GCM Systematic Error Correction and Specification of the Seasonal Mean Pacific–North America Region Atmosphere from Global SSTs

THOMAS M. SMITH AND ROBERT E. LIVEZEY

Climate Prediction Center, NCEP/NWS/NOAA, Camp Springs, Maryland

(Manuscript received and in final form 3 July 1997)*

ABSTRACT

Specifications of 1- and 3-month mean Pacific–North America region 700-hPa heights and U.S. surface temperatures and precipitation, from global sea surface temperatures (SSTs) and the ensemble average output of multiple runs of a general circulation model with the same SSTs prescribed, were explored with canonical correlation analysis. In addition to considerable specification skill, the authors found that 1) systematic errors in SST-forced model variability had substantial linear parts, 2) use of both predictor fields usually enhanced specification performance for the U.S. fields over that for just one of the predictor fields, and 3) skillful specification and model correction of the heights and temperatures were also possible for nonactive or transitional El Niño–Southern Oscillation situations.

1. Introduction

The intent of the work presented here is very pragmatic, namely to develop a quantitative linear description of a certain class of systematic errors in an atmospheric general circulation model (GCM). These are errors in the ensemble mean representation of flow regimes forced by aspects of the global sea surface temperature (SST) field. Given this information it is possible to partially correct these SST-dependent errors, perhaps through model development from new insights into model behavior but certainly through postprocessing of model output.

The demonstration of the feasibility of such a postprocessor is a second objective of this work. This is important because the GCM examined here is representative of the family of models that is used for the second step of the two-tiered process to make operational dynamic seasonal forecasts for the U.S. National Weather Service, namely versions of the National Centers for Environmental Prediction (NCEP) Medium Range Forecast Model (MRF). The first step of this procedure consists of the production of a forecast of global SSTs. This is followed by use of the SST forecast as specified lower boundary conditions to drive multiple integrations of the atmospheric GCM, which are subsequently averaged to form the ensemble forecasts. These steps, as well as the models employed, are described in detail in Kumar et al. (1996).

While the atmospheric GCM examined here has demonstrated (Livezey et al. 1996; hereafter L96) skill in the reproduction of observed conditions when forced by prescribed observed SSTs, especially during episodes of the El Niño–Southern Oscillation (ENSO), it is also clear that this ENSO episode skill is compromised substantially by recurring systematic errors, particularly in biases and positions of features (Livezey et al. 1997; hereafter L97).

As a first attempt to quantitatively diagnose these and other SST-related systematic model errors, we chose the use of canonical correlation analysis (CCA). This technique matches patterns in two different multivariate datasets (which can be referred to as the predictors and predictands) that are most correlated with each other. Moreover, the mathematical operator (called the transformation matrix) relating these predictor and predictand patterns can be used to optimally linearly specify the latter from the former. Its use here was influenced by and evolved naturally from the work of Barnett and Preisendorfer (1987), Barnston (1994), and Barnston and Smith (1996; hereafter BS). More technical discussions of the form in which CCA is exploited here can be found in these papers and in Livezey and Smith (1999a; hereafter LS1).

Our goal was to identify observed patterns whose amplitudes vary coherently with both corresponding modeled patterns and the global SST field that drives

* This article was originally part of a submission received 15 November 1996.

Corresponding author address: Dr. Robert E. Livezey, Climate Prediction Center, W/NP51, Rm. 604, 5200 Auth Rd., Camp Springs, MD 20744.
E-mail: livezey@sgi84.wwb.noaa.gov

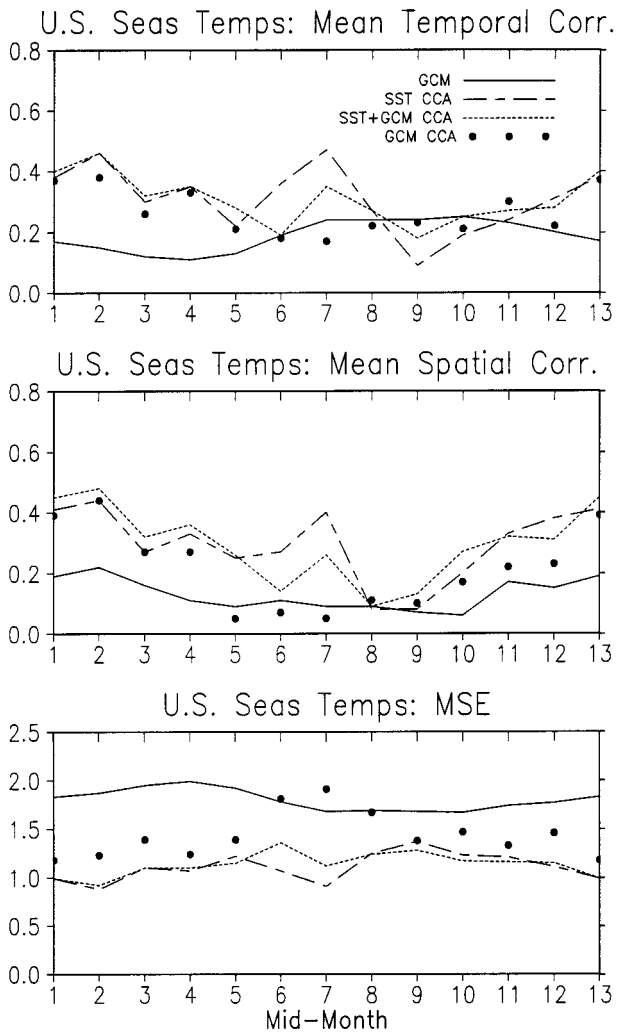

FIG. 1. Seasonalities of average performance measures for specification of 3-month mean U.S. surface temperatures by use of uncorrected ensemble-mean GCM output (solid lines), CCA with only global SSTs as predictors (long dashed lines), and CCA with both global SSTs and model output as predictors (short dashed lines). The top panel is for temporal correlations, the middle for spatial correlations, and the bottom for mean squared errors.

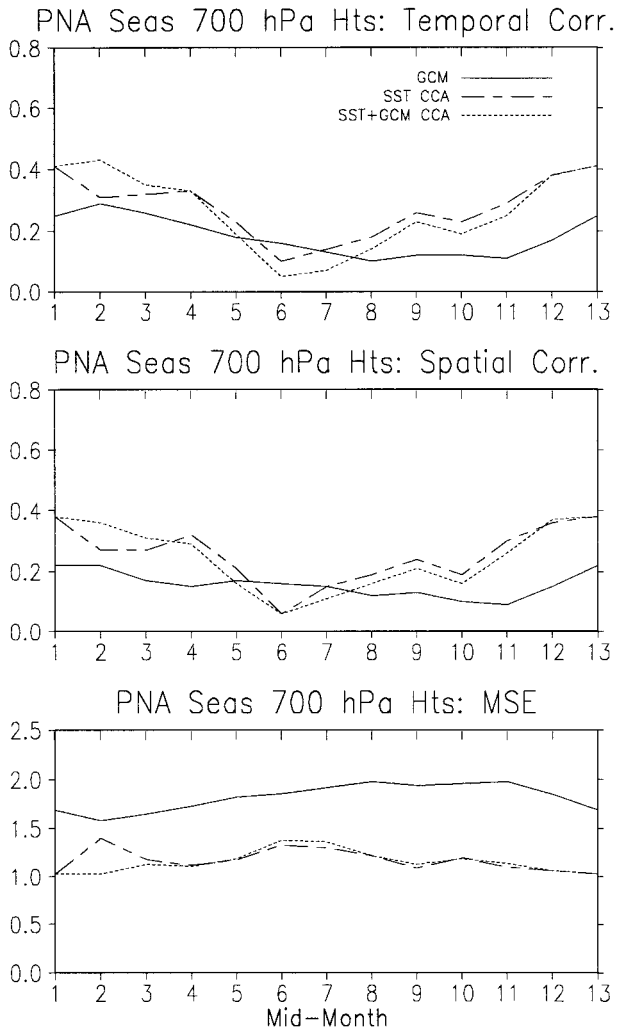

FIG. 2. Same as Fig. 1 except for PNA-region 700-hPa heights.

the model. Differences between corresponding observed and modeled patterns then would represent the error patterns associated with the accompanying SST patterns. Consequently, for the CCAs described here predictors consisted of either model-produced ensemble mean fields (e.g., U.S. surface temperatures) and/or their associated global SST fields, and as the predictand the observed fields corresponding to the model predictor fields. This setup is quite similar to that used by BS when only the global SSTs are used as predictors or those by Graham et al. (1994) and Ward and Navarra (1997) when only the modeled fields are used. The relative merits of these different approaches for the problems addressed here raise a number of additional questions.

The most important of these questions is which combination of predictors is best for the different objectives of this work. Obviously, in order to quantitatively relate systematic model errors in prescribed SST runs to specific SST forcing it is appropriate to use both the global SSTs and the associated modeled fields as predictors, and for this reason alone such two-predictor models will constitute the centerpiece of this paper. In other words, there is a diagnostic advantage to using two predictor fields.

In contrast, the question of which approach will perform best as a model postprocessor (i.e., produce the best forecast) in a forecast environment cannot be unambiguously answered in the context of the analyses presented here. This is because SSTs to drive a suite of GCM runs in such an environment are produced by an independent forecast model (which will make errors), whereas all of the analyses herein utilize observed SSTs (with only observational and analysis errors). There are,

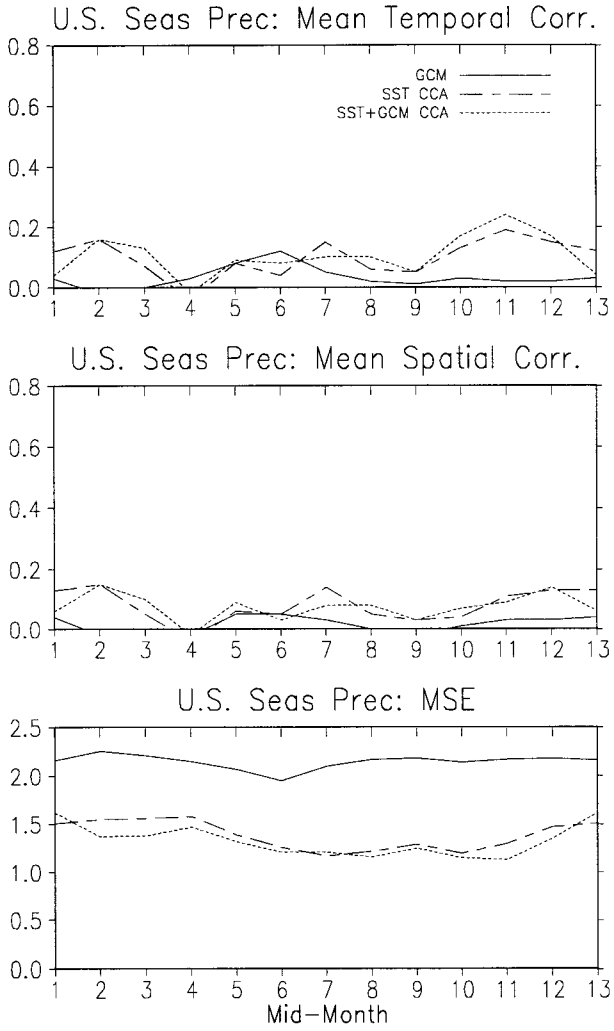

FIG. 3. Same as Fig. 1 except for U.S. precipitation.

