Potential to improve precipitation forecasts in Texas through the incorporation of multiple teleconnections

Liyan Tian,[a*] Zachary Leasor[b] and Steven M. Quiring[b]

[a] *Climate Science Laboratory, Department of Geography, Texas A&M University, College Station, TX, USA*

[b] *Atmospheric Sciences Program, Department of Geography, Ohio State University, Columbus, OH, USA*

[*] *Corresponding author. 803A Eller O&M Building, Texas A&M University, College Station, TX 77843, USA.*

Email: liyantian@tamu.edu

Tele: (979)-587-3366

Abstract

Climate oscillations are one of the primary factors that influence precipitation. This study uses canonical correlation analysis (CCA) to examine how El Niño-Southern Oscillation (ENSO), Atlantic Multidecadal Oscillation, North Atlantic Oscillation, Pacific Decadal Oscillation, and the Pacific-North American pattern influence precipitation in Texas. This study identifies the months, regions, and time lags where the relationships between climate oscillations and precipitation are strongest. Correlation results indicate that ENSO accounts for the greatest amount of precipitation variance in Texas. However, including all five climate oscillations is important and together they account for a greater amount of the variance in precipitation than any individual climate oscillation. Precipitation in southern Texas is more strongly influenced by climate oscillations than other regions in Texas. The CCA results demonstrate that there are statistically significant relationships between the climate oscillations and precipitation at time lags longer than 6 months during the summer and at time lags shorter than 6 months during the winter. Based on the CCA results, a precipitation forecast model was developed for the three climate regions that we defined. In the cases of January, the Heidke Skill Score (HSS) of our model is comparable or higher to those achieved by the Climate Prediction Center (CPC) in each region. For all of the 36 month/region cases (12 months * 3 regions), there are 50% cases that the HSS of our model is comparable or higher to those achieved by the CPC. The results of this study illustrate that including multiple teleconnections can increase forecast skill, and statistical methods are useful for precipitation forecasting at a 0-month lead time.

Keywords:

Climate Oscillations, Precipitation Forecast, Canonical Correlation Analysis, Texas

1. Introduction

Drought is a recurrent natural hazard that arises from a considerable deficiency in precipitation [*Zargar et al.*, 2011]. The occurrence of drought has impacts on agriculture, hydrology, ecosystems, society, and the economy [*Heim*, 2002; *Quiring and Papakryiakou*, 2003; *Zargar et al.*, 2011]. Climate oscillations have been shown to be one of the factors that causes variations in precipitation [*Ning and Bradley*, 2014; *Trenberth*, 2011]. A climate oscillation is defined as a slowly varying change of climate about a mean that recurs with some regularity [*American Meteorological Society (2016)*]. An understanding of the interactions between climate oscillations and precipitation variability is vital for the prediction and mitigation of drought.

A great deal of previous research has focused on how El Niño-Southern Oscillation (ENSO) affects precipitation. For example, *Hunt* [2015] performed a multi-millennial simulation with a coupled global climatic model to investigate extreme rainfall events in the Dust Bowl region, located in the southern Great Plains. This region was characterized by a persistent drought and associated dust storms during the 1930s [*Schubert et al.*, 2004]. *Schubert et al.,* [2004] found that ENSO has a significant impact on the generation of rainfall anomalies at an interannual timescale. In contrast, *Hu and Feng* [2001] analyzed the effects of ENSO on the interannual variations in summer rainfall in the central United States and found that there is no persistent effect of ENSO on the summer rainfall in the central United States. The correlations between summer rainfall and tropical Pacific SSTs were strong during 1871-1916 and 1948-1978, but the relationship was weak during 1917-1947 and 1979-present. There are also studies regarding the impact of other teleconnections on precipitation, such as the Atlantic Multidecadal Oscillation (AMO) [*Schlesinger and*

Ramankutty, 1994], North Atlantic Oscillation (NAO) [*Wallace and Gutzler*, 1981], Pacific Decadal Oscillation (PDO) [*Mantua and Hare*, 2002], and Pacific-North American pattern (PNA) [*Wallace and Gutzler*, 1981]. For example, *Hurrell* [1995] found that changes in the mean circulation patterns over the North Atlantic are accompanied by shifts in storms tracks and synoptic-scale eddy activity. These changes affect the transport and convergence of atmospheric moisture and consequently alter regional precipitation. *Sutton and Hodson* [2005] demonstrated that the boreal summer climate was affected by the AMO on multidecadal timescales during the 20th century. *Leathers et al.* [1991] found that the PNA was highly correlated with regional temperature and precipitation from 1947 to 1982 for the fall, winter, and spring months when the PNA serves as a main mode of North Hemisphere mid-tropospheric variability. *McCabe et al.* [2004] demonstrated that climatic oscillations occurring at the decadal scale such as the AMO and PDO have been found to explain around half of the variance in drought frequency across the United States since the 1900s. While the AMO and PDO are important for explaining precipitation variability when considered by themselves, decadal climate oscillations also tend to modulate the impact that ENSO has on precipitation. *Enfield et al.* [2001] found that the AMO has a significant impact in the Mississippi River basin, but not in the Okeechobee river basin. In Texas, the warm phases of the AMO greatly diminish the well-known positive relationship between ENSO and precipitation during the winter season (DJF). *Schubert et al.* [2016] investigated the relationships between sea surface temperatures (SST) and precipitation variability on a global scale. In North America they found that SST variability in the tropical Pacific is the dominant factor that influences precipitation, with some contribution from Atlantic SSTs. Therefore, at interannual time scales, ENSO is the primary driver of precipitation variability throughout much of North and South America. At decadal time scales, the AMO and PDO are the primary drivers of precipitation variability. *Cook et al.* [2014] investigated the pan-continental droughts in North America over the last Millennium. They defined pan-continental drought as synchronous drought in three regions. The results showed that droughts in the Southwest and Central Plains occur in conjunction with either the Southeast or Northwest during La Niña conditions, while droughts in Central Plains, Northwest, and Southeast are primarily associated with the PDO and AMO.

These studies demonstrate that precipitation variability across space and time is influenced by climate oscillations. However, because the impact of each climate oscillation does not occur in isolation, it is important to analyze the impact that multiple teleconnections jointly have on precipitation variability. *Stevens and Ruscher* [2014] investigated the impact of AMO, NAO, PDO and ENSO on temperature and precipitation in the Apalachicola-Chattachoochee-Flint (ACF) River Basin, which supplies water to Alabama, Georgia, and Florida. Their results showed that each of the sub-basins of the ACF are affected in a unique way by climatic oscillations, and no single climatic oscillation can adequately explain/predict the variations in meteorological conditions. *Wise et al.* [2015] analyzed the associations of cool-season precipitation patterns in the United States with teleconnection interactions, including ENSO, NAO, PNA, East Atlantic pattern (EA) and West Pacific pattern (WP). Their results emphasized the importance of considering multiple climatic oscillations when forecasting the seasonal rainfall variability. *Ning and Bradley* [2014] also studied the relationships between winter precipitation variability and teleconnections over the northeastern United States. Their correlation analysis showed that the first Empirical Orthogonal Function (EOF) pattern is significantly correlated with PNA and PDO, the second EOF pattern is significantly correlated with NAO and AMO, and the third EOF pattern is associated with ENSO, PNA and PDO. Therefore, multiple teleconnections should be considered when analyzing the relationship between climate oscillations and precipitation variability. The aforementioned research has shown that ENSO, NAO, AMO, PNA and PDO are the major climate oscillations that have an impact on precipitation in the United States; therefore, this study will investigate the impacts of the five climate oscillations on precipitation variability in Texas.

