation in its own right. The design of a superensemble is currently being explored that can reduce the RMS error and enhance the anomaly correlation at the same time.

As stated previously, many authors have noted a lower skill for the seasonal forecasts of rainfall as compared to other parameters. A premise that seasonal forecasts of precipitation are influenced by internal dynamics, in contrast to boundary forcing, has been offered as a possible explanation for these lower skills. Since internal dynamics and boundary forcing are not mutually exclusive, i.e., one influences the other in a non-linear system, such a partitioning of explanations is questionable. In a seasonal forecast the first month's boundary forcing can easily generate an internal dynamical variability for the next month and vice

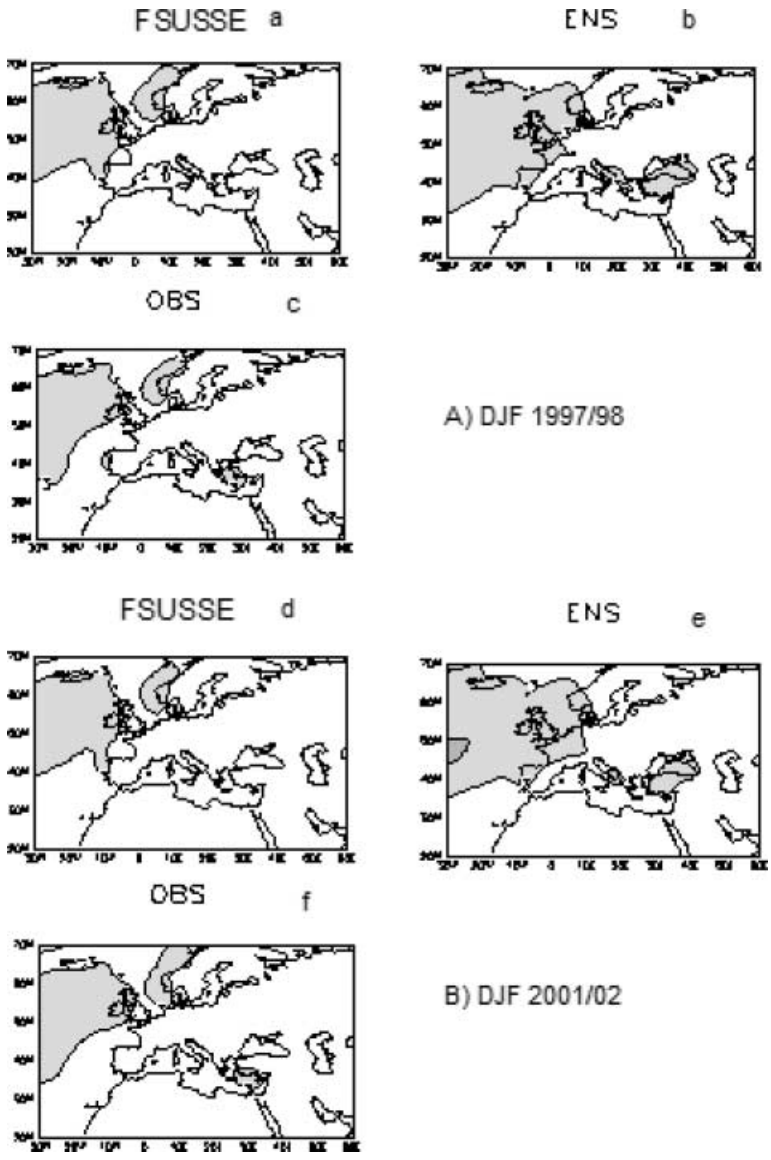

Fig. 11. Seasonal forecasts of precipitation (mm/d) over the Euro-Mediterranean region for (a) DJF 1997/1998 and (b) DJF 2001/2002. Forecasts from the FSUSSE (a and d) and the multimodel ensemble mean of all 13 models (b and e) are shown, along with the observed precipitation from Xie and Arkin (1997) (c and f).

versa. Thus, the reasons for poorer rainfall forecasts still are not apparent.

We shall next address the question of how much improvement in the forecast geographical distribution of seasonal rainfall and the forecast seasonal rainfall anomalies is obtainable from the use of the synthetic superensemble. We will assess that improvement by comparing the synthetic superensemble forecasts and those of the member models to the observed rainfall climatology from the data set of Xie and Arkin (1997).

Typical examples of the seasonal rainfall forecasts over the Euro-Mediterranean region are illustrated in Fig. 11. All these forecasts use a lead-time of 1 month. These are forecasts for months 2, 3 and 4 for December, January and February respectively. An average of these 3 months defines the seasonal rainfall forecasts. The observed seasonal forecasts are shown for the two DJF periods of 1997/1998 and 2001/2002 based on the Xie and Arkin (1997) data sets, along with the forecasts from the synthetic superensemble and the 13-model ensemble mean. The quality of these FSUSSE forecasts is good and matches the high skills (discussed earlier) for the AMIP superensemble. The individual member models' rainfall forecasts were not very impressive (Figs. 9 and 10). Their RMS errors were roughly two or three times larger than those of the synthetic superensemble. The member models carry large systematic errors that render their skill scores rather low, whereas the synthetic superensemble by virtue of its ability to correct such systematic errors can perform 'better'. That 'better' may be no more than the ability of the synthetic superensemble to push its forecasts towards a better climatology. If that is all that the synthetic superensemble does, then one needs to ask about the skill for the seasonal precipitation anomalies, which are central to any climate forecast.

