Meteorol Atmos Phys (2007)
DOI 10.1007/s00703-006-0234-3
Printed in The Netherlands

Meteorology and Atmospheric Physics

[1] Department of Meteorology, The Florida State University, Tallahassee, FL, USA
[2] Department of Physics, University of the West Indies at Mona, Jamaica
[3] Department of Meteorology, Istanbul Technical University, Maslak, Istanbul, Turkey

Improved seasonal climate forecasts for the Caribbean region using the Florida State University Synthetic Superensemble

R. S. Ross[1], **A. Chakraborty**[1], **A. Chen**[2], **L. Stefanova**[1], **S. Sirdas**[1,3], and **T. N. Krishnamurti**[1]

With 20 Figures

Received May 11, 2006; accepted September 27, 2006
Published online: March 21, 2007

Summary

Climate variations in the Caribbean, largely manifest in rainfall activity, have important consequences for the large-scale water budget, natural vegetation, and land use in the region. The wet and dry seasons will be defined, and the important roles played by the El Niño-Southern Oscillation (ENSO) and the North Atlantic Oscillation (NAO) in modulating the rainfall during these seasons will be discussed.

The seasonal climate forecasts in this paper are made by 13 state of the art coupled atmosphere-ocean general circulation models (CGCMs) and by the Florida State University Synthetic Superensemble (FSUSSE), whose forecasts are obtained by a weighted combination of the individual CGCM forecasts based on a training period. The success of the models in simulating the observed 1989–2001 climatology of the various forecast parameters will be examined and linked to the models' success in predicting the seasonal climate for *individual* years. Seasonal forecasts will be examined for precipitation, sea-surface temperature (SST), 2-meter air temperature, and 850 hPa u- and v-wind components during the period 1989–2001. Evaluation metrics include root mean square (RMS) error and Brier skill score. It will be shown that the FSUSSE is superior to the individual CGCMs and their ensemble mean both in simulating the 1989–2001 climatology for the various parameters and in predicting the seasonal climate of the various parameters for *individual* years.

The seasonal climate forecasts of the FSUSSE and of the ensemble mean of the 13 state of the art CGCMs will be evaluated for years (during the period 1989–2001) that have particular ENSO and NAO signals that are known to influence Caribbean weather, particularly the rainfall. It will be shown that the FSUSSE provides superior forecasts of rainfall, SST, 2-meter air temperature, and 850 hPa u- and v-wind components during dry summers that are modulated by negative SOI and/or positive NAO indices. Such summers have become a feature of a twenty-year pattern of drought in the Caribbean region. The results presented in this paper will show that the FSUSSE is a valuable tool for forecasting rainfall and other atmospheric and oceanic variables during such periods of drought.

1. Introduction

Climate variations in the Caribbean are reflected primarily in rainfall activity. These climate shifts have not received the attention accorded to other climatic variations worldwide, such as rainfall oscillations in the Indian monsoon region and in sub-Saharan Africa. Nevertheless, periods of wet years have been followed by periods of dry years throughout the 20th century in the Caribbean region with important consequences for the large-scale water budget, natural vegetation, and land use (Hastenrath, 1985).

The wet and dry seasons in the Caribbean region will be defined in Sect. 2, along with a discussion of the important roles played by the El Niño-Southern Oscillation (ENSO) and the

North Atlantic Oscillation (NAO) in controlling Caribbean region rainfall. In Sect. 3, the methodologies involved in the Florida State University conventional superensemble (FSUSE) and the more advanced Synthetic Superensemble (FSUSSE) will be presented. The FSUSSE climate forecasts are used in this paper. Brief descriptions will be given of the 13 state of the art atmosphere-ocean coupled general circulation models (CGCMs) used in this research. An explanation will be given for the procedures that are used to obtain the seasonal mean values of the various parameters forecast by the 13 CGCMs. These seasonal mean forecasts comprise the model data sets that are used to construct the FSUSSE forecasts. A description of the verifying observed data sets will also be presented.

In Sect. 4, we will examine the success of the individual CGCMs, the ensemble mean of these models, and the FSUSSE in simulating the observed seasonal climatologies of precipitation, sea surface temperature (SST), 2-meter air temperature, and 850 hPa wind for the period 1989–2001. The seasons are defined as January–March (dry season), May–June (early peak of the summer wet season), and September–November (late peak of the summer wet season). The ability of the various models to predict the seasonal climate for *individual* years is enhanced by their ability to simulate the observed seasonal climatology over a set number of years.

In Sect. 5, documentation will be given for the performance of the individual CGCMs, the ensemble mean of these models, and the FSUSSE in predictions of seasonal climate for *individual* years during 1989–2001, considering the variables of precipitation, SST, 2-meter air temperature, and 850 hPa wind. The seasons are defined as in the previous paragraph.

The seasonal climate forecasts of the FSUSSE and the multi-model ensemble mean in relation to the observed patterns of ENSO and NAO for the years 1989–2001 will be explored in Sect. 6. The goal will be to see how well the FSUSSE performs in relation to the multi-model ensemble mean in its seasonal climate forecasts in years that have particular ENSO and NAO signals that are known to influence Caribbean weather, particularly the rainfall. In Sect. 7, a probabilistic forecast evaluation of the precipitation forecasts found in this paper will be presented using the Brier skill score.

Finally, in Sect. 8 the salient findings of this study will be summarized and suggestions for future research will be presented.

2. Caribbean rainfall: Patterns and controls

Rainfall in the Caribbean is characterized by a winter dry season that reaches its peak in January–March and a summer wet season that has an early peak in May–June and a late peak in September–October. This pattern is illustrated for the period 1989–2001 in Fig. 1, where observed rainfall is from CMAP (Xie and Arkin, 1997). There is considerable inter-seasonal variation in precipitation for both the dry and wet seasons. Figure 2 shows such variation for the period 1989–2001.

Rainfall variability in the Caribbean is largely controlled by the combined effects of the El Niño-Southern Oscillation (ENSO) and the North Atlantic Oscillation (NAO) (Giannini et al, 2000; 2001; Enfield and Alfaro, 1999; Rogers, 1988). An El Niño event is associated with below average summer rainfall in year (0) and above average rainfall in the spring of year (+1). The summer rainfall deficit is due to divergent surface wind flow over the Caribbean in response to the eastward shift of the deep convection in the Pacific Ocean. The spring rainfall surplus is due to lagged warming of the tropical Atlantic Ocean. The NAO affects Caribbean rainfall directly through its influence on subsidence and indirectly through its impact on sea surface temperature (SST). In the positive phase of the NAO, pressures in the surface North Atlantic high are anomalously high, producing stronger than average subsidence over the Caribbean, and increased surface wind flow in the region that leads to

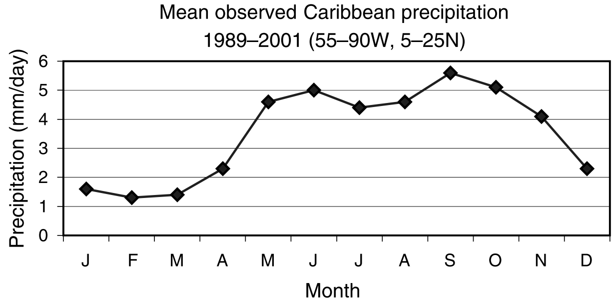


Fig. 1. Mean observed monthly precipitation (mm/day) in the Caribbean (5° N–25° N, 55° W–90° W) for the period 1989–2001 from CMAP data (Xie and Arkin, 1997)

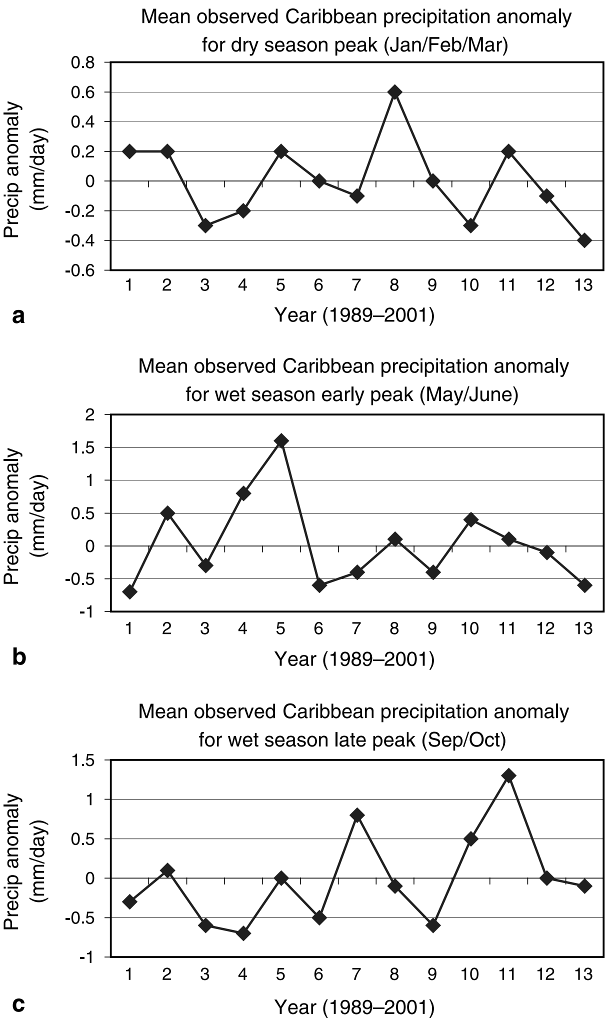


Fig. 2. Observed precipitation anomaly (mm/day) in the Caribbean (5° N–25° N, 55° W–90° W) for (**a**) dry season peak, (**b**) wet season early peak, and (**c**) wet season late peak for each year 1989–2001. Anomaly is defined as departure from the 1989–2001 mean for each period from the CMAP precipitation (Xie and Arkin, 1997)

reduced SSTs. The result is dryer conditions. In the negative phase of the NAO, the north Atlantic high is weaker, the Caribbean experiences reduced subsidence, weaker winds, and warmer SSTs, resulting in wetter conditions.

When the ENSO and NAO oscillations interfere constructively the greatest impact on rainfall occurs. The driest summers occur when a positive NAO phase from the previous winter coincides with El Niño year (0) conditions (divergent surface wind flow, increased subsidence, and cooler than average SSTs). The wettest springs occur when a negative NAO phase from winter is superimposed on El Niño year (+1) conditions (reduced subsidence, and warmer SSTs from both cycles). The wet spring may be lost if El Niño year (+1) conditions are offset by a positive phase in the NAO cycle (warmer than average SSTs negated, and subsidence enhanced). In the past 20 years the frequency of association of El Niño conditions with the positive phase of the NAO has increased, partially explaining the persistence of dryer than average summers during El Niño year (0) and dryer than average springs in El Niño year (+1).

One additional effect of El Niño on Caribbean rainfall may be mentioned. Laing (2004) has pointed out that in the winter dry season during an El Niño year, episodes of heavy rain and flash flooding may be experienced. This results from mid-latitude systems, such as fronts, pre-frontal troughs and upper level lows, tracking well south of their normal tracks and initiating deep convection over the Caribbean region.

3. Methodology, models, and data sets

3.1 Conventional Florida State University Superensemble methodology (FSUSE)

In the conventional Florida State University (FSU) Superensemble technique (FSUSE) (Krishnamurti et al, 1999; 2000a) forecasts from a set of multi-models are used to construct a single consensus forecast. During the training phase of the technique past multi-model forecasts are validated by comparison to the observed (analyzed) fields, and multiple linear regression is used to determine the performance of each model. The regression analysis produces weights that are assigned to each model, and these weights are then used in the forecast phase of the technique to construct the FSUSE forecasts.

In the multiple regression analysis temporal anomalies of the multi-model forecast fields and the observed fields are utilized rather than the actual fields. Thus, in the FSUSE forecast equation the derived weights are multiplied by the corresponding model anomalies. The FSUSE forecast is formulated as

$$S = \overline{O} + \sum_{i=1}^{N} a_i(F_i - \overline{F_i}), \qquad (1)$$

where $\overline{O}$ is the mean observed value during the training period, a_i is the weight assigned to the

i-th model, and F_i and $\overline{F_i}$ are the current forecast and the mean forecast during training for the i-th model, respectively. The summation is over the N number of models used in the ensemble.

In Eq. (1), the weights a_i are obtained by a minimization of the error term G which is defined as

$$G = \sum_{t=1}^{N} (S_t' - O_t')^2, \tag{2}$$

where N is the number of times in the training period and S_t' and O_t' are the FSUSE and observed field anomalies respectively at training time t. A separate FSUSE forecast Eq. (1) is derived for each forecast variable, forecast time, vertical level, and horizontal grid point.

The FSUSE forecast is different from a bias removed ensemble mean, where each weight in Eq. (1) would be $1/N$, with N representing the number of models used in the ensemble. Stefanova and Krishnamurti (2002) showed that the FSUSE forecasts are superior to the bias removed multimodel ensemble mean forecasts. They also showed that even though the FSUSE forecasts are deterministic, the FSUSE forecast equation also carries an equivalent probabilistic forecast.

3.2 Florida State University Synthetic Superensemble (FSUSSE) methodology

This formulation of the Florida State University Synthetic Superensemble (FSUSSE) was shown by Yun et al (2005) to provide better forecasts than the conventional Florida State University Superensemble (FSUSE). In this approach synthetic data sets are used in the superensemble algorithm in place of forecast data sets from the models comprising the ensemble. One set of synthetic data is created corresponding to each model, and this is accomplished by using linear regression with the observed (analyzed) field in empirical orthogonal function (EOF) space.

Observational data can be written as a linear combination of the principal component (PC) and the spatial EOFs as

$$O(x,t) = \sum_{k=1}^{M} P_k(t) \cdot \Phi_k(x), \tag{3}$$

where $P_k(t)$ is the PC time series and $\Phi_k(x)$ represent the EOF component of the k-th mode. In this study M, the number of modes selected, was chosen such that the EOFs explain 99% of the variance in the actual observed data set. The results are independent of choice of M, as long as the EOFs explain at least 95% of the variance.

In a similar way the model data sets are also expanded in terms of PC time series and EOFs as

$$F_i(x,t) = \sum_{k=1}^{M} F_{i,k}(t) \cdot \phi_{i,k}(x), \tag{4}$$

where i denotes the i-th model in the ensemble. M is the number of modes selected, and $F_{i,k}(t)$ and $\phi_{i,k}(x)$ are the PC time series and spatial EOF for the k-th mode, respectively.

A coherent pattern of the forecast and observational data is created by multiple linear regression. The k-th mode of the PC time series of the observations is expressed as a linear combination of the k-th mode of the PC time series of the different models in the ensemble, with an error term formulated as

$$P_k(t) = \sum_{i=1}^{N} \alpha_{i,k} \cdot F_{i,k}(t) + \varepsilon_{i,k}, \tag{5}$$

where i denotes the i-th model in the ensemble, $\alpha_{i,k}$ is the weight assigned to the i-th model for mode k, and $\varepsilon_{i,k}$ is the error term. The α's are estimated from a multiple linear regression that minimizes the error variance $E(\varepsilon^2)$. Once the weights are determined, the regression improved PC time series of the k-th mode for the i-th model is defined as

$$F_{i,k}^{reg}(t) = \alpha_{i,k} \cdot F_{i,k}(t) \tag{6}$$

and the synthetic data for the i-th model is defined as

$$F_i^{reg}(x,t) = \sum_{k=1}^{M} F_{i,k}^{reg}(t) \cdot \Phi_k(x), \tag{7}$$

where $\Phi_k(x)$ are the EOF components of the observation. The synthetic data is generated by Eq. (7) for every model in the ensemble, and this data is then used with the conventional FSU Superensemble methodology (FSUSE) to construct a Synthetic Superensemble forecast (FSUSSE). This forecast was found to be better than the conventional FSUSE forecast (Yun et al, 2005). It will be seen that these forecasts offer a significant improvement over the forecasts of individual models in the ensemble, as well as the forecasts

Table 1. Characteristics of models in the FSU Synthetic Superensemble

Models	Atmospheric component			Oceanic component		
	Model	Res.	IC	Model	Res.	IC
KOR[1,3]	FSU	T63L14	ECMWF with Phy Init[5]	HOPE global	$5° \times 0.5°–5°$ -17L	Coupled Relax Obs SST[6]
KNR[1,4]	FSU	T63L14	ECMWF with Phy Init[5]	HOPE global	$5° \times 0.5°–5°$ -17L	Coupled Relax Obs SST[6]
AOR[2,3]	FSU	T63L14	ECMWF with Phy Init[5]	HOPE global	$5° \times 0.5°–5°$ -17L	Coupled Relax Obs SST[6]
ANR[2,4]	FSU	T63L14	ECMWF with Phy Init[5]	HOPE global	$5° \times 0.5°–5°$ -17L	Coupled Relax Obs SST[6]
CCM3		T63L26	AVN	NCOM slab		SST NCEP/NCAR
POAMA	BAM	T47L17	BAM analysis	ACOM	$5° \times 0.8°–1.5°$ 25L	optimum interp.[10]
CERFACS	ARPEGE	T63L31	ERA40	OPA 8.2	$2° \times 2°$ 31Levs	ERA40
CNRM	ARPEGE	T63L31	ERA40	OPA 8.0	182×152 GP 31 Levs	ERA40
LODYC	IFS	T95L40	ERA40	OPA 8.2	$2° \times 2°$ 31 Levs	ERA40
INGV	ECHAM-4	T42L19	Coupled AMIP-type	OPA 8.1	$2° \times 0.5°–1.5°$ 31 Levs	ERA40
MPI	ECHAM-5	T42L19	Coupled run	MPI-OMI	$2.5° \times 0.5°–2.5°$ 23 Levs	Coupled run
Met Office	ARPEGE	T63L31	ERA40	GloSea OGCM	$1.25° \times 0.3°–0.25°$ 40 Levs	ERA40
ECMWF	IFS	T95L40	ERA40	HOPE-E	$1.4° \times 0.3°–1.4°$ 29 Levs	ERA40

[1] Krishnamurti and Bedi (1988); [2] Grell (1993); [3] Chang (1979); [4] Lancis and Hansen (1974); [5] Krishnamurti et al (1991); [6] LaRow and Krishnamurti (1998); [7] Bureau of Meteorology Unified Atmospheric Model – BAM; [8] Australian Community Ocean Model version 2 (ACOM2). [9] Large et al (1997); [10] Schiller et al (1997)

of the multi-model ensemble mean, when compared to the observed fields.

3.3 Models

The superensemble forecasts presented in this paper are based on 13 state of the art atmosphere-ocean coupled general circulation models (CGCMs). Table 1 provides acronym references for these models, along with their characteristics.

Four versions of the Florida State University Global Spectral Model (FSUGSM) were run based on different combinations of two cumulus parameterization schemes (Arakawa-Schubert and Kuo) and two different radiation parameterization schemes. The Arakawa-Schubert scheme is a simplified form of the original scheme (Arakawa and Schubert, 1974) as formulated by Grell (1993). The Kuo scheme is detailed in Krishnamurti et al (1980), and Krishnamurti and Bedi (1988), and represents a modified version of the scheme as presented by Anthes (1977). The newer radiation scheme is based on band models, while the older scheme is emissivity-absorptivity based (Krishnamurti et al, 2002). The atmospheric component of the FSUGSM was run at the horizontal resolution of T63 with 14 levels in the vertical. The following acronyms represent the four versions of the FSUGSM in Table 1: Kuo scheme with older radiation scheme (KOR); Kuo scheme with newer radiation scheme (KNR); Arakawa-Schubert scheme with older radiation scheme (AOR); Arakawa-Schubert scheme with newer radiation scheme (ANR). The oceanic component of the FSUGSM is derived from the Hamburg Ocean Primitive Equation (HOPE) global model (Wolff et al, 1997) and has a horizontal resolution of 5° longitude by 0.5°–5.0° latitude, with higher resolution over the equatorial latitudes. Initial conditions for the oceanic model were obtained by a comprehensive coupled assimilation in which relaxation was used to obtain initial SSTs from the observed SSTs (Krishnamurti et al, 2002).

The suite of models used in this research also included seven models of the Development of a European Multi-model Ensemble system for seasonal to inTERannual prediction (DEMETER) as described in Palmer et al (2004). The DEMETER data set contains an ensemble of 9 simulations for each model, and these 9 simulations were averaged before a given model's data was entered into the superensemble.

Two additional models were utilized to complete the ensemble of 13 models. These are the coupled version of the Community Climate Model-3 (CCM3) from the National Center for Atmospheric Research (NCAR) and the Predictive Ocean Atmosphere Model for Australia-1 (POAMA1). The atmospheric component of the CCM3 model was run at a resolution of T63 with 18 levels in the vertical. The oceanic data set was formulated in a manner similar to that for the four FSUGSM models, as described above. The POAMA1 model's atmospheric component was run at a resolution of R47 with 17 vertical levels. The model is described in detail in Wang et al (2004).