Only simultaneous relationships (zero lead time) between teleconnections and precipitation were evaluated in the studies described above. However, there can be significant time lags between teleconnections and precipitation. For example, *Redmond and Koch* [1991] analyzed how ENSO and PNA influence precipitation, temperature, and streamflow in the western United States. Their results indicated that June-November ENSO was strongly correlated with October-March precipitation, suggesting that the winter precipitation was related to ENSO at a six-month time lag. *Harshburger et al.* [2002] also demonstrated that the state of ENSO during the fall season can be used to predict winter precipitation in the western U.S. *McCabe and Dettinger* [1999] investigated the relationship between ENSO during fall season and the winter precipitation. Their results indicated that the strength of the correlations between fall ENSO and winter precipitation in the western U.S. varied over space and time during the 20^{th} century. When PDO is negative, the relationship between ENSO and precipitation is strong. When PDO is positive, ENSO and precipitation are weakly correlated. *Brown and Comrie* [2004] studied the impact of fall ENSO on winter precipitation in the western U.S. They found significant correlations between fall ENSO and winter precipitation in the Southwest U.S. Specifically, they found that wet winters tend to follow El Niño events, and dry winters follow La Niña. Our study will also investigate the lagged relationships between multiple teleconnections and precipitation.

The state of Texas frequently experiences drought [*Stahle and Cleaveland*, 1988]. The four most significant droughts in Texas during the last century occurred in 1916-1918, 1925, 1948-1957, and 2010-2011 [*Hoerling et al.*, 2013]. The increased potential evapotranspiration that accompanies the warmer temperatures that are characteristic of Texas create an environment in which drought can occur even with minor precipitation deficits [*Nielsen-Gammon*, 2011]. Droughts in Texas are caused by numerous factors, including natural atmospheric variability (i.e., climate oscillations), land-atmosphere interactions, and thermodynamic conditions [*Fernando et al.*, 2016; *Myoung and Nielsen-Gammon*, 2010; *Seager et al.*, 2014]. This paper investigates the simultaneous and lagged relationships between Texas precipitation and ENSO, NAO, AMO, PNA and PDO. The goals of this paper are to: (1) determine which climate oscillation accounts for the greatest amount of precipitation variance in Texas, (2) identify the regions and months (or seasons) where climate oscillations have the largest impact on precipitation, (3) identify at what time lag the relationship between climate oscillations and precipitation are strongest and (4) forecast precipitation based on multiple climate oscillations and compare with the precipitation forecast from Climate Prediction Center (CPC).

2. Data

2.1 Precipitation

Monthly precipitation data from the PRISM (Parameter-elevation Regressions on Independent Slopes Model) dataset were used in this study (http://www.prism.oregonstate.edu). PRISM was developed by the Spatial Climate Analysis Service at Oregon State University. The gridded data are generated by interpolating meteorological data from approximately 13,000 surface stations and incorporating spatial information including elevation, slope, rain shadows, temperature inversions, and coastal effects [*Daly et al.*, 2002; *Daly et al.*, 2008; *Daly et al.*, 1994]. The monthly PRISM datasets are available at 2.5 arcmin (4 km) resolution from January 1895 to the present. The PRISM dataset is ideal for this study because it provides a long and consistent record [*Mishra and Singh*, 2010]. Data used in this study cover a 110-year period from 1901 to 2010.

2.2 Climatic Oscillations

Five climate oscillations are investigated in this study: ENSO, NAO, AMO, PNA, and PDO. ENSO is the most frequently studied climatic oscillation. During an El Niño event, easterly trade winds weaken or reverse and cause anomalous warming of the ocean surface, changing patterns of meteorological variables such as precipitation [*Stevens and Ruscher*, 2014]. The NINO3.4 SST anomaly is used in this study to represent ENSO conditions. It is based on departures from the three-month running mean of SSTs in the NINO3.4 region. Positive NINO3.4 values are associated with El Niño events, while negative values indicate La Niña events. NINO3.4 SST anomaly data from 1901 to 2010 can be downloaded from the NOAA PSD website (http://www.esrl.noaa.gov/psd/gcos_wgsp/Timeseries/Nino34/). The NINO3.4 SST index is calculated from the Hadley Centre Sea Ice and Sea Surface Temperature data set (HadISST1). It is the area averaged SST from 5°S-5°N and 170°-120°W [*Rayner et al.*, 2003].

The NAO is an atmospheric oscillation in the North Atlantic Ocean. The NAO index from the Climate Research Unit is defined as the normalized pressure difference between a station located in the Azores and a station in Iceland [*Stevens and Ruscher*, 2014]. The NAO index from 1901 to 2010 can be downloaded from the NOAA PSD website (http://www.esrl.noaa.gov/psd/gcos_wgsp/Timeseries/NAO/).

The AMO is a 60-85 year cycle of variable SSTs in the North Atlantic Ocean that has been shown to correlate with precipitation in the United States [*Stevens and Ruscher,* 2014]. The AMO index is calculated using the Kaplan SST as the detrended time series of the area weighted averaged SST over the North Atlantic from 0˚ to 70˚N [*Enfield et al.*, 2001]. The smoothed AMO index from 1901 to 2010 can be downloaded from NOAA PSD (http://www.esrl.noaa.gov/psd/gcos_wgsp/Timeseries/AMO/).

The PNA index indicates the nature of atmospheric circulation in the Northern Hemisphere. A positive phase of the PNA indicates meridional flow with an enhanced jet stream while a negative phase indicates zonal flow [*Henderson and Robinson*, 1994]. The PNA index is calculated using the 500 mb heights from the 20th Century Reanalysis Project Version V2 dataset. Area-averaged geopotential heights from four regions in the Northern Hemisphere are combined for the PNA index [*Barnston and Livezey*, 1987]. The PNA index data from 1901 to 2010 can be downloaded from NOAA PSD (http://www.esrl.noaa.gov/psd/data/20thC_Rean/timeseries/monthly/PNA/).

The PDO is based on monthly SST variability in the North Pacific Ocean. The PDO index is calculated based on the EOF analyses of the monthly SST anomalies in the North Pacific [*Mantua et al.*, 1997]. The PDO index from 1901 to 2010 can be downloaded from NOAA PSD http://www.esrl.noaa.gov/psd/gcos_wgsp/Timeseries/PDO/.

3. Methods

3.1 Precipitation anomalies

Monthly precipitation data were converted into precipitation anomalies using Equation 1,

$$PA_i = P_i - PM_i \qquad (1)$$

where PA_i is the monthly precipitation anomaly, P_i is the monthly precipitation data for the given month, and PM_i is the mean monthly precipitation for the given month.

