# Potential predictability of seasonal precipitation over the United States from canonical ensemble correlation predictions

K.-M. Lau
Climate and Radiation Branch, NASA Goddard Space Flight Center, Greenbelt, Maryland, USA

Kyu-Myong Kim
Science Systems and Applications, Inc., Lanham, Maryland, USA

Samuel S. P. Shen
Department of Mathematical Sciences, University of Alberta, Edmonton, Canada

Received 25 October 2001; revised 2 January 2002; accepted 7 January 2002; published 2 April 2002.

[1] Potential predictability of seasonal precipitation over the US is explored using a new canonical ensemble correlation (CEC) prediction model, which optimally utilizes intrinsic sea surface temperature (SST) variability in major ocean basins. Results show that CEC yields a remarkable (10–20%) increase in baseline prediction skills for seasonal precipitation over the US for all seasons, compared to traditional statistical predictions using global SST. While the tropical Pacific, i.e., El Niño, contributes to the largest share of potential predictability in the southern tier States during boreal winter, the North Pacific and the North Atlantic are responsible for enhanced predictability in the northern Great Plains, Midwest and the southwest US during boreal summer. Overall, CEC significantly reduces the spring-summer predictability barrier over the conterminous US, thereby raising the skill bar for seasonal precipitation predictions. *INDEX TERMS:* 3309 Meteorology and Atmospheric Dynamics: Climatology (1620); 3339 Meteorology and Atmospheric Dynamics: Ocean/atmosphere interactions (0312, 4504); 3354 Meteorology and Atmospheric Dynamics: Precipitation (1854); 3367 Meteorology and Atmospheric Dynamics: Theoretical modeling

## 1. Introduction

[2] It is well known that sea surface temperature (SST) in the tropical Pacific associated with the El Niño is the main cause of enhanced seasonal precipitation forecasting skill for the United States (US) during the boreal winter. However the skill drops dramatically in the spring, reaching a minimum in the warm season. The dramatic reduction in forecast skill from winter to summer through the spring season is known as the "spring predictability barrier", which has been endemic in both statistical and dynamical seasonal forecasts. Recent studies have found significant predictability from tropical and extratropical Pacific SST on warm season precipitation over the upper Great Plains and Atlantic States of the US during El Niño summers [*Wang et al.*, 1999; *Ting and Wang*, 1997; *Ropelewski and Halpert*, 1986]. However, the forecasting skill was still relatively low in summer, even during time of strong SST signal in the tropical Pacific.

[3] It has been suggested that the reduced precipitation predictability in the summer over the US stems from the weaker, and more poleward position of the upper level westerly flow in the northern hemisphere, thus limiting the transmission of tropical SST influence to the US continent [*Lau and Peng*, 1992]. However, SST variability in other ocean basins, especially the extratropics may begin to have an impact on US precipitation in summer [*Higgins et al.*, 2000]. In addition to SST, other factors such as soil moisture, snow cover and vegetation may influence US precipitation predictability for both summer and winter [*Koster et al.*, 2000]. The canonical ensemble correlation (CEC) prediction scheme presented in this paper has been developed with the purpose of systematically exploring potential predictability from various sources and carrying out experimental climate predictions. In this paper, the rudiments of the CEC prediction scheme is introduced and results for US seasonal precipitation prediction are presented.


0094-8276/02/2001GL014263$05.00

## 2. Basic Concepts and Procedure

[4] The basic concept of the CEC is shown schematically in Figure 1. Given an initial climate state, denoted by $\Omega_i$, the real climate system evolves into a climate state $\Omega_f$ in some future time $\Delta t > 0$. Because of changing boundary conditions and the chaotic nature of climate evolution from the initial state, $\Omega_i$ or arbitrarily close-by states, a range of possible future states, within the "event cone", R, may be possible. A given predictor, based on a set of canonical variables, may produce an event cone $P_1$, that maximizes the overlap (or minimize the distance metric) with R. If $P_1$ and R have a large overlap, then $P_1$ is considered a good predictor for $\Omega_f$. Other predictors such as $P_2$ may produce a lesser overlap with R, but it may capture different climatic states spanned by the event cone R, not covered by $P_1$. This situation can be applied to other predictors $P_n$, which span yet other portions of the event cone R. The objective of the CEC is to construct, based on the totality of individual canonical correlation predictions, an "ensemble event cone" that maximizes the overlap with R, taking into account all possible outcomes. In CEC, the number of predictors, represented by $P_n$, is unlimited, and the predictors are not subject to orthogonality constraints.