3.4 Data sets

The various parameters forecast by the 13 CGCMs were available for this study as monthly averages for the period 1989–2001. Seasonal mean values have been used in the study. The climate forecasts were structured to coincide with the unique rainfall seasons in the Caribbean as described in Sect. 2 (winter dry season peak in January–March, summer wet season early peak in May–June and late peak in September–November). To get the January–March seasonal forecasts, the year was divided into four blocks of three months each (January–March, April–June, July–September, October–December). This yielded 52 seasonal forecasts for each of the 13 models (4 seasons × 13 years) of which 13 seasonal forecasts for each model were obvi-

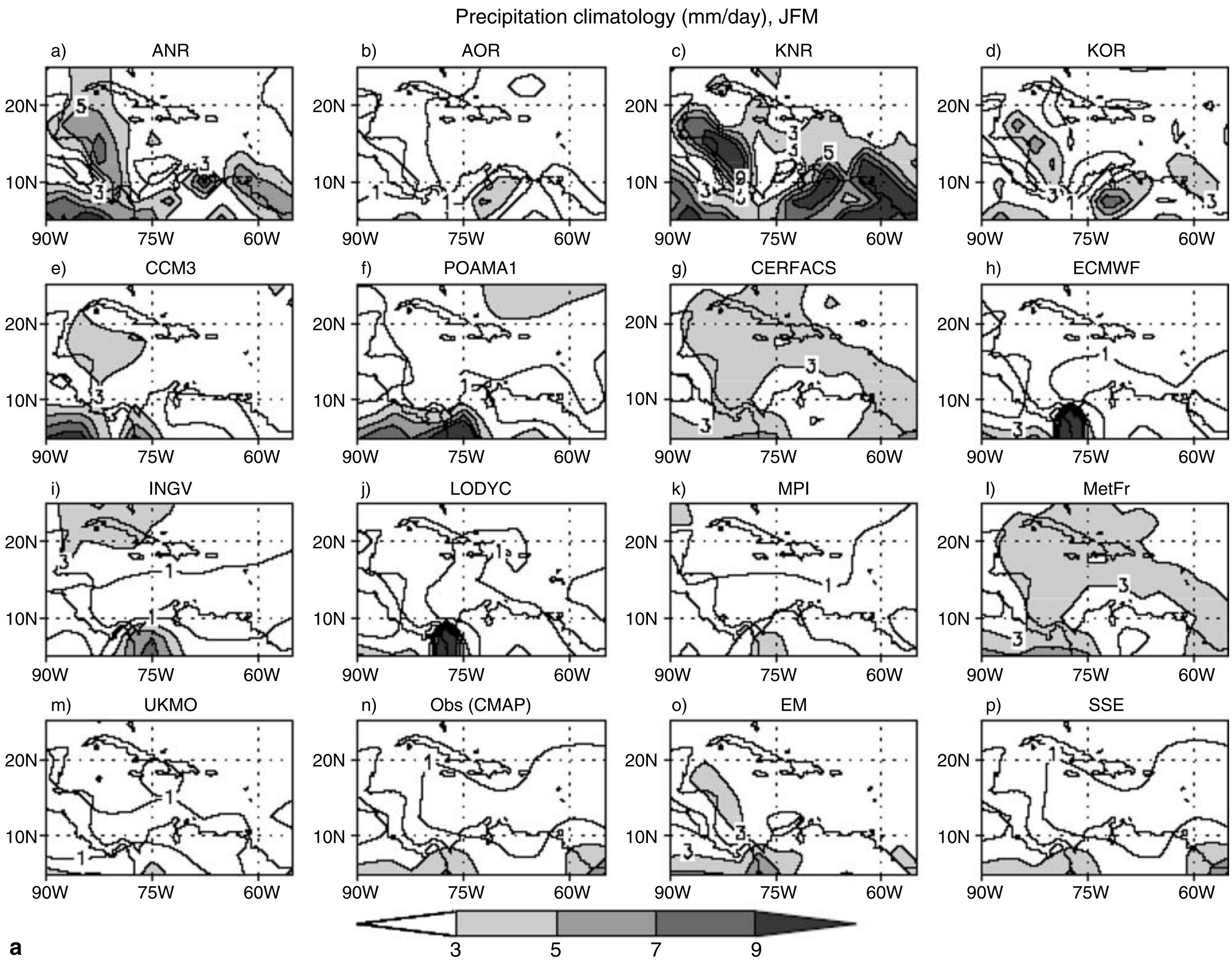


Fig. 3a. Precipitation climatology (mm/day) for January–March 1989–2001 for each of the 13 models in the multi-model ensemble, the multi-model ensemble mean (EM), and the FSU Synthetic Superensemble (SSE). The observed precipitation climatology is from CMAP data (Xie and Arkin, 1997). (**b**) Error in the precipitation climatology

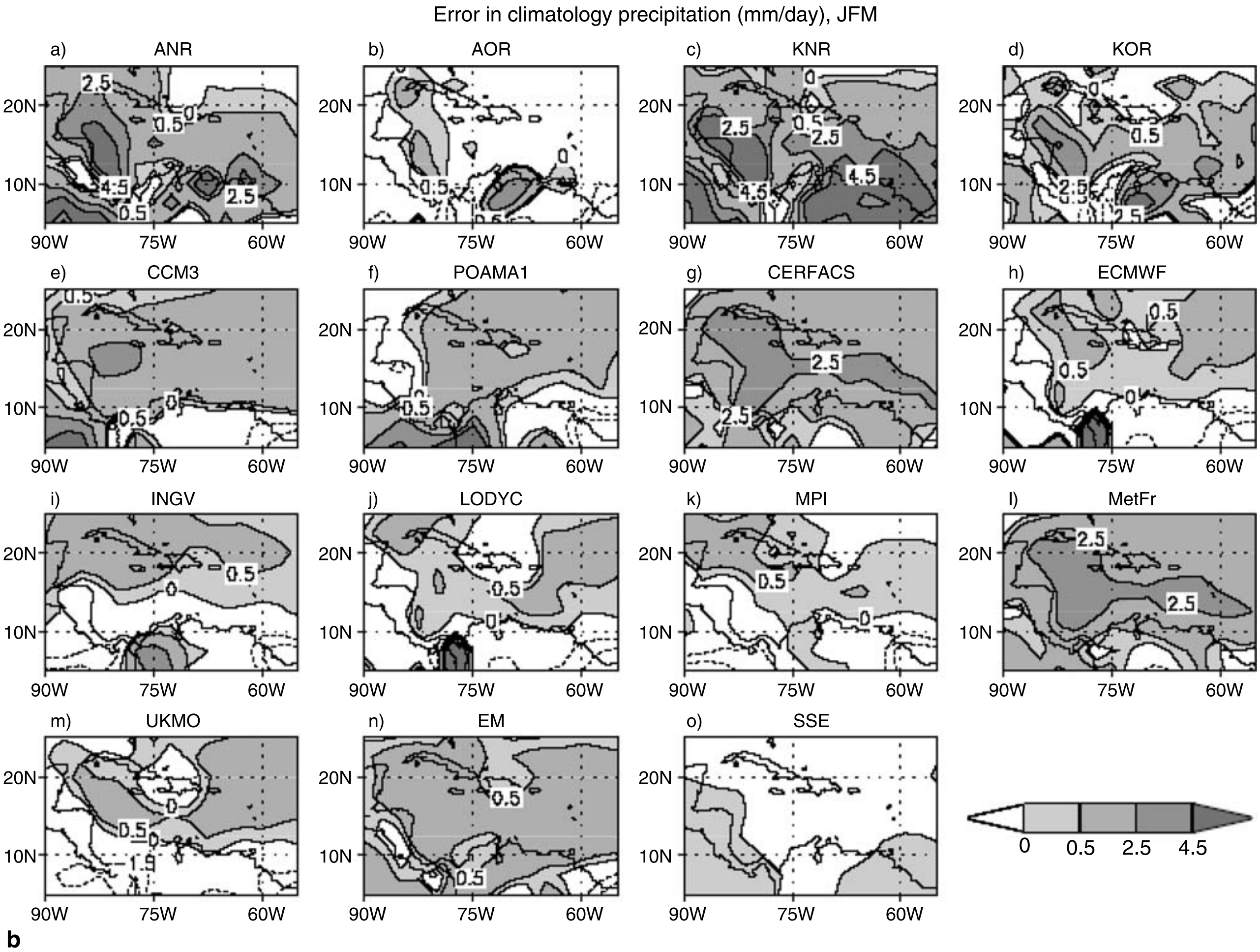


Fig. 3 (*continued*)

ously for January–March. To get the September–November seasonal forecasts, the year was divided into monthly blocks of March–May, June–August, September–November, and December–February. Again, 13 of the 52 seasonal forecasts generated for each model were for September–November. Finally, to get the May–June "seasonal" forecasts, the year was divided into 6 monthly blocks (January–February, March–April, May–June, July–August, September–October, November–December). This yielded 78 "seasonal" forecasts for each of the 13 models (6 "seasons" $\times$ 13 years) of which 13 of these forecasts for each model were obviously for May–June. FSUGCM models, as well as the POAMA1 and CCM3 models, were initialized on the first day of each month of the year and the forecasts were carried out to 3 months. For these models the first month's forecast following the initialization was used in constructing the seasonal forecasts; hence, there was no "lead time" in these forecasts. DEMETER models were initialized on the first day of November, February, May, and August and the forecasts were carried out to 6 months. The first month's forecast following the initialization was not used; hence, the seasonal forecasts from these models are termed "one month lead time" forecasts. For the DEMETER models the January–March seasonal forecasts are averages of months 3 and 4 from the November initialization and month 2 from the February initialization. The May–June seasonal forecasts are averages of month 4 from the February initialization and month 2 from the May initialization. The September–November seasonal forecasts are averages of months 2, 3, and 4 from the August initialization. The observational data sets for precipitation are from CMAP (Xie and Arkin, 1997) while all other variables are from the ECMWF reanalysis, except for sea-surface temperature (SST), which is obtained according to the technique described by Reynolds et al (2002). All data sets were interpolated to a common $2.5^\circ \times 2.5^\circ$ horizontal grid prior to the construction of the FSUSSE forecasts. The Caribbean domain is defined as the region 90° W–55° W and 5° N–25° N.

In order to construct the FSUSSE forecasts for a particular year, a cross validation technique developed by Deque (1997) was used. The year being forecast (divided into 4 seasons to treat the September–November and January–March periods in this study, and divided into 6 "seasons" to treat the May–June period in this study) was excluded from the data set, and the weights of the FSUSSE were computed based on the performance of the models during the rest of the years, defined as the training period. These weights were used with the individual models in the ensemble to construct the superensemble forecasts by Eq. (1).

4. Model seasonal climatology for 1989–2001

Before attempting to predict the seasonal climate for a particular year using the CGCMs, it is important to determine how well such models perform in generating the seasonal climatology over the 13-year period of the study. Sperber and Palmer (1996) found that predictions of seasonal climate and its interannual variability in a CGCM are better if the model's long-term climatology is close to observation. Krishnamurti et al (2000b) showed that models with a better rainfall

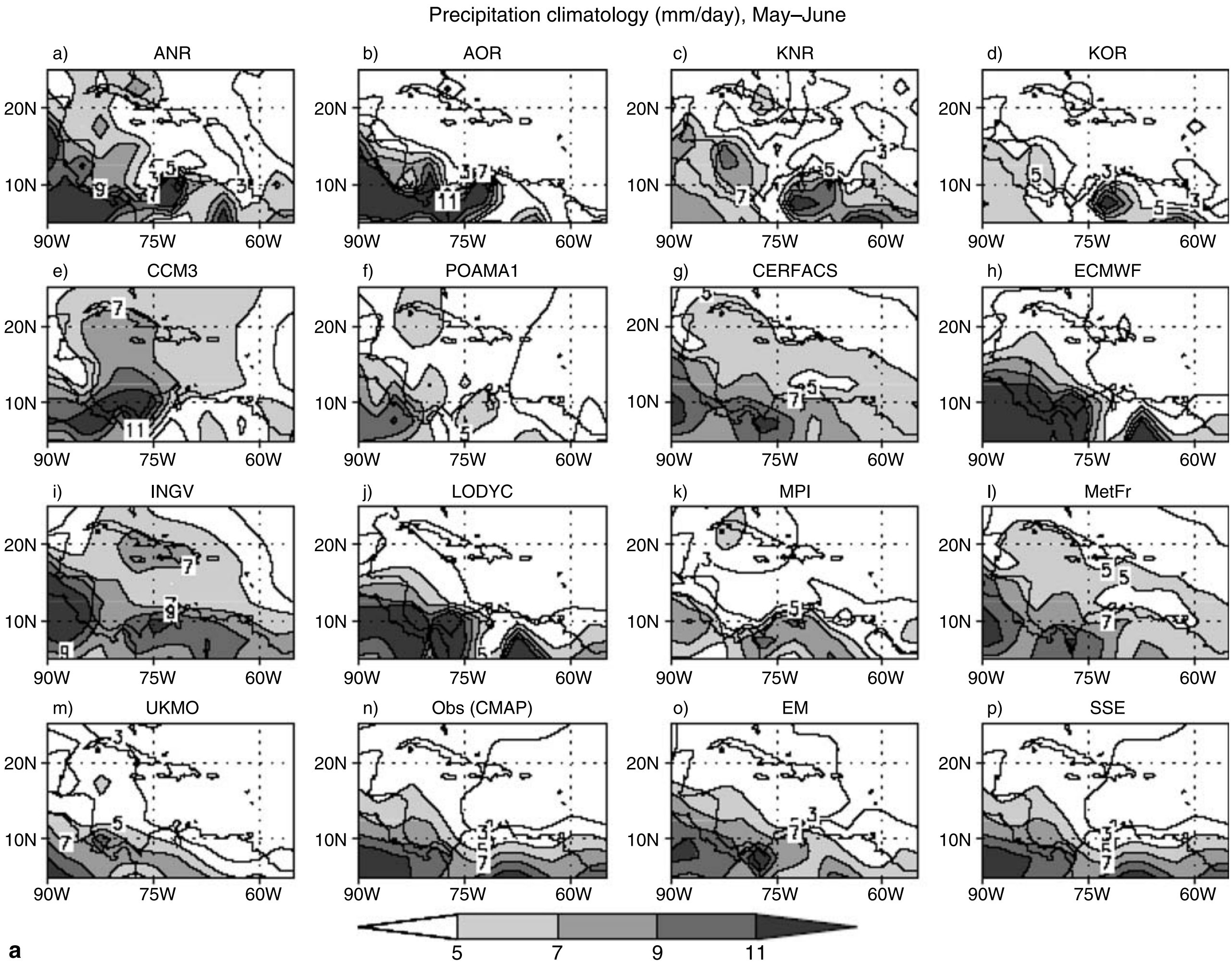


Fig. 4a. Precipitation climatology (mm/day) for May–June 1989–2001 for each of the 13 models in the multi-model ensemble, the multi-model ensemble mean (EM), and the FSU Synthetic Superensemble (SSE). The observed precipitation climatology is from CMAP data (Xie and Arkin, 1997). (**b**) Error in the precipitation climatology (mm/day) based on a comparison of the model climatology to the CMAP observed precipitation climatology shown in (**a**)

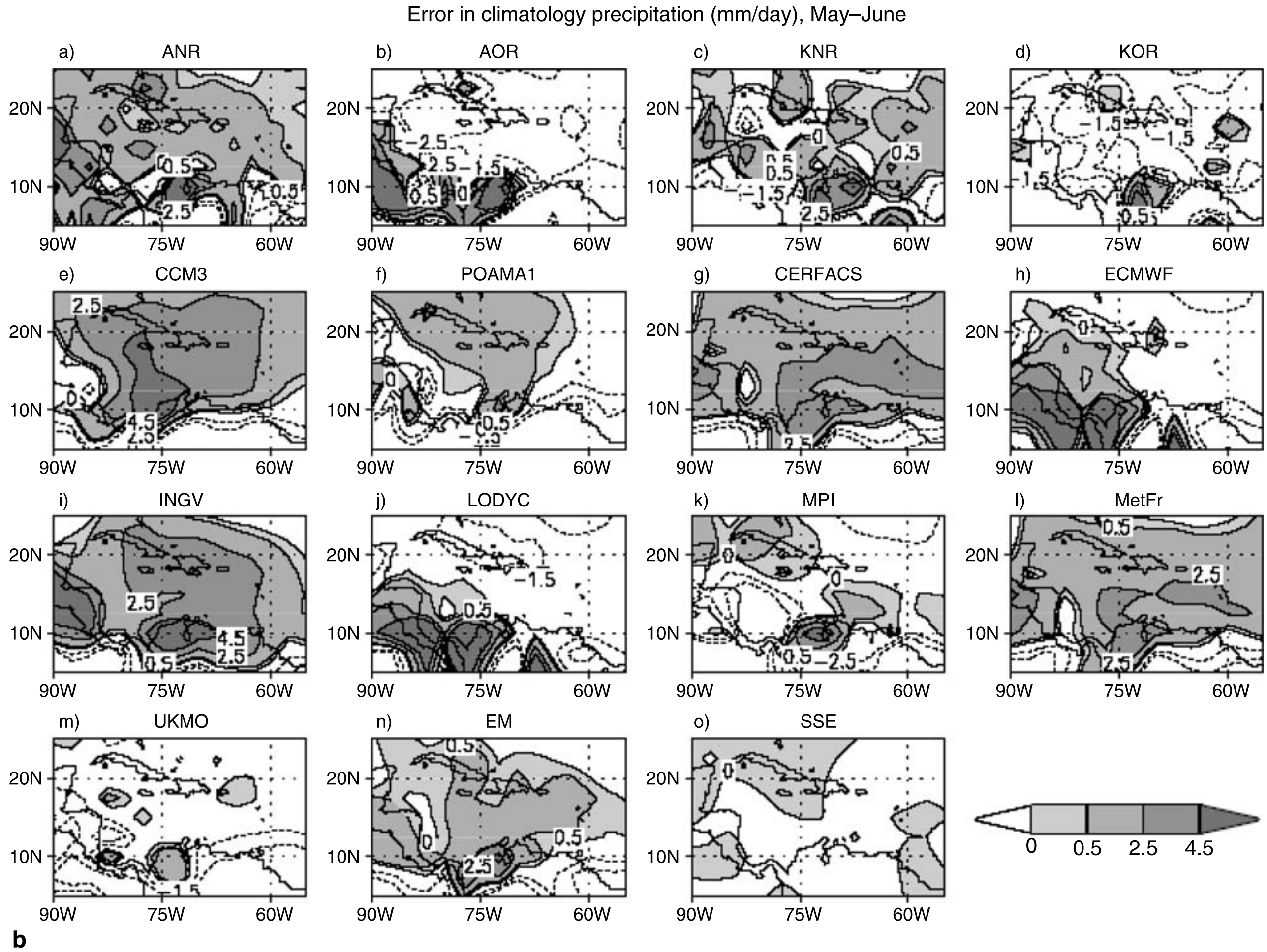


Fig. 4 (*continued*)

climatology generally have less systematic errors in rainfall predictions and better ability to simulate interannual variations in rainfall. Before examining seasonal climate forecasts for individual years in the Caribbean area, we will first examine the various models' success in simulating the observed 13-year seasonal climatology with regard to precipitation, sea surface temperature (SST), 2-meter temperature, and wind at the 850 hPa level. Individual model, multi-model ensemble mean, and FSUSSE climatologies will be considered. The seasons considered are the dry season (JFM), the early peak of the summer wet season (MJ), and the late peak of the summer wet season (SON).

4.1 Precipitation

The precipitation climatology of all the models for January–March (the Caribbean dry season) based on the period 1989–2001, along with the observed climatology from the data of Xie and Arkin (1997), is shown in Fig. 3a. Figure 3b shows the "error" in each model's climatology based on a comparison to the observed Xie and Arkin climatology. The observed map in Fig. 3a (n) shows the very dry conditions over the Caribbean Sea, with precipitation largely limited to the region south of 10° N in association with the Intertropical Convergence Zone (ITCZ). These figures show that all models overestimate the precipitation during the Caribbean dry season in comparison to observation, especially models ANR, KNR, KOR, CERFACS, and MetFr. The multi-model ensemble mean (Fig. 3a (o) and 3b (n)) also overestimates the precipitation, but generally to a lesser extent than individual models in the ensemble. The FSUSSE represents the dry season climatology extremely well in comparison to observation (Fig. 3a (p) and 3b (o)), suggesting that the dry season rainfall forecast for a given year obtained from the FSUSSE might be more reliable than for any other model in the ensemble and their ensemble mean, based on the findings of Sperber and Palmer (1996), and Krishnamurti et al (2000b) cited above.

The precipitation climatology for May–June (Caribbean summer wet season, early peak) is shown in Fig. 4a, with the "error" in the various models' climatology depicted in Fig. 4b. The observed map in Fig. 4a (n) reveals an approximate three-fold increase in precipitation over the Caribbean Sea in comparison to the January–March climatology (Fig. 3a (n)), along with a doubling of precipitation south of 10° N and a northward migration of the region of precipitation into Central America in comparison to the January–March climatology. Comparison of Fig. 4a and 4b shows that a number of models are too wet (e.g., ANR, CERFACS, INGV) and a number are too dry (e.g., KOR, POAMA1) in comparison to observation. Of the individual models the UKMO generally has the best simulation of the observed climatology (Fig. 4a (m) and 4b (m)). The multi-model ensemble mean also reproduces the observed climatology well, although values are too high over the Caribbean Sea and too low over the South American continent near the equator (Fig. 4a (o) and 4b (n)). The FSUSSE represents the early peak in the summer wet season rainfall climatology extremely well in comparison to observation (Fig. 4a (p)

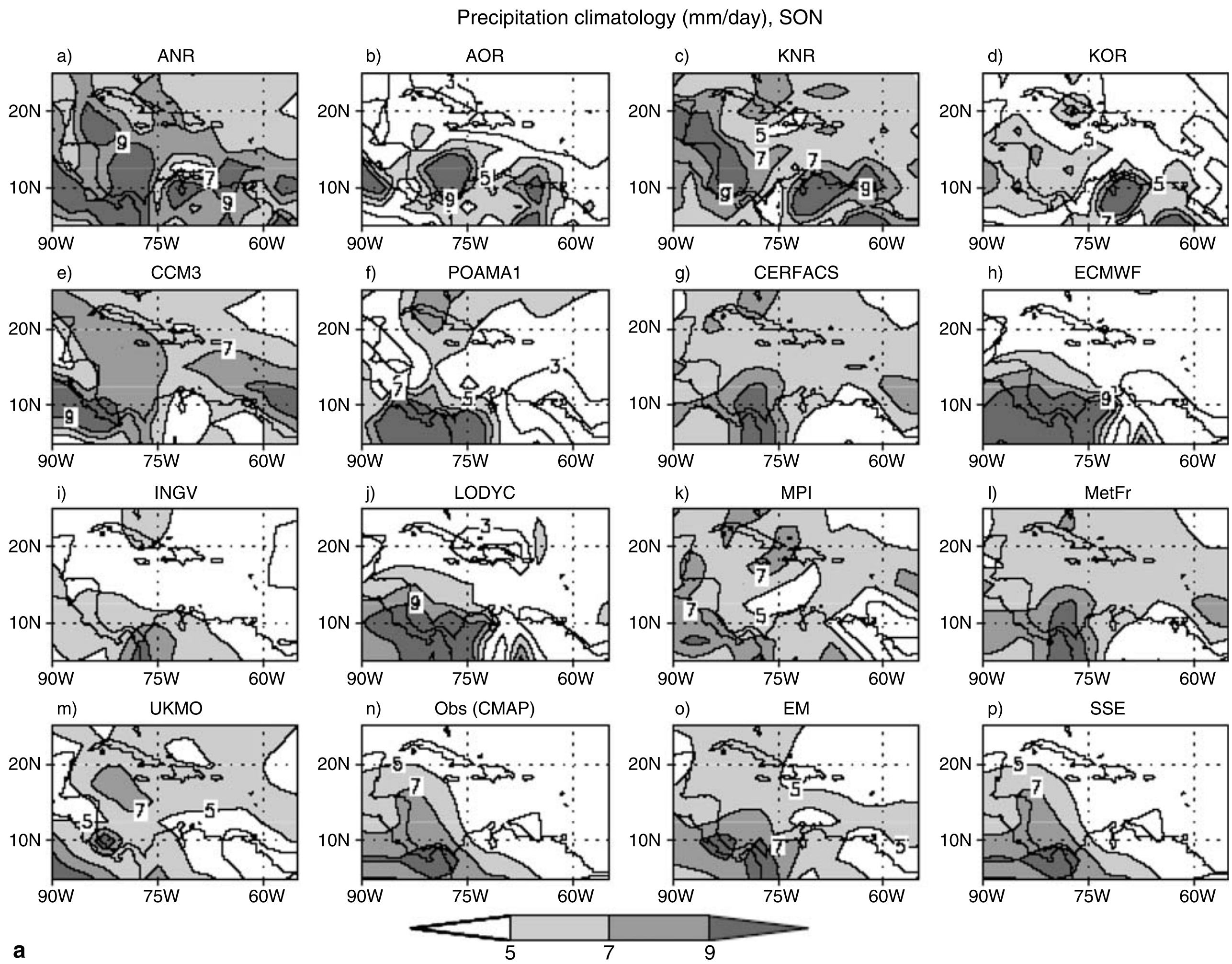


Fig. 5a. Precipitation climatology (mm/day) for September–November 1989–2001 for each of the 13 models in the multi-model ensemble, the multi-model ensemble mean (EM), and the FSU Synthetic Superensemble (SSE). The observed precipitation climatology is from CMAP data (Xie and Arkin, 1997). (**b**) Error in the precipitation climatology (mm/day) based on a comparison of the model climatology to the CMAP observed precipitation climatology shown in (**a**)