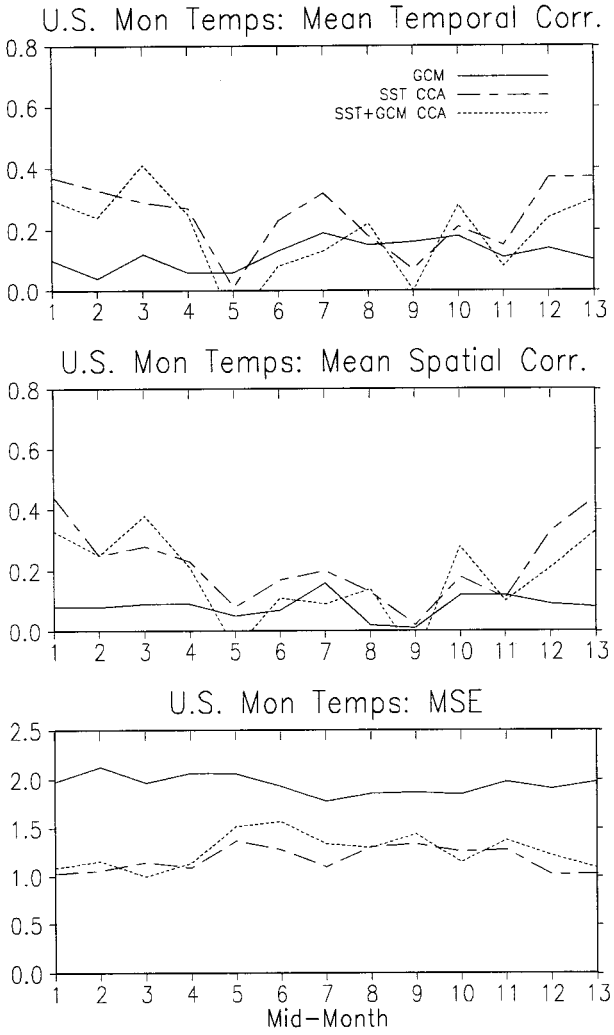

FIG. 4. Same as Fig. 1 except for 1-month mean U.S. surface temperatures.

however, some observations that can be made beforehand to shed some light on this issue.

First, there are two reasons to believe specifications of the observed fields by just their corresponding model fields will be less successful than the other two predictor possibilities examined here, both related to the fact that the raw material for the development of the CCA relationships is an ensemble of long prescribed SST runs. These model runs do not reproduce all ocean–atmosphere covariability, particularly that portion where the ocean is driven by the atmosphere, but very likely also a portion where the opposite is true. We will return to this point in the following paragraph. Additionally, without the SST field as a predictor it would be more difficult for the CCA to discriminate pattern matchups between the model and observations that are forced by underlying SST anomalies and those that randomly and separately occur through internal variability. One way to mitigate this problem would be to average a large enough number of runs for the ensemble mean to effectively remove all model internal variability. The 13-member ensemble used here filters much of the unforced noise, but not all.

On the other hand, it is difficult to argue in advance whether use of only global SST predictors would produce inferior results to use of two predictor fields. If comparable effectiveness can be achieved by use of just global SSTs, then one might question the need for a suite of GCM integrations for the second step of the two-tier forecast process. In this instance, a statistical specification of the forecast fields from the forecast SSTs would be all that was needed. Despite not being sure about the relative performance in practice of one- and two-predictor CCA that both use global SSTs, our expectation was that there would be differences (not necessarily corrupt) in the leading pattern matchups.

The reason for this expectation was our belief that inclusion of the model output would enhance simultaneous SST–atmosphere relationships where the atmo-

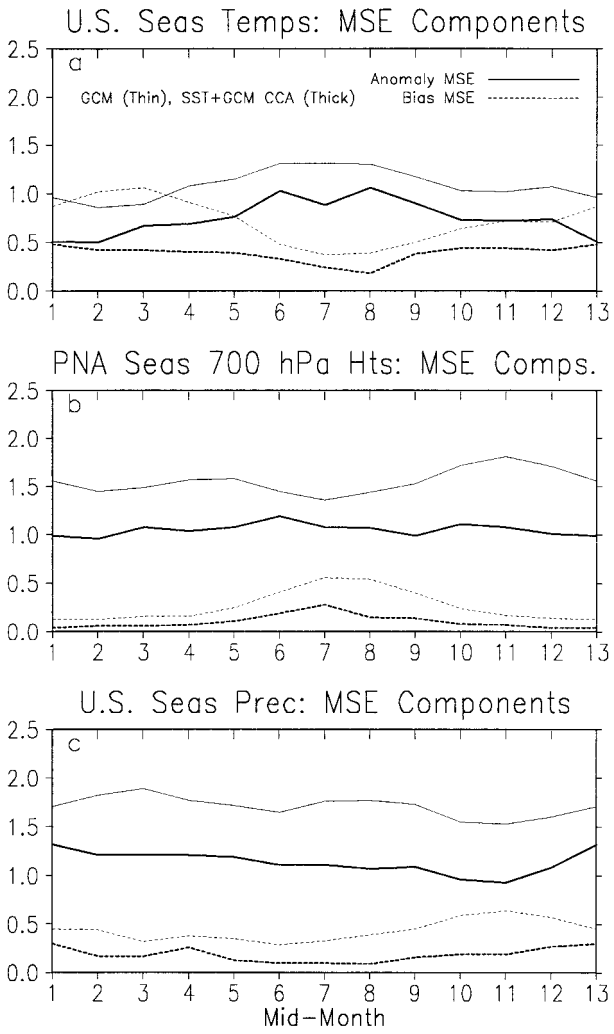

FIG. 5. Seasonalities of components of average mean-squared error for specification of 3-month mean (a) U.S. surface temperatures, (b) PNA-region 700-hPa heights, and (c) U.S. precipitation. Thin lines are for uncorrected ensemble-mean GCM output and thick lines for CCA with both global SSTs and model output as predictors, while solid lines are for anomaly squared errors and dashed lines bias squared errors.

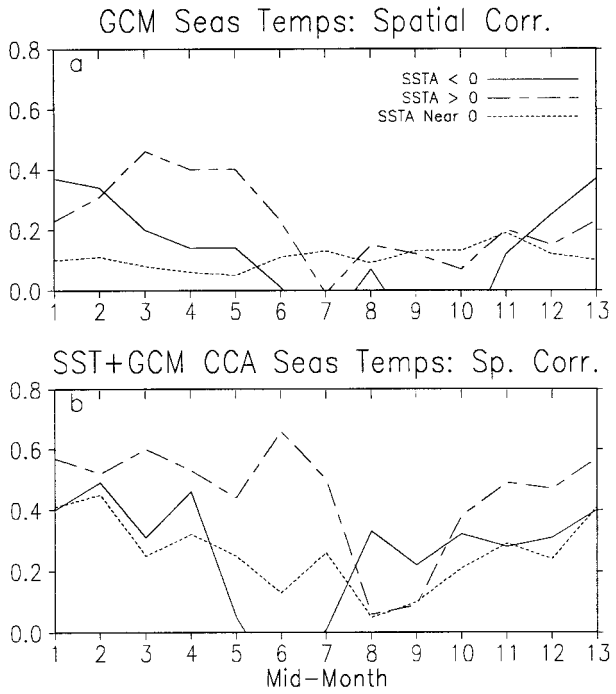

FIG. 6. Seasonalities of mean spatial correlations for specification of 3-month mean U.S. temperatures by use of (a) uncorrected ensemble-mean GCM output, and (b) CCA with both global SSTs and model output as predictors, for cases with below normal SSTs in a key area (see text) in the central equatorial Pacific Ocean (solid lines), above normal (long dashed lines), and near normal (short dashed lines).

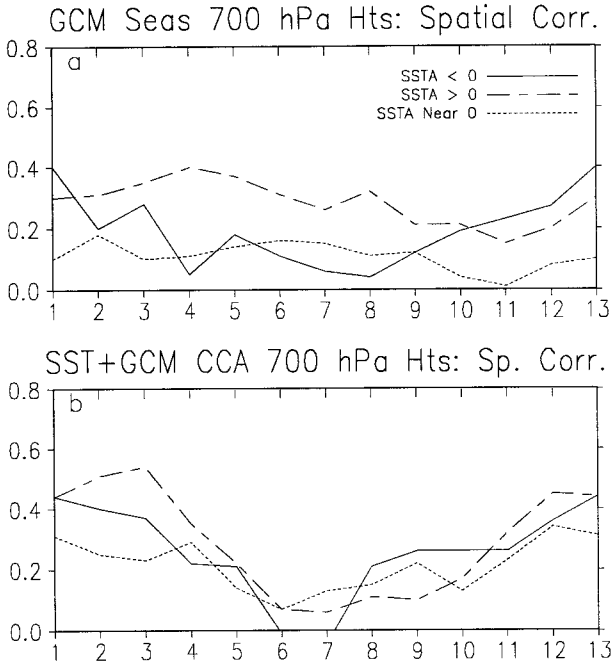

FIG. 7. Same as Fig. 6 except for PNA-region 700-hPa heights.

sphere was dominantly slave to the ocean at the expense of relationships where the opposite was the case. This is because GCM integrations with prescribed SSTs do not contain information about situations where the ocean mostly follows the atmosphere. Thus a CCA that uses both fields as predictors should have a better chance of detecting SST-driven atmospheric signals but should relegate other ocean–atmosphere relationships to higher-order modes. If it is found below that there is comparable specification skill for both postprocessors over the study period, then it is likely that the two-predictor model will perform at least as well as the version using only global SSTs in operational practice. This is because only the SST-driven signals currently can be exploited in