3.2 Empirical Orthogonal Functions

Empirical Orthogonal Functions (EOFs) were used to identify regions in Texas that have similar precipitation. EOFs are commonly used for regionalization because they can effectively reduce dimensionality and extract patterns [*Hannachi et al.*, 2007; *Lorenz*, 1956; *Navarra and Simoncini*, 2010]. EOF analysis was performed using the gridded monthly precipitation anomalies from January 1901 to December 2010 at all locations in Texas. The first step of EOF analysis is to calculate correlation coefficients among all variables.

A VARIMAX (orthogonal) rotation method was applied because it simplifies the structure of the resultant patterns by forcing the value of the loading coefficients towards zero or ± 1 [*Hannachi et al.*, 2007]. The VARIMAX rotation technique is a popular method used in climate regionalization studies because the rotation tends to produce more spatially coherent regions [*White et al.,* 1991]. An unrotated EOF is primarily used as a data reduction technique and is not appropriate for climate regionalization [*Yarnal*, 1993]. After rotation, each grid cell was assigned to the factor on which they had the highest loadings.

The first three factors were retained and collectively they explain more than 85% of the variance.

Figure 1 shows the precipitation regions identified using the rotated EOFs. Region 1 is located in northeastern Texas and it experiences the greatest precipitation variability with a 48.45 mm standard deviation. Region 2 is located in northwestern Texas and it experiences the least precipitation variability with a 25.58 mm standard deviation. Region 3 is located in southern Texas and its standard deviation is 40.63 mm. The regional precipitation anomalies shows that there is no significant trend in regional precipitation over the study period. Mean annual precipitation in Texas has a distinct east-to-west gradient. Based on the 1971 to 2000 normals [*Committee*, 2006], the mean annual precipitation is highest in eastern Texas (~1500 mm) and lowest in western Texas (~300 mm).

3.3 Canonical Correlation Analysis

Canonical correlation analysis (CCA) was used to analyze the relationships between precipitation and the five teleconnections. Simultaneous (no lag) and lagged relationships (1, 3, 6, 12, and 24-month lags) were evaluated using monthly and seasonal data. Seasons were defined using the normal climatological convention of winter (DJF), spring (MAM), summer (JJA), and fall (SON).

CCA is a linear multivariate approach used to compare two sets of data, independent and dependent, with each set composed of multiple arrays of variables [*Thompson*, 2005]. CCA attempts to find relationships between a set of predictor variables and a set of predicted variables. The linear combinations represent the weight of at least two variables from the respective set, therefore creating the two variant arrays (U_1 & V_1) seen in Equation

2, in which *x* represents the precipitation anomalies, *y* represents the climate oscillations, *a* represents the coefficients for precipitation, and *b* represents the coefficients of the climate oscillations [*Borga*, 2001; *Stevens and Ruscher*, 2014].

$$U_1 = a_1x_1 + a_2x_2 + \cdots a_nx_n$$

$$V_1 = b_1y_1 + b_2y_2 + \cdots b_ny_n \tag{2}$$

The loading matrices calculated using Equation 2 produce canonical loadings, which are linear correlations between the variables and the variate. The loadings are used to calculate the canonical cross loadings that determine the linear correlation between the independent variable and dependent variable. The canonical cross loadings of the climate oscillations are estimated using the correlation coefficient in Equation 3 where S_{xx} and S_{yy} are variance-covariance matrices of the respective variable and S_{xy} and S_{yx} and the covariance matrices of precipitation and the climate oscillations [*Stevens and Ruscher*, 2014].

$$r_c b = [S_{yy}]^{-1}[S_{xy}][S_{xx}]^{-1}[S_{yx}]b \tag{3}$$

In this study, the dependent variable set is precipitation anomalies at different lags and the independent set is the five climatic oscillations. The canonical loadings and cross loadings are used to understand the relationships, while the canonical correlation values and proportion of variance explained in the dependent variables by the independent variate are used to examine the overall strength of each analysis. This approach allows us to simultaneously examine the impacts of climatic oscillations on precipitation variations [*Stevens and Ruscher*, 2014].

CCA provides information about (1) the varying effects of climate oscillations in different regions, and (2) how the strength of the relationships change for each time lag.

Canonical roots that are not statistically significant at the 95% confidence level were eliminated based upon the methods used by *Stevens and Ruscher* [2014].

3.4 Monthly Precipitation Prediction

A CCA-based linear regression model was developed to evaluate whether climate oscillations can be used to produce skillful monthly forecasts of precipitation in Texas. The linear regression model uses weights for each of the climate oscillations calculated as the dividend between the canonical loadings and the dependent and independent arrays. The CCA-based forecast model was built using data from 1901-1980 and evaluated using data from 1981 to 2010.

The Heidke Skill Score (HSS) was used to evaluate the skill of the precipitation forecast and to facilitate comparison to the skill of the CPC monthly precipitation forecast. The HSS was calculated based on observed and predicted precipitation values from 1981-2010 which were grouped into three percentile ranges based upon their distribution; below normal, average, and above normal. This was done to standardize the precipitation predictions in a manner that is consistent with the methodology used by CPC. Since the CPC precipitation forecast skill scores are based on observed and predicted precipitation data from 1981 to 2010 [*CPC,* 2016a], the skill score of the CCA-based model was also calculated using precipitation data from 1981 to 2010. The HSS values were calculated using Equation 4,

$$HSS = \frac{(NC-E)}{(T-E)} \tag{4}$$

Where *NC* is the number of correct forecasts, *T* is the total number of forecasts, and *E* is the number of forecasts expected to verify based upon climatology.

4. Results

4.1 CCA Results

Figure 2 shows the simultaneous correlations (no lag) between each climate oscillation and precipitation for each month in each Texas regions. Only correlations that are statistically significant at the 95% confidence level are shown. ENSO has the most statistically significant correlations with precipitation, followed by PNA, PDO, NAO, and AMO. There are a total of 36 month/region combinations (12 months * 3 regions) and there is a statistically significant correlation between ENSO and precipitation in 19 of the 36 cases (53%). There is a statistically significant correlation between PNA and precipitation in 28% of these combinations. PDO, NAO, and AMO have statistically significant correlations in 22%, 11%, and 0% of these 36 combinations, respectively.

Figure 3 shows how the correlations between multiple climate oscillations and precipitation vary by month and region. Correlations were calculated for the following combinations of climate oscillations: ENSO, ENSO/PNA, ENSO/PNA/PDO, ENSO/PNA/PDO/NAO, and ENSO/PNA/PDO/NAO/AMO. Most of the statistically significant correlations occur during the winter months and the number of significant correlations increases as additional climate oscillations are included. Even the AMO, which did not have any statistically significant correlations during the univariate analysis, helped to explain more of the variance in precipitation when included with other climate oscillations. Not surprisingly, our results show that the inclusion of additional climate oscillations is helpful for explaining precipitation variability in Texas.

Next, the dependent cross loadings were calculated as the correlation between the observed dependent variable (i.e., precipitation) and the opposite canonical variate, which is the linear combination of the five climate oscillations (Figure 4). Similar to the correlations, most of the significant cross loadings are observed during the winter months. Additionally, most of the significant cross loadings during the winter occurred at shorter time lags, while there were more significant cross loadings at longer time lags during the summer months. Specifically, in 39 out of 43 cases, the cross loadings during October to March of the following year occurred at less than 3-months lags. In 27 out of 38 cases, the cross loadings during April to September occurred at no less than 3-months lags (Figure 4).