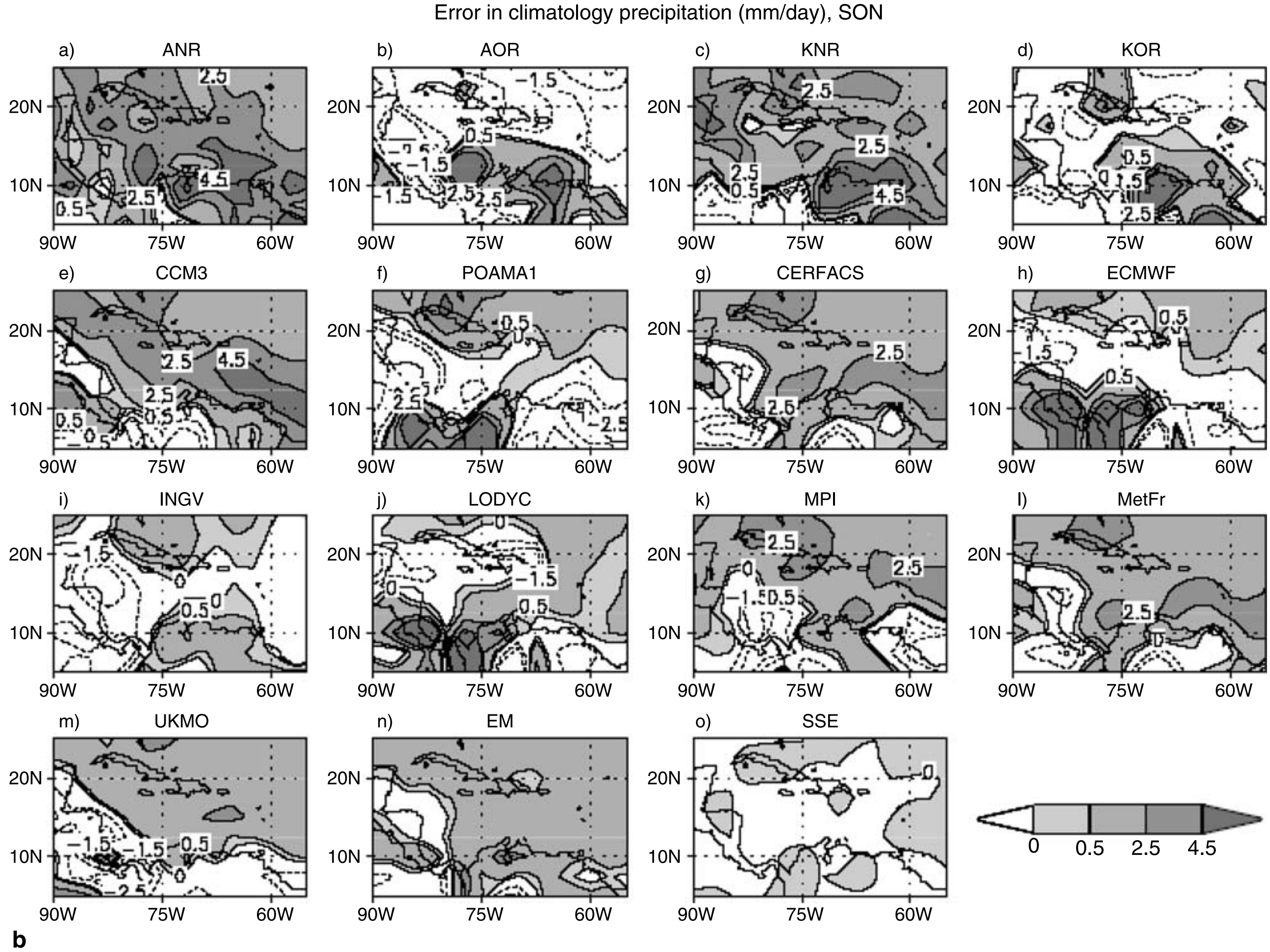


Fig. 5 (*continued*)

and 4b (o)), suggesting that the May–June seasonal rainfall forecast for a given year obtained from the FSUSSE might be more reliable than for any other model in the ensemble and their ensemble mean, just as we stated above with regard to the dry season rainfall forecasts.

The precipitation climatology for the late peak of the summer wet season (September–November) is shown in Fig. 5a, with the "error" in the various models' climatology depicted in Fig. 5b. The observed map in Fig. 5a (n) shows that the precipitation amounts have increased slightly over the Caribbean Sea, while amounts have decreased somewhat over northeastern South America, with farther northward migration of the precipitation region over Central America, all in comparison to the May–June climatology. Just as seen in the May–June climatology, some models are too wet in comparison to observation, e.g., ANR, KNR, CCM3, while one model in particular, INGV, is too dry. Among individual models in the ensemble, the ECMWF model probably comes closest to duplicating the observed precipitation climatology. The multi-model ensemble mean is generally too wet, particularly over the eastern portion of the Caribbean region (Fig. 5a (o) and 5b (n)). As with the dry season and the early peak of the summer wet season previously discussed, the FSUSSE represents the precipitation climatology of the late peak of the summer wet season extremely well, in terms of overall pattern (Fig. 5a (p)) and magnitude of error (Fig. 5b (o)). This suggests that the September–November seasonal rainfall forecast for a given year obtained from the FSUSSE might be more reliable than any other model in the ensemble and their ensemble mean, based on the results of Sperber and Palmer (1996), and Krishnamurti et al (2000b) discussed above.

4.2 Sea-surface temperature (SST)

The SST climatology for all models for the Caribbean dry season of January–March based on the period 1989–2001, together with the observed climatology from Reynolds et al (2002), is shown in Fig. 6a. Figure 6b shows the "error" in each model's SST climatology based on a comparison to the observed climatology. The observed map in Fig. 6a (n) shows an axis of warm ocean water extending east–west across the Caribbean with the warmest water in excess of 300 K located in the western Caribbean. Several individual model climatologies capture this general pattern (e.g., AOR, CERFACS, INGV, MetFr, UKMO). The multi-model ensemble mean (Fig. 6a (o)) depicts the general pattern but underestimates the area extent of the western Caribbean SST maximum. The FSUSSE shows an SST climatology (Fig. 6a (p)) that is in very close agreement with the observed climatology (Fig. 6a (n)). The error maps in Fig. 6b show that individual models may have SST climatologies that depart by as much as 3–4 K from the observed values. The RMS error in the SST climatology for each model, as compared to the observed climatology, is shown in each panel of Fig. 6b. Individual model values range from 0.99 (INGV) to 2.78 (POAMA1). The multi-model ensemble mean has an RMS error of 1.05. The errors in the FSUSSE climatology are all less than 0.5 K, and the low RMS error value of

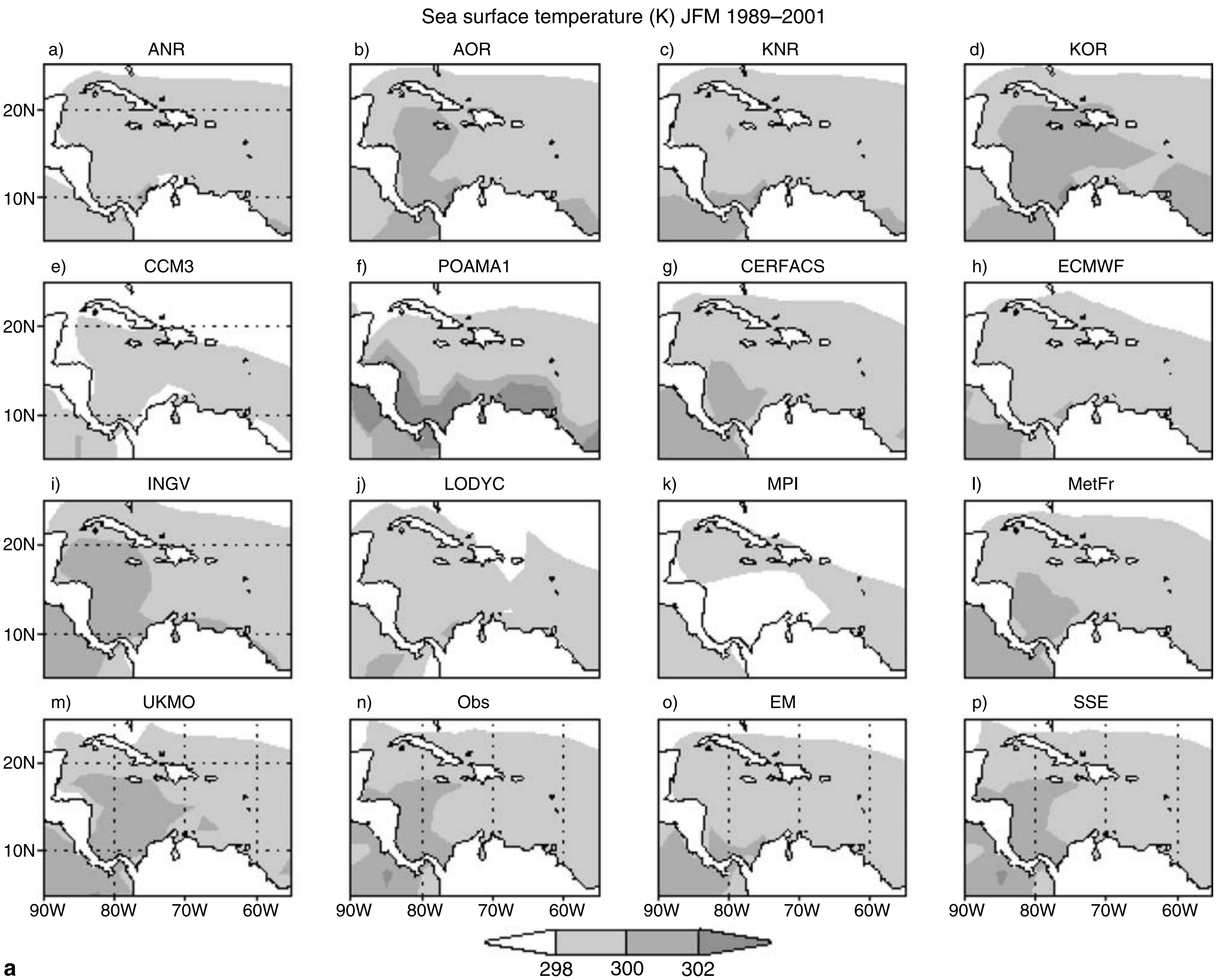


Fig. 6a. Sea-surface temperature climatology (K) for January–March 1989–2001 for each of the 13 models in the multi-model ensemble, the multi-model ensemble mean (EM), and the FSU Synthetic Superensemble (SSE). The observed sea-surface temperature climatology is from Reynolds et al (2002). (**b**) Error in the sea-surface temperature climatology (K) based on a comparison of the model climatology to the observed sea-surface temperature climatology from Reynolds et al (2002)

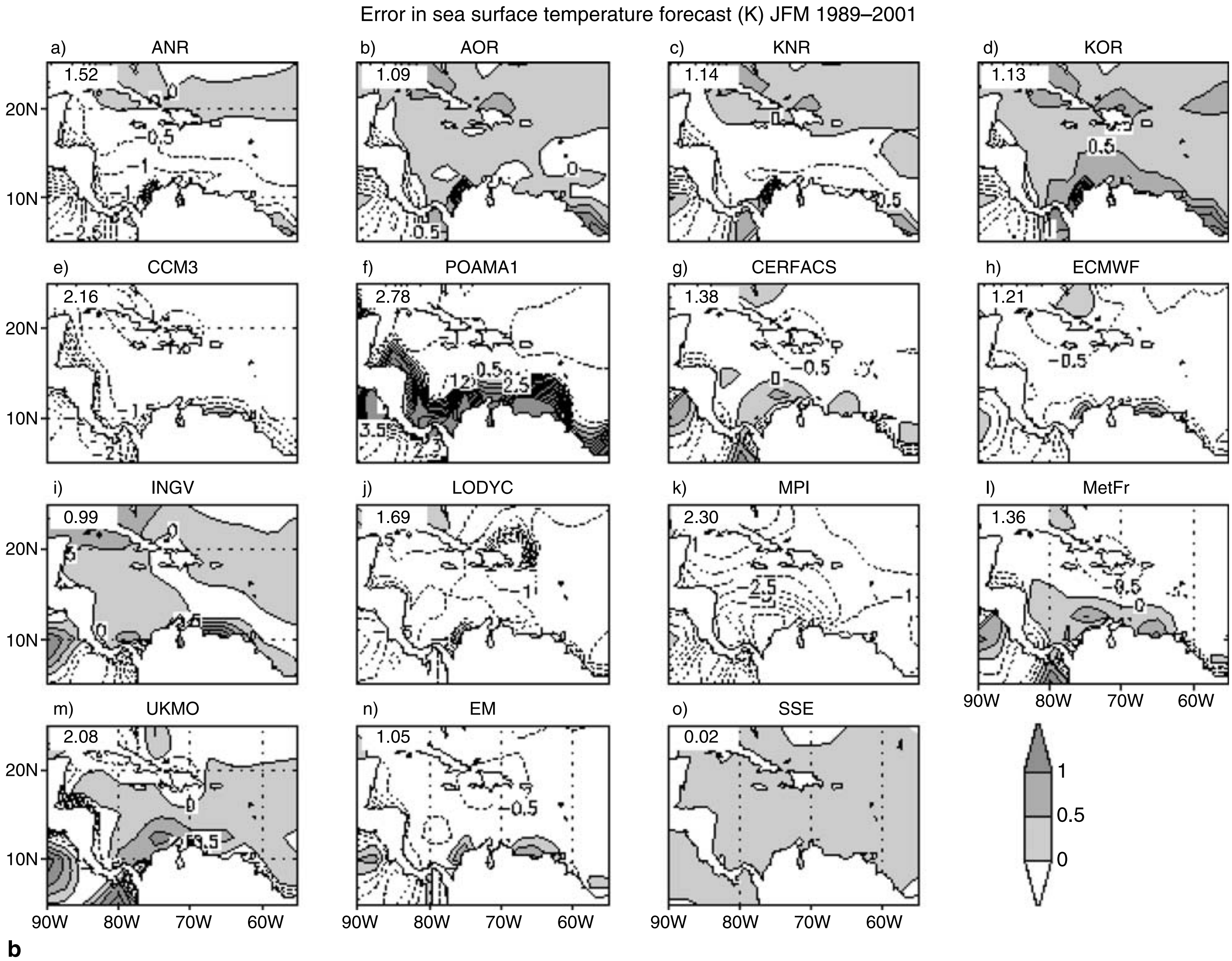


Fig. 6 (*continued*)

0.02 is quite impressive. Based on the depictions of Fig. 6, the FSUSSE clearly has by far the best representation of the SST climatology for the Caribbean dry season. The FSUSSE predictions of SST for the dry season of a given year would be expected to have a higher probability of verification than the predictions of other models in the ensemble and their ensemble mean, by the same line of reasoning as was stated for rainfall predictions in Sect. 4a above.

The SST climatology for May–June, the early peak of the Caribbean wet season, is shown in Fig. 7a, while the "error" in the various models' climatology is depicted in Fig. 7b. The observed SST climatology in Fig. 7a (n) shows that the Caribbean SSTs have warmed by about 2 K over the January–March period, and the region of maximum SST in the western Caribbean has expanded. In Fig. 7a, only the FSUSSE has an excellent representation of the observed climatology. This is born out by the RMS errors seen in Fig. 7b, where the FSUSSE has an RMS error of only 0.01. The multi-model ensemble mean is considerably worse in its prediction, with an RMS error of 1.36. Of the individual models the best model is the KOR (RMS error of 0.88), and the worst model is the POAMA1 (RMS error of 2.62). With an excellent SST climatology, the FSUSSE would have the highest probability of yielding the best SST forecast for the May–June period over the Caribbean for a particular year.

The results of the SST climatology predictions for the late peak of the summer wet season in the Caribbean (September–November) are shown in Fig. 8a and b. Figure 8a shows that warm SSTs cover the entire Caribbean with a maximum in excess of 302 K in the northwestern Caribbean.

As with the previous seasons, the FSUSSE has a near perfect representation of the observed climatology. Figure 8b demonstrates that the errors in the FSUSSE forecast are all less than 0.5 K and that the RMS errors are a miniscule 0.01 K. The multi-model ensemble mean does not perform well with an RMS error value of 1.46. The best individual model performance is by the ANR (RMS error of 1.33), while the worst performance is by the CCM3 (RMS error of 3.11). Based on these results, the FSUSSE would be expected to have the highest probability of producing the best SST forecast for the September–October period over the Caribbean in comparison to the individual models comprising the ensemble and their ensemble mean.

4.3 2-meter temperature

The climatology of air temperature at 2 meters from the 7 DEMETER models plus the POAMA1 model for the Caribbean dry season of January–March based on the period 1989–2001, along with the observed climatology from the ECMWF reanalysis, is depicted in Fig. 9a. Figure 9b shows the "error" in each model's climatology based on a comparison to the climatology from the ECMWF reanalysis. Note that 2-meter temperature forecasts were not available from the suite of 4 FSU models (KOR, KNR, AOR, ANR) and the CCM3 model. The observed climatology in Fig. 9a (i) shows temperatures in excess of 298 K across the Caribbean, while the warmest

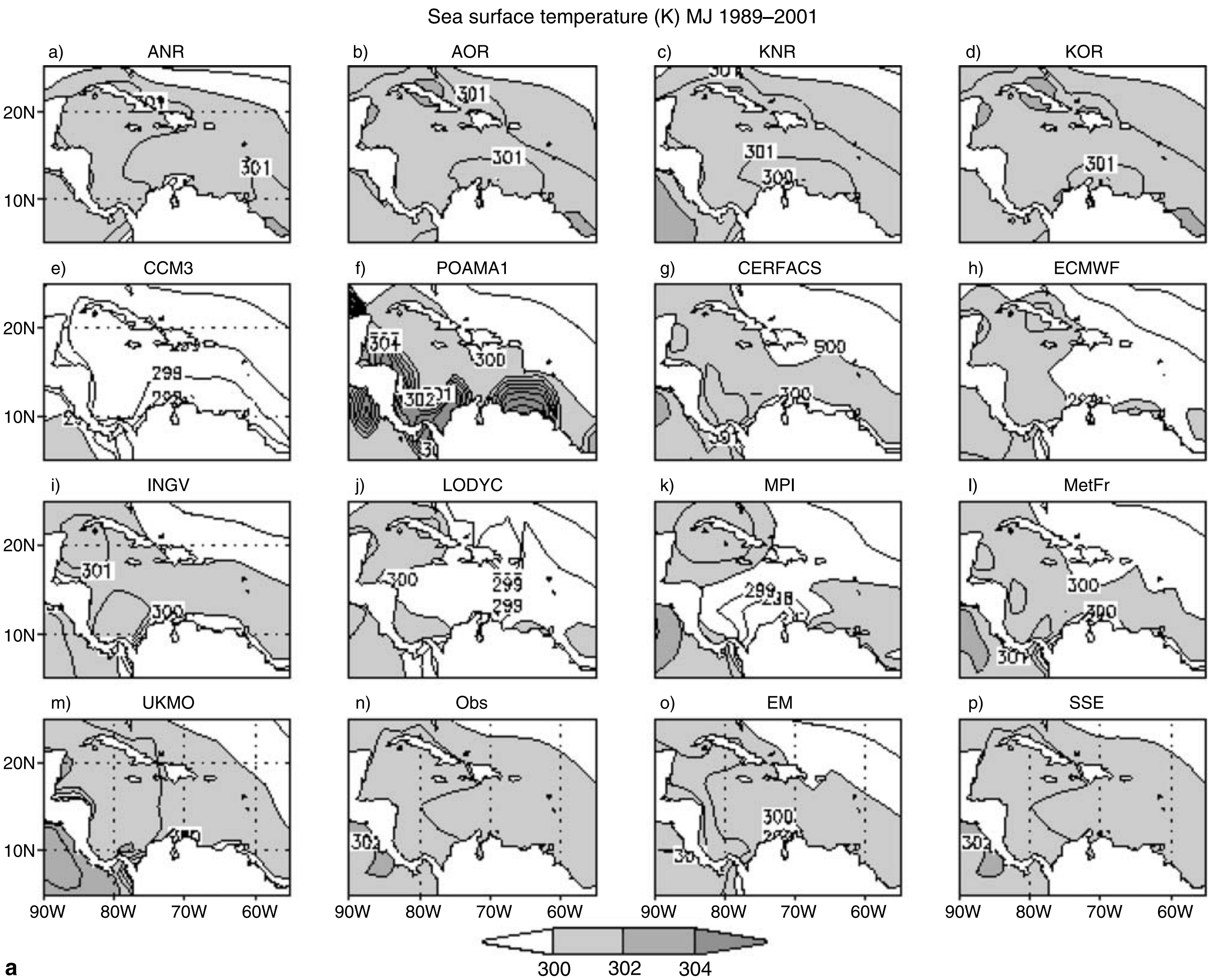


Fig. 7a. Sea-surface temperature climatology (K) for May–June 1989–2001 for each of the 13 models in the multi-model ensemble, the multi-model ensemble mean (EM), and the FSU Synthetic Superensemble (SSE). The observed sea-surface temperature climatology is from Reynolds et al (2002). (**b**) Error in the sea-surface temperature climatology (K) based on a comparison of the model climatology to the observed sea-surface temperature climatology from Reynolds et al (2002)