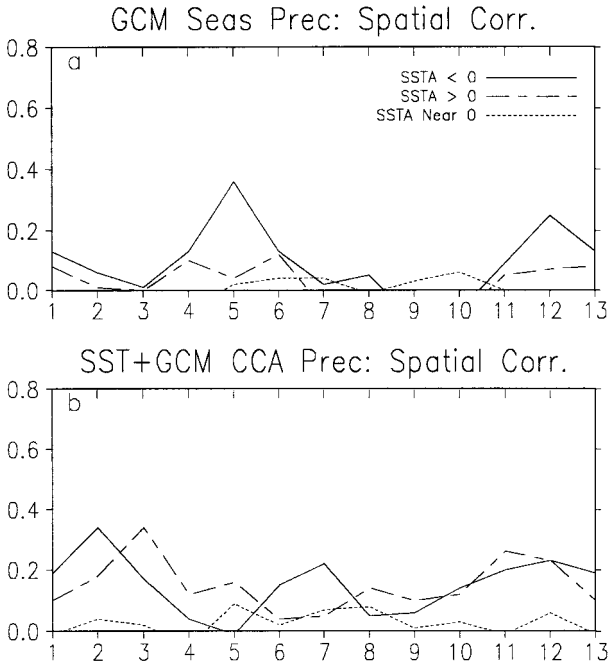

FIG. 8. Same as Fig. 6 except for U.S. precipitation.

operational prediction, and they will receive more emphasis in the two-predictor postprocessor. Another reason to continue use of the GCM is its ability to suggest scenarios in forecast situations that are not well represented in the historical database.

Thus, separate CCAs have been performed using only global SSTs, only modeled fields, and both fields together as predictors, and specification performance compared. Datasets, experimental strategies, and performance evaluation techniques are all described in section 2, followed by detailed results of the specification performance analyses in section 3, including the discussion of several notable successes and failures of the correction techniques in a cross-validated test. The paper concludes with a summary and additional comments in section 4.

2. Methods

a. Data

The model and observational fields that were selected for correction and specification respectively are 700-hPa heights on an approximately equal-area, 163-point grid [a subset of the Barnston and Livezey (1987) grid] encompassing the eastern Pacific–North America (PNA) region (25°N to the pole and 30°W to the date line) and surface temperatures and precipitation for 327 U.S. climate divisions. The records initially available for processing were monthly means spanning the full 43-yr period from 1950 to 1992 along with the data for January and February 1993. In the case of the GCM, monthly mean data were ensemble averages of 13 independent model integrations all forced by prescribed global SSTs reconstructed from the observational records by Smith et al. (1996) from 1950–81 and the SSTs of Reynolds and Smith (1994) thereafter. The global SST analyses

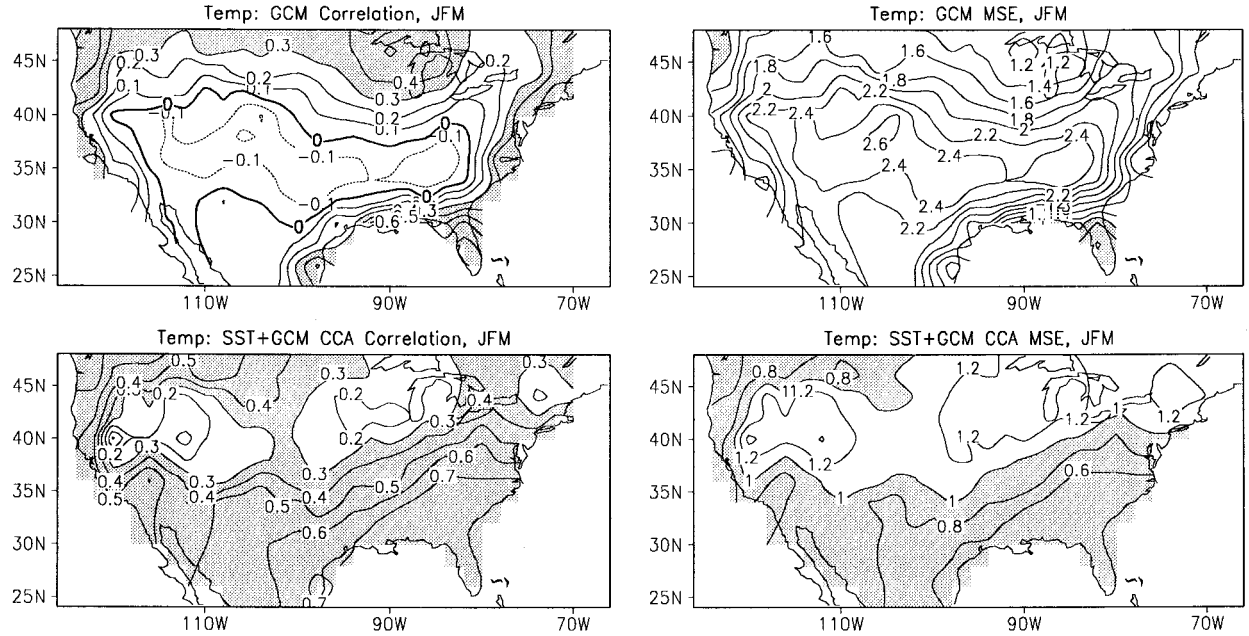

FIG. 9. Temporal correlations (left panels) and mean squared errors (right panels) for specification of Jan–Mar (JFM) U.S. surface temperatures by use of uncorrected ensemble-mean GCM output (top panels) and CCA with both global SSTs and model output as predictors (bottom panels).

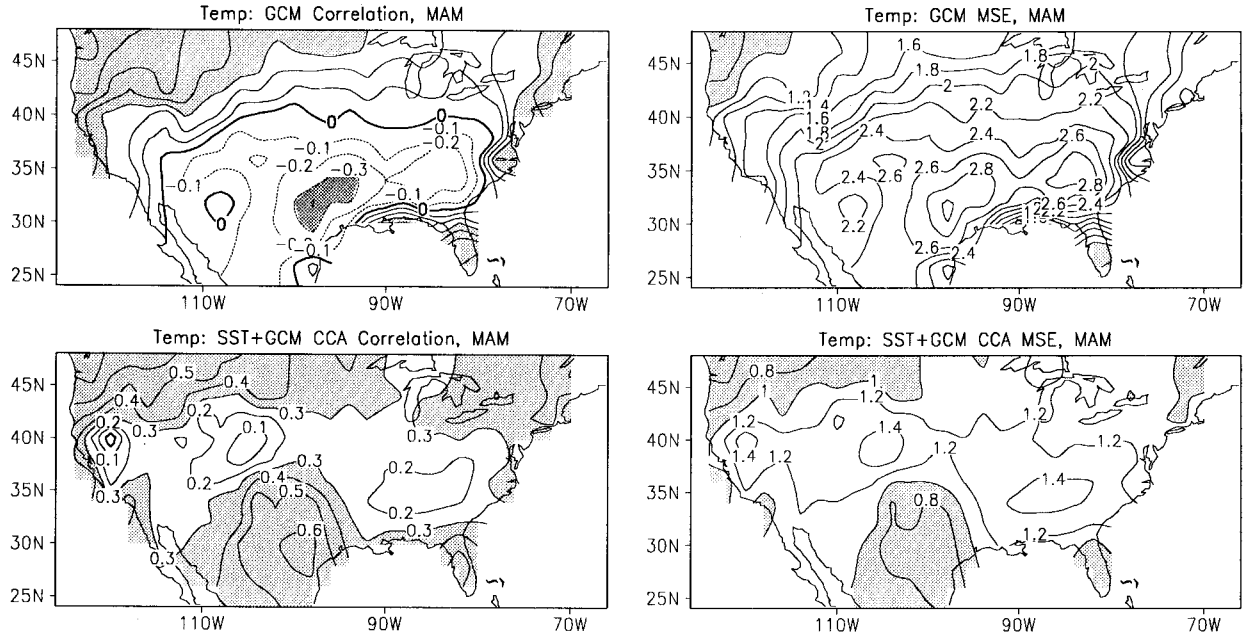

FIG. 10. Same as Fig. 9 except for Mar–May (MAM).

of Smith et al. (1996) averaged to a 6° lat × 6° long grid, with 796 points, comprised the last dataset used here. All sets were essentially the same as those used and described in L96 and L97.

Seasonality in the first two moments was removed from all records (including those from 1993) by standardization at each gridpoint with sample means and standard deviations computed for each month from the 1950–1992 records. This standardization had the effect of weighting all data points equally in subsequent principal component (PCA) and CCA analyses. Three-month means were formed by simple averages of the standardized anomalies for each month. The effect of this step was to weight all months equally, regardless of their prestandardization relative variance, in eigenvector and other analyses of multimonth mean data.

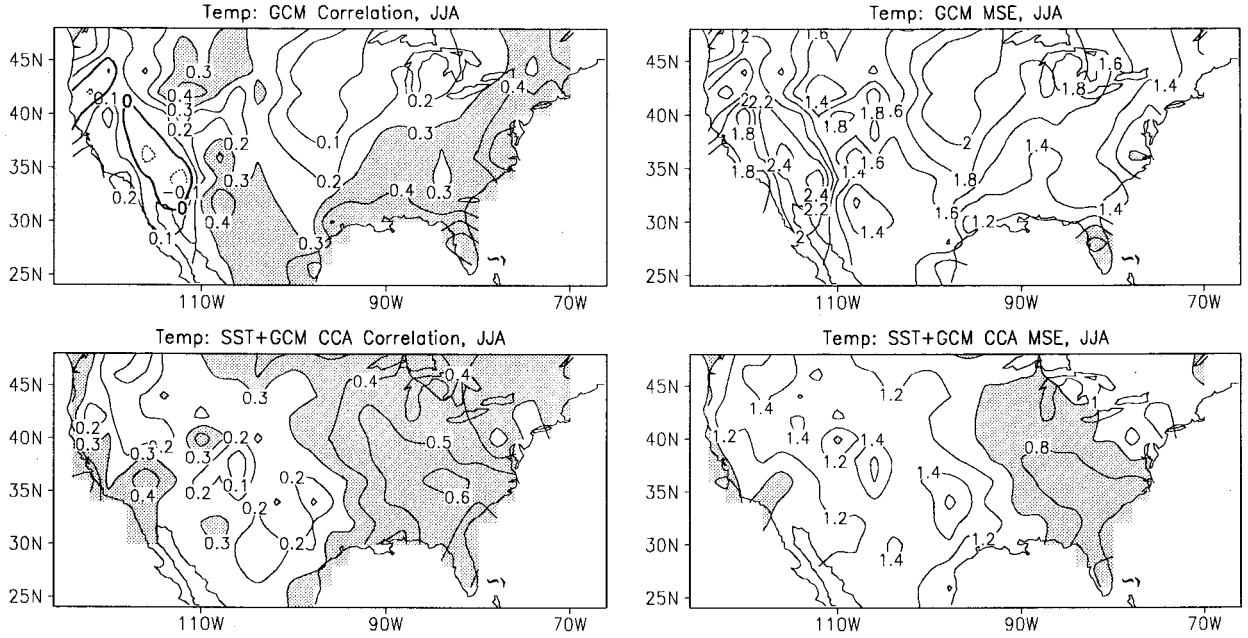

FIG. 11. Same as Fig. 9 except for JJA.