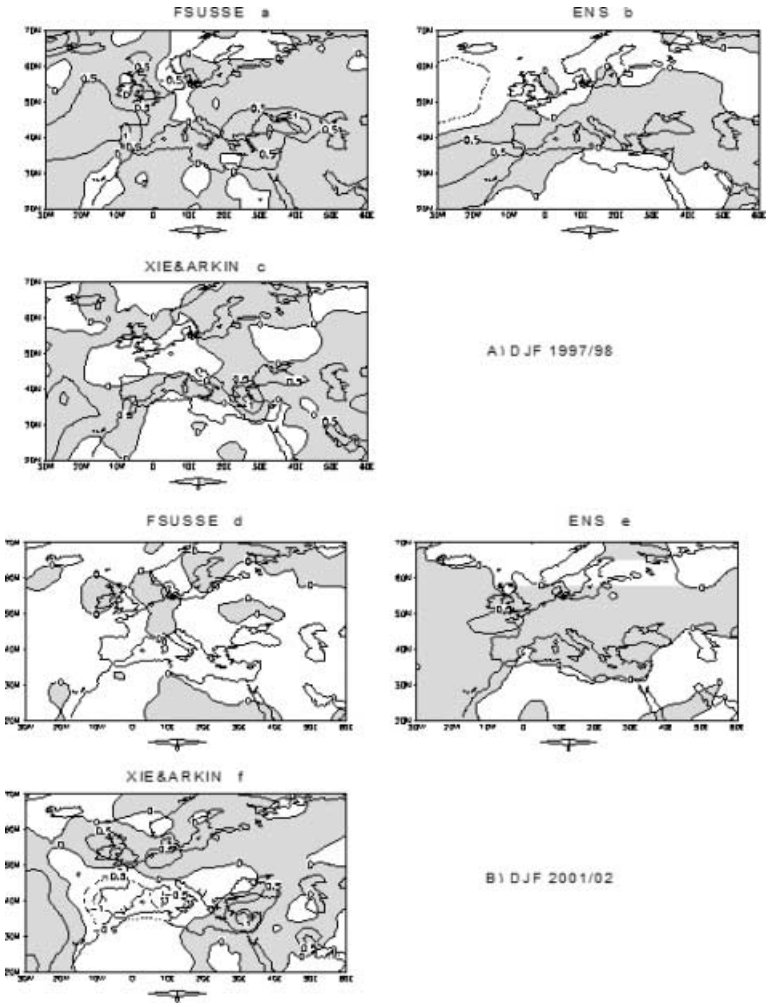

Fig. 12. Seasonal precipitation anomaly forecasts (mm/d) for (a) DJF 1997/1998 and (b) DJF 2001/2002. Forecasts are shown for the FSUSSE (a and d), the multimodel ensemble mean of 13 models (b and e), while observed anomalies are from Xie and Arkin (1997) (c and f).

Figure 12 illustrates examples of precipitation anomaly forecasts for DJF 1997/1998 and DJF 2001/2002. The anomalies are calculated with respect to the observed and the forecast seasonal mean values to define the observed or the forecast anomalies, respectively. Anomaly correlations for these forecast anomalies are generally of the order of 0.2–0.4 over the Euro-Mediterranean domain. Thus, the overall skill of forecasts for the seasonal anomalies is small. Those skills were only slightly better for the synthetic superensemble. Qualitatively, the seasonal rainfall forecast anomaly fields seen in Fig. 12 do seem to provide some useful guidance on where to expect above or below normal rainfall a season in advance. During DJF 1997/1998 (but not during DJF2001/2002), the FSUSSE successfully depicted a belt of above normal seasonal rainfall over the Euro-Mediterranean extending from west to east along 40N. The FSUSSE failed to capture the enhanced positive rainfall anomalies north of the Euro-Mediterranean region. Below normal values occur over the central Euro-Mediterranean Sea, with heavy rainy areas to the north, east and west of that region. Overall, little to no improvement in seasonal precipitation anomaly forecasts for the two selected seasons was obtained from use of the FSUSSE.

7. Probabilistic measure of precipitation forecast skill

The equitable threat score and bias score will be used to provide a probabilistic measure of precipitation forecast skill. These skill scores are based on comparing regions of forecast versus observed precipitation for certain threshold values. Following Schneider et al. (1996) the equitable threat score is defined as,

$$\text{EquitableThreatScore} = \frac{H - \left(F \times \frac{O}{N}\right)}{F + O - H - \left(F \times \frac{O}{N}\right)} (0 \leq ETS \leq 1), \tag{7.1}$$