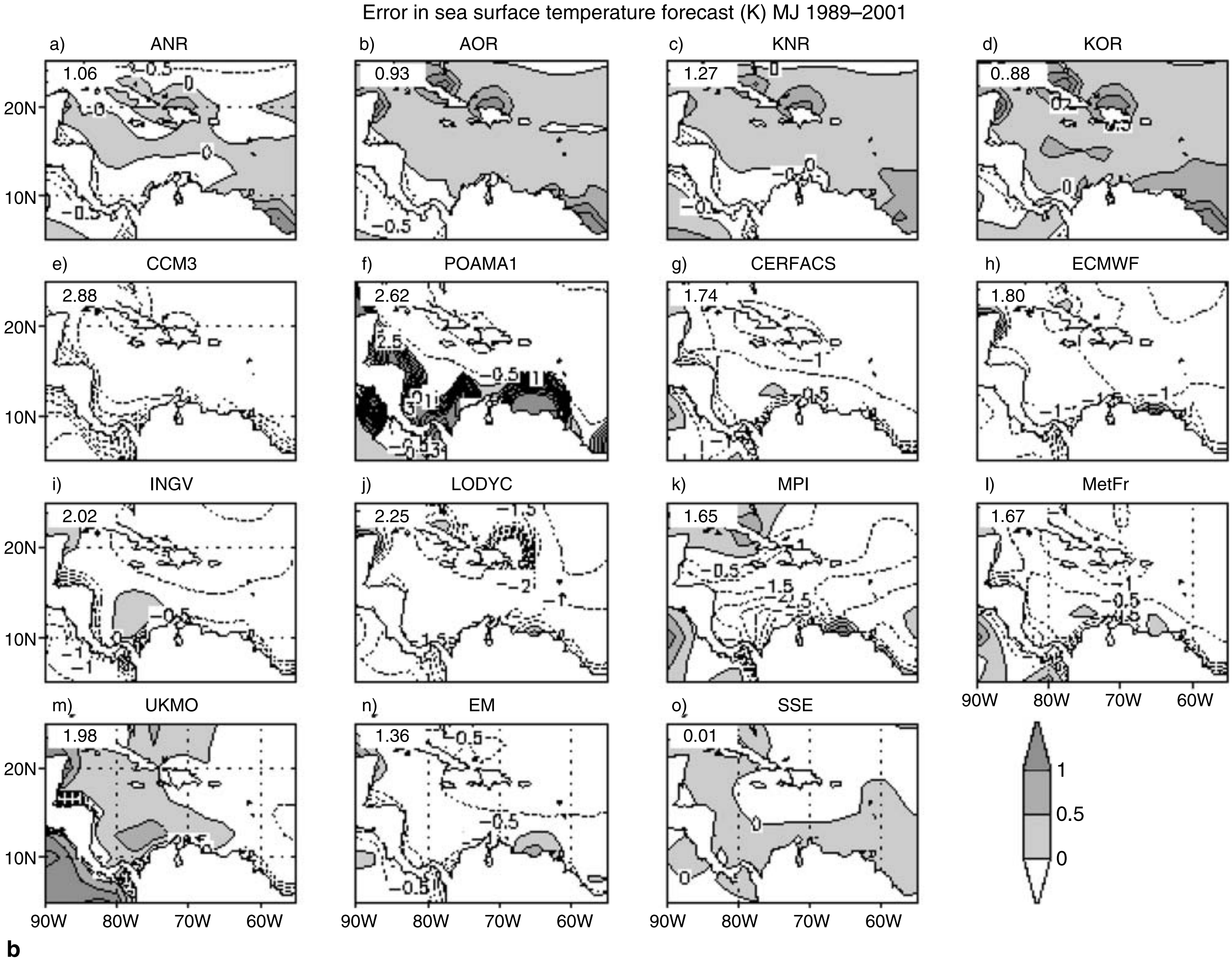


Fig. 7 (*continued*)

temperatures in excess of 300 K are found in the eastern Pacific and in a small region of northern Brazil. A number of models capture this general pattern (e.g., CERFACS, MeteoFrance, UKMO), while POAMA1 is too warm along 10° N (Fig. 9a (a) and 9b (a)) and MPI is too cold throughout most of the region (Fig. 9a (f) and 9b (f)). Both the ensemble mean and the FSUSSE capture the general temperature pattern well, as seen in Fig. 9a, but Fig. 9b reveals that the FSUSSE has the lowest RMS error (0.04) in comparison to the ensemble mean (RMS error of 0.82) and all the models comprising the ensemble mean (RMS errors ranging from 0.71 to 2.32). This excellent performance by the FSUSSE in duplicating the observed 2 meter temperature climatology raises confidence in the model's ability to provide the best forecast of this variable for a particular year's dry season in comparison to all the other models and to the ensemble mean of those models.

The observed 2 meter temperature climatology in Fig. 10a (i) shows that by the early peak of the Caribbean summer wet season in May–June, air temperatures have warmed over the entire Caribbean to values in excess of 300 K. The only individual model to come close to duplicating this climatology is the UKMO (Fig. 10a (h)). The ensemble mean (Fig. 10a (j)) also does a credible job with the 2 meter temperature climatology but is too cold in parts of the eastern Caribbean. The FSUSSE (Fig. 10a (k)) does an excellent job in duplicating the temperature climatology, a fact that is quantified in Fig. 10b (j) where this model's errors are all seen to be less that 1 K, with an RMS error of only 0.02. In this figure, other individual models are seen to have much larger positive and negative errors, and the ensemble mean is seen to have larger errors over Brazil. Outside of the FSUSSE the better models are the INGV (RMS error of 0.87) and the CERFACS, UKMO, and ensemble mean (all with RMS errors of 0.90). The worst performance is by the LODYC, which is too cold throughout and with an RMS error of 1.62 (see Fig. 10a (e) and 10b (e)). The superior performance of the FSUSSE again provides confidence that this model will have the greatest probability of giving the best forecast of the May–June 2 meter temperatures for a particular year across the Caribbean in comparison to the ensemble mean and the models comprising that mean.

The FSUSSE is also seen to have the best representation of the 2 meter temperature for the late peak in the Caribbean summer wet season (September–November) in Fig. 11a and b. Generally the individual models seem to have their best performance during this period, as compared to the periods of January–March and May–June. CERFACS, ECMWF, INGV, Meteo-France, and UKMO all reproduce the climatology well in comparison to the observed. The ensemble mean also has a good representation of the climatology (see Fig. 11a (j) and 11b (i)). But again, the FSUSSE has the best 2 meter temperature representation with errors of less than 1 K and an RMS error of only 0.02 (see Fig. 11a (k) and 11b (j)). This model would be expected to produce the best 2-meter temperature forecasts for September–November for a particu-

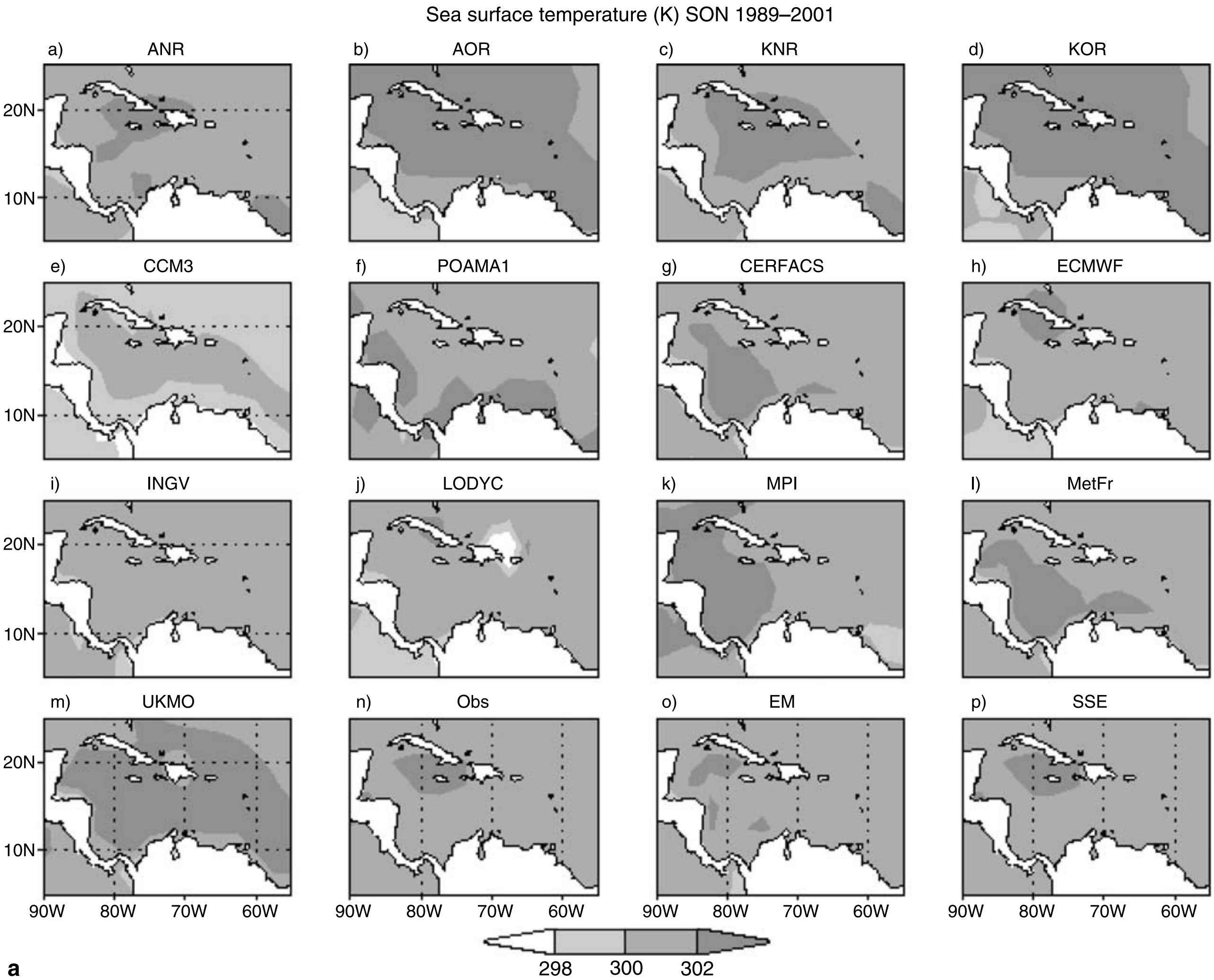


Fig. 8a. Sea-surface temperature climatology (K) for September–November 1989–2001 for each of the 13 models in the multi-model ensemble, the multi-model ensemble mean (EM), and the FSU Synthetic Superensemble (SSE). The observed sea-surface temperature climatology is from Reynolds et al (2002). (**b**) Error in the sea-surface temperature climatology (K) based on a comparison of the model climatology to the observed sea-surface temperature climatology from Reynolds et al (2002)

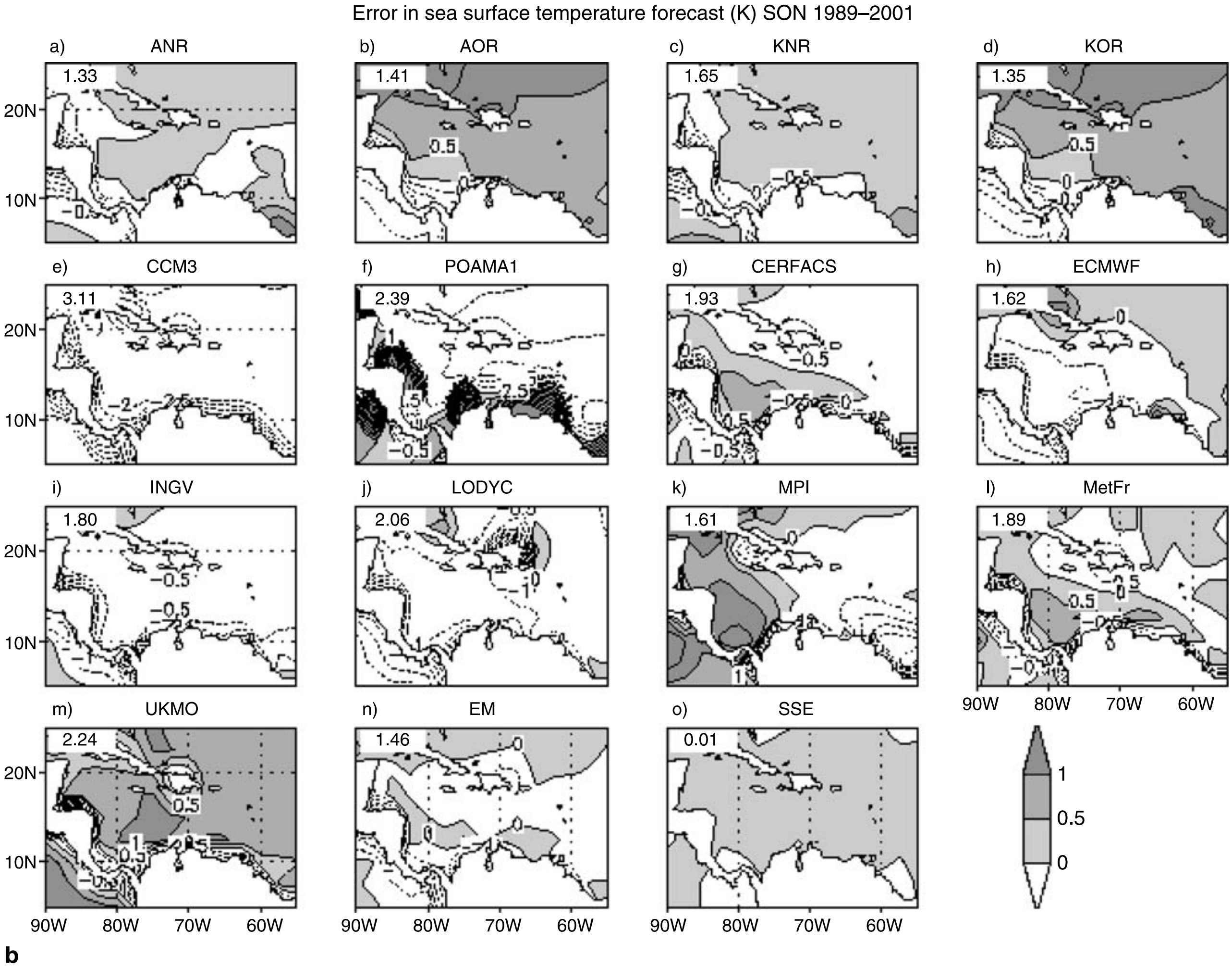
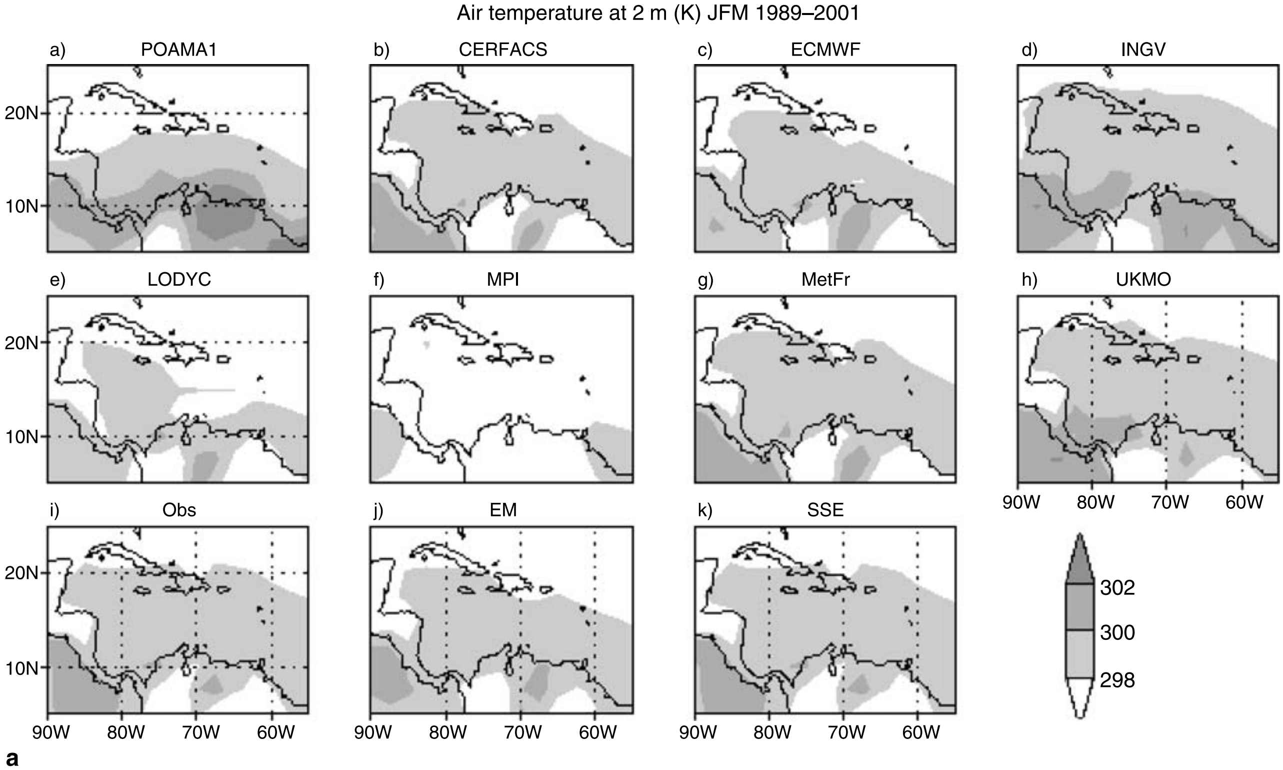


Fig. 8 (*continued*)

lar year in comparison to the ensemble mean and the models comprising that mean.

4.4 850 hPa wind

The climatology of 850 hPa wind for all the models, along with the observed climatology from the ECMWF reanalysis, is shown in Figs. 12 and 13. This climatology is presented for the *u*- and *v*-components of the wind, and for brevity only the climatology for September–November will be shown here. The September–November climatology is representative of the climatologies for January–March and May–June, as will be discussed. Figure 12a depicts the observed and forecast climatologies of the *u*-component of the wind at 850 hPa, while Fig. 12b shows the "error" in the forecast *u*-component climatology for each of the models. The observed climatology in Fig. 12a (n) shows easterly winds across the entire region with a maximum in wind speed of 10 m/s in the western Caribbean. Among individual models the ECMWF and LODYC (Fig. 12a (h) and (j)) provide good representations of the observed climatology. The ensemble mean (Fig. 12a (o)) underestimates the wind maximum, while the FSUSSE (Fig. 12a (p)) provides a near perfect representation of the observed climatology. The error maps in Fig. 12b confirm the qualitative inspection of Fig. 12a, with the FSUSSE having errors less than 1 m/s and an overall RMS error of 0.05. The good performance of the ECMWF and LODYC models is confirmed (RMS errors of 0.97 and 0.89, respectively). The ensemble mean is seen to underestimate the wind maximum by 2 m/s and has an overall RMS error of 1.11 (Fig. 12b (n)). The worst performance is by the POAMA1 model



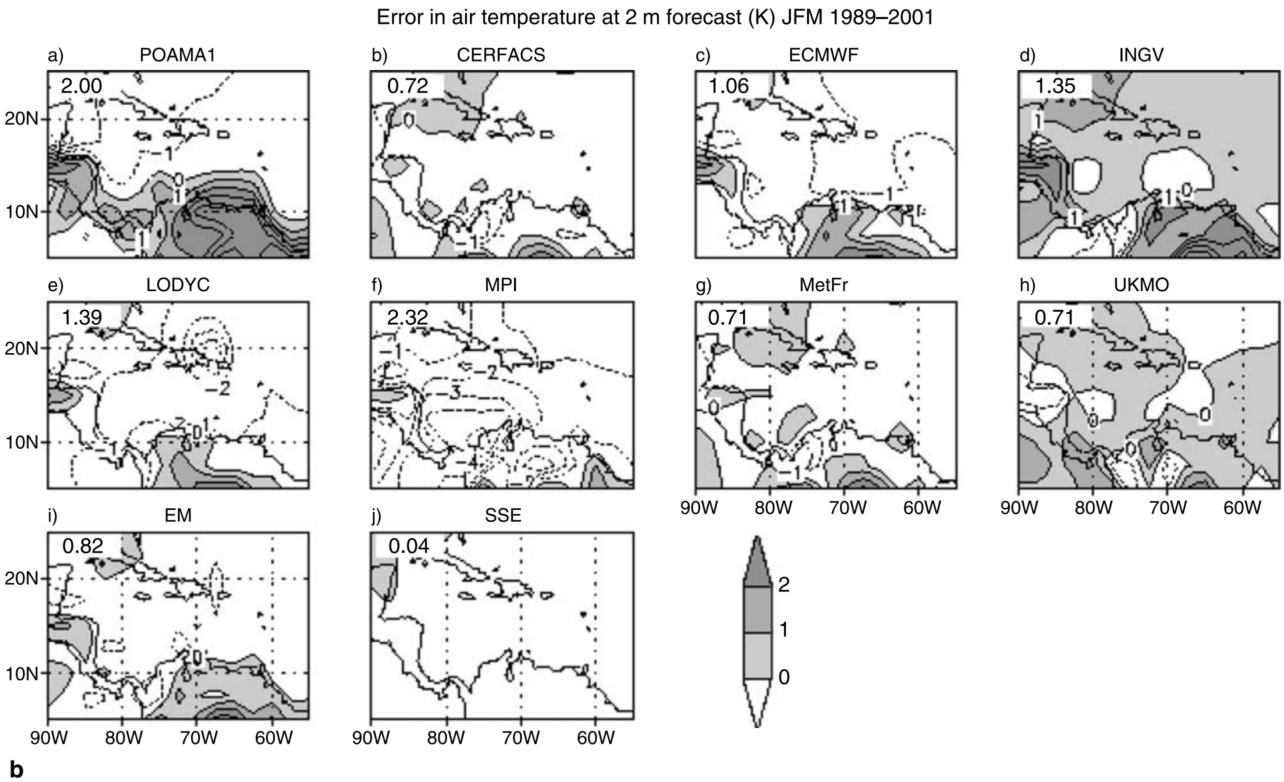


Fig. 9a. Climatology of 2 meter air temperature (K) for January–March 1989–2001 for each of the 8 models in the multi-model ensemble, the multi-model ensemble mean (EM), and the FSU Synthetic Superensemble (SSE). The observed 2-meter air temperature is from the ECMWF reanalysis. (**b**) Error in the 2-meter air temperature climatology (K) based on a comparison of the model climatology to the observed 2-meter air temperature climatology from the ECMWF reanalysis shown in (**a**)