[5] Specifically, the CEC prediction model is based on linear regressions that maximize the correlation between the weighted integral of SST and precipitation fields including an equal-area correction factor [*Kim and North*, 1998; *Shen et al.*, 2001]. In this study, the monthly anomaly data of SST and precipitation are obtained by removing the climatology and linear trend and normalized by the sample standard deviation at each grid box. Hereafter only detrended data are used. The matrix for the correlation eigen-problem is solved in the EOF space to obtain the maximum correlation between the canonical correlation variables of SST and precipitation [*Barnett and Preisendorfer*, 1987]. The predicted precipitation field, $P(t + \Delta t)$, is then expressed in terms of the canonical variables of $SST(t)$, where $t$ denotes time and $\Delta t$ is the forecast lead time. To maximally extract precursory signals in global SST, the world ocean is partitioned into non-

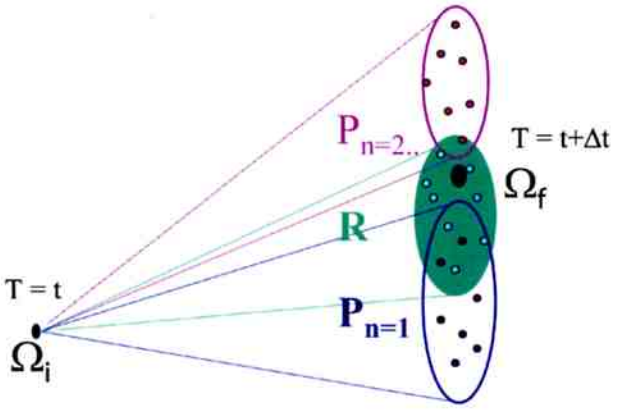


**Figure 1.** Schematic diagram illustrating the basic concepts of Canonical Ensemble Correlation (CEC) prediction.

overlapping sectors, and separate forecasts each using up to 6 canonical variables, are made based on different ocean sectors. This is a crucial step, because predictions from individual sectors allows the natural SST variability in that sector to be fully utilized in prediction. The CEC forecast is then obtained at each grid box as a weighted average of the individual forecasts. Here, the CEC has an advantage over traditional Canonical Correlation Analysis (CCA) prediction which employs direct CCA from global SST, where intrinsic ocean variability outside of the tropical Pacific is often obscured by the dominant El Niño signals.

[6] The precipitation data used for this study are derived from optimal interpolation from over 17,000 stations in the Global Historical Climatological Network Version 2 and the Climate Anomaly Monitoring System for the period 1951–1999 [*Chen et al.*, 2001]. The data cover the global land with a spatial resolution of 2.5 degrees latitude-longitude. In this work, only data over the US continent are used. The SST data are obtained from the US National Center for Environmental Prediction for the same period with a spatial resolution of 2° × 2° latitude-longitude [*Smith et al.*, 1996]. To reduce small-scale noise, the SST data are further averaged to boxes of 4° × 6° latitude-longitude.

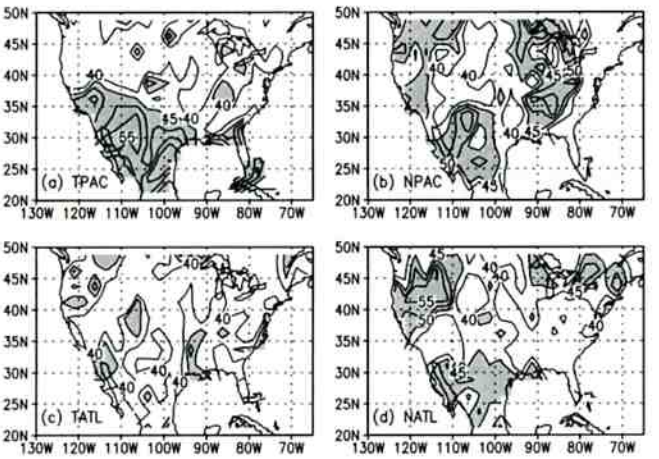


**Figure 2.** Three-category hit score (%) for DJF precipitation prediction derived from SST anomalies from (a) Tropical Pacific, (b) North Pacific, (c) Tropical Atlantic, and (d) North Atlantic. A hit score of 33% or less indicates the absence of prediction skill. At 5% and 1% significance levels, the greater than 33% hit scores are approximately 45% and 48% respectively. The area with the hit score greater than or equal to 45% is shaded.