4.2 Precipitation Forecast Results

As described above, a CCA-based linear regression model was developed to evaluate whether the climate oscillations can be used to produce skillful monthly forecasts of precipitation in Texas. Figure 4 shows that most of the statistically significant cross loadings occurred during the winter. Therefore, January was selected to build the CCA-based regression model for the three regions. To compare the CCA-based forecast skill to that of the CPC, the CCA-based regression model was built using all five climate oscillations at a 0-month lead. That is, climate oscillations from December are used to forecast January precipitation. A second regression model was also built using only ENSO. The skill of this model will be compared to that of the CCA-based model that uses all 5 climate oscillations. This will show the relative value of including additional oscillations.

Table 1 shows how the performance of these models varies by month and region. Both models perform best in region 3. The model that uses all five climate oscillations has an R^2 of 0.48 and an MAE of 22.46 mm. The ENSO-only model has an R^2 of 0.41 and an MAE of 24.25 mm. The regression model with all five climate oscillations has a little bit better performance than the regression model with only ENSO because the regression model with five teleconnections explained more variance of the precipitation.

The precipitation forecasts are least skillful in region 2. In fact, both the ENSO-only and five variable model are not statistically significant. The performance of the regression model with all five teleconnections for region 1 is better than the regression model with only ENSO, even though the regression model with all five teleconnections only has an R^2 of 0.15. The ENSO-only model is not statistically significant. The comparison of the two types of regression models suggests that using a prediction model based solely on the five teleconnections can produce somewhat better predictions of precipitation in some regions in Texas than the model only based on ENSO. The various performances of the regression models for the three regions in Texas is related with the cross loadings between the multiple teleconnections and precipitation. The cross loadings for region 3 are highest in January, followed by the cross loadings for region 1 and 2 (Figure 4). The performance of the regression model is best for region 3, followed by region 1, while region 2 has the worst performance. The errors of the regression models for all regions are high. This indicates that the CCA model cannot accurately predict the magnitude of the precipitation anomalies. However, it is not uncommon that the skill of monthly to seasonal forecasts is relatively low [*McCabe and Dettinger,* 1999*; Barnston et al.,* 1996]. Therefore, in the next section of the paper we compare the climate oscillation-based forecasts developed in this paper to the CPC forecasts.

4.3 Comparing Forecast Skill to CPC

The CPC provides monthly precipitation forecasts at a 0-month lead. The 0-month lead of precipitation forecast is created and updated the last day of the month for the following month. Therefore, all data in the initial month are used to predict precipitation in the subsequent month. Our precipitation forecast is similar with this type of CPC monthly precipitation forecast. Both the CCA-based forecast model and the CPC forecast model utilize all antecedent precipitation data from the first month to predict precipitation in the following month. The difference between CCA-based forecast and the CPC forecast is the methodology. The CCA-based forecast utilizes a regression model that includes the five teleconnections. The CPC forecast is primarily based on a dynamical model [*CPC,* 2016b]. The dynamical model uses a set of current precipitation observations and equations describing the physical behavior of the precipitation system to predict the precipitation in a short time future. Then, the predicted precipitation data are used as the initial condition for a subsequent prediction for the next time-step until the future prediction time is reached. The CPC reports the Heidke Skill Score for various regions (Figure 5). Because the regions that are used by CPC do not match the regions that were defined in this study using EOF analysis and the years used by the CPC do not match the years of our study, it is difficult to directly compare forecast skill. Therefore, we have presented the results in two different ways, a direct comparison and an indirect comparison. The direct comparison evaluates the forecast skill from 2005 to 2010. The indirect comparison evaluates the CCA forecast skill from 2000 to 2010 and the CPC forecast skill during 2005 to 2015, so that there is a larger sample size of predictions even though the years may not match.

Table 2 displays the results of the direct comparison during 2005-2010 for the three regions in Texas. Since the regions used by CPC do not match the regions that were defined using EOF analysis in this study, several corresponding CPC regions were used in this comparison. Region 1 defined by EOF in this study approximately includes Region 60 and

Region 61 from the CPC. Region 2 defined by EOF in this study approximately includes Region 54, Region 55, Region 64, and Region 65 from the CPC. Region 3 defined by EOF in this study approximately includes Region 62 and Region 63 from the CPC. For Region 1, the HSS for the CCA model is higher than the HSS for the CPC in Regions 60, but lower than the CPC in Regions 61. The HSS of the CCA model is lower than the average skill score for Regions 60-61. For Region 2, the HSS for the CCA model is higher than the HSS for the CPC in Regions 64 and Region 65, but lower than the CPC in Regions 54 and Region 55. The HSS of the CCA model is higher than the average skill score for these four regions. In Region 3, the HSS for the CCA model is higher than the HSS for the CPC in Regions 62, but lower than the CPC in Regions 63. The HSS of the CCA model is higher than the average skill score for Regions 62-63. Since these scores may be affected by the smaller sample size of six years, an indirect comparison of forecast skill was also performed (Table 2).

As the sample size increases, the forecast evaluation should approach the true skill and become more stable. The HSS of the indirect comparison is similar to the direct comparison. The HSS of the CCA model in Region 1 is higher than the HSS from the CPC in Region 60 but lower than the HSS of the CPC in Region 61. For Region 2, the HSS of the CCA model is higher than the HSS of the CPC regions. In Region 3, the HSS of the CCA model is higher than the HSS of the CPC in Region 62 but lower than the HSS of the CPC in Region 63. However, the HSS for the CCA model is higher than the average HSS for the CPC regions for all cases. One limitation of the indirect comparison is that the years used to assess forecast skill are not same for the CPC and the CCA. However, these results support what was found in the direct comparison and suggest that the skill of the CCA model is equivalent or better than the CPC in most locations and timescales.

The HSS of the CCA and CPC models were also compared for all other months. The results show that in 18 out of 36 cases (12 months * 3 regions) the HSS for the CCA model is comparable or higher than the average HSS for the CPC regions in the direct comparison. In the indirect comparison, HSS for the CCA model is higher than the average HSS for the CPC regions in only 13 of 36 cases. Even though in less than 50% cases the HSS of the CCA model is better than the CPC forecast, the results of this study can be useful for precipitation forecasting at a 0-month lead time during months when the performance of CCA model is better than CPC forecast.

5. Discussion and Conclusion

Correlations between the five climate oscillations and precipitation indicated that ENSO accounts for the greatest amount of precipitation variance in Texas. Many previous studies have also shown that ENSO is the most important factor that affects precipitation variability [*Barnston et al.,* 1996*; Dai and Wigley,* 2000*; Ropelewski and Halpert,* 1996]. However, across nearly every month and region, the correlations between the climate oscillations and precipitation variability were stronger when the combined impact of multiple teleconnections was considered. This result is consistent with previous studies such as *Stevens and Ruscher* [2014], *Wise et al.* [2015], and *Ning and Bradley* [2014]. *Stevens and Ruscher* [2014] indicated that the sub-basins of the ACF are affected differently by multiple climatic oscillations, and no particular climatic oscillation can explain surface meteorological variation. *Wise et al.* [2015] also emphasized the importance of considering multiple climatic oscillations when forecasting the seasonal rainfall variability.