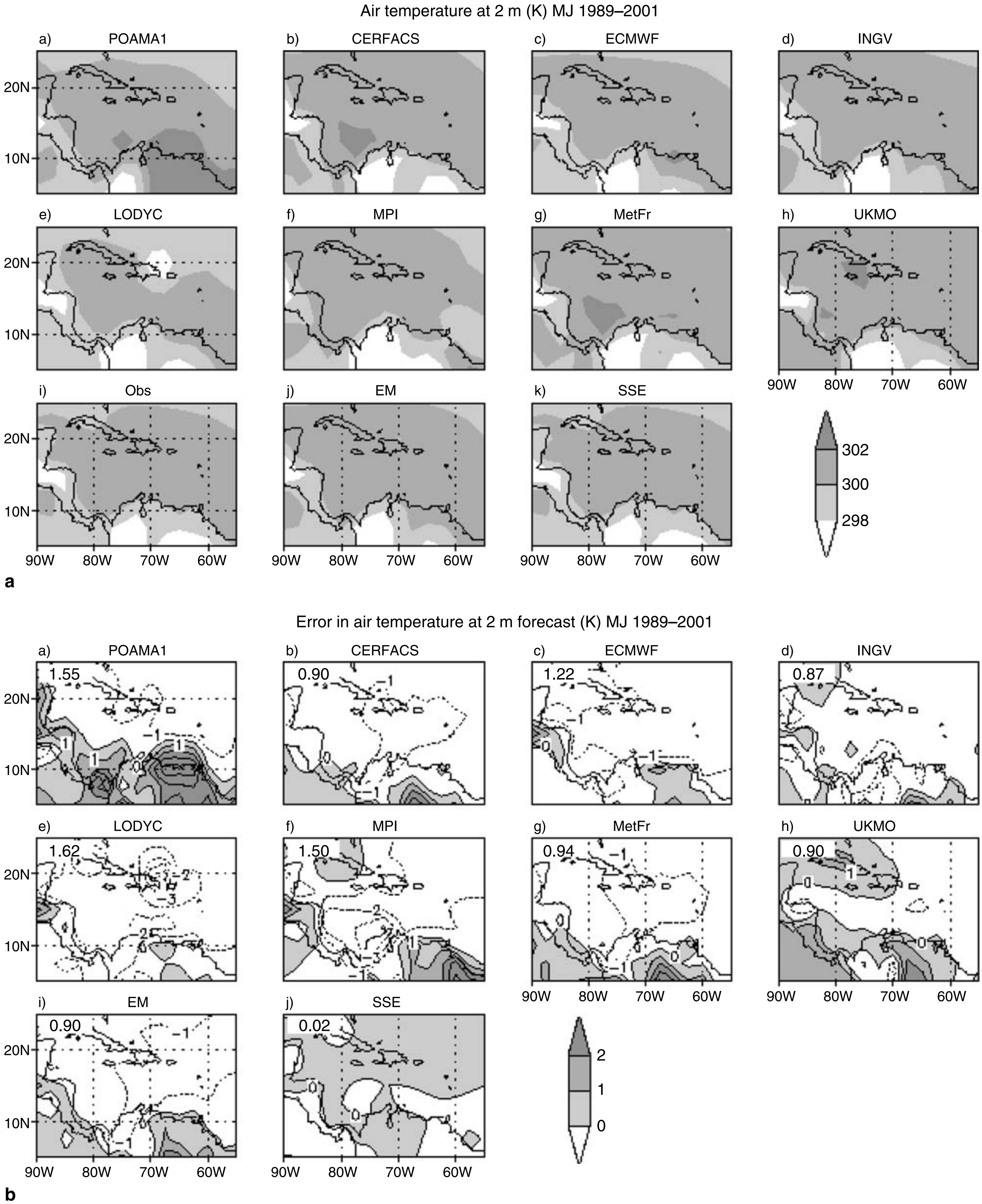


Fig. 10a. Climatology of 2-meter air temperature (K) for May–June 1989–2001 for each of the 8 models in the multi-model ensemble, the multi-model ensemble mean (EM), and the FSU Synthetic Superensemble (SSE). The observed 2 meter air temperature climatology is from the ECMWF reanalysis. (**b**) Error in the 2-meter air temperature climatology (K) based on a comparison of the model climatology to the observed 2-meter air temperature climatology from the ECMWF reanalysis shown in (**a**)

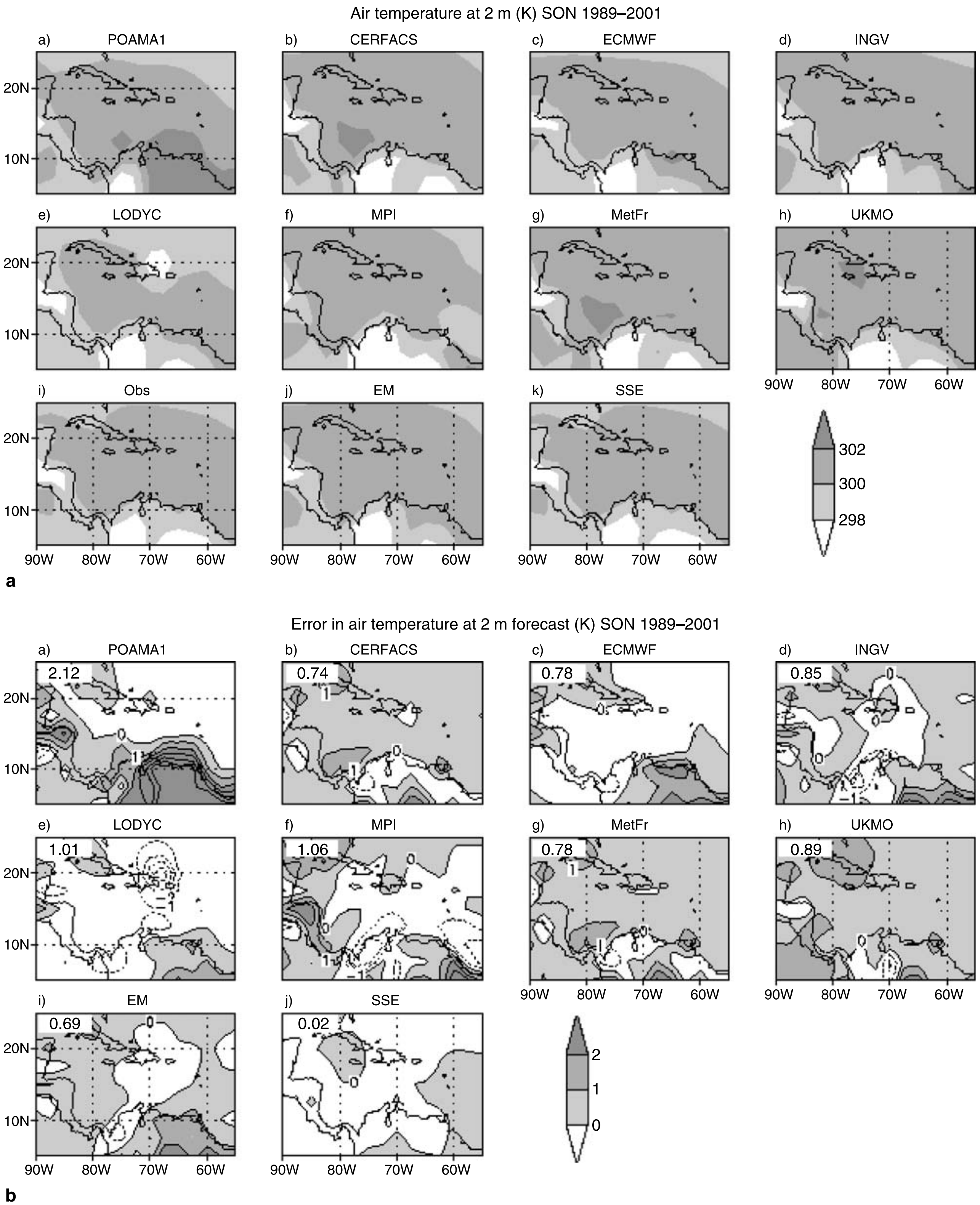


Fig. 11a. Climatology of 2 meter air temperature (K) for September–November 1989–2001 for each of the 8 models in the multi-model ensemble, the multi-model ensemble mean (EM), and the FSU Synthetic Superensemble (SSE). The observed 2-meter air temperature climatology is from the ECMWF reanalysis. (**b**) Error in the 2-meter air temperature climatology (K) based on a comparison of the model climatology to the observed 2-meter air temperature climatology from the ECMWF reanalysis shown in (**a**)

with an underestimate of the wind maximum of 5 m/s and an overall RMS error of 2.52.

The results of the u-component climatological depictions for January–March and May–June will not be shown, as stated. However, the results are quite similar to those for September–November. In January–March the FSUSSE has an overall RMS error of only 0.05, the ensemble mean has an RMS error of 1.08, and the average RMS error for the individual models is 1.85. In May–June the FSUSSE again has an overall RMS error of 0.05, the ensemble mean has an RMS error of 1.84, and the average RMS error for the individual models is 2.54.

Figure 13a depicts the observed and forecast climatologies of the v-component of the wind at 850 hPa for September–November, while Fig. 13b shows the "error" in the forecast v-component climatology of each of the models. The observed climatology in Fig. 13a (n) shows southerly winds over the eastern two-thirds of the Caribbean, with northerly winds in the extreme western Caribbean, a pattern that indicates a trough axis is present in the western Caribbean in the easterly flow (see Fig. 12a). Most of the individual models capture this essential pattern with varying degrees of success, with the ANR (Fig. 13a (a)) showing the best qualitative result

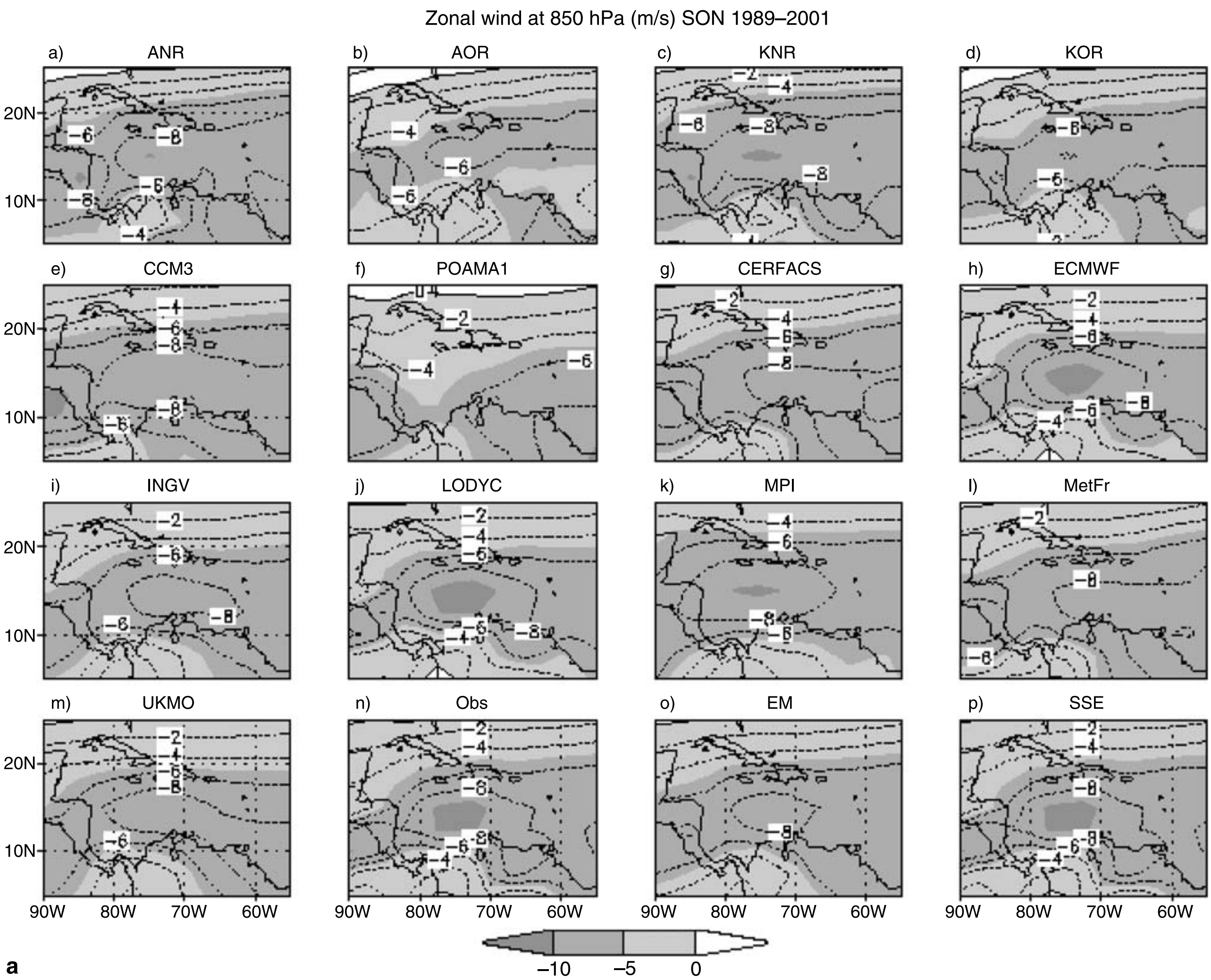


Fig. 12a. Climatology of the u-component of the wind at 850 hPa (m/s) for September–November 1989–2001 for each of the 13 models in the multi-model ensemble, the multi-model ensemble mean (EM), and the FSU Synthetic Superensemble (SSE). The observed climatology is from the ECMWF reanalysis. (**b**) Error in the climatology of the u-component of the wind at 850 hPa (m/s) based on a comparison of the model climatology to the observed climatology of the u-component of the 850 hPa wind from the ECMWF reanalysis shown in (**a**)

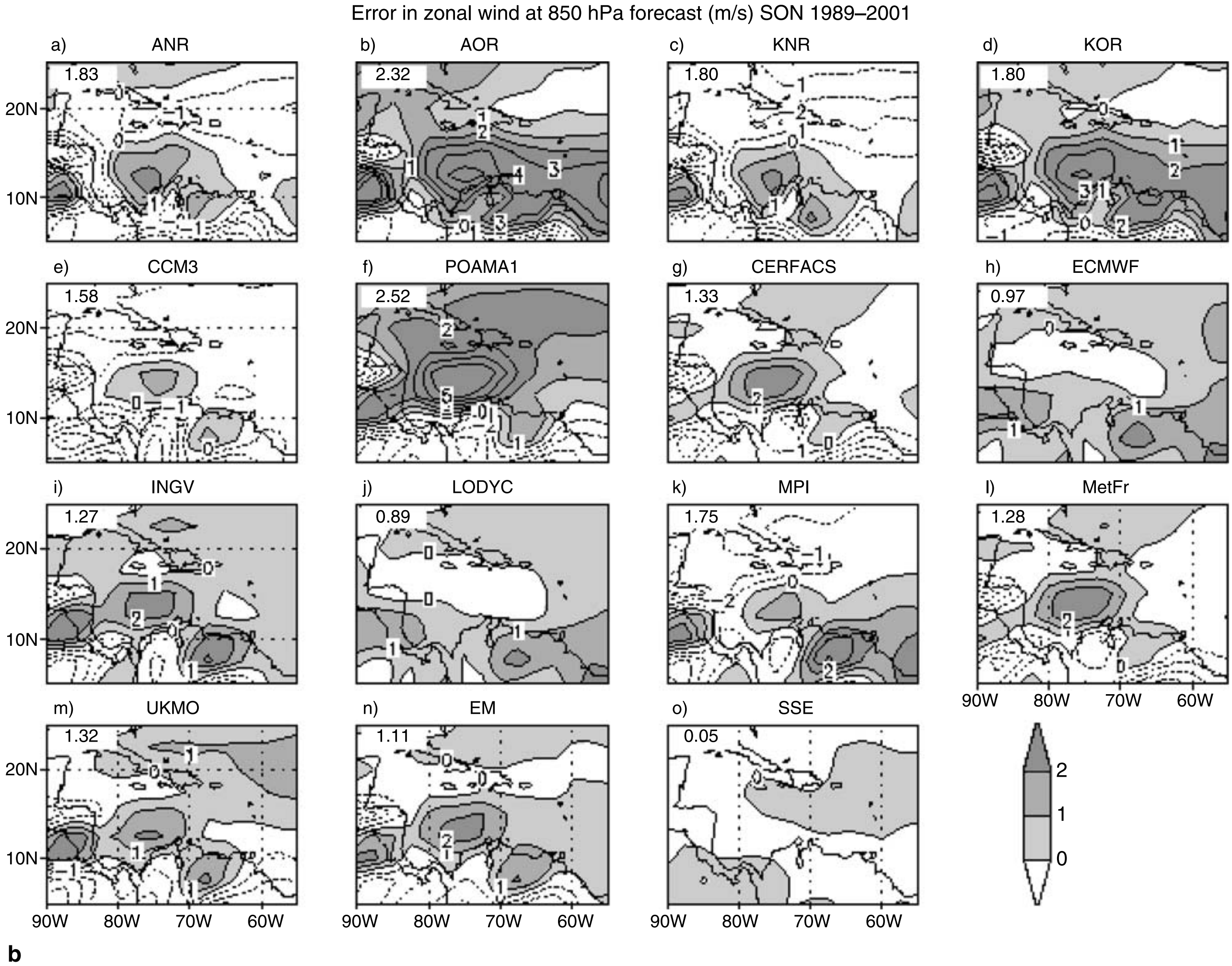


Fig. 12 (*continued*)

and the INGV (Fig. 13a (i)) showing the worst qualitative result. The ensemble mean (Fig. 13a (o)) also captures the basic pattern although the northerly winds in the western Caribbean are underestimated. The FSUSSE (Fig. 13a (p)) has a nearly perfect representation of the observed climatology. The error maps in Fig. 13b confirm the qualitative patterns seen in Fig. 13a, with the FSUSSE having errors less than 1 m/s and an overall RMS error of 0.11. The good performance of the ANR model (Fig. 13b (a)) is confirmed with an RMS error of 0.73, while the poor performance of the INGV is confirmed with errors up to 3 m/s and an RMS error of 1.56. The ensemble mean (Fig. 13b (n)) has larger errors than the FSUSSE (values in excess of ± 1 m/s) and an RMS error of 0.67.

The results of the climatological depictions of the v-component of the wind for January–March and May–June will not be shown, but the results are very similar to those for September–November. In January–March the FSUSSE has an overall RMS error of only 0.05, the ensemble mean has an RMS error of 0.79, and the average RMS error for the individual models is 1.33. In May–June the FSUSSE has an even lower RMS error of 0.03, the ensemble mean has an RMS error of 0.90, and the average RMS error for the individual models is 1.21.

The performance of the FSUSSE in establishing the best representation of the Caribbean 850 hPa u- and v-wind component climatologies in comparison to the ensemble mean and the individual models comprising the ensemble mean, indicates that this model will have the highest probability of providing the best 850 hPa u- and v-wind component climatological forecasts for a particular year for all three seasons considered in this paper based on the results of Sperber and Palmer (1996).

5. Seasonal predictions for individual years during 1989–2001

In this section of the paper, we will examine Caribbean region seasonal climate forecasts of precipitation, sea-surface temperature (SST), 2-meter temperature, and wind at the 850 hPa level for individual years. Forecasts from the individual models comprising the multi-model ensemble, the multi-model ensemble mean, and the FSUSSE will be considered.

5.1 Precipitation

In Fig. 14, the forecasts that are the foundation of the seasonal precipitation climatologies shown in Figs. 3–5 are depicted for each year during the period 1989–2001. Model forecasts are displayed as the multi-model ensemble means for each of the following: 4 FSU models, 7 DEMETER models, 13 models (FSU + DEMETER + POAMA1 + CCM3). The observed seasonal precipitation values for each year from Xie and Arkin (1997) are also shown, as are the FSUSSE forecasts for each year using the cross-validation technique of Deque (1997). Time series of precipitation are presented for January–March