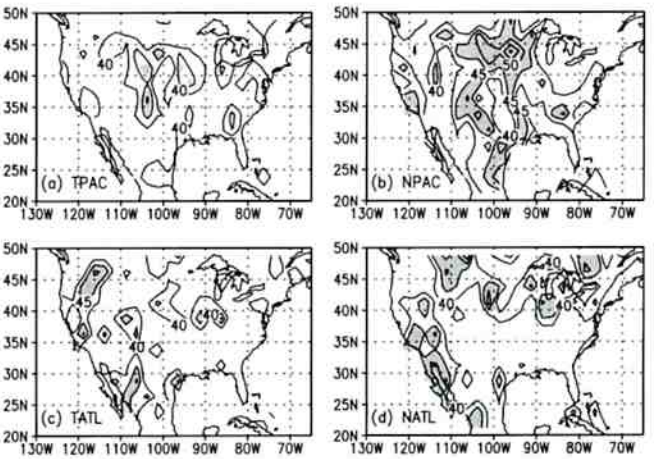


**Figure 3.** Same as in Figure 2, except for JJA.

[7] For validation, we use a bootstrapping techniques, in which the EOFs and canonical correlation of $SST(t)$ and $P(t + \Delta t)$ are computed using 48 years, with the forecasting year taken out. The climatologies, statistics, and weights are re-computed for each 48 years. As a result, 49 hindcasts can be obtained. To avoid spurious skill and to ensure the robustness of the statistics, we have varied the training periods in a range from 48 to 40 years, and conducted cross-validations of the forecasts [cf. *Barnston and Van den Dool*, 1993; *Van den Dool*, 1987]. The results presented here are largely independent of the training period and the cross-validation technique used. In this paper, we focused only on the zero-lag predictions i.e., specification of rainfall from simultaneous SST. Predictability of individual forecasts with $\Delta t > 0$ and further applications of CEC are reported elsewhere. The zero-lag "prediction" represents the potential predictability for precipitation given perfect knowledge of the simultaneous SST field. In practice, the zero-lag CEC prediction can be used in conjunction with a two-tier forecast scheme, in which the SST is predicted by an ocean model or a coupled model.

## 3. Evaluation of Potential Predictability

[8] To evaluate the forecast skill, a three-category "hit" score is used [*Wang et al.*, 1999]. For each grid box, the observed precipitation values for a given season in 49 years are sorted in an ascending order. Three categories are formed according to the first third (below normal), the middle third (normal), and the last third percentiles (above normal). If the forecast and observed precipitations are in the same category, the forecast is a "hit". The forecasting skill is the hit rate, which is the number of correct forecasts divided by the total number of forecasts, i.e., 49. For a no-skill random forecast, the expected hit score is 33.33%. For a sample size of 49 years the hit rate of 45% (48%) is significantly different from a random forecast at the 5% (1%) significant level. Other skill scores have been tested. The results reported here are robust and independent of the choice of skill scores.

[9] As a test of the CEC, the hit scores for US wintertime (DJF) and summertime (JJA) predictions, based on SST in individual ocean basins, have been computed. The ocean basins are the tropical Pacific (TPAC, 30°S–30°N), the North Pacific (NPAC, north of 30°N), the tropical Atlantic (TATL, 30°S–30°N), the North Atlantic (NATL, north of 30°N), and the Indian Ocean (IND, north of 30°S). We have also computed the skill score using the global ocean, i.e., all ocean basins. The all-ocean skill score is comparable to that computed from TPAC, because of the dominance of the ENSO signal in an all-ocean SST EOF decomposition (See discussion for Figure 6). In all the results shown, six dominant EOF modes are used. The skill scores vary only slightly if more EOFs have been used. Figure 2 shows the DJF forecast results