where

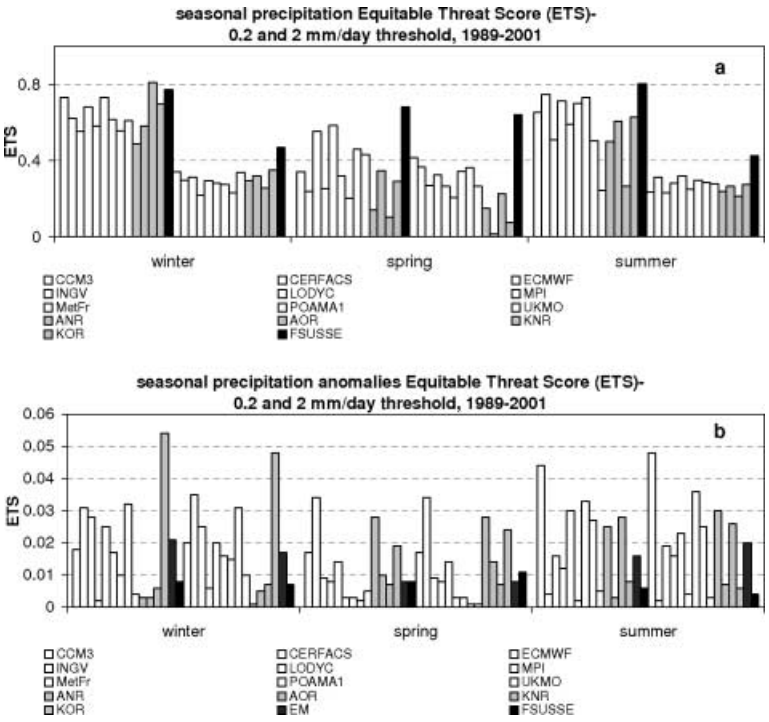

Fig. 13. Equitable threat score for seasonal precipitation (a) and seasonal positive precipitation anomalies (b) for the 13 models of the multimodel ensemble, their ensemble mean, and the FSUSSE. Threshold values are 0.2 and 2 mm/d.

F = the number of grid boxes that forecast more than the threshold

O = the number of grid boxes that observe more than the threshold

H = the number of grid boxes that correctly forecast more than the threshold

$\left(F \times \frac{O}{N}\right)$ = the expected number of correct forecasts due to chance

N = the total number of grid boxes inside the verification domain

The equitable threat score is one measure of probabilistic forecast skill. The equitable threat score can vary from a small negative number to 1.0, where 1.0 represents a perfect forecast. This is basically the ratio of the correct forecast area to the total area of the forecast and observed precipitation. The model gets penalized for forecasting rain in the wrong place as well as for not forecasting rain in the right place. Thus, the model with the highest score is generally the model with the best forecast skill.

Figure 13 shows the equitable threat score for seasonal precipitation forecasts (part (a) of figure) and for seasonal precipitation forecast anomalies (part (b) of figure) for the 13 models of the multimodel ensemble, their ensemble mean, and the FSUSSE for the 0.2 and 2.0 mm/d thresholds. Fig. 14 has an identical depiction, but for the bias score (defined below). In these figures the first fourteen bars represent the 0.2 mm/d threshold, and the second fourteen bars represent the 2 mm/d threshold for each of the seasons (winter, spring, and summer). The time periods for each season are as follows:

WINTER (DJF) — 1 DEC 1988 - 29 FEB 2001

SPRING (MAM) — 1 MAR 1989 - 31 MAY 2001

SUMMER (JJA) — 1 JUN 1989 - 31 AUG 2001

The equitable threat score for precipitation forecasts, and the bias score to be discussed below, were calculated for 1989–2001 over the region 30W–60E and 20N–70N (Fig. 1). The FSUSSE was found to be better than the models comprising the multimodel ensemble with respect to the equitable threat score, as seen in Fig. 13a. In the winter season (DJF) the FSUSSE equitable threat score value is 0.77 (0.47) for a threshold of 0.2 (2.0) mm/d. KNR, CCM3 and MPI models show values of approximately 0.7 (0.3) for 0.2 (2.0) mm/d. The values for the POAMA, AOR and ANR models decrease to about 0.50 (0.27), with those models generally having the lowest equitable threat score values for all the models and for 0.2 (2.0) mm/d threshold. In spring (Fig. 13a), the FSUSSE equitable threat score is 0.68 at a threshold of 0.2 mm/d. Although this value drops to 0.64 by a threshold of 2 mm/d, the FSUSSE maintains the highest score in comparison to all models and for two thresholds. CERFAC, INGV, and MetFr model values are just below those of the FSUSSE, with the POAMA and KNR models having the lowest values for two thresholds. The FSUSSE also has the best equitable threat scores in summer (Fig. 13a). For this season the CERFACS and MetFr models have better scores than the other models. In terms of equitable threat scores (Fig. 13a) it can be seen that the FSUSSE typically has the best ETS values for seasonal precipitation forecasts, especially for precipitation amounts over 2.0 mm for all seasons. CCM3, CERFAC and MPI seem to be better than other ensemble models especially under