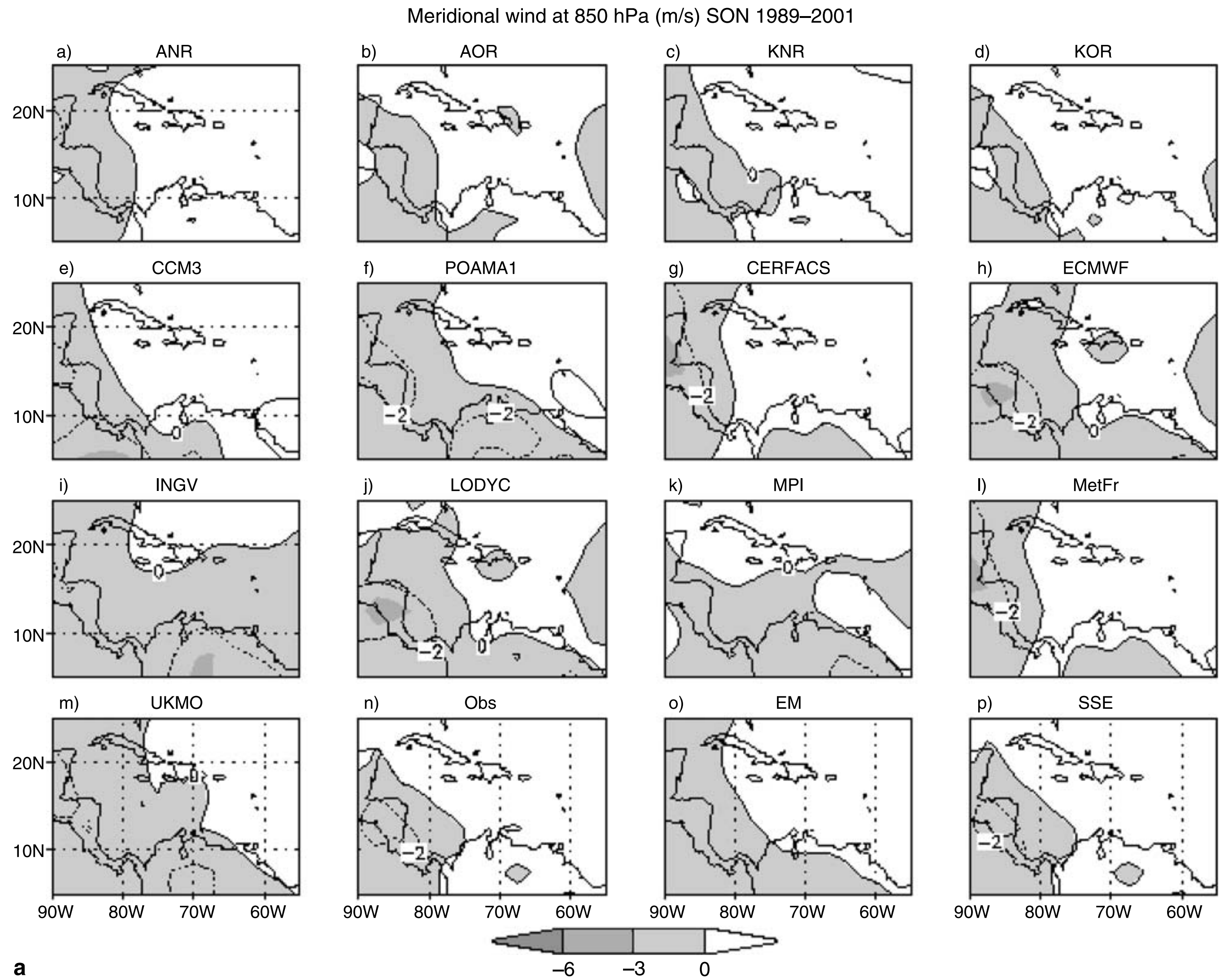
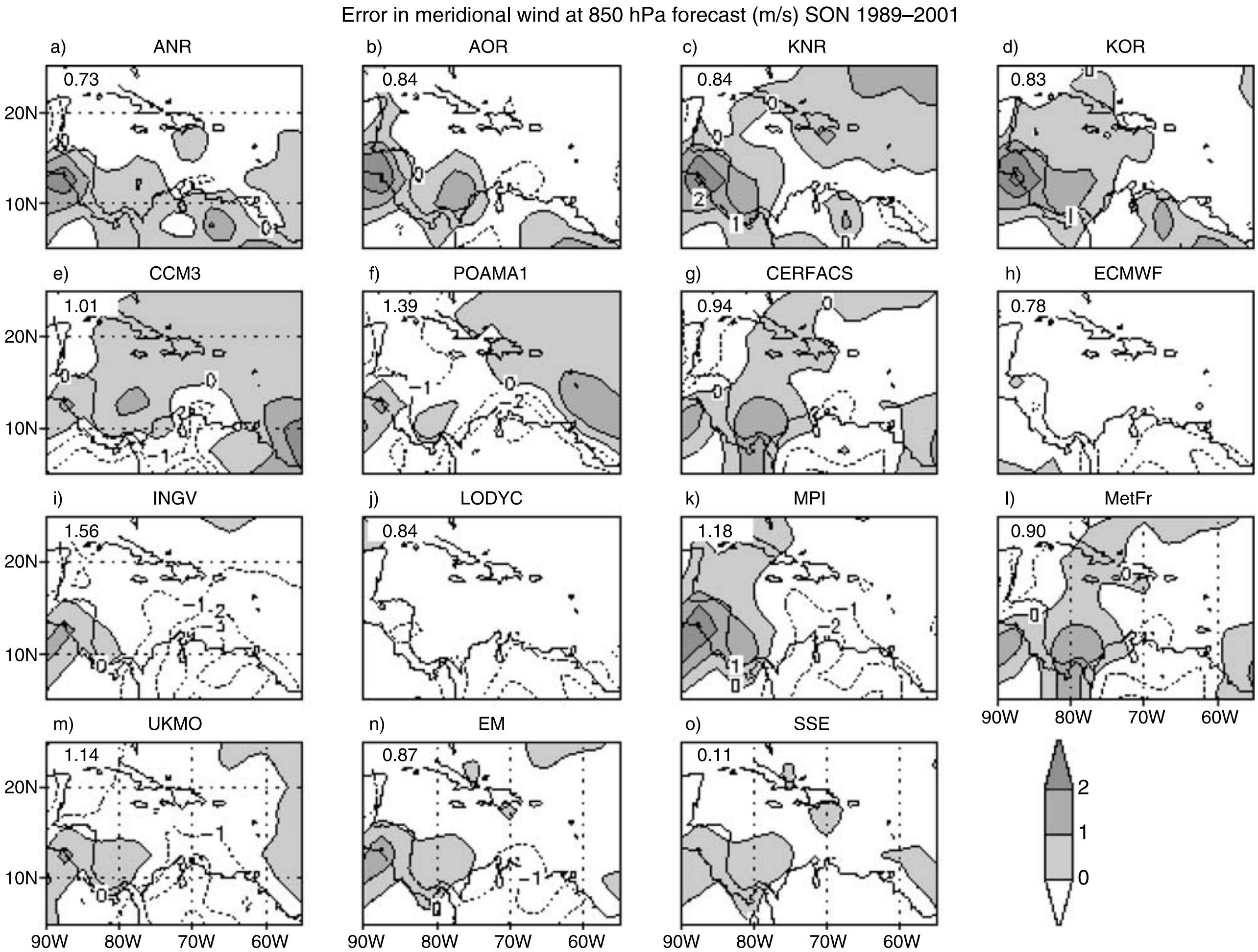


Fig. 13a. Climatology of the v-component of the wind at 850 hPa (m/s) for September–November 1989–2001 for each of the 13 models in the multi-model ensemble, the multi-model ensemble mean (EM), and the FSU Synthetic Superensemble (SSE). The observed climatology is from the ECMWF reanalysis. (**b**) Error in the climatology of the v-component of the wind at 850 hPa (m/s) based on a comparison of the model climatology to the observed climatology of the v-component of the 850 hPa wind from the ECMWF reanalysis shown in (**a**)



b

Fig. 13 (*continued*)

(Fig. 14a), September–November (Fig. 14b), and May–June (Fig. 14c).

The January–March depiction (Fig. 14a) confirms that the models comprising the ensemble over-forecast the dry season precipitation in the Caribbean region (see Fig. 3a and b). This is true for the FSU ensemble mean, the DEMETER ensemble mean, and the 13-model ensemble mean for every year in the sequence. The FSUSSE precipitation forecasts are very close to the observed precipitation for all years, except for 1996 (too dry) and 1991 and 2001 (too wet). Even though the FSUSSE seasonal forecasts for a particular year will not always be correct in terms of specifying an unusually wet or dry season, these results show that the FSUSSE seasonal forecasts are *more likely* to be correct than other models in the ensemble, or the multi-model ensemble mean. This confirms the point made in Sect. 4 of this paper, i.e., according to Sperber and Palmer (1996) a model that correctly represents the seasonal climatology of a region is more likely to provide the best seasonal forecast for a particular year.

The May–June time series (Fig. 14c) confirms that the models comprising the ensemble may either under-forecast or over-forecast the summer wet season early peak in precipitation in the Caribbean region in comparison to observation (see Fig. 4a and b). The FSU ensemble mean tends to under-forecast the precipitation, while the DEMETER ensemble mean tends to over-forecast the precipitation. The 13-model ensemble mean is just as likely to under-forecast or to over-forecast precipitation from year to year during the early summer wet period. The FSUSSE

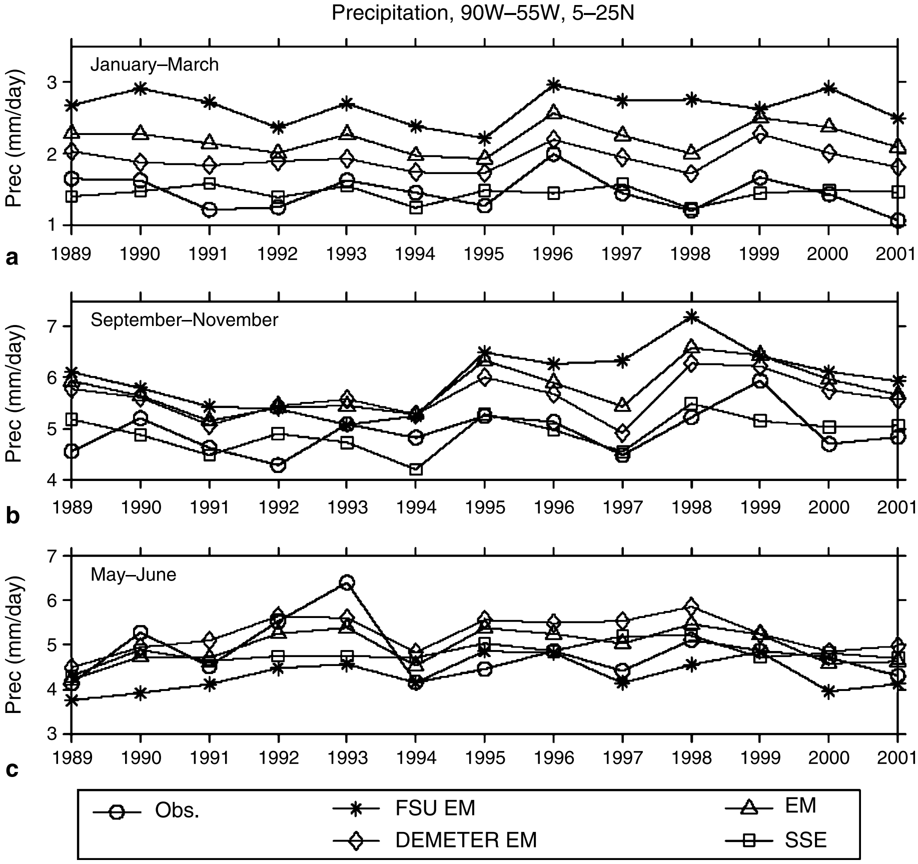


Fig. 14. Precipitation forecasts (mm/day) from the FSU multi-model ensemble mean, the DEMETER multi-model ensemble mean, the 13 model ensemble mean (EM), and the FSU Synthetic Superensemble (SSE), along with the observed precipitation from the CMAP data (Xie and Arkin, 1997) for the years 1989–2001: (**a**) dry season of January–March, (**b**) summer wet season late peak of September–November, (**c**) summer wet season early peak of May–June

does not have the clear advantage over the other models that was seen for January–March (Fig. 14a), but it still has some advantage. The FSUSSE is too dry in 1992–1993 and too wet in 1994–1995, and 1997. The FSUSSE is still *more likely* to give the best precipitation forecast for a given summer wet season early peak (May–June) in comparison to the other models and their ensemble mean, because the FSUSSE has the best prediction of the May–June seasonal climatology (Fig. 4a and b).

The depiction of the summer wet season late peak in precipitation during September–November (Fig. 14b) confirms that the models comprising the ensemble tend to over-forecast the precipitation in comparison to observation (see Fig. 5a and b). In Fig. 14b, the FSU ensemble mean, the DEMETER ensemble mean, and the 13-model ensemble mean all consistently over-forecast the precipitation in comparison to observation. The FSUSSE forecasts align very closely with the observed precipitation except for 1994 and 1999 (too dry) and 1989 and 1992 (too wet). From these results one can again conclude that the FSUSSE is *more likely* to give the

Table 2. Performance of FSUSSE and ensemble means for precipitation forecasts during 1989–2001 (13 years): dry season (January–February–March); summer wet season late peak (September–October–November); summer wet season early peak (May–June). Hit = No. "perfect" forecasts, wet (dry) = no. forecasts too wet (dry), best = no. forecasts best in comparison to other models. H + W + D = 13 for a given model. best column > 13 if models are tied for best

	Hit	Best	Wet	Dry
January–February–March 1989–2001				
FSUSSE	4	12	4	5
13-Model EM	0	0	13	0
FSU EM	0	0	13	0
DEMETER EM	0	1	13	0
September–October–November 1989–2001				
FSUSSE	5	10	4	4
13-Model EM	0	1	13	0
FSU EM	1	2	12	0
DEMETER EM	0	3	13	0
May–June 1989–2001				
FSUSSE	6	6	4	3
13-Model EM	3	3	7	3
FSU EM	4	6	2	7
DEMETER EM	2	4	9	2

best forecast of precipitation for the summer wet season late peak for a given year in comparison to the other models and their ensemble mean, although this will not necessarily be true for every year. This is because the FSUSSE provides the best prediction of the September–November seasonal climatology (Fig. 5a and b).

Table 2 concisely summarizes the forecast results shown in Fig. 14. In the dry season the advantage of the FSUSSE over the other models is overwhelming. For the 13 forecast years the FSUSSE has a "perfect" forecast in 4 years and is the best model in 12 of the 13 years. When its forecasts are not "perfect" there is an equal division between too wet (4 years) and too dry (5 years). The other models depicted are too wet in all 13 years, revealing a striking systematic error. In the late peak of the summer wet season the FSUSSE maintains its advantage over the other models, with a "perfect" forecast in 5 years and the best forecast in 10 of the 13 years. Its forecasts that are too wet or dry are evenly divided, with 4 wet forecasts and 4 dry forecasts. The other models are again too wet in 13 out of 13 years, except for the FSU ensemble mean which has a "perfect" forecast in one year. In the early peak of the summer wet season the FSUSSE has more competition from the other models. Its forecasts are the best in 6 of the 13 years, but other models also show significant "best" performance, e.g., the FSU ensemble mean is best in 6 of the 13 years. Nevertheless, the FSUSSE has a "perfect" forecast in 6 of the 13 years, and has little bias for wet or dry, with 4 years being too wet and 3 years being too dry. Among the other models the bias towards too wet is not as extreme as for the other two seasons, and the FSU ensemble mean actually reverses its bias from too wet to too dry (7 of 13 years too dry).

The RMS errors in precipitation forecasts in mm/day are shown in Fig. 15 for each year 1989–2001 for the seasons of January–March, May–June, and September–November. Each one of the 13 models comprising the multi-model ensemble is represented, along with the 13-model ensemble mean, and the FSUSSE. Figure 15a shows results for the Caribbean region dry season of January–March. The FSUSSE is seen to have the lowest RMS errors in comparison to the individual models in the multi-model ensemble and to the 13-model ensemble mean for all years except 1998. For the summer wet season late peak of September–November shown in Fig. 15b, the FSUSSE again outperforms its competitors for all years except 1999. Finally, for the summer wet season early peak of May–June seen in Fig. 15c, the FSUSSE outperforms its competitors for all years in the period 1989–2001.

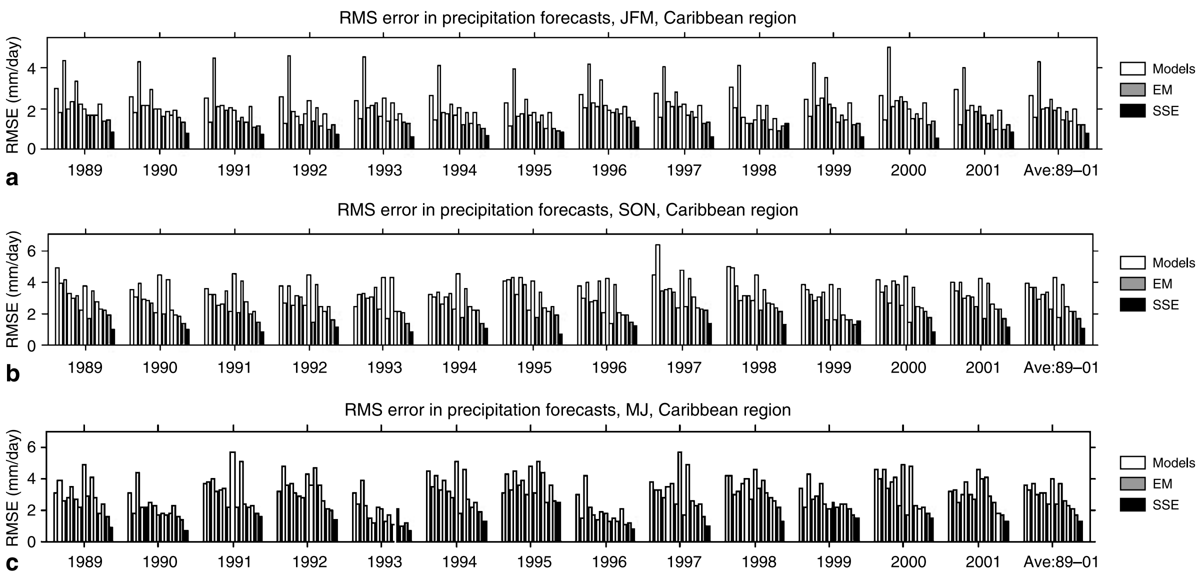


Fig. 15. RMS error (mm/day) in Caribbean region precipitation forecasts by season for each year during the 13 year period of 1989 through 2001, along with the 13 year mean RMS error in precipitation forecasts. The RMS error is shown for each of the 13 models in the multi-model ensemble, for the 13 model ensemble mean, and for the FSUSSE: (**a**) dry season of January–March, (**b**) summer wet season late peak of September–November, (**c**) summer wet season early peak of May–June

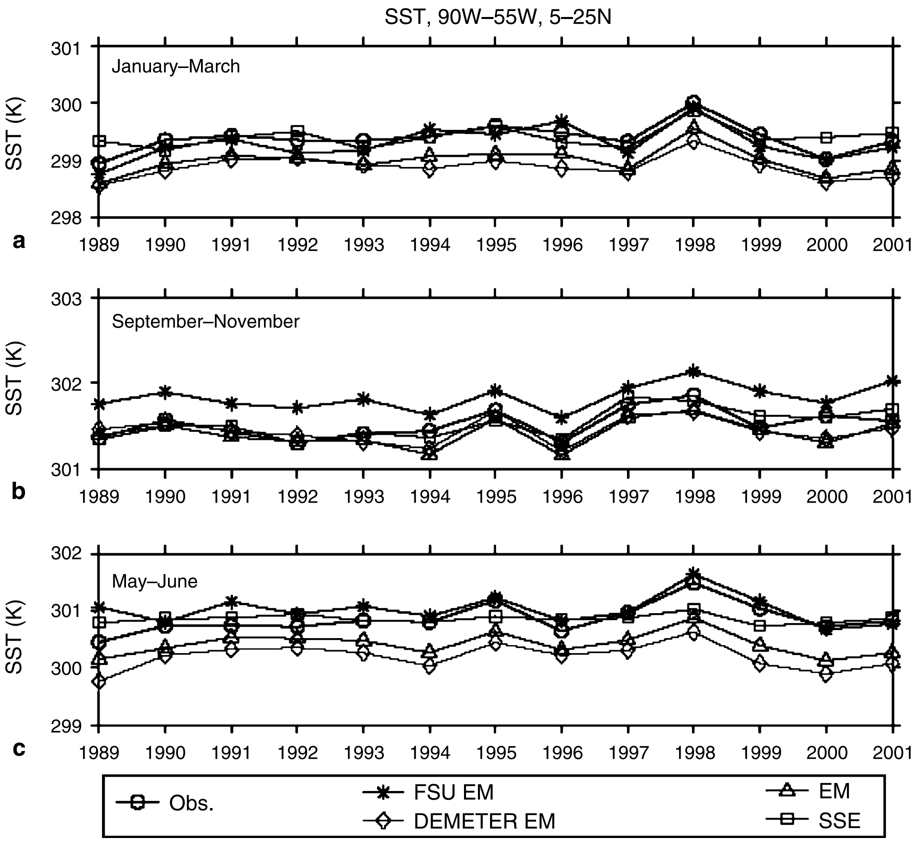


Fig. 16. SST forecasts (K) from the FSU multi-model ensemble mean, the DEMETER multi-model ensemble mean, the 13 model ensemble mean (EM), and the FSU Synthetic Superensemble (SSE), along with the observed SST from Reynolds et al (2002) for the years 1989–2001: (**a**) dry season of January–March, (**b**) summer wet season late peak of September–November, (**c**) summer wet season early peak of May–June

5.2 Sea-surface temperature (SST)

In Fig. 16, the forecasts that are the foundation of the seasonal SST climatologies shown in Figs. 6–8 are depicted for each year during the period 1989–2001. Model forecasts are displayed for the multi-model ensemble means for each of the following: 4 FSU models, 7 DEMETER models, 13 models (FSU + DEMETER + POAMA1 + CCM3). The observed seasonal SST values for each year from Reynolds et al (2002) are also shown, as are the FSUSSE forecasts for each year using the cross-validation technique of Deque (1997). Time series of SST are presented for January–March (Fig. 16a), September–November (Fig. 16b), and May–June (Fig. 16c).

The January–March depiction (Fig. 16a) confirms that the FSUSSE seasonal forecasts of SST for this period are superior to the forecasts of other models (See Fig. 6a and b), as represented by the three ensemble means that are depicted. The FSUSSE SST forecasts are very close to the observed SSTs for all years, except for 1989 and 2000 (both too warm). The 13 member ensemble mean (EM) and the DEMETER ensemble mean are too cold for all years. The FSU ensemble mean performs better than the other ensemble means, but in years when it doesn't come close to the observed value, it tends also to be too cold. Because of its ability to best reproduce the observed SST climatology for January–March (Fig. 6a and b), the FSUSSE is seen to be more likely to give the best January–March SST forecast for a particular year in comparison to the other models.

The May–June time series (Fig. 16c) shows that the FSUSSE and the FSU ensemble mean generally have better predictions of this season's SST values for each of the years 1989–2001 in comparison to the other models which are represented by the DEMETER ensemble mean and the 13 member ensemble mean (EM). The FSUSSE SST predictions are very close to the observed except for the years 1995, 1998, and 1999 (too cold) and 1989 and 1992 (too warm). The DEMETER ensemble mean and the 13 member ensemble mean are noticeably too cold for every year in the sequence. Because of its superior prediction of the SST climatology for

May–June (Fig. 7a and b) the FSUSSE is seen in Fig. 16c to be *more likely* to give the best seasonal SST prediction for a particular year in comparison to the other models.

The September–November time series (Fig. 16b) shows that the FSUSSE has a nearly perfect seasonal forecast of SST for all years during the period 1989–2001 except for 1995 (slightly too cold) and 1999 and 2001 (slightly too warm). The DEMETER and 13 model ensemble means are very competitive with the FSUSSE, but tend to be too cold for several of the years. The FSU ensemble mean is noticeably too warm for all years. Because of its superior ability to reproduce the September–November SST climatology (Fig. 8a and b) the FSUSSE has the highest probability of making the best September–November seasonal SST forecast for a particular year, as documented by Fig. 16b.

Table 3 concisely summarizes the SST forecast results shown in Fig. 16. In the dry season of January–March the FSUSSE has a clear advantage over the other models. For the 13 forecast years the FSUSSE has a "perfect" forecast in 6 of the years and is the best model in 11 of the years. Its bias is evenly divided between SST forecasts that are too warm (4 years) and too cold (3 years). The FSU ensemble mean is the FSUSSEs closest competitor, with a "perfect" forecast in 5 of the years, and the best forecast in 8 of the years. However, the FSU ensemble mean has a cold bias, with 7 years producing SST's that are too cold, and only one year producing ocean temperatures that are too warm. Both the DEMETER ensemble mean and the 13 model ensemble mean fail to make a "perfect" forecast, both fail to have the best forecast for any year, and both have a consistent cold bias (SST's too cold in all 13 forecast years). In the late peak of the summer wet season during September–November the FSUSSE maintains its advantage over the other models but has more competition than for the period of January–March. The FSUSSE has a "perfect" forecast in 8 of the years, is best in 10 of the years, and its bias is evenly divided between SST forecasts that are too warm (3 years) and too cold (2 years). The competition comes from the DEMETER ensemble mean ("perfect" forecast in 5 years, best forecast in 5 years, but with a bias toward too cold, i.e., 6 years too cold and only 2 years too warm) and the 13 model ensemble mean ("perfect" forecast in 6 years, best forecast in 5 years, but with a definite cold bias, i.e., 7 years too cold and no years too warm). Finally, for the early peak of the summer wet season during May–June, the FSUSSE has strong competition from the FSU ensemble mean. The FSUSSE has "perfect" forecasts in 5 of the years and has the best forecast in 9 of the years, while the FSU ensemble mean has "perfect" forecasts in 4 of the years and has the best forecast in 9 of the years. However, the FSUSSE forecasts have biases that are more evenly balanced than the FSU ensemble mean (FSUSSE has forecasts that are too warm in 5 years and too cold in 3 years, while the FSU ensemble mean has forecasts that are too warm in 9 years, with no forecasts that are too cold). Overall, Table 3 confirms the superior performance of the FSUSSE in comparison to the other models in the accuracy of its seasonal forecasts of SST for individual years during this 13 year study.