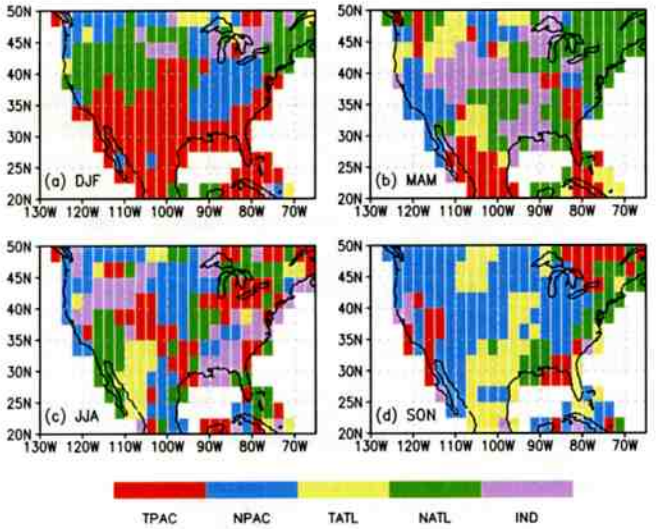


**Figure 4.** The "influence function" on US precipitation by SST from dominant variability in different ocean basins. The color indicates the most important influence from the corresponding ocean basins. For example, the red region is most influenced by the Tropical Pacific: (a) Season DJF, (b) season MAM, (c) season JJA, and (d) SON.

using the SST from TPAC, NPAC, TATL, and NATL. Figure 2a indicates that TPAC has the overall highest score and the most spatially coherent score pattern, concentrating in southwestern US and Mexico, and southeastern US. The NPAC (Figure 2b) contributes significant scores (>45%) in the west and southwest of US and the Great Lakes and Ohio Valley. The TATL (Figure 2c) and IND (not shown) appear to have the least skill scores, compared to the other ocean basins while the NATL (Figure 2d) is responsible for the high hit rates in the Pacific Northwest, northeast and southwest US. The skill score for JJA from individual ocean basin, as shown in Figure 3, is much reduced and less organized, with the exception of the NPAC, which appears to possess prediction skill in a region stretching from the Gulf Coast of Texas to the northern Great Plains and the Midwest.

[10] To evaluate the influence of each ocean basin on precipitation prediction over different regions of the US, each grid box is identified with the ocean basin of "maximal" influence, based on the highest temporal correlation between predicted and observed precipitation. Figure 4 shows the distribution of the "influence function" for US precipitation predictability for all four seasons. During DJF (Figure 4a), it is clear that the TPAC has the strongest influence across the southern states, spanning the southwest, Mexico, the Gulf Coast, the southeast and the eastern seaboard. The TPAC influence reaches up to the mountain states and central US. The NPAC has the strongest influence in the Ohio Valley and the northwest, while the NATL controls the northeastern seaboard, Northern California, Idaho, and Montana. During MAM (Figure 4b), the influence of the TPAC reduces substantially, while the NATL gains influence in the northeast and along the East Coast. Other regions appear to have comparable, but generally weaker influences (relative to the wintertime) from different ocean basins. The previously noted lower skill score in JJA is also reflected in the rather disorganized pattern of the influence function all over the US (Figure 4c), with perhaps the exception of northern Great Plains which is subject to the strongest influence from the North Pacific. The JJA pattern suggests a lack of a single dominant SST-related forcing mechanism for US summertime precipitation variability. In SON (Figure 4d), the dominant influence from the NPAC emerges over the Pacific Northwest, the central mountain and southwest states, and the Northern Great Plains/Midwest region. Elsewhere, the TATL appears to have gained influence relative to the other ocean basins. It is clear from the foregoing results that El Niño effect, through SST in the TPAC, is not always the major contributor to rainfall signal over the US, especially in the warm seasons.

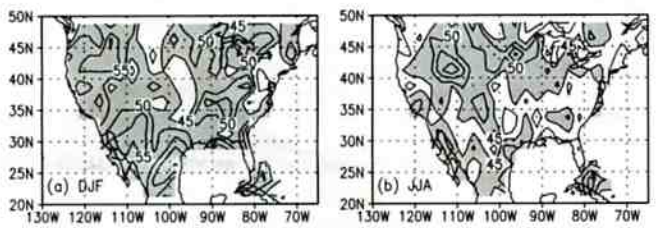


**Figure 5.** The spatial distribution of the CEC precipitation skill score over the US for (a) DJF and (b) JJA. The area with the hit score greater than or equal to 45% is shaded.