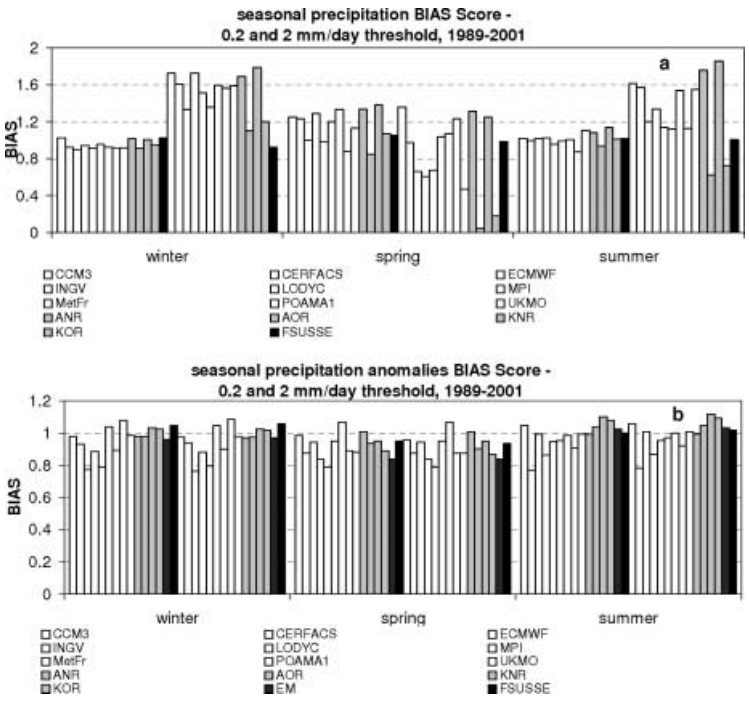

Fig. 14. Bias score for seasonal precipitation (a) and seasonal precipitation anomalies (b) for the 13 models of the multimodel ensemble, their ensemble mean, and the FSUSSE. Threshold values are 0.2 and 2 mm/d.

2.0 mm for winter and summer. LODYC, ECMWF and UKMO are better than other models for spring. Also MetFr and INGV are better than other models for summer seasons. All the models show decreasing skill with increasing precipitation threshold, with almost a steady fall in skill above 0.2 mm. Fig. 13b shows that most models have ETS values under 0.03 for seasonal forecasts of positive anomalies. Especially for spring these values are close to zero, which defines the worst ETS value. CERFACS, CCM3 and KOR have better ETS values for both 0.2 and 2 mm/d thresholds. The FSUSSE values are worse than most of the individual models and usually worse than values for the multimodel ensemble mean. The reason for this performance is not immediately obvious, and further research is being conducted on this question. The FSUSSE has its highest ETS value for the 2.0 mm/d threshold in spring, but this value is surpassed by several other models. In summary, the FSUSSE performs well in terms of ETS for forecast precipitation but not for forecast precipitation anomalies.

Next, we will consider the bias score, which is defined as,

$$\text{Bias} = \frac{N_\text{f}}{N_\text{o}} \tag{7.2}$$

where

N_f = number of grid points where event is forecasted.

N_o = number of grid points where event is observed.

This score does not comment specifically on the skill of a model in forecasting the exact placement of precipitation, but it does give an indication of whether a model is consistently over-forecasting or under-forecasting the area extent of precipitation. The best model in this regard is the one that has a bias score closest to 1.0, which means that the model is not over-forecasting or under-forecasting the area extent of precipitation. If the model has a bias score greater than 1.0, it is over-predicting the area extent of precipitation, and if the score is less than 1.0 it is under-predicting that area extent.

Precipitation bias scores for thresholds of 0.2 and 2.0 mm/d are shown in Fig. 14 for the FSUSSE, the 13 individual models of the multimodel ensemble, and their ensemble mean. It can be seen that the FSUSSE maintains scores that are very close to the perfect score of 1.0 throughout, while the competitors for the most part have values significantly larger than 1 (over-forecast of the area coverage of precipitation). In winter (Fig. 14a) the FSUSSE maintains a value close to 1.0 for two threshold levels, as do the AOR and KOR models. All other models show bias scores that grow very large beyond a threshold of 0.2 mm/d, with a number of models showing scores larger than 1.4 by a threshold of 2 mm/d. In spring (Fig. 14a) the FSUSSE again has the better bias scores, with values close to 1.0 for two thresholds. The CERFAC and MPI models have bias scores that ranging between 0.8 and 1.2. The ECMWF, INGV, LODYC and UKMO models all under forecast the area extent of precipitation at thresholds beyond 0.2 mm/d, while the MetFr, CCM3 and POAMA models over estimate the area extent of precipitation at these thresholds. In summer (Fig. 14a) the FSUSSE again has the best bias scores, and the POAMA, MPI and LODYC models also maintain scores that are close to the ideal of 1.0. The ANR, KNR, CERFACS and ECMWF models all develop very large bias scores with values greater than 1.6 by a threshold of 2.0 mm/d. All these graphs show a striking tendency for the DEMETER, CCM3 and POAMA models to over-forecast the area coverage of