The RMS errors in SST forecasts (K) are shown in Fig. 17 for each year 1989–2001 for the seasons of January–March, May–June, and September–November. Each of the 13 models

Table 3. Performance of FSUSSE and ensemble means for SST forecasts during 1989–2001 (13 years): dry season (January–February–March); summer wet season late peak (September–October–November); summer wet season early peak (May–June). Hit = no. "perfect" forecasts, warm (cold) = no. forecasts too warm (cold), best = no. forecasts best in comparison to other models. H + W + C = 13 for a given model. Best column > 13 if models are tied for best

	Hit	Best	Warm	Cold
January–February–March 1989–2001				
FSUSSE	6	11	4	3
13-Model EM	0	0	0	13
FSU EM	5	8	1	7
DEMETER EM	0	0	0	13
September–October–November 1989–2001				
FSUSSE	8	10	3	2
13-Model EM	6	5	0	7
FSU EM	0	0	13	0
DEMETER EM	5	5	2	6
May–June 1989–2001				
FSUSSE	5	9	5	3
13-Model EM	0	3	0	13
FSU EM	4	9	9	0
DEMETER EM	0	0	0	13

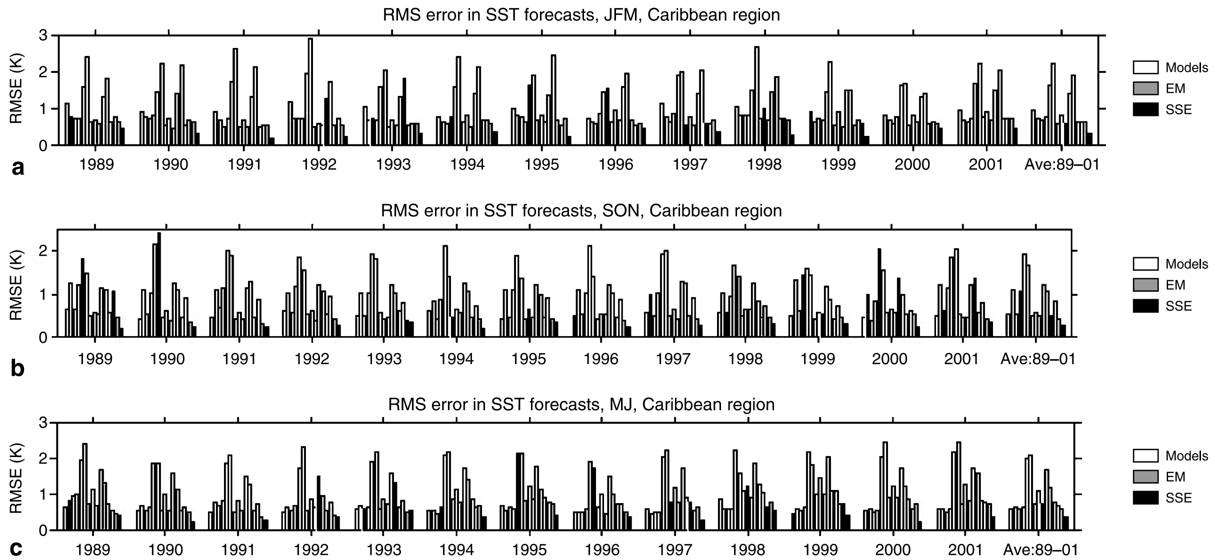


Fig. 17. RMS error (K) in Caribbean region SST forecasts by season for each year during the 13 year period of 1989 through 2001, along with the 13 year mean RMS error in SST forecasts. The RMS error is shown for each of the 13 models in the multi-model ensemble, for the 13 model ensemble mean, and for the FSUSSE: (**a**) dry season of January–March, (**b**) summer wet season late peak of September–November, (**c**) summer wet season early peak of May–June

comprising the multi-model ensemble is represented, along with the 13 model ensemble mean, and the FSUSSE. Figure 17a shows results for the Caribbean region dry season of January–March. The FSUSSE is seen to have the lowest RMS errors in SST in comparison to the individual models in the multi-model ensemble and to the 13 model ensemble mean for all 13 years. For the summer wet season late peak of September–November shown in Fig. 17b, the FSUSSE again has the lowest RMS errors for all of the years. Finally, for the summer wet season early peak of May–June seen in Fig. 17c, the FSUSSE has the lowest RMS errors in comparison to its competitors in all years except 1993, when its RMS errors are slightly larger than for the ensemble

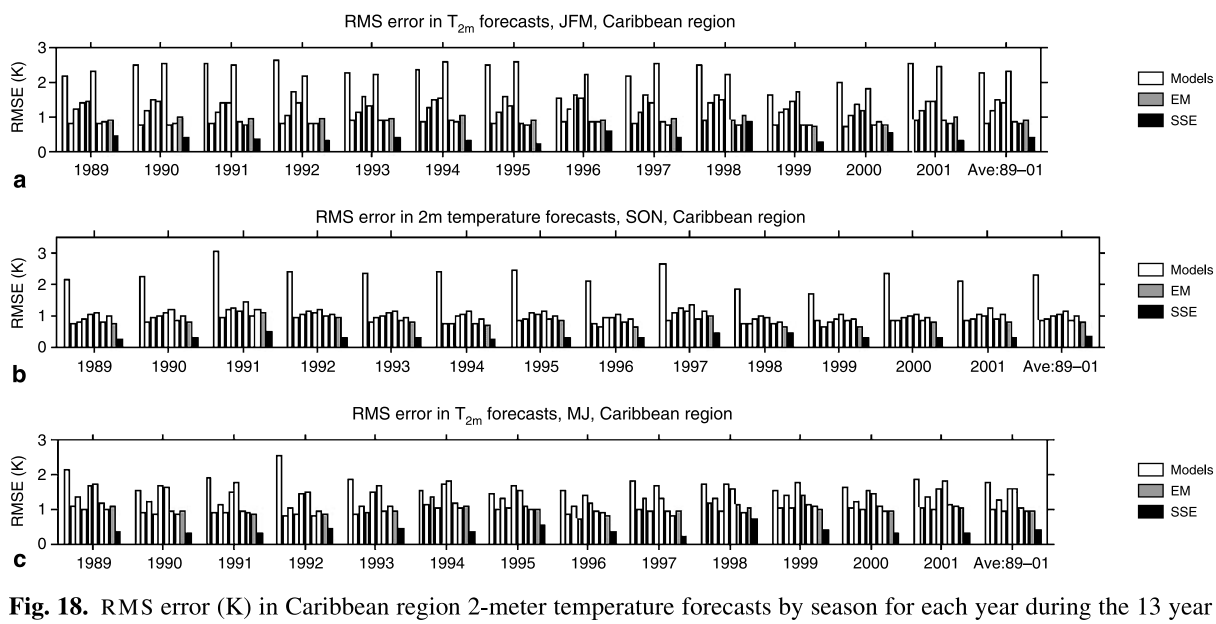


Fig. 18. RMS error (K) in Caribbean region 2-meter temperature forecasts by season for each year during the 13 year period of 1989 through 2001, along with the 13 year mean RMS error in 2-meter temperature forecasts. The RMS error is shown for each of the 8 models in the multi-model ensemble, for the 8 model ensemble mean, and for the FSUSSE: (**a**) dry season of January–March, (**b**) summer wet season late peak of September–November, (**c**) summer wet season early peak of May–June

mean, but still smaller than for each of the 13 models comprising the ensemble.

5.3 2-meter air temperature

For 2-meter temperature, the individual year performance by the 8 models comprising the ensemble mean, the ensemble mean of those 8 models, and the FSUSSE is represented by the RMS errors (K) for the years 1989–2001 in Fig. 18. Figure 18a shows results for the Caribbean region dry season of January–March. The FSUSSE is seen to have the smallest RMS errors for all years except 1998, when the FSUSSE RMS errors are larger than one of the individual models in the ensemble mean, although its errors are still smaller than for the 8-member ensemble mean. For the summer wet season late peak of September–November shown in Fig. 18b, the FSUSSE has smaller RMS errors than all 8 of the individual models comprising the ensemble and their ensemble mean for all years in the sequence. The same universally superior performance is seen for the FSUSSE forecasts for the summer wet season early peak of May–June seen in Fig. 18c. In Sect. 4c (Figs. 9–11), the FSUSSE was seen to produce the best 2-meter temperature climatology for all three seasons during the period 1989–2001 and it was stated that because of that performance the FSUSSE was more likely to produce the best 2-meter temperature forecast in comparison to the other models for a particular year. That statement has been verified for the individual year forecasts during the period 1989–2001 based on the RMS errors presented in Fig. 18.

5.4 850 hPa wind

The individual year performance of each of the 13 models, their ensemble mean, and the FSUSSE in predicting the seasonal climatology of the 850 hPa wind will be represented by the RMS errors (m/s) in the *u*-component (Fig. 19) and the *v*-component (Fig. 20) for each of the three seasons treated in this research. Figure 19a (*u*-component) and Fig. 20a (*v*-component) show the RMS errors for the Caribbean dry season of January–March for the years 1989–2001, along with the 13-year mean RMS errors. For the *u*-component the FSUSSE has the smallest RMS error in comparison to that of the individual models and their ensemble mean in 10 of the 13 years (all years except 1990, 1995, and 1998). For the *v*-component the FSUSSE has the smallest RMS error in all years except for 1998. Figure 19b (*u*-component) and Fig. 20b (*v*-component) show the RMS errors for the Caribbean summer wet

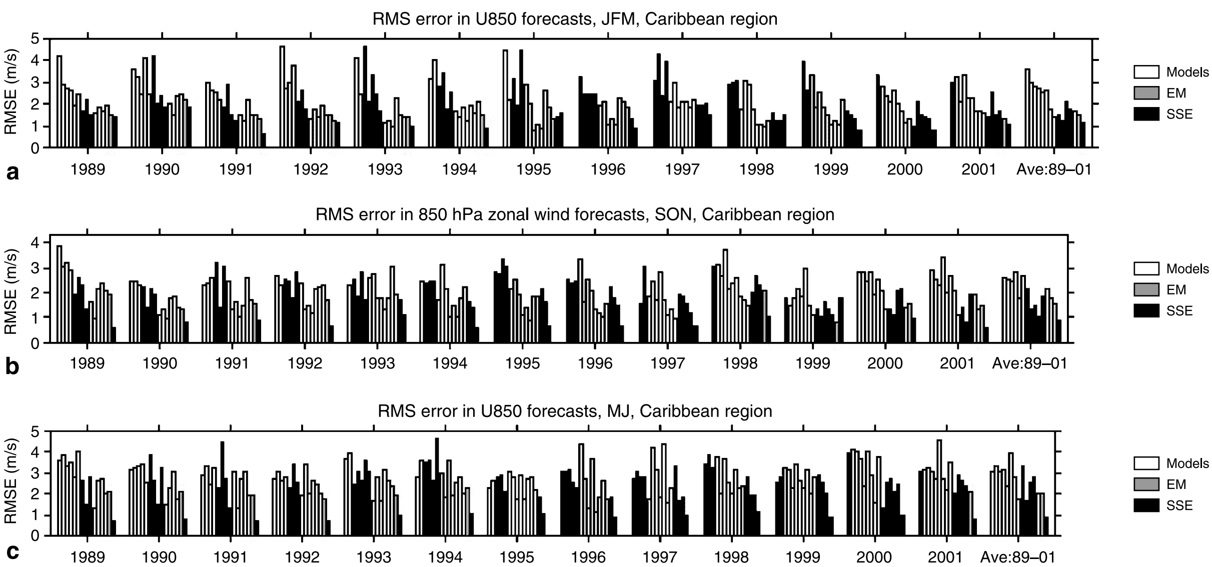


Fig. 19. RMS error (m/s) in Caribbean region 850 hPa *u*-component wind forecasts by season for each year during the 13 year period of 1989 through 2001, along with the 13 year mean RMS error in 850 hPa *u*-component wind forecasts. The RMS error is shown for each of the 13 models in the multi-model ensemble, for the 13 model ensemble mean, and for the FSUSSE: (**a**) dry season of January–March, (**b**) summer wet season late peak of September–November, (**c**) summer wet season early peak of May–June

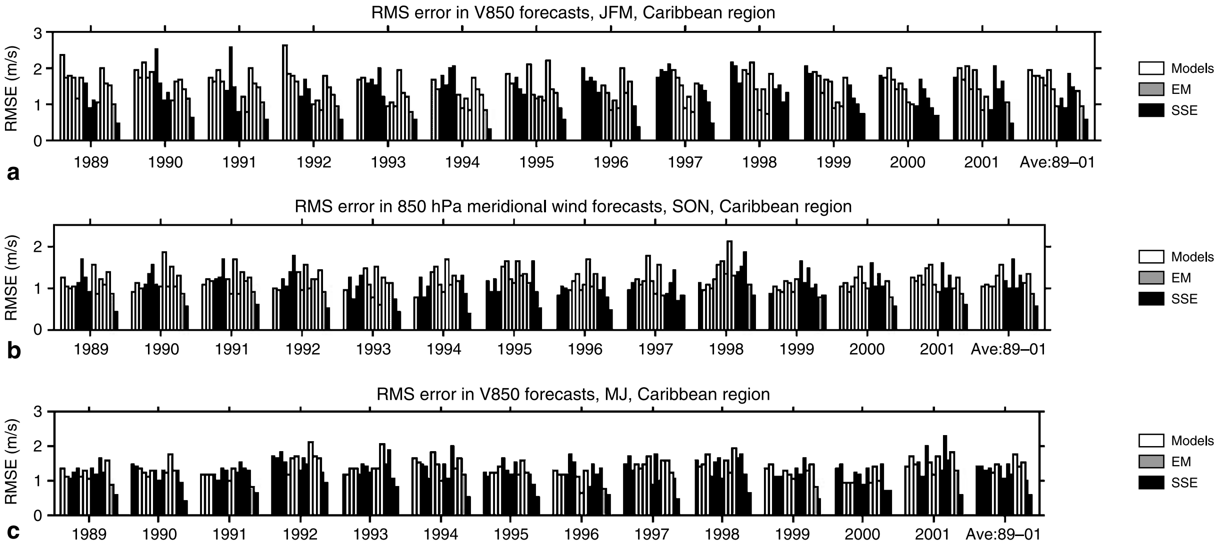


Fig. 20. RMS error (m/s) in Caribbean region 850 hPa v-component wind forecasts by season for each year during the 13 year period of 1989 through 2001, along with the 13 year mean RMS error in 850 hPa v-component wind forecasts. The RMS error is shown for each of the 13 models in the multi-model ensemble, for the 13 model ensemble mean, and for the FSUSSE: (**a**) dry season of January–March, (**b**) summer wet season late peak of September–November, (**c**) summer wet season early peak of May–June

season late peak of September–November. For the u-component the FSUSSE has the smallest RMS error in all years except for 1999. For the v-component the FSUSSE has the smallest RMS error in 11 of the 13 years (all years except 1997 and 1999). Figure 19c (u-component) and Fig. 20c (v-component) depict the RMS errors for the Caribbean summer wet season early peak of May–June. For the u-component the FSUSSE has the smallest RMS error in all 13 years, and for the v-component the FSUSSE has the smallest RMS error in all years except for 2000. These results verify the claim made in Sect. 4d of this paper, i.e., because the FSUSSE has the best representation of the 850 hPa u- and v-wind component climatologies in comparison to the individual models and their ensemble mean, the FSUSSE has the highest probability of providing the best 850 hPa u- and v-wind component climatological forecasts for a particular year for all three seasons studied here based on the results of Sperber and Palmer (1996).

6. Relationship of the seasonal climate forecasts for individual years to ENSO and NAO

In Sect. 4, we examined the success of the FSUSSE and the various climate forecast models, along with the ensemble mean of those forecast models, in depicting the observed climate of the Caribbean region during the period 1989–2001 for the parameters of rainfall, SST, 2-meter air temperature, and the u- and v-wind components at 850 hPa. The climatologies of those parameters were considered for the periods of January–March (winter dry season), May–June (early peak of summer wet season), and September–November (late peak of summer wet season). In Sect. 5, we studied the success of the FSUSSE and the climate forecast models in predicting the same parameters and for the same seasons, but for the *individual* years during the period 1989–2001. In this section of the paper we will explore the various climate forecasts in relation to the observed patterns of ENSO and NAO occurrence during the years 1989–2001. The goal will be to see how well the models perform in their climate forecasts in years that have particular ENSO and NAO patterns that are known to influence Caribbean weather, particularly the rainfall.

In Sect. 2, it was pointed out that an El Niño year, which has a Southern Oscillation Index (SOI) <0, will produce below average summer rainfall over the Caribbean, as will a year with a positive NAO. It was explained that this is due to years with positive NAO having reduced SSTs

Table 4. Negative SOI and positive NAO indices in comparison to CMAP observed (Obs) dry May–June (MJ) and September–November (SON) periods during 1989–2001, along with the RMS errors (mm/day) from the predictions of precipitation by the FSUSSE (SE) and the 13-model ensemble mean (EM) for these periods. The column-mean RMS errors (mm/day) are shown in the last row of the table

Year	SOI<0	NAO>0	Obs dry MJ	Obs dry SON	SE RMS dry MJ	EM RMS dry MJ	SE RMS dry SON	EM RMS dry SON
1989		X	X	X	0.8	1.8	1.0	2.0
1990		X						
1991	X	X	X	X	0.5	1.7	1.0	1.5
1992	X	X		X			1.1	1.7
1993	X	X						
1994	X	X	X	X	1.5	2.0	1.2	1.6
1995	X	X	X		2.3	2.5		
1996								
1997	X	X	X	X	0.9	1.3	1.1	2.1
1998		X						
1999		X						
2000		X						
2001	X		X		1.2	1.8		
Mean RMS					1.2	1.9	1.1	1.8

in the Caribbean, and years with negative SOI having divergent surface wind flow in that region. Relationships between SOI and NAO and Caribbean spring rainfall were also discussed. In this study only the summer rainfall relationship to SOI and NAO will be examined because of the way the months were constructed to define the seasons. We are using May–June to define the early peak in the summer wet season and September–November to define the late peak in the summer wet season. There is no specific definition of the spring season in the monthly groupings employed in this study.

Table 4 shows the years with SOI<0 and NAO>0 during the period 1989–2001, along with a tabulation of the years with below average precipitation (dry) (from Fig. 2) in May–June and September–November. For these dry MJs and SONs the RMS errors in the seasonal rainfall forecasts of the FSUSSE and the 13-model ensemble mean are shown. Several important conclusions may be drawn from this table. First, a strong relationship between a negative SOI and a dry summer is confirmed, since 6 of 7 years with negative SOI's had either a dry MJ or SON. A relationship between a positive NAO and a dry summer is also confirmed, although not as pronounced as for the SOI. Of the 11 years with a positive NAO, 6 had either a dry MJ or SON. A strong relationship between the correspondence of both a negative SOI and a positive NAO and a dry summer is confirmed, with 5 of 6 years fitting this pattern. It should also be noted that without exception all dry MJs and all dry SONs had either a negative SOI or a positive NAO, or both. Finally, it is obvious that the FSUSSE had a superior forecast to the 13-model ensemble mean for the dry summers. For all 7 dry MJs, the FSUSSE had a smaller RMSE than the ensemble mean, with a mean RMSE of 1.2 mm/day in comparison to the mean RMSE of 1.9 mm/day for the ensemble mean. For all 6 dry SONs, the FSUSSE had a smaller RMSE than the ensemble mean, with a mean RMSE of 1.1 mm/day in comparison to a mean RMSE of 1.8 mm/day for the ensemble mean. These numbers represent significant differences in rainfall predictions. For example, the difference in RMSE of 1.1 versus 1.8 mm/day in SON, represents a difference in rainfall, averaged over the Caribbean region, of approximately 2.50 inches during the three-month period. Stated another way, the FSUSSE eliminates 2.50 inches of the rainfall error averaged over the Caribbean in SON in comparison to the ensemble mean. It is clear that the FSUSSE gives more accurate rainfall predictions over the Caribbean during the dry summers that are modulated by the influences of SOI and NAO.

Next we will examine the relationship between a positive NAO and the occurrence of colder than average SSTs in the Caribbean during the period 1989–2001. Table 5 shows the years with ob-

Table 5. Observed (Obs) cold SSTs for May–June (MJ) and September–November (SON) periods in years with a positive NAO index during1989–2001, along with the RMS errors (K) from the predictions of SST by the FSUSSE (SE) and the 13-model ensemble mean (EM) for these periods. The column-mean RMS errors (K) are shown in the last row of the table. The observed SSTs are from Reynolds et al (2002)

Year	NAO > 0	Obs cold SST MJ NAO > 0	Obs cold SST SON NAO > 0	SE RMS cold SST MJ NAO > 0	EM RMS cold SST MJ NAO > 0	SE RMS cold SST SON NAO > 0	EM RMS cold SST SON NAO > 0
1989	X	X	X	0.3	0.4	0.2	0.4
1990	X	X		0.2	0.4		
1991	X	X	X	0.2	0.3	0.2	0.3
1992	X	X	X	0.3	0.4	0.2	0.4
1993	X	X		0.5	0.4		
1994	X	X		0.4	0.7		
1995	X						
1996							
1997	X						
1998	X						
1999	X		X			0.2	0.4
2000	X	X		0.2	0.7		
2001							
Mean RMS				0.3	0.5	0.2	0.4

served positive NAO along with an indication of which of those years had observed colder than average SSTs in the May–June and September–November periods (from Fig. 16). The table also shows the RMS errors in the SST predictions of the FSUSSE and the 13-model ensemble mean for the MJ and SON periods of the years with positive NAO and observed colder that average SSTs. Several important conclusions may be drawn from this table. First, there is a relationship between positive NAO and observed colder than average SSTs in the Caribbean, but the relationship is stronger in MJ than in SON. Table 5 reveals that of the 11 years that had a positive NAO, 7 of those years had colder than average SSTs in MJ, while only 4 of those years had colder than average SSTs in SON. Table 5 reveals that the FSUSSE had superior SST forecasts to the 13-model ensemble mean for the colder than average SST cases in the MJ and SON periods of years with a positive NAO. The FSUSSE has smaller RMS errors than the 13-model ensemble mean in 6 of the 7 MJ cases (FSUSSE has a larger RMSE only in 1993), with a mean RMSE of 0.3 K in comparison to the mean RMSE of 0.5 K for the ensemble mean. The FSUSSE has smaller RMS errors than the ensemble mean in all 4 of the SON cases, with a mean RMSE of 0.2 K in comparison to the mean RMSE of 0.4 K for the ensemble mean. It is clear that the FSUSSE provides more accurate SST forecasts over the Caribbean during cold SST episodes during the summer that are modulated by a positive NAO.