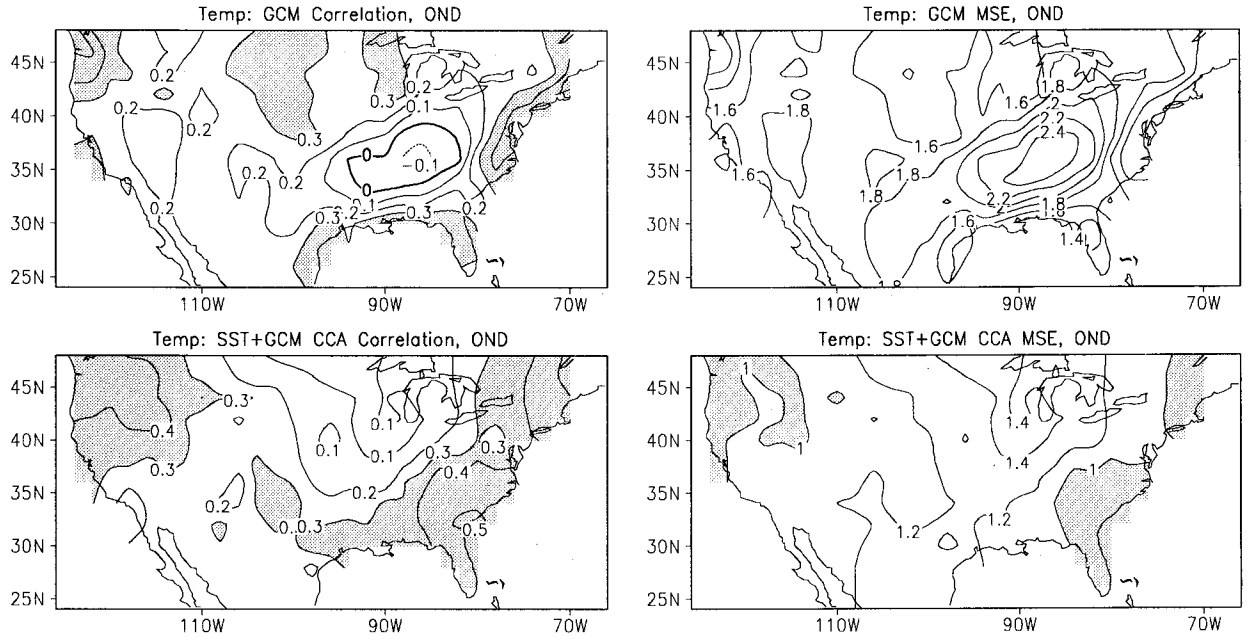

FIG. 12. Same as Fig. 9 except for Oct–Dec (OND).

b. The cross-validated CCA

All CCA empirical modeling was performed using cross validation to reduce artificial skill from overfitting. The cross validation was performed in the same manner as in Barnston (1994) and BS: for analysis of each month or season the data for one year were withheld, *all* calculations (including restandardization) were performed without it, and the performance of the resulting CCA relationships evaluated with the reserved data. This process was then repeated for every year in the record. In effect, each CCA problem (field–month or field–season) consisted of 43 separate CCAs evaluated on 43 separate independent cases. For more background on procedures and properties of cross validation refer to Michaelson (1987) and Livezey (1995).

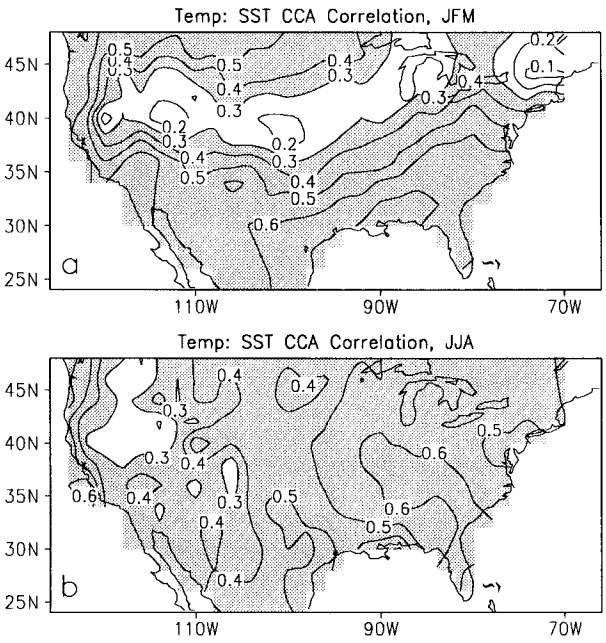

FIG. 13. Temporal correlations for specification of (a) JFM and (b) JJA U.S. surface temperatures by use of CCA with only global SSTs as predictors.

As described in section 1, three related but different sets of linear relationships (with variants) were studied with CCA, both of which can be thought of as simultaneous specifications of a particular atmospheric field. The six target fields (or predictands) separately specified for 12 periods over the year were the monthly and seasonal mean PNA region 700-hPa heights and U.S. surface temperature and precipitation, standardized and processed as described in the previous section. The first set of CCAs used only the global SSTs as simultaneous predictors, the second set only the model field corresponding to the target predictand field, while the third set used both (e.g., the global SST and the model U.S. precipitation fields for December were used to specify the December U.S. precipitation fields). The first set of analyses repeated the specification work of BS with some key differences and were compared to their results (see LS1), while the other two sets were new specification analyses that also constituted diagnoses (as well as formal vehicles for correction) of SST-dependent model systematic errors. Two variants of the combined SST and GCM predictor fields analyses were also examined. In the first of these associated model output, 700-hPa heights were added to the other predictors to

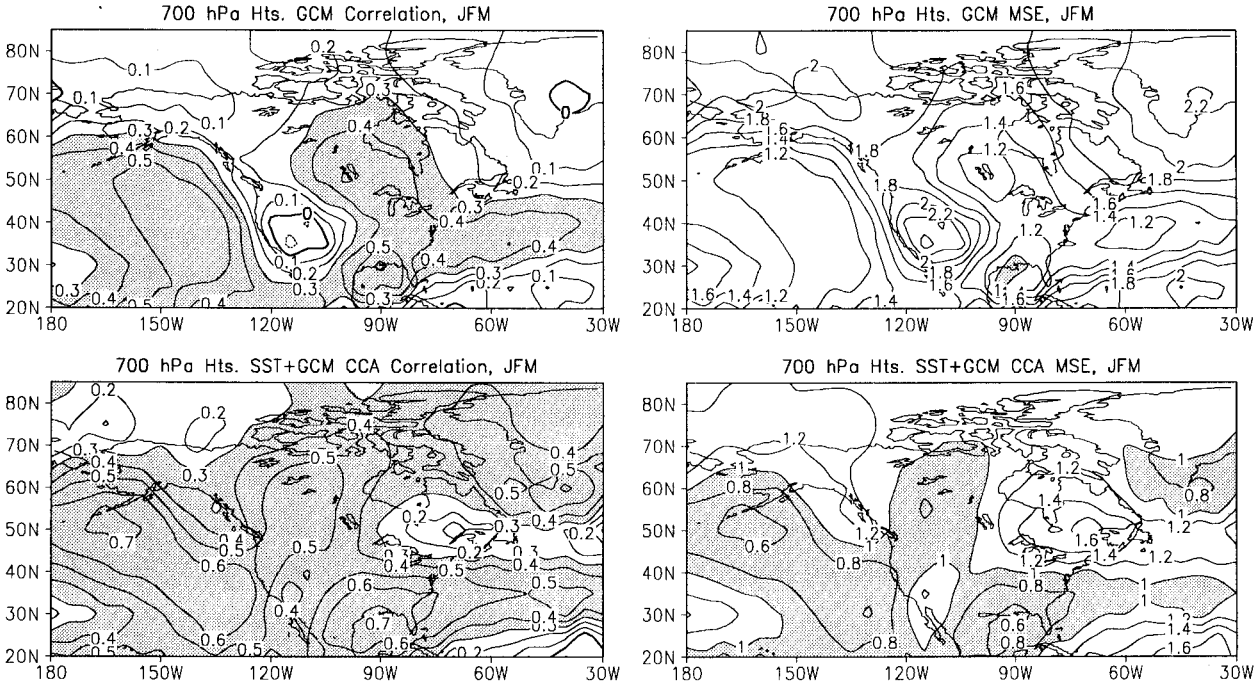

FIG. 14. Same as Fig. 9 except for JFM PNA region 700-hPa heights.

specify U.S. surface temperature or precipitation, and in the second a temporal sequence of four nonoverlapping global SST fields ending in the one coincident to the predictand field were used instead of just the concurrent one.