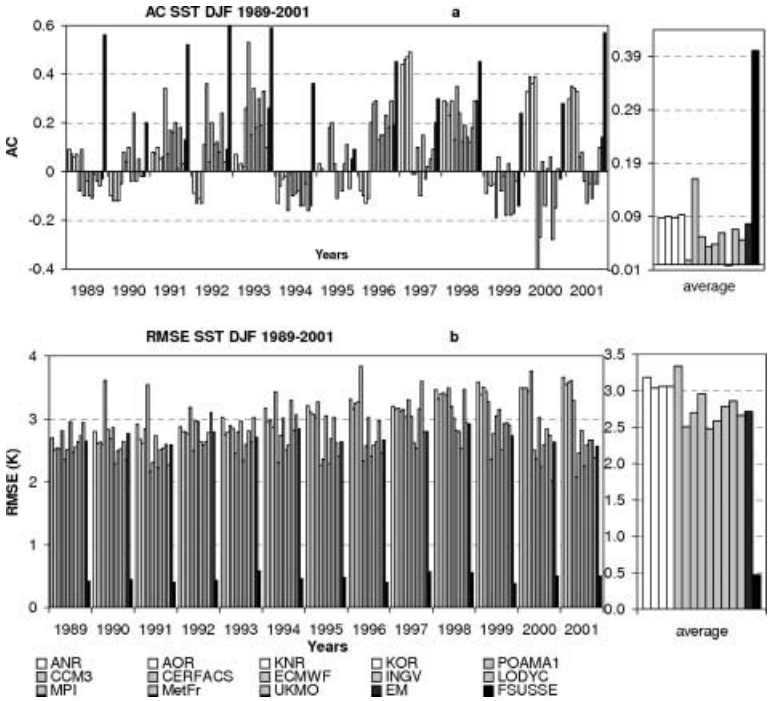

Fig. 15. (a) Anomaly correlation between forecasted anomalies and observed anomalies of SST for individual models in the ensemble, the multimodel ensemble mean, and the FSUSSE for DJF 1989–2001. (b) Root mean square (RMS) error in the forecasts of sea surface temperature anomalies (SST) (K) by the individual models in the ensemble, the multimodel ensemble mean, and the FSUSSE for DJF 1989–2001.

precipitation amounts that are over 2.0 mm/d, and an equally striking ability of the FSUSSE to correct this tendency based on the training period. The FSUSSE repeatedly outperforms all other models, with its curve most closely following the 1.0 line. For precipitation amounts above 0.2 mm, the FSUSSE was clearly the better performer. The improvement in precipitation bias of the FSUSSE over the ECMWF and MPI models can be best seen in the cool seasons, especially during the winter period. This improvement is likely due to the improved resolution of the FSUSSE model. KOR, MPI and ECMWF have almost similar bias scores especially for precipitation amounts with a threshold of 0.2 mm/d during winter. UKMO, ECMWF and LODYC models have better bias scores for spring. LODYC, ECMWF, MPI and INGV models have better bias scores for summer.

Bias scores for seasonal precipitation forecast anomalies are almost 1.0 for all models during winter, spring and summer (Fig. 14b). On average, it appears that the multimodel ensemble mean and the FSUSSE show the best bias scores for 0.2 and 2 mm/d. The FSUSSE model shows the same biases for heavier or lighter precipitation amounts, with the FSUSSE showing improvement in these biases over the four FSU coupled models, ANR, KOR, KNR and AOR

8. Seasonal prediction of sea surface temperature (SST) and surface air temperature

Figures 15 and 16 show the anomaly correlations and RMS errors in the predictions of SST and surface air temperature, respectively, by the 13 models comprising the ensemble, the multimodel ensemble mean, and the FSUSSE for DJF 1989–2001. The anomaly correlations and the RMS errors are both error measures for the SST and surface air temperature anomalies. That is, both metrics are based on anomalies for the model forecast fields versus anomalies for the observed fields. These figures show that the FSUSSE is very successful at reducing RMS errors in the anomalies and in increasing the anomaly correlation scores, in relation to the other models and their ensemble mean, for both SST and surface air temperature forecasts. FSU coupled models have the lowest anomaly correlations for both SST and surface air temperature. The remaining models in the ensemble have higher anomaly correlation scores for SST and surface temperature than the FSU coupled models but not generally as high as the multimodel ensemble mean. The FSUSSE anomaly correlation scores are higher than the ensemble mean scores for both SST and surface air temperature. Fig. 15b depicts the RMS errors for the SST anomaly forecasts, while Fig. 16b shows RMS errors for surface air temperature anomaly forecasts. The FSU coupled models have the highest RMS errors, with the remaining models in the ensemble showing lower values for both SST and surface air temperature anomaly forecasts. The multimodel ensemble mean RMS errors generally represent an improvement over the individual models in the ensemble, but the FSUSSE shows the lowest RMS error values. CERFACS and INGV show better RMS error values than other models for winter. FSUSSE has the smallest RMS errors for SST and surface air temperature anomalies in comparison to all other models. Overall, the FSUSSE appears to represent a very promising technique for seasonal forecasts of SST and surface air temperature anomalies.