Finally, using Table 6 we will consider the performance of the FSUSSE and the multi-model ensemble mean in predicting the 2-meter air temperature and the *u*- and *v*-wind components at 850 hPa during dry MJ and SON periods for the years 1989–2001. These dry early and late sum-

Table 6. Mean RMS errors in the predictions of the FSUSSE (SE) and the 8-model ensemble mean (EM) for temperature at 2-meters and the 13-model ensemble mean (EM) for *u*-component and *v*-component of the wind at 850 hPa for CMAP observed dry May–June (MJ) and September–November (SON) periods during 1989–2001. The years during which the dry conditions in MJ and SON occurred are shown in Table 4. The mean RMS error for 2-meter temperature is in units of K, and the units for the mean RMS error for the wind components are m/s

Parameter	SE RMS dry MJ	EM RMS dry MJ	SE RMS dry SON	EM RMS dry SON
T 2-meter	0.3	1.0	0.3	0.8
u-wind 850 hPa	0.9	2.0	0.7	1.6
v-wind 850 hPa	0.5	1.0	0.5	0.8

mer periods, respectively, are tabulated in Table 4. All occurred with either negative SOI or positive NAO, or both. For 2-meter temperature the multi-model ensemble mean is composed of 8 models, and for the *u*- and *v*-wind components the multi-model ensemble mean is composed of 13 models. Table 6 shows the RMS errors for the predictions of the FSUSSE and the multi-model ensemble mean averaged over all of the observed dry MJs (6) and dry SONs (5). It is evident that the FSUSSE has markedly smaller RMS errors in comparison to the ensemble mean for all three parameters and for both dry MJ and dry SON periods. It should be noted that the individual year RMS errors (that are averaged to determine the values in the table) are smaller in the FSUSSE than in the ensemble mean in every instance, except for the *v*-wind component during SON of 1997 (0.8 m/s for FSUSSE and 0.6 for the ensemble mean). These results show that the FSUSSE provides more accurate forecasts of near surface air temperature and 850 hPa wind over the Caribbean during dry conditions in early and late summer that occur in conjunction with negative SOI and/or positive NAO phases. All in all, the FSUSSE provides superior forecasts of rainfall, SST, 2-meter air temperature, and 850 hPa wind during dry summers that are modulated by negative SOI and/or positive NAO. It was pointed out in Sect. 2 of this paper that dry summers associated with negative SOI and positive NAO have become a feature of a twenty-year pattern of drought in the Caribbean region. The results cited here have shown the FSUSSE to be a valuable tool for forecasting rainfall and other atmospheric and oceanic variables during such periods of drought in the Caribbean.

7. Probabilistic forecast evaluation

The Brier score (Brier, 1950) was designed to quantify the performance of a probabilistic forecast of a determined event. The Appendix provides a full treatment of the mathematics of the Brier score, which includes a modern formulation of this score by Wilks (1995), and a decomposition of the Brier score into three terms, reliability, resolution, and uncertainty, as proposed by Murphy (1973). The Appendix also presents the formulation of this score as the *Brier skill score*, which is the formulation used in this paper for all calculations, following Stefanova and Krishnamurti (2002). For Brier skill scores, a value of 1 indicates a perfect forecast, a value of 0 indicates a forecast that is equal in skill to a forecast of climatology, and a negative value denotes a forecast that has less skill than a forecast of climatology. The reliability term evaluates the statistical accuracy of the forecast, i.e., a perfectly reliable forecast is one for which the observed conditional frequency is equal to the forecast probability. The resolution term considers the distance between the forecast frequency and the unconditional climatological frequency, and thereby measures the ability of the forecasts to distinguish between different regimes. The uncertainty term is a measure of the variability of the system and is not influenced by the forecasts. In this paper we will use the Brier skill score to verify the probabilistic skill of the FSUSSE and the 13-model ensemble mean for rainfall over the Caribbean region. We will focus on the summer wet season's late peak in September–October–November (SON). The Brier skill score will be presented as an average over the Caribbean domain (90° W–55° W, 5° N–25° N) and over the months of SON during the 13-year period 1989–2001 (Table 7), and as an average over the same region and months, but for the years 1991, 1994, and 1997 when conditions were dry, in association with SOI <0 and NAO >0 (Table 8). In both cases, rainfall amounts will be broken into categories of 0–5 mm/day and 5–10 mm/day.

Tables 7 and 8 both show superior Brier skill scores for the FSUSSE in comparison to the 13-model ensemble mean for all comparisons. Both tables indicate higher scores for the reliability term than for the resolution term, indicating that the forecasts from both models have greater sta-

Table 7. Brier skill scores for the Florida State University Synthetic Superensemble (FSUSSE) and the multi-model ensemble mean (ENSM) averaged over the Caribbean domain (90° W–55° W, 5° N–25° N) and over the months of September–October–November during the 13-year period 1989–2001, for rainfall bins of 0–5 mm/day and 5–10 mm/day

Model	0–5			5–10		
	B	B_{rel}	B_{res}	B	B_{rel}	B_{res}
FSUSSE	0.57	0.98	0.59	0.47	0.97	0.50
ENSM	0.18	0.81	0.36	0.08	0.91	0.17

Table 8. Brier skill scores for the Florida State University Synthetic Superensemble (FSUSSE) and the multi-model ensemble mean (ENSM) averaged over the Caribbean domain (90° W–55° W, 5° N–25° N) and over the months of September–October–November for the years 1991, 1994, and 1997 when conditions were dry in association with SOI < 0 and NAO > 0, for rainfall bins of 0–5 mm/day and 5–10 mm/day

Model	0–5			5–10		
	B	B_{rel}	B_{res}	B	B_{rel}	B_{res}
FSUSSE	0.67	0.96	0.70	0.52	0.97	0.54
ENSM	0.24	0.78	0.47	0.06	0.86	0.20

tistical accuracy and lesser ability to distinguish between different precipitation regimes. In both tables scores for both models are generally higher for the smaller rainfall amounts (0–5 mm/day) than for the larger rainfall amounts (5–10 mm/day) except for the reliability scores for both models in the dry years of 1991, 1994, and 1997 (Table 8). The advantage of the FSUSSE over the ensemble mean in terms of Brier skill scores is generally greater in the dry years shown in Table 8 (e.g., advantage of 0.43 for overall Brier skill score in Table 8 for 0–5 mm/day versus 0.39 for this score in Table 7, and advantage of 0.46 for overall Brier skill score in Table 8 for 5–10 mm/day versus 0.39 for this score in Table 7). This result indicates that the FSUSSE forecasts have greater probabilistic skill, and are therefore a very useful tool for climate prediction, in years that have dry summer wet seasons modulated by negative SOI index and positive NAO index.

8. Conclusions

This paper has evaluated the success of the FSUSSE and 13 state of the art CGCMs in making seasonal forecasts of precipitation, SST, 2-meter air temperature, and 850 hPa u- and v-wind components for the dry season of JFM, the early peak of the summer wet season in MJ, and the late peak of the summer wet season in SON over the Caribbean region during the years 1989–2001. The major conclusions are given below.

Conclusions (1) through (4) pertain to the models' simulation of the observed Caribbean seasonal climate for the period 1989–2001.

(1) The FSUSSE simulation of 1989–2001 rainfall climatology more closely matches the observed climatology in comparison to the 13 CGCMs and their ensemble mean. These models overestimate the JFM and SON precipitation and are equally likely to overestimate or underestimate the MJ precipitation.

(2) For the 1989–2001 simulation of SST climatology the FSUSSE more closely matches the observed climatology in comparison to the 13 CGCMs and their ensemble mean. For JFM, area mean RMS errors are: 0.99 K to 2.78 K for the CGCMs, 1.05 K for the ensemble mean, 0.02 K for the FSUSSE. In MJ the RMS errors are: 0.88 K to 2.62 K for the CGCMs, 1.36 K for the ensemble mean, 0.01 K for the FSUSSE. In SON the RMS errors are: 1.33 K to 3.11 K for the CGCMs, 1.46 K for the ensemble mean, 0.01 K for the FSUSSE.

(3) The FSUSSE simulation of 1989–2001 2-meter air temperature climatology more closely matches the observed climatology in comparison to the 8 CGCMs used in forecasting this variable and their ensemble mean. For JFM the area mean RMS errors are: 0.71 K to 2.32 K for the CGCMs, 0.82 K for the ensemble mean, 0.04 K for the FSUSSE. In MJ the RMS errors are: 0.87 K to 1.62 K for the CGCMs, 0.90 K for the ensemble mean, 0.02 K for the FSUSSE. In SON the RMS errors are: 0.74 K to 2.12 K for the CGCMs, 0.69 K for the ensemble mean, 0.02 for the FSUSSE.

(4) For the 1989–2001 simulation of 850 hPa u-component and v-component wind climatology the FSUSSE is in better agreement with the observed climatology in comparison to the 13 CGCMs and their ensemble mean. The results for SON are representative of the results for JFM and MJ. For the u-component of the wind in SON the FSUSSE predictions have wind errors in the easterly flow of less than 1 m/s and an area mean RMS error of 0.05 m/s. The ensemble mean underestimates the maximum in the easterly flow by 2 m/s and has an area mean RMS error of 1.11 m/s. RMS errors among the CGCMs range from 0.89 m/s to 2.52 m/s. For the v-component of the wind in SON the FSUSSE has the best representation of the trough in the easterly flow located in the western Caribbean.

The FSUSSE wind errors are less than 1 m/s with an area mean RMS error of 0.11 m/s. The ensemble mean has an area mean RMS error of 0.67 m/s, while the individual CGCMs have area mean RMS errors ranging from 0.73 m/s to 1.56 m/s.

Conclusions (5) through (8) pertain to the models' seasonal climate predictions for individual years during the period 1989–2001 in the Caribbean region.

(5) Table 2 (based on Fig. 14) concisely summarizes the performance of the various models for individual year seasonal forecasts of rainfall. In the dry season (JFM) the advantage of the FSUSSE is overwhelming with 4 "perfect" forecasts, the best forecast in 12 of the 13 years, and little bias toward too wet or too dry. All other models show the striking systematic error of being too wet in all 13 years. In the early peak of the summer wet season (MJ) the FSUSSE has more competition from other models but still maintains both best forecast and "perfect" forecast status in 6 of the 13 years, with little bias. Other models show less bias toward being too wet as compared to JFM. In the late peak of the summer wet season (SON) the FSUSSE regains its advantage over other models with 5 "perfect" forecasts, the best forecast in 10 of the 13 years, and with little bias. Other models show a wet bias in all 13 years, except for one year in the FSU ensemble mean forecasts.

(6) Table 3 (based on Fig. 16) summarizes the performance of the various models for individual year seasonal forecasts of SST. In JFM the FSUSSE is the model of choice with 6 "perfect" forecasts, the best forecast in 11 of the 13 years, and with little bias. The DEMETER and 13-model ensemble means have a striking cold bias with SSTs being too cold in all 13 years. In MJ the FSUSSE does not have the clear advantage over other models as in JFM. It has a "perfect" forecast in 5 years, the best forecast in 9 years, and with little bias. The FSU ensemble mean is competitive with the FSUSSE with 4 "perfect" forecasts, the best forecast in 9 years, but with a warm bias with SSTs being too warm in 9 years. In SON the FSUSSE again is the clear model of choice with 8 "perfect" forecasts, the best forecast in 10 of the 13 years, and with little bias. The DEMETER and 13-model ensemble means both demonstrate a definite cold bias.

(7) For individual year predictions of 2-meter air temperature the FSUSSE has the smallest RMS errors in comparison to all 8 CGCMs and their ensemble mean for all years and all seasons (JFM, MJ, SON), except for one model in JFM of 1998.

(8) For individual year predictions of u-component and v-component of the wind at 850 hPa the FSUSSE has the smallest RMS errors for both wind components in almost all of the years and seasons (JFM, MJ, SON), when compared to the 13 CGCMs and their ensemble mean. Exceptions are: JFM predictions of u-component in 1990, 1995, and 1998; JFM prediction of v-component in 1998; MJ prediction of v-component in 2000; SON prediction of u-component in 1999; SON predictions of v-component in 1997 and 1999.

Conclusions (9)–(14) pertain to the relationship of the seasonal forecasts for individual years during the period 1989–2001 over the Caribbean to the observed ENSO and NAO phenomena. Conclusions relate to how well the models perform in their climate forecasts in years that have particular ENSO and NAO patterns that are known to influence Caribbean weather, particularly the rainfall.

(9) A strong relationship between a negative SOI and a dry summer was confirmed. A relationship between a positive NAO and a dry summer was also confirmed, although not as strong as for the SOI relationship. A strong relationship between the correspondence of both a negative SOI and a positive NAO and a dry summer was confirmed. These statements are based on looking first at years with negative SOI and and/or positive NAO and then checking to see if the summer was dry. Turning it around and looking first for dry summers, it was found that all such summers had either a negative SOI, or a positive NAO, or both.

(10) The FSUSSE gives more accurate rainfall predictions during the dry summers that are

modulated by the influences of SOI and NAO in comparison to the 13-model ensemble mean based on RMS error values.

(11) There is a relationship between positive NAO and observed colder than average SSTs in the summer, but the relationship is stronger in early summer (MJ) than in late summer (SON).

(12) The FSUSSE has superior SST forecasts to the 13-model ensemble mean for the colder than average SST cases in MJ and SON of years with a positive NAO, based on RMS error values.

(13) The FSUSSE provides more accurate forecasts of 2-meter air temperature than the 8-model ensemble mean during dry summers that occur with negative SOI and/or positive NAO phases.

(14) The FSUSSE provides more accurate forecasts of 850 hPa u- and v-wind components than the 13-model ensemble mean during dry summers that occur with negative SOI and/or positive NAO phases.

Conclusion (15) relates to the probabilistic forecast skill of the FSUSSE.

(15) The FSUSSE was shown to have higher Brier skill scores in relation to the 13-model ensemble mean for September–October–November, the late peak in the summer wet season, for the 13 year period 1989–2001 and for the dry years of 1991, 1994, and 1997. The precipitation during these three years was modulated by negative SOI index and positive NAO index. The results indicate that the FSUSSE is a very useful forecast tool for seasonal climate forecasts during dry summers that are modulated by circulation patterns in the Pacific and Atlantic oceans. Such dry summers have contributed to the current protracted drought in the Caribbean region.

Overall, the FSUSSE seasonal climate forecasts have been shown to be the most dependable in comparison to the other models used in this research for all variables examined and for all seasons studied, particularly with regard to forecasts of dry summers that are related to SOI and NAO signals, which has been a major focus of this paper. The conclusions drawn here should be tested by additional years of data. In future research the Caribbean region could be broken down into sub-regions, e.g., the western versus the eastern region, or the northern versus the southern region, as relationships between the climate and large-scale controls are known to vary geographically. The climate forecasts might well be categorized by different atmospheric circulation patterns, in a synoptic climatology approach, to better explore the connections between circulation patterns and climatic differences, and to test the models' ability to forecast seasonal climate under different circulation regimes. Finally, the FSUSSE model should be applied to real time seasonal climate forecasting in the Caribbean region.

Appendix: Mathematical expressions for Brier skill score

Consider an event ε that either happens at realization k or does not [$o(k)=1$ if ε occurred, $o(k)=0$ if it did not] and is forecast to occur with probability $f(k)$. Wilks (1995) defined the Brier score as

$$b = \frac{1}{n}\sum_{k=1}^{n}[f(k) - o(k)]^2,$$

where the index k refers to the forecast-observation pairs and n is the total number of such pairs within the dataset. The lowest possible value of the Brier score is zero, which is achieved with a perfect deterministic forecast.

Assume the probabilistic forecast for ε is done within I discrete categories y_i. The frequency with which forecasts of y_i are issued is $p(y_i)$. The frequency within a category y_i forecast with which the event ε actually occurs is the conditional frequency $\bar{o}_i = p[o(k) = 1|y_i]$. A reliability diagram is a plot of $\bar{o}_i$ versus y_i accompanied by the forecast frequency distribution $p(y_i)$ versus y_i. For a perfect forecast, the reliability diagram would be a line at an angle of 45°.

Murphy (1973) has shown that it is useful to decompose the Brier score into three terms:

$$b = \underbrace{\sum_{i=1}^{I} p(y_i)(y_i - \bar{o}_i)^2}_{\text{reliability}} - \underbrace{\sum_{i=1}^{I} p(y_i)(\bar{o}_i - \bar{o})^2}_{\text{resolution}} + \underbrace{\bar{o}(1-\bar{o})}_{\text{uncertainty}}$$

$$= b_{\text{rel}} - b_{\text{res}} + b_{\text{unc}},$$

where $\bar{o} = (1/n)\sum_{k=1}^{n} o(k)$ is the unconditional mean frequency of occurrence of the event ε. The reliability term evaluates the statistical accuracy of the forecast – a perfectly reliable forecast is one for which the observed conditional frequency $\bar{o}_i$ is equal to the forecast probability (i.e., over all forecasts for y-percent chance of ε, ε will occur in y percent of the times).

The resolution term addresses the distance between the forecast frequency and the unconditional climatological frequency. Forecasts that are always close to the climatological frequency exhibit good reliability because the forecast frequency matches the observed frequency, but show poor resolution because they are not able to distinguish between different regimes.

The uncertainty term is a measure of the variability of the system and is not influenced by the forecast.

This paper presents the results as Brier *skill* scores. Such skill scores are calculated with respect to a reference forecast as

$$B = (b - b_{\text{ref}})/(b_{\text{perf}} - b_{\text{ref}}) = 1 - (b/b_{\text{ref}}),$$

where b_{perf} is 0.

If a climatological forecast ($b_{\text{res}} = 0, b_{\text{rel}} = 0$) is taken as a reference,

$$B = 1 - (b/b_{\text{unc}}),$$

$$B_{\text{rel}} = 1 - (b_{\text{rel}}/b_{\text{unc}}),$$

$$B_{\text{res}} = b_{\text{res}}/b_{\text{unc}}.$$

For a perfect forecast system,

$$B = B_{\text{rel}} = B_{\text{res}} = 1.$$

For a climatological forecast,

$$B = B_{\text{rel}} = B_{\text{res}} = 0.$$

Acknowledgement

This work was supported by NSF grant ATM-0419618 and NASA grant NAGS-13563.

References

Anthes RA (1977) A cumulus parameterization scheme utilizing a one-dimensional cloud model. Mon Wea Rev 105: 270–286

Arakawa A, Schubert WH (1974) Interaction of a cumulus cloud ensemble with the large-scale environment, part I. J Atmos Sci 31: 674–701

Brier GW (1950) Verification of forecasts expressed in terms of probabilities. Mon Wea Rev 78: 1–3

Deque M (1997) Ensemble size for numerical seasonal forecasts. Tellus 49A: 74–78

Enfield DB, Alfaro EJ (1999) The dependence of Caribbean rainfall on the interaction of the tropical Atlantic and Pacific Oceans. J Climate 12: 2093–2103

Giannini A, Cane MA, Kushnir Y (2001) Interdecadal changes in the ENSO teleconnection to the Caribbean region and the North Atlantic Oscillation. J Climate 14: 2867–2879

Giannini A, Kushnir Y, Cane MA (2000) Interannual variability of Caribbean rainfall, ENSO, and the Atlantic Ocean. J Climate 13: 297–311

Grell GA (1993) Prognostic evaluation of assumptions used by cumulus parameterization. Mon Wea Rev 121: 764–787

Hastenrath S (1985) Climate and circulation of the tropics. Dordrecht: D. Reidel, 455 pp

Krishnamurti TN, Bedi HS (1988) Cumulus parameterization and rainfall rates III. Mon Wea Rev 116: 583–599

Krishnamurti TN, Kishtawal CM, LaRow TE, Bachiochi DR, Zhang Z, Williford CE, Gadgil S, Surendran S (1999) Improved weather and seasonal climate forecasts from multimodel superensemble. Science 285: 1548–1550

Krishnamurti TN, Kishtawal CM, Shin DW, Williford CE (2000a) Multimodel superensemble forecasts for weather and seasonal climate. J Climate: 13: 4196–4216

Krishnamurti TN, Kishtawal CM, Zhang Z, LaRow TE, Bachiochi DR, Williford CE, Gadgil S, Surendran S (2000b) Improving tropical precipitation forecasts from a multi analysis superensemble. J Climate 13: 4217–4227

Krishnamurti TN, Ramanathan Y, Pan HL, Pasch RJ, Molinari J (1980) Cumulus parameterization and rainfall rates I. Mon Wea Rev 108: 465–472

Krishnamurti TN, Stefanova L, Chakraborty A, Kumar TSVV, Cocke S, Bachiochi D, Mackey B (2002) Seasonal forecasts of precipitation anomalies for North America and Asian monsoons. J Meteor Soc Japan 80: 1415–1426

Laing AG (2004) Cases of heavy precipitation and flash floods in the Caribbean during El Niño winters. J Hydrometeorology 5: 577–594

Murphy AH (1973) A new vector partition of the probability score. J Appl Meteor 12: 595–600

Palmer TN et al (2004) Development of a European Multi-Model Ensemble System for seasonal to inter-annual prediction (DEMETER). Bull Amer Meteor Soc 85: 853–872

Reynolds RW, Rayner NA, Smith TM, Stokes DC, Wang W (2002) An improved in situ and satellite SST analysis for climate. J Climate 15: 1609–1625

Rogers JC (1988) Precipitation variability over the Caribbean and tropical Americas associated with the Southern Oscillation. J Climate 1: 172–182

Sperber KR, Palmer TN (1996) Interannual tropical rainfall variability in general circulation model simulations associated with the Atmospheric Model Intercomparison Project. J Climate 9: 2727–2750

Stefanova L, Krishnamurti TN (2002) Interpretation of seasonal climate forecast using Brier skill score, FSU superensemble, and the AMIP-I data set. J Climate 15: 537–544

Wang G, Alves O, Zhong A, Smith N, Schiller A, Meyers G, Tseitkin F, Godfrey S (2004) POAMA: an Australian ocean-atmosphere model for climate prediction. 15th Symp. on Global Change and Climate Variations, 84th Annual Conf. of the American Meteorological Society, Seattle, WA, USA

Wilks DS (1995) Statistical methods in the atmospheric sciences. San Diego, CA: Academic Press, 467 pp

Wolff JO, Maier-Reimer E, Legutke S (1997) The Hamburg Ocean Primitive Equation Model HOPE. DKRZ Technical Report No. 13 (Modellberatungsgruppe, ed.), 98 pp

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

Yun WT, Stefanova L, Mitra AK, Kumar TSVV, Dewar W, Krishnamurti TN (2005) Multimodel synthetic superensemble algorithm for seasonal climate prediction using DEMETER forecasts. Tellus 57A: 280–289

Corresponding author's address: Robert S. Ross, Department of Meteorology, The Florida State University, Tallahassee, FL 32306, USA (E-mail: bross@coven.met.fsu.edu)