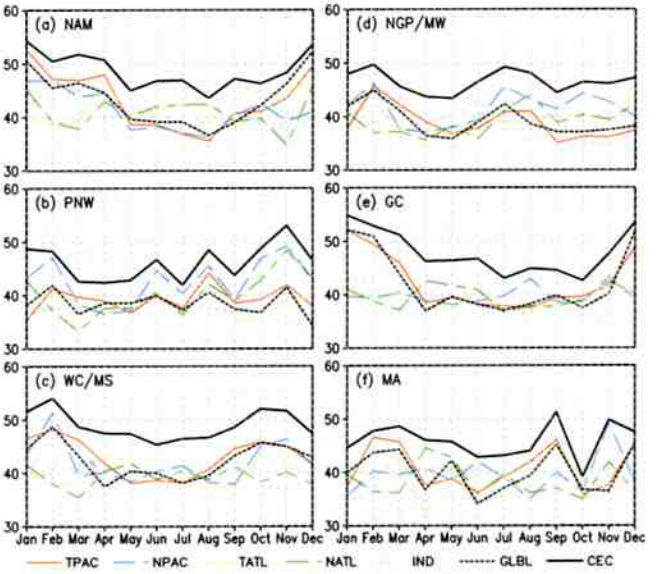


**Figure 6.** The seasonal cycle of the mean seasonal forecast skill for the selected regions: (a) North American Monsoon, (b) Pacific Northwest, (c) West Coast and Mountain States, (d) Great Plains and Midwest, (e) Gulf Coast, and (f) Mid- Atlantic coast. The thick solid line indicates the CEC forecast. The forecasts from the five individual ocean basins and global ocean are as indicated by different colored lines, e.g., the red solid line represents the skill from the TPAC.

[11] The CEC forecast is obtained from each individual ocean-basin forecast by assigning an appropriate weight for each forecast at every grid point. In this paper, we will show results for the simplest version of the CEC forecast, which is obtained by assigning a weight of unity to the most skillful forecast and zero to the rest, based on the 48-year training period. The result of this CEC is not too different from those based on the super-ensemble approach [*Krishnamurti et al.*, 2000] with forecast weights proportional to the regression coefficients. From a comparison of Figure 5 and Figures 2 and 3, it is clear that the CEC forecasts raises the skill score in all regions, relative to the forecasts from individual ocean basins regardless of the season. In DJF, the skill score increases substantially in the Pacific Northwest and the Great Lakes/Ohio Valley, most due to the inclusion of SST signal from the NPAC and the NATL (see Figure 2). In JJA, areas with hit scores significantly greater than random forecast skill increase substantially, especially in northern tier states and in the southwest. The increased score in JJA is mostly derived from SST signal from

the NPAC and NATL. Note that the 49-year mean CEC skill scores for DJF and JJA are generally higher and cover more areas than previous studies based on ENSO years only [*Wang et al.*, 1999].

[12] The increase in skill score by the CEC is robust and applicable to all regions and all seasons, as evident in the averaged skill score over six representative regions of the US, i.e., the North American Monsoon region (NAM: 105–115°W, 20–32°N), the Pacific North West (PNW: 120–127°W, 40–50°N), West Coast/ Mountain States (WC/MS: 105–125°W, 32–40°N), Northern Great Plain/Midwest (NGP/MW: 87–105°W, 37–45°N), the Gulf Coast (GC: 80–100°W, 25–32°N), and the Mid-Atlantic (MA: 70–77°W, 32–42°N). Shown in Figure 6 are the 49-year mean skill scores for CEC, and for the five individual basins and the global ocean for three-month means running throughout the entire annual cycle, averaged over the six regions. For all regions, regardless of the time of the year, there is a substantial increase, ranging from 10–20%, in the CEC skill score compared to those from the global ocean and from individual basins. Notice that the skills for the global ocean follow closely those of the TPAC. The increase in CEC skill is most notable in the spring and summer, thus greatly reducing the spring predictability barrier. In regions, such as NAM and the GC, the increase is only modest during the boreal winter, presumably because all the predictable SST signal is due to El Niño, which is already maximally extracted from the tropical Pacific. However in other regions such as WC/MS, PNW, and NGP/MW, the wintertime skill scores are also substantially increased.

[13] Most interesting, the NGP/MW regions show a skill score of approximately 50% for both summer and winter. In the one-season lag forecast (not shown), the skill score for this region is actually higher in the summer than in the winter, mainly due to the impact of the North Pacific SST. This result is consistent with the recent findings [*Lau and Weng*, 2000a, 2000b], which showed that enhanced summertime precipitation in the northern Great Plains and Midwest may be related to the occurrence of global monsoon modes, which are sustained by air-sea interaction over the North Pacific.