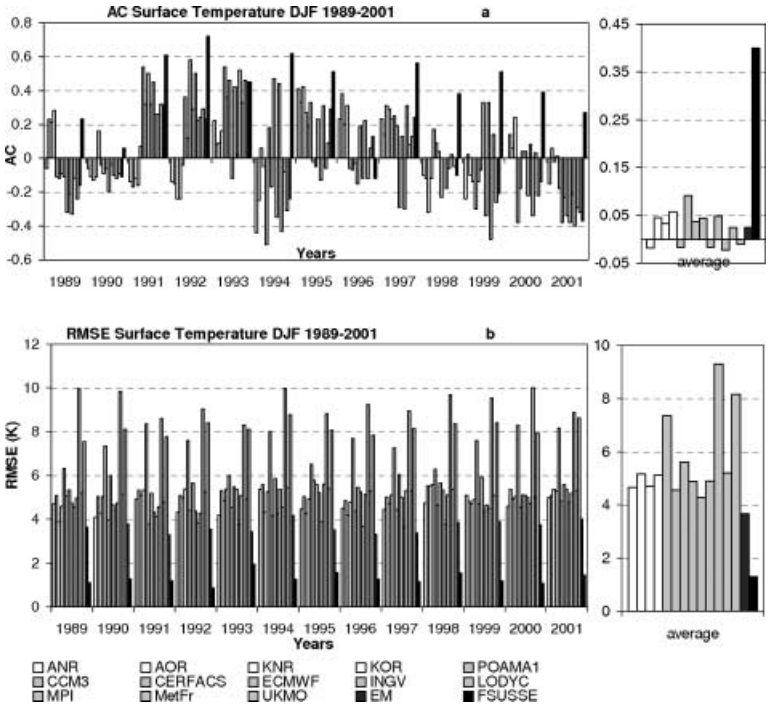

Fig. 16. (a) Anomaly correlation between forecasted anomalies and observed anomalies of surface air temperature for individual models in the ensemble, the multimodel ensemble mean, and the FSUSSE for DJF 1989–2001. (b) Root mean square (RMS) error in the forecasts of surface air temperature anomalies (K) by the individual models in the ensemble, the multimodel ensemble mean, and the FSUSSE for DJF 1989–2001.

9. Concluding remarks

The biggest contribution of this study is the pronounced reduction of RMS errors, along with a slight improvement in the anomaly correlation scores, for seasonal forecasts of precipitation over the Euro-Mediterranean region. The participating models in this superensemble exercise included all of the best available coupled atmosphere-ocean models from Europe, as well as models from Australian, NCAR and a suite of FSU models. This study includes the largest collection of data sets for seasonal climate forecasts. The FSUSSE used in this study is based on a multiple linear regression of weights between observed and simulated PCs, together with observed EOFs. The superensemble combines multimodel forecast runs to arrive at an improved product that typically outperforms the participating member models and their ensemble mean. This is true in terms of RMS errors in forecast anomalies of precipitation but not for the anomaly correlation coefficient for these forecast anomalies. The suite of FSU coupled models deployed for this synthetic superensemble was first independently studied to assess its integrity for simulating seasonal mean climatological features, including precipitation. It was noted that the overall performance of these individual models is comparable to the best among the European suite of models (DEMETER).

Euro-Mediterranean region precipitation skills (RMS errors) for seasonal forecasts show that member model skills are almost two times lower compared to those of the synthetic superensemble. The superensemble is designed for the least-square minimization of the RMS error. Thus, the result is not surprising. The expression for the anomaly correlation entails products of domain sums and is not that straightforward and amenable to improvement using a minimization principle. There are limited improvements in the anomaly correlations of precipitation forecast anomalies over seasonal timescales by the FSUSSE; the anomaly correlations for the synthetic superensemble are indeed higher than those of most member models and the ensemble mean for fall and spring (considering the average over the period 1989–2001). But this is not the case for winter and summer (see Fig. 10).

This paper has also investigated the skill of the FSUSSE as a probabilistic forecast, comparing its forecasts with that of the multimodel ensemble over the Euro-Mediterranean region. The equitable threat and bias scores are calculated for different 'thresholds' of precipitation in mm/d: 0.2 and 2.0 for winter, spring, and summer. The ETS and bias scores of the FSUSSE are better than for most of the other models and their ensemble mean for seasonal precipitation forecasts. However, for seasonal positive precipitation anomaly forecasts, the equitable threat scores of the FSUSSE are lower than for most of the other models and their ensemble mean, although the bias scores are still competitive. While evaluating these results, it should be kept in mind that the data used covered a period of only 13 yr and this may cast some doubts on the long-term robustness of the result. In addition, the choice of such a large region may also mix up results on final prediction scores between Northern Europe, where strong anomalies of precipitation and surface temperature are present during winter in connection with the major general circulation anomalies, and North Africa where precipitation anomalies are

in general lower. This may lead to problems in the interpretation of the results.

The FSUSSE forecasts of SST and surface air temperature for the season considered (winter) were found to be excellent. For these parameters, both the AC coefficients and the RMS errors as applied to the forecast anomalies were found to be far superior to the corresponding values for the models in the multimodel ensemble and their mean for all years 1989–2001.

The construction of a superensemble for seasonal prediction in the Euro-Mediterranean region proved to be worthwhile. The skills were improved compared to those of the member-coupled models, in the ways discussed in the previous paragraphs in this concluding section of the paper. The higher skills for the synthetic superensemble seem to arise for two reasons: the PC time-series and the EOF selection do seem to filter out some of the higher frequency noise that tends to degrade the skills. The spatial structures (EOFs) for model forecasts are projected based on the observed part of the training phase. The time evaluation of the seasonal anomalies in model forecasts are handled better due to the use of weights determined from the statistics of principal component time-series data of the models and the observational counterpart during the training phase. The final superensemble weights are determined from synthetic data sets thus constructed.

10. Acknowledgments

This research was supported by NSF grants ATM-0108741 and ATM-0311858, NASA grant NAG5-13563, NOAA grant NAG16GP1365, and FSU Research Foundation grant 1338-831-45. DEMETER data used in this study have been provided by ECMWF through the ECMWF data server.

References

Alves, O., Wang, G., Zhong, A., Smith, N., Warren, G., and co-authors. 2002. POAMA: Bureau of Meteorology operational coupled model seasonal forecast system. *Proc. ECMWF Workshop on the Role of the Upper Ocean in Medium and Extended Range Forecasting, Reading*, United Kingdom, ECMWF, 22–32.