Using this knowledge, CCA was applied to identify the months, regions, and time lags where the relationships between teleconnections and precipitation are the strongest. Dependent cross loadings were used to provide a means for quantifying the relationship between the five combined teleconnections and the precipitation anomalies at various time lags. The results of the CCA analysis were generally in agreement with the correlation results. The strongest canonical cross loadings occurred during the winter and there were more time lags that were statistically significant during the winter. These results agree with studies such as *Hu and Feng* [2001] and *Leathers et al.* [1991] which suggest that teleconnections have a stronger impact on North American precipitation during the fall, winter, and spring. Statistically significant relationships were found for longer time lags (> 6 months) during the summer months, while most of the statistically significant relationships were found at shorter time lags (< 6 months) during the winter. These findings are supported by previous studies that observed the strongest relationships between precipitation and teleconnections during the winter months [*Leathers et al.*, 1991; *Ning and Bradley*, 2014; *Sutton and Hodson*, 2005; *Wise et al.*, 2015]. There were differences in the strengths of the canonical loadings and the performance of the CCA forecasts across the three regions of Texas used in this study. The differences in performance suggests that our EOF-based regionalization successfully identified three climatically distinct regions.

A CCA-based forecast model was developed using five climate oscillations. The model was shown to have forecast skill that was similar, and in some cases, better than the CPC. While the monthly forecasts for the CPC generally use dynamical models for precipitation prediction, the results of this study suggest that statistical methods could improve the quality of forecasts, particularly during situations when the dynamical model performs poorly. Since one of the objectives of this paper was to determine the value of considering multiple teleconnections, the results of the CCA-based model was compared to a model using only ENSO. The results show that the CCA-based model was slightly better than the model using only ENSO. The correlations between the teleconnections and precipitation shows the CCA-based model can explain more of precipitation variance. Additionally, the p-values of the CCA-based model are statistically significant at a 95% confidence level in regions 1 and 3, indicating that the model predictions are significantly different than a forecast utilizing solely the mean precipitation (climatology forecast). The ENSO-based model is only statistically significant in region 3. However, the errors of CCA-based model are higher than the model using only ENSO in some regions. Generally, the impacts of teleconnections are strongest in Region 3, which is located in the southern Texas. ENSO is the primary factor influencing the precipitation in Texas and its impacts in Region 3 is stronger than in other regions. This result is consistent with the study of *Stevens and Ruscher* [2014]. *Stevens and Ruscher* [2014] found that the southern part of the ACF basin is influenced by ENSO more strongly than other parts of the basin. The impacts of ENSO in the southern United States are likely related to the subtropical jet stream. During El Niño events, the strengthened subtropical jet shifts the winter storm tracks to the south and this brings more energy and moisture in the region [*Redmond and Koch*, 1991; *Wise et al.*, 2015]. Therefore, in El Niño years, the Southwest, Southeast, and Great Plains in the United States tend to be wetter than normal. While, in La Niña years, these regions are dryer than normal [*Wise et al.*, 2015]. Overall, using multiple teleconnections is valuable for explaining and predicting precipitation patterns in Texas. The relative importance of these teleconnections varies by region, month, and time lag. The results presented here suggests that the CCA-based model using only five teleconnections is able to adequately forecast precipitation variability in Texas.

Further research will evaluate whether including additional teleconnections can improve the accuracy of precipitation forecasts. In addition, it may also be useful to explore other statistical modeling approaches such as weighted multiple linear regression model using canonical weights to improving the forecasts. Finally, the skill of the CCA forecast model was evaluated over a multi-year period. It may be more helpful to evaluate how forecast skill changes during years when there are strong ENSO events. It is likely that the skill of the model varies significantly over time and that it is strongest during ENSO events and that the skill weakens when there are not strong remote forcings.

Texas is a region where there are relatively strong relationships between teleconnections and precipitation, particularly ENSO. However, the CCA analysis employed in this paper can be applied to diagnose the impacts of multiple teleconnections on precipitation variability in other regions around the world. While the CCA-based model can effectively predict precipitation with skill comparable to the CPC, climate oscillations only explain around half of the precipitation variability. While the purpose of this study was to observe the impact teleconnections have on precipitation at various time lags, the seasonal forecasting of precipitation could improve with the additional consideration of variables not related to teleconnections. Antecedent temperature, precipitation, and soil moisture could help to improve the forecast. Land-based hydrological processes also have influence on precipitation variability [*Koster and Suarez,* 1995*; Koster et al.,* 2000]. *Koster* *and Suarez* [1995] investigated the impacts of sea surface temperatures and land surface hydrological state to the annual and seasonal precipitation variability. They found that the land surface's impacts on the precipitation variability is greatest during summer when the precipitation processes are very sensitive to surface conditions. *Koster et al.* [2000] indicated precipitation anomalies can be amplified by land surface processes. A positive precipitation anomaly can lead to an evaporation anomaly through land-atmospheric feedback, which in turn leads to additional precipitation through water recycling. Since evaporation is related with soil moisture and temperature, soil moisture and temperature can be used to improve the precipitation forecast. These types of studies are useful for examining other areas which could improve precipitation forecasts, while this study focuses primarily on identifying the strength and nature of the relationship between precipitation and various teleconnections in Texas.

Acknowledgements:

We gratefully acknowledge the PRISM data from the Spatial Climate Analysis Service at Oregon State University, and climate oscillations indices from NOAA Physical Sciences Division (PSD).

References:

American Meteorological Society, cited 2016: "Climatic Cycle". Glossary of Meteorology. [Available online at http://glossary.ametsoc.org/wiki/Climatic cycle]

Barnston AG, Livezey RE. 1987. Classification, seasonality and persistence of low-frequency atmospheric circulation patterns. *Monthly Weather Review* **115:** 1083–1126. doi:10.1175/1520-0493(1987)115<1083:CSAPOL>2.0.CO;2.

Barnston AG, Thiao W and Kumar V. 1996. Long-lead forecasts of seasonal precipitation in Africa using CCA. *Weather and Forecasting* **11:**506-520. doi:10.1175/1520-0434(1996)011<0506:LLFOSP>2.0.CO;2.

Borga M. 2001. Canonical correlation: a tutorial. *On line tutorial <u>http://people.imt.liu.se/magnus/cca</u>* **4**: 5.

Brown DP, Comrie AC. 2004. A winter precipitation 'dipole' in the western United States associated with multidecadal ENSO variability. *Geophysical Research Letters* **31**. doi:10.1029/2003GL018726.

Cook BI, Smerdon JE, Seager R, Cook ER. 2014. Pan-continental droughts in North America over the last millennium. *Journal of Climate* **27**: 383-397. doi:10.1175/JCLI-D-13-00100.1.

CPC. 2016a. http://www.cpc.ncep.noaa.gov/products/verification/summary/index.php?page=tutorial

CPC. 2016b. http://www.cpc.ncep.noaa.gov/products/predictions/30day/

Dai A and Wigley TML. 2000. Global patterns of ENSO-induced precipitation. *Geophysical Research Letters* **27:** 1283-1286. doi:10.1029/1999GL011140.