The mathematics and application of CCA at each iteration of the cross validation are described in LS1. Briefly, the steps were restandardization of the gridded data without the withheld year, predictor field weighting, separate PCAs of the predictor and predictand covari-

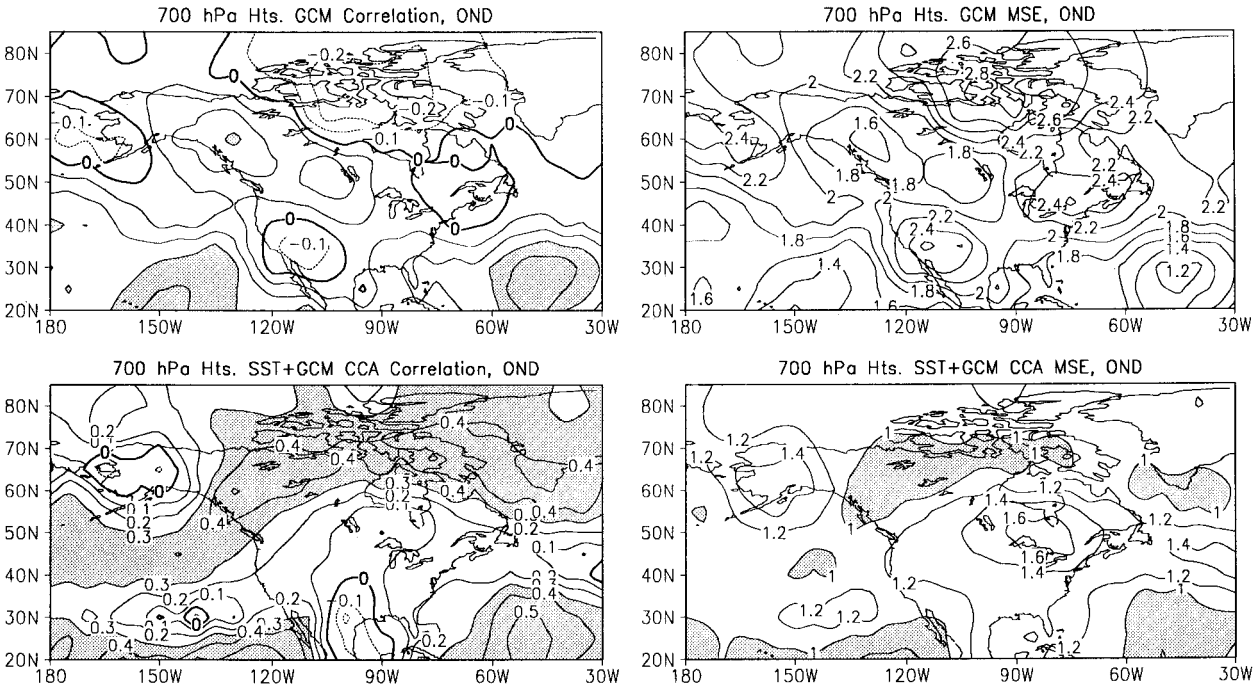

FIG. 15. Same as Fig. 14 except for OND.

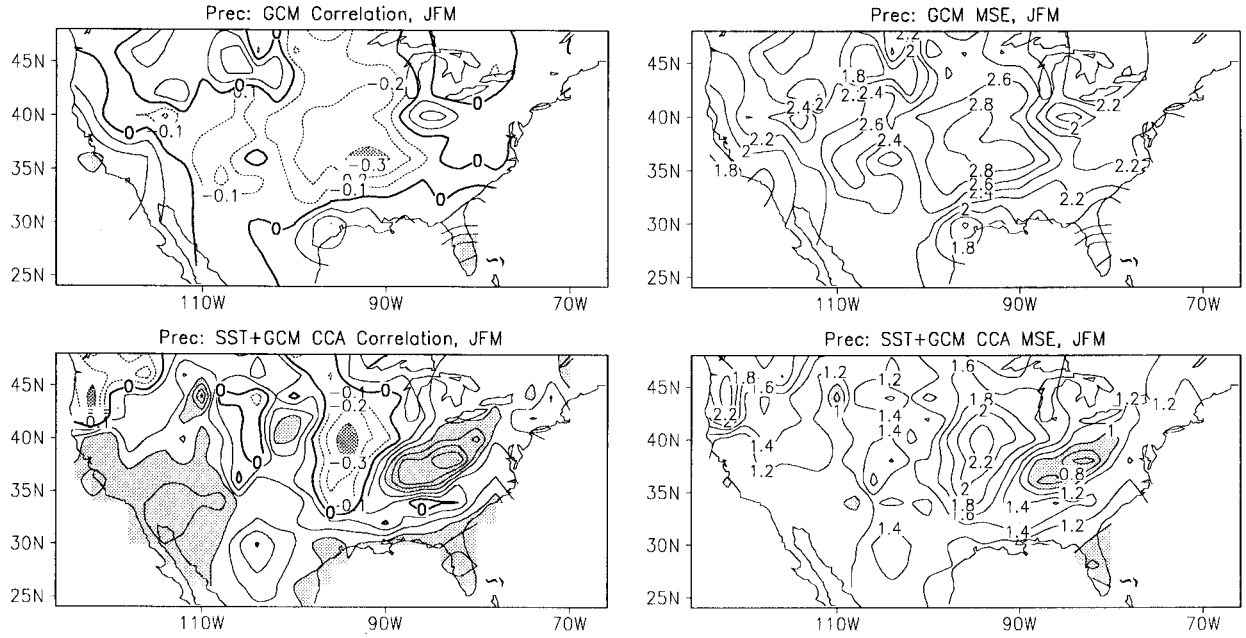

FIG. 16. Same as Fig. 9 except for JFM U.S. precipitation.

ance matrices, and CCA between truncated sets of predictor and predictand time series. A specification of the reserved predictand field was then made by restandardization of the reserved data with the climatologies of the complementary datasets (i.e., the other years), projection of the resulting predictor data on the leading CCA predictor modes, and use of these amplitudes and the transformation matrix to generate the specified field. The resulting sets of specified fields were then compared to the observations using the measures described in section 2d. Differences in the procedure to those in previous studies were in the assignment of the predictor weights and in the number of predictor and predictand PCs retained for CCA. The effect of the former did not turn out to be particularly important but that of the latter did. These methodological differences and their impact are described in LS1.

For all PCA here, climate division data have not been weighted, while SST data have been weighted with the cosine of the latitude (proportional to the area of a grid element) rather than the square root of the cosine. Thus, in these PCAs high-latitude SST grid points perhaps carried less weight than was desirable rather than more (see LS1). Additionally, the total variance of the SST data was reduced to less than 796, because variance was not redistributed among grid points to conserve the total. As discussed below, sensitivity to these weights was examined as part of our experimental strategy. They were varied by uniform further percentage reduction of variance of the global SST data.

c. Experimental strategy

Even though cross validation was used to reduce artificial specification skill, there were still a large number of choices and variants for CCA models and thereby a number of additional sources of artificial skill external to the cross validations. Specifically, monthly and seasonal mean specifications with three combinations of predictor fields (and other predictor field variants) with a wide range of predictor and predictand truncations were computed. Additionally, specifications from both global SSTs and model output were tried with a variety of interfield predictor weights.

To minimize vulnerability to additional overfitting, all of these choices were first made a priori. Initially, seasonal-mean specifications with combined, equally weighted global SST and model output predictor fields with fixed truncations were computed. For each variable, predictor and predictand truncations were then systematically varied to maximize specification performance season by season. These truncations were then adjusted so that they varied smoothly over the annual cycle. The final truncation sets (one each for PNA-region heights and U.S. surface temperature and precipitation) were then fixed in all subsequent sensitivity experiments, including the monthly mean specifications, the use of other combinations of predictor fields, and the use of different interpredictor weights.

d. Performance evaluation

Three measures were principally used to assess performance of the many versions of CCA specification/correction for each variable and season or month. 1) The temporal correlation at a data point between specification and observation, 2) the spatial *anomaly* correlation (as defined in L96), and 3) the squared error

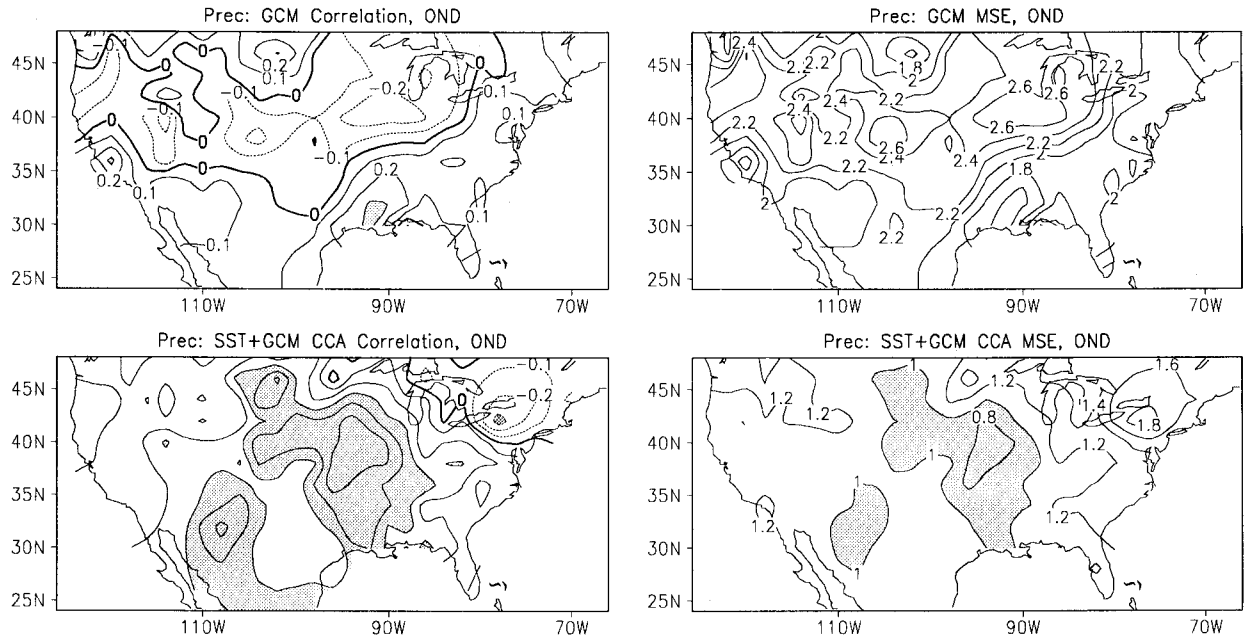

FIG. 17. Same as Fig. 16 except for OND.

between specified and observed maps of standardized data at a particular time. The map squared error was further decomposed into the sum of the squared error in the map means and the squared errors in the deviations from the map means.

The time correlations are presented below both in the form of maps and averaged over all grid points and the map measures presented both as time series and averaged over all time periods. The grand average time correlations and map measures are also graphed by season or month to depict the seasonality in specification/correction performance.

Last, the map measures are alternatively averaged over subsets of years for which it has been shown (L97) that there are strong teleconnections between moderate to large SST anomalies of both signs in the central equatorial Pacific Ocean and the North American atmosphere. The criterion for selection of a year was an SST anomaly with magnitude greater than or equal to 0.8°C in the area between 150°W to the date line and 5°N–5°S. Months in which this was the case are listed in Table 1 in L97. The reason this stratification of spatial scores was examined was to check to what extent ENSO-like episodes contributed to successful correction/specification.