Chakraborty, A., Krishnamurti, T. N. and Gnanaseelan, C. 2006. *Prediction of the Diurnal Cycle using a Multimodel Superensemble. Part II: Clouds Dept. of Meteorology*, Florida State University, Tallahassee, FL 32306 Report Number: 2006-7. March 22, 2006.

Frei, C. and C. Schar, A. 1998. Precipitation climatology of the Alps from high-resolution rain gauge observations. *Int. J. Climatol.* **18**(8), 873–900.

Gadgil, S. and Sajani, S. 1998. Monsoon precipitation in the AMIP runs. *Climate Dynamics* **14**(9), 659–689.

Gates, W. L., Boyle, J. S., Covey, C., Dease, C. G., Doutriaux, C. M., and co-authors. 1999. An overview of the results of the Atmospheric Model Intercomparison Project (AMIP I). *Bull. Am. Meteorol. Soc.* **80**(1), 29–55.

Grell, G. A. 1993. Prognostic evaluation of assumptions used by Cumulus parameterizations. *Mon. Wea. Rev.* **121**(3), 764–787.

Hartmann, H. C., Pagano, T. C., Sorooshian, S. and Bales, R. 2002. Confidence builders - Evaluating seasonal climate forecasts from user perspectives. *Bull. Am. Meteorol. Soc.* **83**(5), 683.

Kanamitsu, M., Kumar, A., Juang, H. M. H., Schemm, J. K., Wang, W. Q., and co-authors. 2002. NCEP dynamical seasonal forecast system 2000. *Bull. Amer. Meteor. Soc.* **83**(7), 1019–1037.

Kharin, V. V. and Zwiers, F. W. 2002. Climate predictions with multimodel ensembles. *J. Clim.* **15**, 793–799.

Kiehl, J. T., Hack, J. J., Bonan, G. B., Boville, B. A., Williamson, D. L., and co-authors 1998. The National Center for Atmospheric Research community climate model: CCM3. *J. Climate* **11**, 1131–1149.

Krishnamurti, T. N. and Bedi, H. S. 1988. Cumulus Parameterization and Rainfall Rates .3. *Mon. Wea. Rev.* **116**(3), 583–599.

Krishnamurti, T. N., Xue, J. S., Bedi, H. S., Ingles, K. and Oosterhof, D. 1991. Physical initialization for numerical weather prediction over the Tropics. *Tellus Series A-Dynamic Meteorology and Oceanography* **43**(4), 53–81.

Krishnamurti, T. N., Sinha, M. .C., Jha, B. and Mohanty, U. C. 1998. A study of south Asian monsoon energetics. *J. Atmos. Sci.* **55**(15), 2530–2548.

Krishnamurti, T. N., Kishtawal, C. M., LaRow, T. E., Bachiochi, D. R., Zhang, Z., and co-authors. 1999. Improved weather and seasonal climate forecasts from multimodel superensemble. *Science* **285**, 1548–1550.

Krishnamurti, T. N., Kishtawal, C. M., Zhang, Z., LaRow, T., Bachiochi, D., and co-authors. 2000a. Multi-model ensemble forecasts for weather and seasonal climate. *J. Climate* **13**(23), 4196–4216.

Krishnamurti, T. N., Bachiochi, D., LaRow, T., Jha, B., Tewari, M., and co-authors. 2000b. Coupled atmosphere-ocean modeling of the El Niño of 1997-98. *J. Climate* **13**(14), 2428–2459.

Krishnamurti, T. N., Stefanova, L., Chakraborty, A., Kumar, T. S. V. V., Cocke, S., and co-authors. 2002. Seasonal forecasts of precipitation anomalies for North American and Asian monsoons. *J. Meteorol. Soc. Japan* **80**(6), 1415–1426.

Lacis, A. A. and Hansen, J. E. 1974. Parameterization for absorption of solar–radiation in earth's atmosphere. *J. Atmos. Sci.* **31**(1), 118–133.

LaRow, T. E. and Krishnamurti, T. N. 1998. Initial conditions and ENSO prediction using a coupled ocean-atmosphere model. *Tellus* **50A**, 76–94.

Latif, M. 1987. Tropical Ocean circulation experiments. *J. Phys. Oceanogr.* **17**(2), 246–263.

Lorenz, E. N. 1963. Deterministic non-periodic flow. *J. Atmos. Sci.* **20**, 130–141.

Mann, M. E. 2001. Large-scale temperature patterns in past centuries: Implications for North American climate change. *Hum. Ecol. Risk Assess.* **7**(5), 1247–1254.

Mason, S. J., Goddard, L., Graham, N. E., Yulaeva, E., Sun, L. Q. and co-authors. 1999. The IRI seasonal climate prediction system and the 1997/98 El Niño event. *Bull. Am. Meteorol. Soc.* **80**(9), 1853–1873.

McPhaden, M. J., Busalacchi, A. J., Cheney, R., Donguy, J. R., Gage, K. S. and co-authors. 1998. The tropical ocean global atmosphere observing system: A decade of progress. *J. Geophys. Res.-Oceans* **103**(C7), 14169–14240.

Molteni, F., Buizza, R., Palmer, T. N. and Petroliagis, T. 1996. The ECMWF ensemble prediction system: Methodology and validation. Quart. *J. Roy. Meteor. Soc.* **122**, 73–119.