Daly C, Neilson RP, Phillips DL. 1994. A statistical-topographic model for mapping climatological precipitation over mountainous terrain. *Journal of Applied Meteorology* **33**: 140-158. doi:10.1175/1520-0450(1994)033<0140:ASTMFM>2.0.CO;2.

Daly C, Gibson WP, Taylor GH, Johnson GL, Pasteris P. 2002. A knowledge-based approach to the statistical mapping of climate. *Climate research* **22**: 99-113.

Daly C, Halbleib M, Smith JI, Gibson WP, Doggett MK, Taylor GH, Curtis J, Pasteris PP. 2008. Physiographically sensitive mapping of climatological temperature and precipitation across the conterminous United States. *International Journal of Climatology* **28**: 2031-2064. doi:10.1002/joc.1688.

Enfield DB, Mestas-Nuñez AM, Trimble PJ. 2001. The Atlantic multidecadal oscillation and its relation to rainfall and river flows in the continental US. *Geophysical Research Letters* **28**: 2077-2080. doi:10.1029/2000GL012745.

Fernando DN, Mo KC, Fu R, Pu B, Bowerman A, Scanlon BR, Solis RS, Yin L, Mace RE, Mioduszewski JR. 2016. What caused the spring intensification and winter demise of the 2011 drought over Texas? *Climate Dynamics* 1-14. doi:10.1007/s00382-016-3014-x.

Hannachi A, Jolliffe I, Stephenson D. 2007. Empirical orthogonal functions and related techniques in atmospheric science: A review. *International Journal of Climatology* **27**: 1119-1152. doi:10.1002/joc.1499.

Harshburger B, Ye H, Dzialoski J. 2002. Observational evidence of the influence of Pacific SSTs on winter precipitation and spring stream discharge in Idaho. *Journal of Hydrology* **264**: 157-169.

Heim Jr RR. 2002. A review of twentieth-century drought indices used in the United States. *Bulletin of the American Meteorological Society* **83**: 1149-1165. doi:10.1175/1520-0477(2002)083<1149:AROTDI>2.3.CO;2.

Henderson KG, Robinson PJ. 1994. Relationships between the pacific/north american teleconnection patterns and precipitation events in the south-eastern USA. *International Journal of Climatology* **14**: 307-323. doi:10.1002/joc.3370140305.

Hoerling M, Kumar A, Dole R, Nielsen-Gammon JW, Eischeid J, Perlwitz J, Quan X, Zhang T, Pegion P, Chen M. 2013. Anatomy of an extreme event. *Journal of Climate* **26**: 2811-2832. doi:10.1175/JCLI-D-12-00270.1.

Hu Q, Feng S. 2001. Variations of teleconnection of ENSO and interannual variation in summer rainfall in the central United States. *Journal of Climate* **14**: 2469-2480. doi:10.1175/1520-0442(2001)014<2469:VOTOEA>2.0.CO;2.

Hunt B. 2016. Rainfall variability and predictability issues for North America. *Climate Dynamics* **46:** 2067-2085. doi:10.1007/s00382-015-2690-2.

Hurrell JW. 1995. Decadal trends in the north atlantic oscillation: regional temperatures and precipitation. *Science* **269**: 676-679. doi:10.1126/science.269.5224.676.

Koster RD and Suarez MJ. 1995. Relative contributions of land and ocean processes to precipitation variability. *Journal of Geophysical Research: Atmospheres* **100:** 13775-13790. doi:10.1029/95JD00176.

Koster RD, Suarez MJ and Heiser M. 2000. Variance and predictability of precipitation at seasonal-to-interannual timescales. *Journal of hydrometeorology* **1:** 26-46. doi:10.1175/1525-7541(2000)001<0026:VAPOPA>2.0.CO;2.

Leathers DJ, Yarnal B, Palecki MA. 1991. The Pacific/North American teleconnection pattern and United States climate. Part I: Regional temperature and precipitation associations. *Journal of Climate* **4**: 517-528. doi:10.1175/1520-0442(1991)004<0517:TPATPA>2.0.CO;2.

Lorenz EN. 1956. Empirical orthogonal functions and statistical weather prediction.

Mantua NJ, Hare SR. 2002. The Pacific decadal oscillation. *Journal of Oceanography* **58**: 35-44.

Mantua NJ, Hare SR, Zhang Y, Wallace JM, Francis RC. 1997. A Pacific interdecadal climate oscillation with impacts on salmon production. *Bulletin of the American Meteorological Society* **78**: 1069-1079. doi:10.1175/1520-0477(1997)078<1069:APICOW>2.0.CO;2.

McCabe GJ, Dettinger MD. 1999. Decadal variations in the strength of ENSO teleconnections with precipitation in the western United States. *International Journal of Climatology* **19**: 1399-1410. doi:10.1002/(SICI)1097-0088(19991115)19:13<1399::AID-JOC457>3.0.CO;2-A.

McCabe GJ, Palecki MA, Betancourt JL. 2004. Pacific and Atlantic Ocean influences on multidecadal drought frequency in the United States. *Proceedings of the National Academy of Sciences* **101**: 4136-4141. doi:10.1073/pnas.0306738101.

Mishra AK, Singh VP. 2010. A review of drought concepts. *Journal of Hydrology* **391**: 202-216.

Myoung B, Nielsen-Gammon JW. 2010. The convective instability pathway to warm season drought in Texas. Part I: The role of convective inhibition and its modulation by soil moisture. *Journal of Climate* **23**: 4461-4473. doi:10.1175/2010JCLI2946.1.

Navarra A, Simoncini V. 2010. *A guide to empirical orthogonal functions for climate data analysis*. Springer Science & Business Media.

Nielsen-Gammon JW. 2011. The changing climate of Texas. *The impact of global warming on Texas*: 39-68.

Ning L, Bradley RS. 2014. Winter precipitation variability and corresponding teleconnections over the northeastern United States. *Journal of Geophysical Research: Atmospheres* **119**: 7931-7945. doi:10.1002/2014JD021591.

Quiring SM, Papakryiakou TN. 2003. An evaluation of agricultural drought indices for the Canadian prairies. *Agricultural and Forest Meteorology* **118**: 49-62.

Rayner N, Parker DE, Horton E, Folland C, Alexander L, Rowell D, Kent E, Kaplan A. 2003. Global analyses of sea surface temperature, sea ice, and night marine air temperature since the late nineteenth century. *Journal of Geophysical Research: Atmospheres* **108**: 4407. doi:10.1029/2002JD002670.

Redmond KT, Koch RW. 1991. Surface climate and streamflow variability in the western United States and their relationship to large-scale circulation indices. *Water Resource Research* **27**: 2381-2399. doi:10.1029/91WR00690.

Ropelewski CF and Halpert MS. 1996. Quantifying southern oscillation-precipitation relationships. *Journal of climate* **9:** 1043-1059.

Schlesinger ME, Ramankutty N. 1994. An oscillation in the global climate system of period 65-70 years. *Nature* **367**: 723-726. doi:10.1038/367723a0.