3. Performance analyses

The results of the performance analyses for all three variable fields studied, PNA-region 700-hPa heights and U.S. surface temperature and precipitation, will be discussed simultaneously to contrast various aspects of their specification/correction performances.

a. Empirical modeling choices

The experimental strategy outlined in section 2c started with seasonal mean CCAs with two equally weighted predictors, namely global SSTs and the GCM-produced field corresponding to the observed predictand. The first objective was to find sets of predictor and predictand PC truncations for these empirical models that optimized specification/correction performance and varied smoothly from season to season for each variable field. These results were reported in LS1 (their Table 1) and *all* subsequent CCAs used these truncations. Monthly mean analyses employed the truncations selected for the corresponding seasonal mean CCAs centered on each month.

Differences in specification/correction performance between two-predictor CCAs and those that use only global SSTs will be the subject of much of the remainder of this paper, so they will not be mentioned in this subsection. However, it was determined that the results were relatively insensitive to a broad range of departures from equal weights (total variances) for the two predictor fields. In the rest of the paper results are presented for two-predictor CCAs of U.S. surface temperature or precipitation in which the GCM-produced field carried about 2.4 times the variance of the SST field and for PNA-region 700-hPa heights about 1.1 times. For the former pair performances were slightly better overall when the variances of the predictor fields were matched.

Otherwise it was found that the addition of either the GCM's 700-hPa heights as a predictor for U.S. surface temperature and precipitation or observed global SSTs at several lead times for all three variables did not no-

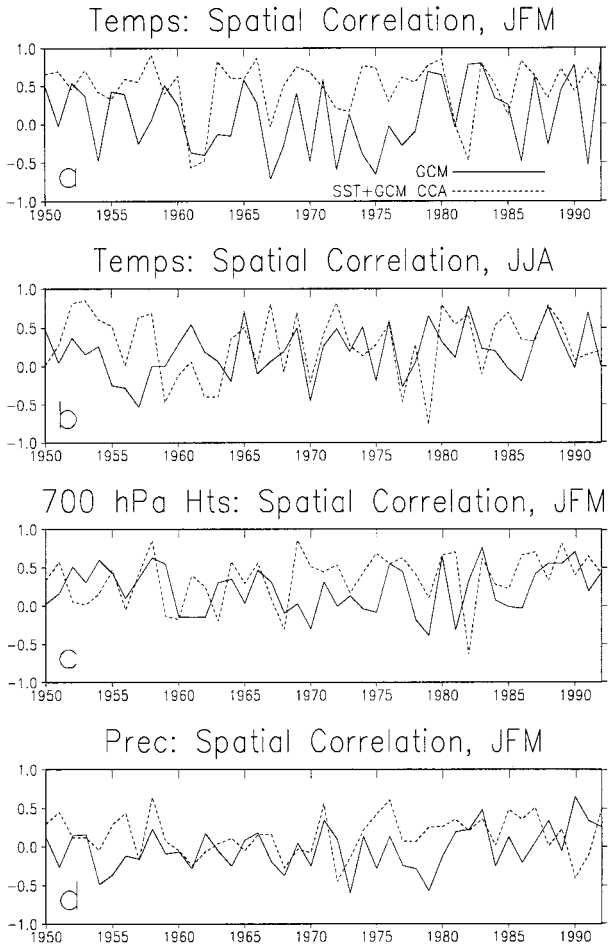

FIG. 18. Time series of spatial correlations for specification of (a) JFM U.S. surface temperatures, (b) JJA U.S. surface temperatures, (c) JFM PNA region 700-hPa heights, and (d) JFM U.S. precipitation, by use of uncorrected ensemble-mean GCM output (solid lines) and CCA with both global SSTs and model output as predictors (dashed lines).

ticeably enhance performance, and that CCA of monthly mean rather than seasonal mean data led to degradation in specification/correction performance. A few selected results for monthly means will be presented below.

b. Performance variability

1) SEASON TO SEASON

The seasonality of grand averages of the three main measures of specification performance are presented in Figs. 1–4. The figures are, respectively, for seasonal mean U.S. surface temperature, PNA-region 700-hPa heights, and U.S. precipitation, and monthly mean U.S. surface temperature. Each figure has a separate panel for spatial averages of temporal correlations, time averages of anomaly correlations, and mean-squared error, and each panel has separate curves for the uncorrected GCM maps and global SST predictor and two-predictor CCA corrections. Only Fig. 1 also contains curves for CCA corrections from just GCM-produced U.S. surface temperatures to exemplify the degradation in specification/correction performance when global SSTs are withheld as predictors (see the discussion in section 1). Basically expectations about the performance of specifications from model data only were confirmed. In the remainder of the paper we will only focus on the other two predictor combinations. Any generalizations made below about the figures do not necessarily refer to this additional set of one-predictor corrections in Fig. 1.

The most important result depicted in the figures is the considerable advantage in performance (by all measures) of the CCA specifications over the raw model output for the overwhelming majority of months/seasons. The only instances for seasonal specification in which this advantage is not present in the correlations are for late summer–early fall for temperature (Fig. 1) and late spring–early summer for 700-hPa heights (Fig. 2) and precipitation (Fig. 3). Indeed, for precipitation it would be difficult to argue for the existence of any overall specification skill without CCA correction. This is underlined by the fact that the mean-squared error of uncorrected GCM seasonal-mean precipitation uniformly exceeds the expected value of 2.0 for random specifications.

A somewhat more subtle, but still important, result of the aggregated performance analysis is that seasonal-mean temperature and precipitation corrections utilizing both the global SST and model-output fields as predictors more often than not outperform those utilizing only the SSTs (Figs. 1 and 3). Notable exceptions are for the May–July and June–August (JJA) periods in Fig. 1. The opposite is the case for seasonal-mean 700-hPa height (Fig. 2) and all examined monthly mean (e.g., Fig. 4) specifications.

As argued in section 1, a plausible explanation for superiority of some one-predictor specifications over two-predictor ones is that a greater proportion of covariance in these cases came from the atmosphere driving the ocean compared to the other cases. For instance, the monthly mean data used here more strongly reflect quasi-stationary large-scale circulation regimes (like blocking) to which mid- and high-latitude oceans are most responsive. Likewise, in contrast to the U.S. domain for seasonal surface temperature and precipitation data, the domain for 700-hPa heights encompasses a substantial part of the extratropical Atlantic and Pacific oceans that some of the structures it depicts can directly influence.

If these ideas have some validity then the full advantage of some of the specifications that only employ global SSTs as predictors will not be realized in NCEP's monthly and seasonal forecast operations. This is because coupled GCM prediction (of the ocean or atmosphere) currently has negligible skill in the extratropics at operational lead times, thereby necessitating the two-tier forecast process described in the introduction. This

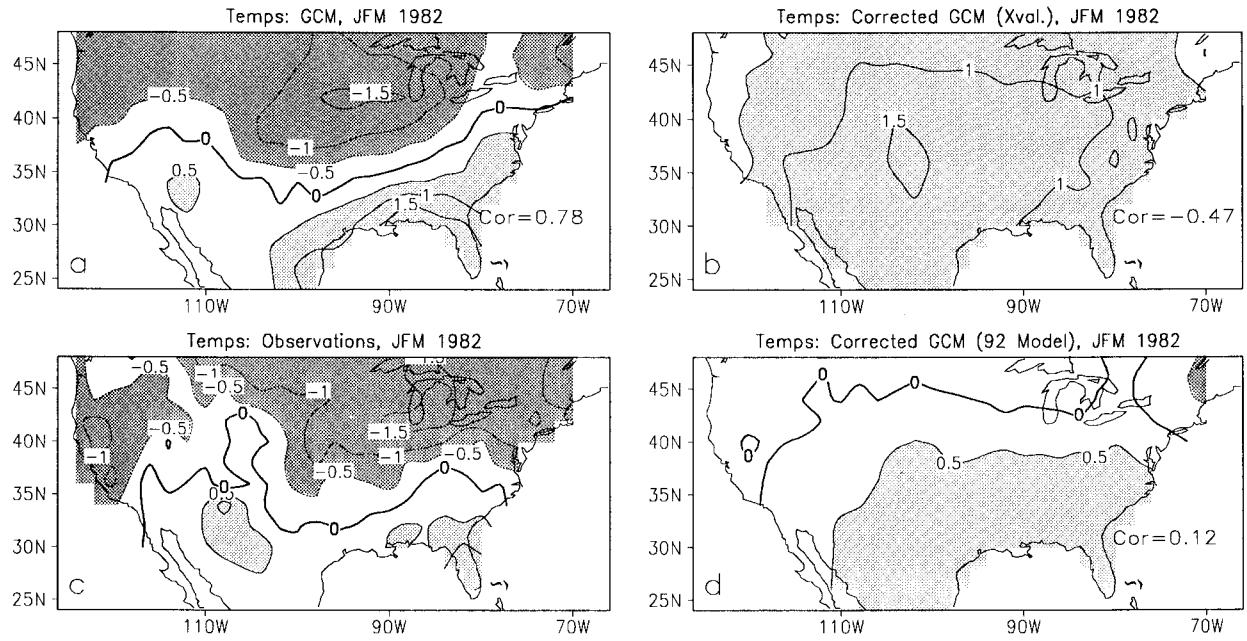

FIG. 19. Standardized U.S. surface temperatures for JFM 1982: (a) uncorrected ensemble-mean GCM output, (b) independently specified by use of CCA with both global SSTs and model output as predictors, (c) observed, and (d) dependently specified (cross-validated two-predictor CCA with 1992 excluded). Spatial correlations with observed shown in lower right corner where appropriate.

process precludes any possibility of exploitation of ocean–atmosphere relationships in which the atmosphere is dominant. Thus the relative merit of the one- and two-predictor specifications in Figs. 2 and 4 for CPC operations is not clear.

Another feature of season to season variation in aggregated correlation scores (the top two panels in each of Figs. 1–4) is the similarity between corresponding average temporal and spatial correlation curves. The principal exception to this occurs for uncorrected model seasonal-mean U.S. temperatures for the warm half of the year: average temporal correlations substantially exceed average spatial correlations. This behavior is consistent with a situation where the GCM reproduced the gross features of a large-scale temperature trend but was unable to capture case-to-case variations in spatial structure of the temperature anomalies.