Moron, V., Navarra, A., Ward, M. N. and Roeckner, E. 1998. Skill and reproducibility of seasonal rainfall patterns in the tropics in ECHAM-4 GCM simulations with prescribed SST. *Clim. Dyn.* **14**(2), 83–100.

Mullen, S. L. and Baumhefner, D. P. 1994. Monte Carlo simulations of explosive cyclogenesis. *Mon. Wea. Rev.* **122**, 1548–1567.

Neelin, J. D., Battisti, D. S., Hirst, A. C., Jin, F. F., Wakata, Y. and co-authors. 1998. ENSO theory. *J. Geophys. Res.-Oceans* **103**(C7), 14261–14290.

Palmer, T. N., Alessandri, A., Andersen, U., Cantelaube, P., Davey, M., Delecluse, P., Deque, M., Diez, E., Doblas-Reyes, F. J., Feddersen, H., Graham, R., Gualdi, S., Gueremy, J.-F., Hagedorn, R., Hoshen, M., Keenlyside, N., Latif, M., Lazar, A., Maisonnave, E., Marletto, V., Morse, A. P., Orfila, B., Rogel, P., Terres, J.-M. and Thomson, M .C. 2004. Development of a European Multi Model Ensemble System for Seasonal to Inter-Annual Prediction (DEMETER). *BAMS* **85**(6), 853–872.

Pavan, V. and Doblas-Reyes, F. J. 2000. Multi-model seasonal hindcasts over the Euro-Atlantic: skill scores and dynamic features. *Clim. Dyn.* **16**(8), 611–625.

Pielke, R. and Carbone, R. E. 2002. Weather impacts, forecasts, and policy - An integrated perspective. *Bull. Am. Meteorol. Soc.* **83**(3), 393.

Quadrelli, R., Pavan, V. and Molteni, F. 2001. Wintertime variability of Mediterranean precipitation and its links with large-scale circulation anomalies. *Clim. Dyn.* **17**(5-6), 457–466.

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

Rodó, X., Baert, E. and Comín, F. A. 1997. Variations in seasonal rainfall in Southern Europe during the present century: relationships with the North Atlantic Oscillation and the El Niño-Southern Oscillation. *Clim. Dyn.* **13**(4), 275–284.

Schneider, R. S., Junker, N. W., Eckert, M. T. and Considine, T. M. 1996. The performance of the 29 km Meso-Eta model in support of forecasting at the Hydrometeorological Prediction Center. Preprints, 15th Conf. on Weather Analysis and Forecasting, AMS, Norfolk, VA.

Spencer, R. W. 1993. Global Oceanic precipitation from the MSU during 1979-91 and comparisons to other climatologies. *J. Climate* **6**, 1301–1326.

Sperber, K. R. and Palmer, T. N. 1996. Interannual tropical rainfall variability in general circulation model simulations associated with the atmospheric model intercomparison project. *J. Climate* **9**(11), 2727–2750.

Stefanova, L. and Krishnamurti, T. N. 2002. Interpretation of seasonal climate forecasts using Brier skill score, FSU superensemble, and the AMIP-1 data set. *J. Climate* **15**, 537–544.

Thomson, M. C., Palmer, T., Morse, A. P., Cresswell, M., Conner, S. J. 2000. Forecasting disease risk with seasonal climate predictions. *LANCET*, **355**(9214), 1559–1560.

Toth, Z. and Kalnay, E. 1997. Ensemble forecasting at NCEP and the breeding method. *Mon. Wea. Rev.* **125**, 3297–3319.

World Climate Research Program (WCRP). 1985. International Implementation Plan for TOGA, ITPO Document #1, (first edition), World Meteorological Organization W.M.O., Geneva.

Xie, P., Rudolf, B., Schneider, U. and Arkin, P. A. 1996. Gaugebased monthly analysis of global land precipitation from 1971–1994. *J. Geophys. Res.* **101**(D14), 19023–19034.

Xie, P. and Arkin, P. A. 1997. Global precipitation: A 17-year monthly analysis based on gauge observations, satellite estimates and numerical model outputs. *Bull. Amer. Met. Soc.* **78**(11), 2539–2558.

Xoplaki, E., Luterbacher, J., Burkard, R., Patrikas, I. and Maheras, P. 2000. Connection between the large-scale 500 hPa geopotential height fields and precipitation over Greece during wintertime. *Clim. Res.* **14**(2), 129–146.

Yun, W. T., Stefanova, L. and Krishnamurti, T. N. 2003. Improvement of the multimodel superensemble technique for seasonal forecasts. *J. Clim.*, **16**(22), 3834–3840.

Yun, W. T., Stefanova, L., Mitra, A. K., Vijaya Kumar, T. S. V., Dewar, W., and co-authors. 2005. A multi-model superensemble algorithm for seasonal climate prediction using DEMETER forecasts. *Tellus* **57A**, 280–289.

Zebiak, S. E. and Cane, M. A. 1987. A Model El Niño Southern Oscillation. *Mon. Wea. Rev.* **115**(10), 2262–2278.

Zhong, A., Coleman, R., Smith, N., Naughton, M., Rikus, L., and co-authors. 2001. Ten year AMIP 1 climatology from versions of the BMRC atmospheric model. BMRC Research Report No 83, Bureau of Meteorology Research Centre, Melbourne, Australia, **34**.