Schubert SD, Stewart RE, Wang H, Barlow M, Berbery EH, Cai W, Hoerling MP, Kanikicharla KK, Koster RD, Lyon B. 2016. Global Meteorological Drought: A Synthesis of Current Understanding with a Focus on SST Drivers of Precipitation Deficits. *Journal of Climate* **29**: 3989-4019. doi:10.1175/JCLI-D-15-0452.1.

Schubert SD, Suarez MJ, Pegion PJ, Koster RD, Bacmeister JT. 2004. On the cause of the 1930s Dust Bowl. *Science* **303**: 1855-1859. doi:10.1126/science.1095048.

Seager R, Goddard L, Nakamura J, Henderson N, Lee DE. 2014. Dynamical Causes of the 2010/11 Texas–Northern Mexico Drought. *J Hydrometeorology* **15**: 39-68. doi:10.1175/JHM-D-13-024.1.

Stahle DW, Cleaveland MK. 1988. Texas drought history reconstructed and analyzed from 1698 to 1980. *Journal of Climate* **1**: 59-74. doi:10.1175/1520-0442(1988)001<0059:TDHRAA>2.0.CO;2.

Stevens KA, Ruscher PH. 2014. Large scale climate oscillations and mesoscale surface meteorological variability in the Apalachicola-Chattahoochee-Flint River Basin. *Journal of Hydrology* **517**: 700-714.

Sutton RT, Hodson DL. 2005. Atlantic Ocean forcing of North American and European summer climate. *Science* **309**: 115-118. doi:10.1126/science.1109496.

Texas Rainwater Harvesting Evaluation Committee. 2006. *Rainwater Harvesting Potential and Guidelines for Texas: Report to the 80th Legislature*. Texas Water Development Board.

Thompson B. 2005. Canonical correlation analysis. *Encyclopedia of statistics in behavioral science.*

Trenberth KE. 2011. Changes in precipitation with climate change. *Climate Research* **47**: 123-138. doi:10.3354/cr00953.

Wallace JM, Gutzler DS. 1981. Teleconnections in the geopotential height field during the Northern Hemisphere winter. *Monthly Weather Review* **109**: 784-812. doi:10.1175/1520-0493(1981)109<0784:TITGHF>2.0.CO;2.

White D, Richman M and Yarnal B. 1991. Climate regionalization and rotation of principal components. *International Journal of Climatology* **11:** 1-25. doi:10.1002/joc.3370110102.

Wise EK, Wrzesien ML, Dannenberg MP, McGinnis DL. 2015. Cool-Season Precipitation Patterns Associated with Teleconnection Interactions in the United States. *Journal of Applied Meteorology and Climatology* **54**: 494-505. doi:10.1175/JAMC-D-14-0040.1.

Yarnal B. 1993. *Synoptic climatology in environmental analysis: a primer*. Belhaven.

Zargar A, Sadiq R, Naser B, Khan FI. 2011. A review of drought indices. *Environmental Reviews* **19**: 333-349.

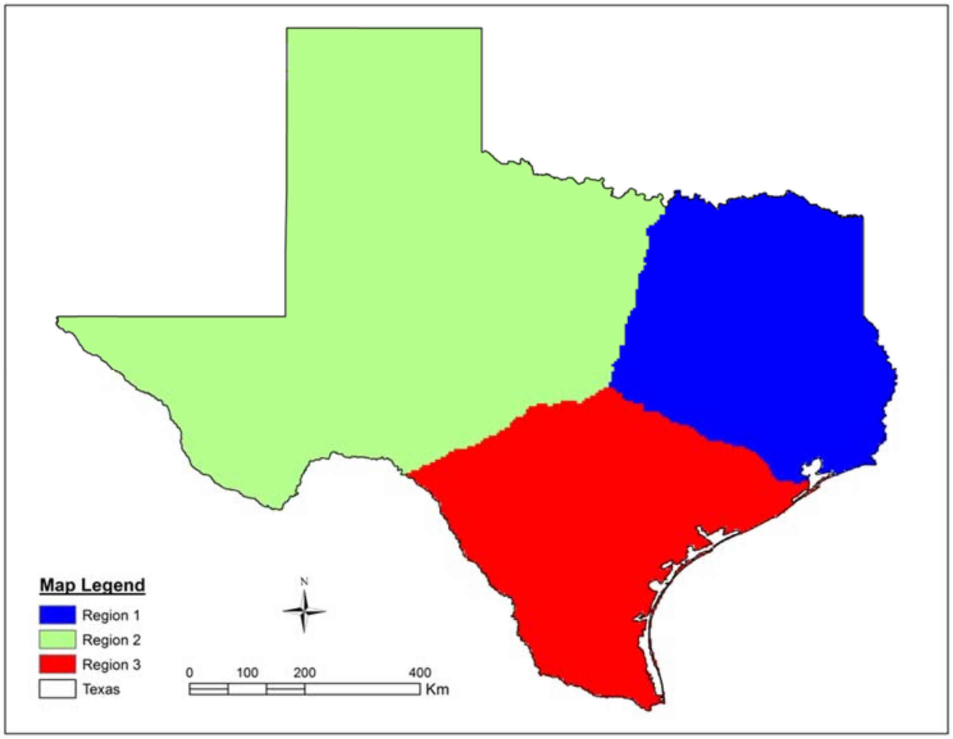

Figure 1. *Texas precipitation regions identified using a VARIMAX EOF analysis based on the first 3 EOFs.*

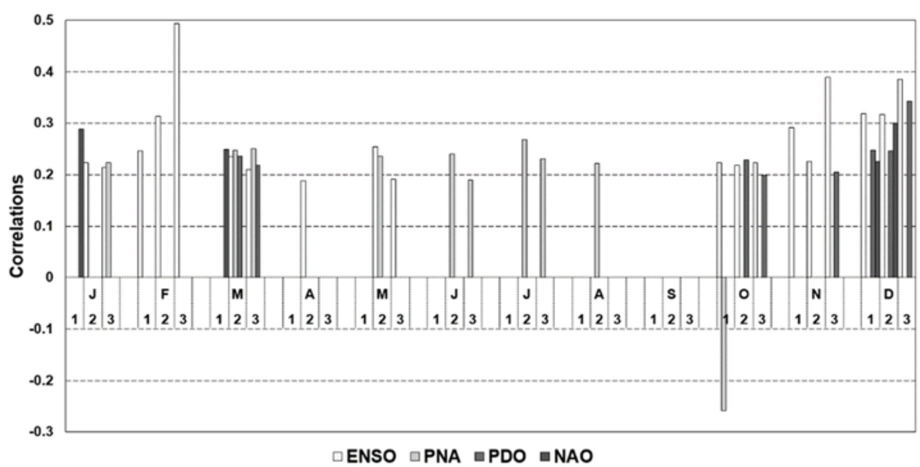

Figure 2. *Correlations between each teleconnection and monthly precipitation anomalies from 1901-2010 in each region (no lag). Only the correlations that are statistically significant at a 95% confidence level are shown.*