Finally, recall that the squared errors of maps were decomposed into squared errors of map averages and squared errors of departures from map averages (which we will refer to as squared anomaly errors). Both types of error were substantially reduced in all of the specification/corrections examined in this section (see Fig. 5 for the seasonal-mean, two-predictor CCAs). Of somewhat more interest are the seasonalities and relative magnitudes of the two types of squared differences.

For example, the only instances in which squared map-average errors are comparable to squared anomaly errors are for U.S. surface temperature maps (uncorrected and corrected) in the cold half of the year (Fig. 5a). These are partially attributable to the strong ENSO-dependent model biases described in L97. Overall, temperature maps have the greatest map-average squared errors, the smallest squared anomaly errors, and the most pronounced seasonality for the two errors. Otherwise, the PNA-region 700-hPa height maps have overall the smallest squared map-average errors (Fig. 5b) and the U.S. precipitation maps the greatest squared anomaly errors (Fig. 5b). The two types of squared errors generally vary out of phase from season to season for all three variable/domains.

2) ENSO

The stratification of average spatial correlations and mean squared errors according to the sign and magnitude of the SST anomaly in the central equatorial Pacific Ocean (see section 2d for details) produced several unanticipated results. These are well exemplified by seasonal-mean, two-predictor CCAs, so stratified average spatial correlations for these specifications are graphed in Figs. 6–8. Discussion of the detailed behavior of the curves for positive and negative SST anomalies will be avoided because of implied sampling error of the plotted quantities as a result of small sample sizes.

It was, of course, anticipated that noticeable performance gains from correction/specification would be realized for ENSO-like situations and this was uniformly observed (Figs. 6–8). After all, these are the cases for which the raw GCM output started with demonstrable simulation skill, whereas for cases with near normal central equatorial Pacific SSTs, the GCM demonstrated little or no skill (Figs. 6a, 7a, and 8a). Nevertheless, CCA corrections also substantially enhance perfor-

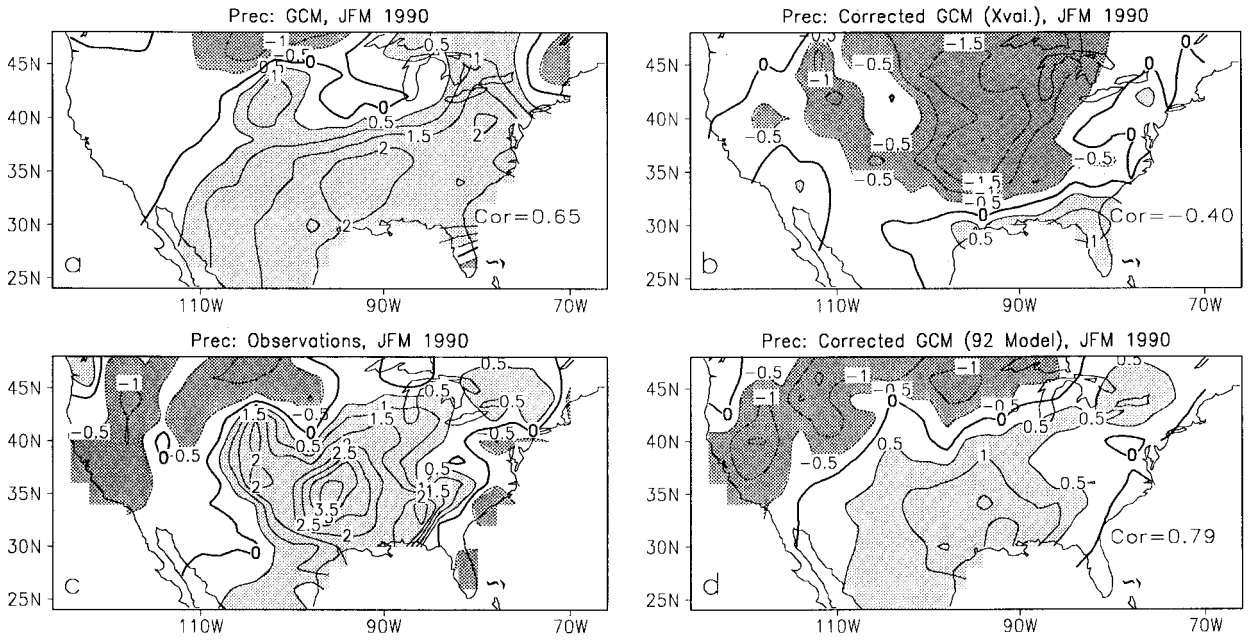

FIG. 20. Same as Fig. 19 except for JFM U.S. precipitation.

mances for cases with neutral-like ENSO conditions, at least for U.S. surface temperatures and PNA-region 700-hPa heights. Thus, it is likely that there are additional signals besides ENSO between the global ocean and the atmosphere that are exploited by CCA. We have unambiguously identified two such relationships that are discussed thoroughly in a companion paper (Livezey and Smith 1999b, this volume; hereafter LS2). In the case of U.S. precipitation, average specification performances of the raw GCM output for near-normal SSTs in the key area are absolutely (especially in consideration of the reduced sample size) indistinguishable from that for random specifications throughout the year (Fig. 8a and a practically uniform mean-squared error of 2.0).

The other unanticipated result of our stratified performance analyses (not shown) is that the average mean-squared errors for positive or negative SST anomaly cases rarely are less and more often than not are somewhat greater than those for near-normal cases, but that these differences between ENSO and non-ENSO situations mostly disappear after CCA corrections.

3) SPATIAL

A selection of maps of uncorrected and corrected temporal correlations and mean-squared errors are shown for successful specifications to underline a number of points. Three such points are illustrated by the sets of four maps for U.S. surface temperatures in four different seasons (Figs. 9–12). First, at least for this variable, CCA correction has resulted over much of the year in large increases in the area where successful specifications can be potentially made. These gains would not have been possible without the presence of substantial systematic errors in the GCM. Second, because performance quality is nonuniformly distributed, there are regions in which these specifications will be very reliable (e.g., the southeastern United States in the winter, Fig. 9). Third, the differences between the maps of performance measures for uncorrected and corrected GCM output largely reflect ENSO-dependent systematic errors described by L97, but not exclusively. For example, the performance charts for the summer season (Fig. 11) do not reflect any documented ENSO signal with which the authors are familiar, but are linked to much lower frequency variability discussed in LS2.

Generally the patterns of performance quality for successful one- and two-predictor specifications are not particularly different from each other. Good examples are the temporal correlation maps for winter temperatures in Figs. 9 and 13a, where overall the latter has modestly larger values than the former. However, counter examples are provided by the charts for summer temperatures in Figs. 11 and 13b, in which the gains of the former over the latter are substantially greater in the western United States. This is one of two instances in which the one-predictor specification for 3-month mean temperatures was substantially more successful than the two-predictor correction. As noted earlier one cannot be sure that this difference in performance would carry over in an operational forecast setting.

For completeness and to emphasize the points already made, charts corresponding to those in Figs. 9 and 12 are presented for PNA-region 700-hPa heights and U.S.

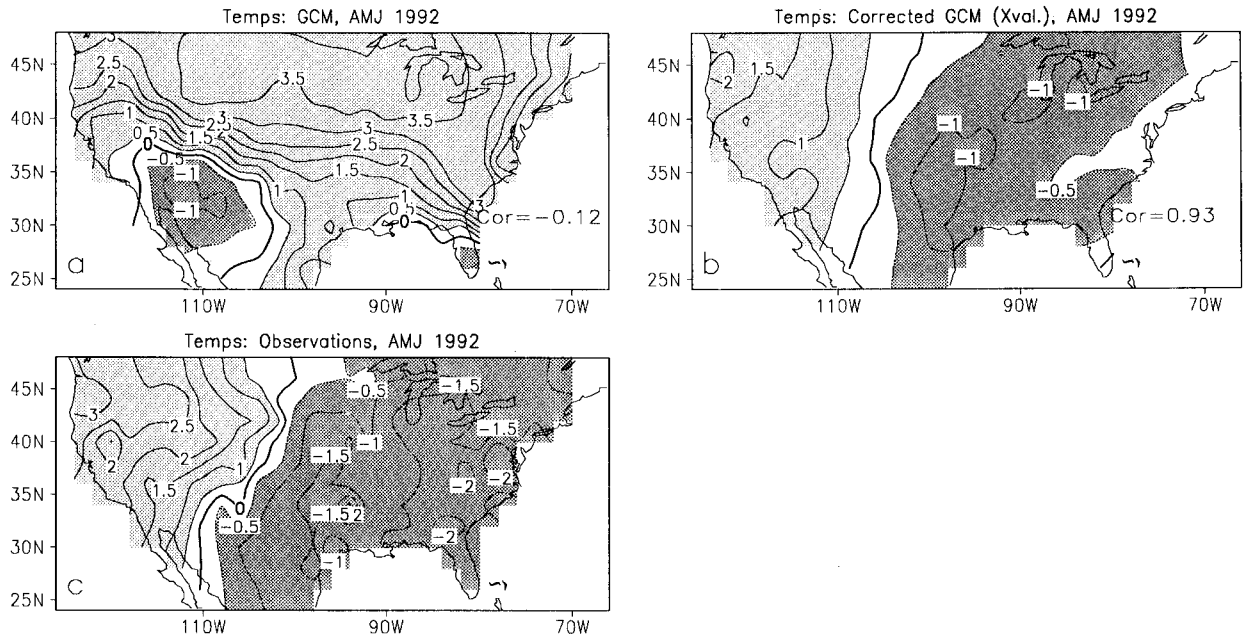

FIG. 21. Same as Fig. 19 except for AMJ U.S. surface temperatures and no (d) panel.

precipitation in Figs. 14 and 15 and Figs. 16 and 17, respectively. The differences in these cases between uncorrected and corrected maps dominantly reflect the ENSO-associated systematic errors described at length in L97.

4) TEMPORAL

To round out the discussion of performance variability, time series of anomaly correlations for uncorrected GCM output and two-predictor CCA specification are presented in the four panels of Fig. 18 for the same variables/seasons for which maps of temporal correlations were presented in Figs. 9, 11, 14, and 16, respectively. Some of the secular variations in relative performance present in the time series are interesting but difficult to interpret physically. Also evident are some notable failures of the cross-validated CCAs in which uncorrected model fields with anomaly correlations well in excess of 0.5 are replaced with ''corrected'' maps with moderately negative anomaly correlations with the observations.