## 4. Potential Application of CEC

[14] Results of the CEC forecast model for seasonal precipitation prediction have shown a remarkable across-the-board increase in prediction skill for major regions of the US regardless of the time of the year. Further increase in skill scores is conceivable by stratifying the data according to phases of major climate events such as ENSO and/or other major decadal climate signals. It is worth noting that the CEC skill reported here, averaged over 49 forecasts, without stratifying, is comparable to, or better than the prediction skill of the previous study for ENSO events [*Wang et al.*, 1999]. When the skill scores are stratified according to El Niño and La Niña, results (not shown) indicate that additional improvement in forecast skills (>60–70% hit rates) can be achieved in the NGP/MW and NAM Region. Another significance of the CEC forecast is its implicit use of the nonlinear interaction among the SSTs over different ocean basins and precipitation over the US. The nonlinearity is reflected in the forecast results since the CEC forecast from all oceans is far better than the sum of the forecasts from individual ocean basins and the forecast from global SST. We note that predictability may also be further increased by including soil moisture, snow cover, and other regional data that provide additional information independent of large scale SST. However, because of the use of information not available in real forecasts and validation, the actual forecast skills are expected to be lower than the potential predictability estimated here. How much of the CEC potential predictability can be harnessed in actual precipitation forecasts using either empirical means or dynamical models is a major challenge for future research.

[15] **Acknowledgments.** This work was supported by the NASA Global Model and Data Analysis Program, the US National Research Council, and Canada Natural Sciences and Engineering Research Council. Dr. P. Xie provided the gridded monthly land precipitation data.

## References

Barnett, T. P., and R. Preisendorfer, 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, 1987.

Barnston, A. G., and H. M. Van den Dool, A degeneracy in cross-validated skill in regression-based forecasts, *J. Climate*, *6*, 963–977, 1993.

Chen, M., P. Xie, J. E. Janowiak, and P. A. Arkin, Global land precipitation: A 50-year monthly analysis based on gauge observations, *J Hydrometeorology*, submitted, 2001.

Higgins, R. W., A. Leetmaa, Y. Xue, and A. Barnston, Dominant factors influencing the seasonal predictability of U.S. precipitation and surface air temperature, *J. Climate*, *13*, 3994–4017, 2000.

Kim, K.-Y., and G. R. North, EOF-based linear prediction algorithm: Theory, *J. Climate*, *11*, 3046–3056, 1998.

Koster, R. D., M. J. Suarez, and M. Heiser, Variance and predictability of precipitation at seasonal-to-interannual timescales, *J. Hydrometeor.*, *1*, 26–46, 2000.

Krishnamurti, T. N., D. W. Shin, and C. E. Williford, Improving tropical precipitation forecasts from a multianalysis superensemble, *J. Climate*, *13*, 4217–4227, 2000.

Lau, K. M., and L. Peng, Dynamics of atmospheric teleconnections during the northern summer, *J. Climate*, *5*, 140–158, 1992.

Lau, K. M., and H. Weng, Teleconnection linking summertime precipitation variability over North America and East Asia, *CLIVAR Exchanges*, *5*, 18–20, 2000.

Lau, K. M., and H. Weng, Remote forcing of US summertime droughts and floods by the Asian Monsoon?, *GEWEX News*, *10*(May Issue), 4–6, 2000.

Ropelewski, C. F., and M. S. Halpert, North American precipitation and temperature patterns associated with El Nino/Southern Oscillation (ENSO), *Mon. Wea. Rev.*, *114*, 2352–2362, 1986.

Shen, S. S. P., K.-M. Lau, K.-M. Kim, and G. Li, A canonical ensemble correlation prediction model for seasonal precipitation anomaly, *NASA/ TM-2001-209989*, 53 pp., 2001.

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

Ting, M. F., and H. Wang, Summertime US precipitation variability and its relation to Pacific sea surface temperature, *J. Climate*, *10*, 1853–1873, 1997.

Van den Dool, H. M., A bias in skill in forecasts based on analogues and antilogues, *J. Climate. Appl. Meteor.*, *26*, 1278–1281, 1987.

Wang, H., M. F. Ting, and M. Ji, Prediction of seasonal mean United States precipitation based on El Nino sea surface temperatures, *Geohphys. Res. Lett.*, *26*, 1341–1344, 1999.

K.-M. Lau, Climate and Radiation Branch, NASA Goddard Space Flight Center, Greenbelt, Maryland 20771, USA. (lau@climate.gsfc.nasa.gov)

K.-M. Kim, Science Systems and Applications, Inc., Lanham, Maryland 20706, USA. (kmkim@climate.gsfc.nasa.gov)

S. S. P. Shen, Department of Mathematical Sciences, University of Alberta, Edmonton, T6G 2G1, Canada. (samshen@climate.gsfc.nasa.gov)