Two of the most prominent of these failures, January–March (JFM) 1982 temperature (Fig. 18a) and 1990 precipitation (Fig. 18d) specifications, are illustrated in Figs. 19 and 20, respectively. In addition to the raw GCM, observed, and cross-validated CCA-produced fields, dependently fitted fields are also displayed. For the latter, we arbitrarily chose the set with 1992 withheld from the 43 sets available to us through the cross-validation experiment.

Note for each case the outstanding match between the GCM and observations and the poor match between the cross-validated specification (which had no prior knowledge of the case) and the observations. In contrast the dependently fitted JFM 1990 precipitation specification (Fig. 20d) did match the observations well, but the dependently fitted JFM 1982 temperature specification (Fig. 19d) did not, despite the fact that it also had prior knowledge of its respective case outcome.

It is difficult to resolve the cause of these or other specification failures. For the JFM 1982 temperature specification the GCM may have duplicated the structure of the observed pattern by accident. Because there are relatively few degrees of freedom in the interannual variability of both the model and observational seasonal temperature fields, chance matchups would have occurred from time to time. Because seasonal precipitation fields have more degrees of freedom than temperature fields, a more likely explanation than chance for the GCM's successful simulation of JFM 1990 precipitation and the correction's lack of success is undersampling of the situation. This impression is reinforced by the outstanding representation of the case by the dependently fitted CCA. However, a warning is necessary here: for the signal analysis work presented in LS2 performance score inflation for CCA on the entire dataset without cross validation was very large. Thus, the success of the dependently fitted specification in Fig. 20d may be just the result of overfitting.

Examples of situations quite opposite to those exemplified by Figs. 19 and 20 are presented in Figs. 21 and 22. These are instances when the GCM's performance suggests no skill whatsoever, but the cross-validated CCA specifications are extremely successful. Here, the reasons for success are clear; both periods

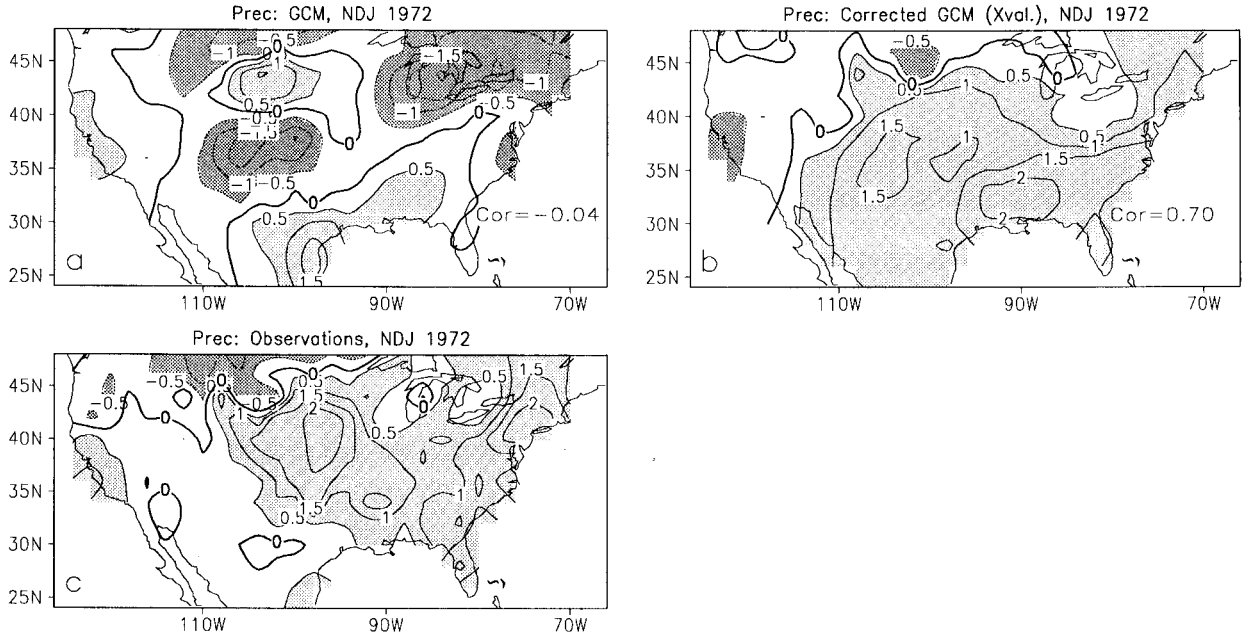

FIG. 22. Same as Fig. 19 except for NDJ U.S. precipitation and no (d) panel.

represented were during warm ENSO episodes and the raw GCM and observed fields to a large extent exhibit the composite patterns for these periods associated with moderate to strong positive SST anomalies in the central equatorial Pacific Ocean (L96). For example, the uncorrected GCM solution for April–June 1992 temperature (Fig. 21a) reflects the model's tendency to hold onto the characteristic ENSO warm-episode winter pattern too long into the spring.

4. Concluding remarks

A large ensemble of long prescribed SST–GCM integrations and the global SSTs that drove them have been matched with monthly and seasonally averaged observed PNA-region 700-hPa heights and U.S. surface temperatures and precipitation. These data have been employed with CCA to linearly explore systematic errors in SST-forced, ensemble-mean responses of the model, and the feasibility of enhancing the empirical specification of these observed fields from SSTs through use of the model output as well. The latter is equivalent to statistical correction of the model fields. Our principal findings were

- Systematic errors in SST-forced model variability have large correctable linear components.
- The linear specification of U.S. surface temperatures and precipitation by either the prescribed SST model simulations of the fields or the global SSTs used in the GCM runs can generally be improved by combination of the two predictor fields.
- Detectable systematic errors of the forced GCM in PNA-region 700-hPa heights and U.S. surface temperatures exist and are linearly correctable for cases with a nonactive or transitional ENSO as well as for cases with strong ENSO episodes. In the companion paper (LS2) two sources for this specification skill from other than ENSO are examined in some detail.

The approach used here can be applied whenever a sufficient ensemble of long prescribed SST GCM simulations and corresponding observations are available. Indeed, given these datasets, our method or appropriate alternatives should be pursued in each model specification/correction problem for which the ENSO or other SST-forced signals are likely to be important sources of interannual variability. Opportunities to do this collaboratively will be sought.

A final point is that the choice here of CCA implies the representation of at least some relationships that are likely nonlinear (L97) as linear approximations. A number of techniques are available to address this problem, but all are hampered by the inevitable trade-off between treatment of the nonlinearity and the increased sampling error of nonlinear methods. Several empirical approaches, both crude and sophisticated, have been tried to evaluate whether positive gains from this tradeoff are possible with existing datasets, but the results so far have been disappointing. Ultimately, the best solution to this problem is the expansion of key datasets, like SSTs, sea level pressures, and surface temperatures and precipitation as far back prior to 1950 as possible. The successful extension of comprehensive spatial analyses of these crucial variables will require the use of sophisticated reconstruction methods (e.g., Kaplan et al. 1997 or Smith et al. 1998), which, in turn, will be enormously

facilitated by acquisition and digitization of currently unexploited data sources.

REFERENCES

Barnett, T. P., and R. Preisendorfer, 1987: Origins and levels of monthly and seasonal forecast skill for United States surface air temperatures determined by canonical correlation analysis. *Mon. Wea. Rev.,* **115,** 1825–1850.

Barnston, A. G., 1994: Linear statistical short-term climate predictive skill in the Northern Hemisphere. *J. Climate,* **7,** 1513–1564.

——, and R. E. Livezey, 1987: Classification, seasonality, and persistence of low-frequency atmospheric circulation patterns. *Mon. Wea. Rev.,* **115,** 1083–1126.

——, and T. M. Smith, 1996: Specification and prediction of global surface temperature and precipitation from global SST using CCA. *J. Climate,* **9,** 2660–2697.

Graham, N., T. P. Barnett, R. Wilde, M. Ponater, and S. Schubert, 1994: On the roles of tropical and midlatitude SSTs in forcing interannual to interdecadal variability in the winter Northern Hemisphere circulation. *J. Climate,* **7,** 1416–1441.

Kaplan, A., Y. Kushnir, M. A. Cane, and M. B. Blumenthal, 1997: Reduced space optimal analysis for historical datasets: 136 years of Atlantic sea surface temperatures. *J. Geophys. Res.,* **102,** 27 835–27 860.

Kumar, A., M. Hoerling, M. Ji, A. Leetmaa, and P. Sardeshmukh, 1996: Assessing a GCM's suitability for making seasonal predictions. *J. Climate,* **9,** 115–129.

Livezey, R. E., 1995: The evaluation of forecasts. *Analysis of Climate Variability: Applications of Statistical Techniques,* H. von Storch and A. Navarra, Eds., Springer-Verlag, 177–195.

——, and T. M. Smith, 1999a: Considerations for use of the Barnett and Preisendorfer (1987) algorithm for canonical correlation analysis of climate variations. *J. Climate,* **12,** 303–305.

——, and ——, 1999b: Covariability of aspects of North American climate with global sea surface temperatures on interannual to interdecadal timescales. *J. Climate,* **12,** 289–302.

——, M. Masutani, and M. Ji, 1996: SST-forced seasonal simulation and prediction skill for versions of the NCEP/MRF model. *Bull. Amer. Meteor. Soc.,* **77,** 507–517.

——, ——, A. Leetmaa, H. Rui, M. Ji, and A. Kumar, 1997: Teleconnective response of the Pacific/North American region atmosphere to large central equatorial Pacific SST anomalies. *J. Climate,* **10,** 1787–1820.

Michaelson, J., 1987: Cross-validation in statistical climate forecast models. *J. Climate Appl. Meteor.,* **26,** 1589–1600.

Reynolds, R. W., and T. M. Smith, 1994: Improved global sea surface temperature analysis using optimum interpolation. *J. Climate,* **7,** 929–948.

Smith, T. M., R. W. Reynolds, R. E. Livezey, and D. C. Stokes, 1996: Reconstruction of historical sea surface temperatures using empirical orthogonal functions. *J. Climate,* **9,** 1403–1420.

——, R. E. Livezey, and S. S. Shen, 1998: An improved method for analyzing sparse and irregularly distributed SST data on a regular grid: The tropical Pacific Ocean. *J. Climate,* **11,** 1717–1729.

Ward, M. N., and A. Navarra, 1997: Pattern analysis of SST-forced variability in ensemble GCM simulations: Examples over Europe and the tropical Pacific. *J. Climate,* **10,** 2210–2220.