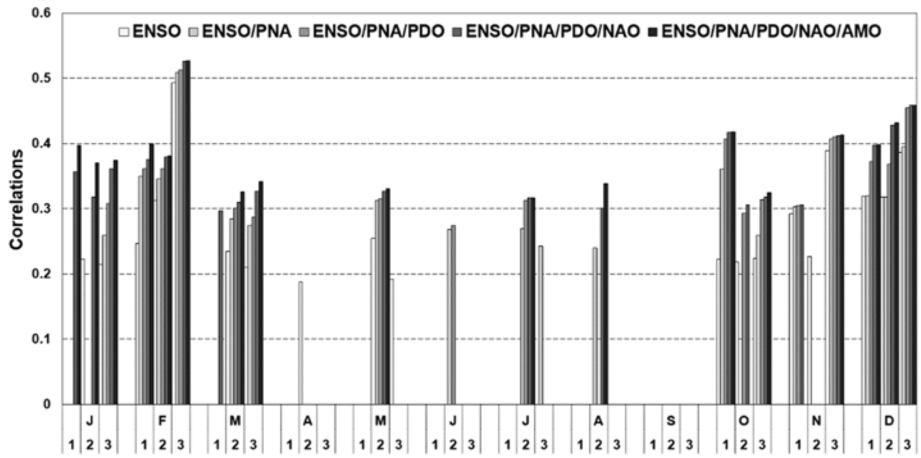

Figure 3. *Correlations between multiple teleconnections and monthly precipitation anomalies from 1901-2010 in each region (no lag). Only the correlations that are statistically significant at a 95% confidence level are shown.*

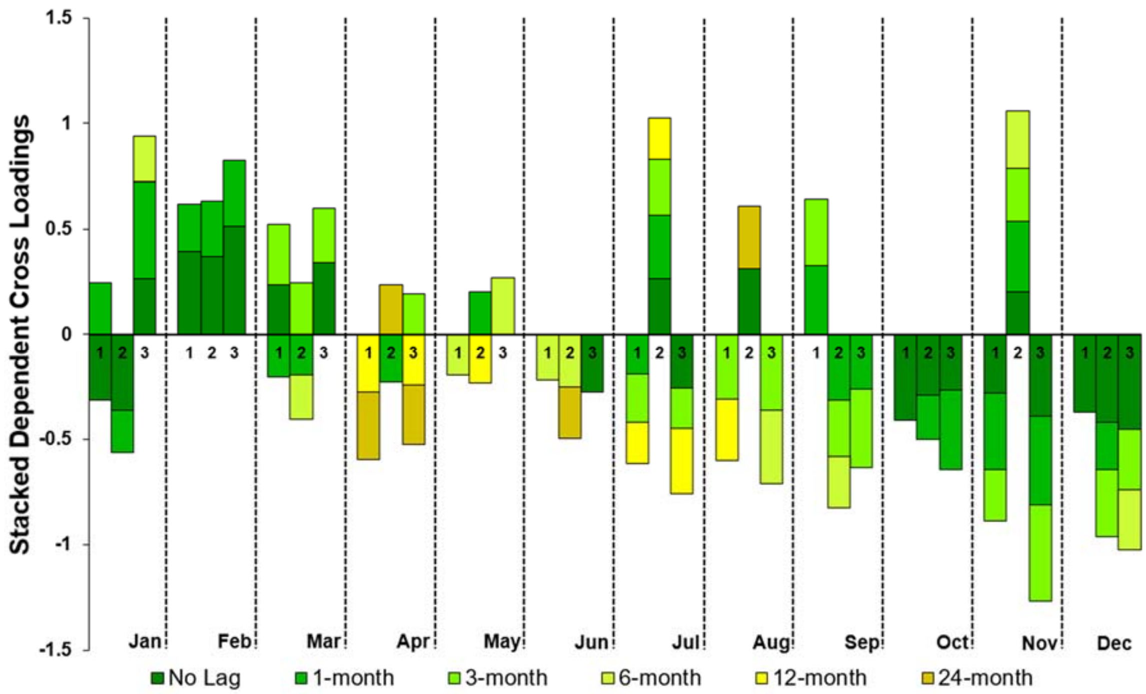

Figure 4. *Clustered stacked bar chart showing dependent cross loadings from the CCA for all teleconnections and monthly precipitation for all regions at various time lags (0 to 24 months). Only the dependent canonical cross loadings that are statistically significant at the 95% confidence level are shown.*

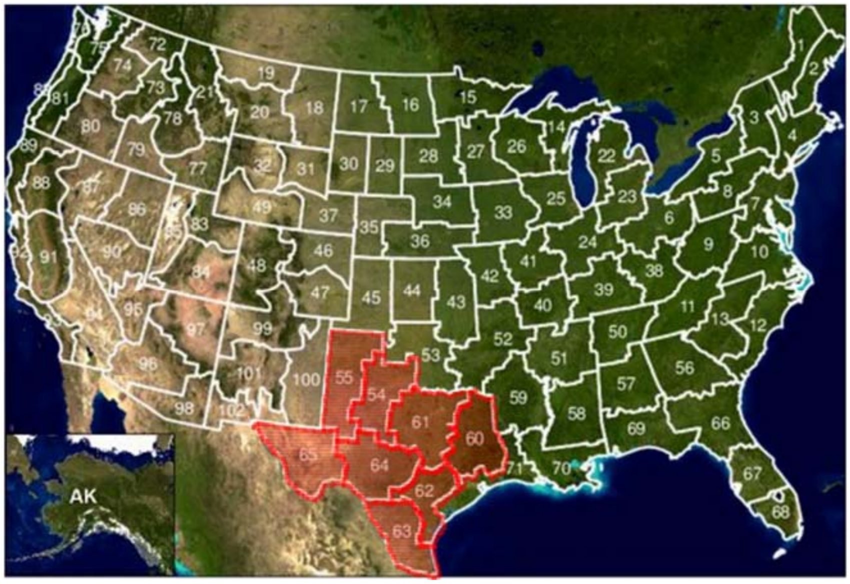

Figure 5. *Map of CPC climate divisions with regions used in this study highlighted in red (http://www.vwt.ncep.noaa.gov/)*

Table 1. *Model evaluation statistics for the precipitation forecasts (0-month lead) of January in the three regions based on the 1981-2010 forecast evaluation period.*

	All five teleconnections				ENSO			
	R^2	P	RMSE(mm)	MAE(mm)	R^2	P	RMSE(mm)	MAE(mm)
Region 1	0.15	0.03	49.11	39.25	0.07	0.16	44.96	36.47
Region 2	0.08	0.14	24.55	18.67	0.05	0.23	25.24	18.45
Region 3	0.48	0.00	28.76	22.46	0.41	0.00	28.55	24.25

Table 2. *Comparison of Heidke Skill Scores between the CCA-based precipitation forecasts model and the CPC for the month of January in three regions (0-month lead)*

HSS 2005-2010 Direct Comparison				**HSS 2000-2010 (CCA) and 2005-2015 (CPC) Indirect Comparison**			
CCA		*CPC*		*CCA*		*CPC*	
Region 1	0.08	Region 60	-0.25	Region 1	0.31	Region 60	-0.18
		Region 61	0.50			Region 61	0.36
		Average	0.13			Average	0.09
Region 2	0.67	Region 54	0.75	Region 2	0.44	Region 54	0.36
		Region 55	0.75			Region 55	0.36
		Region 64	0.50			Region 64	0.36
		Region 65	0.50			Region 65	0.32
		Average	0.63			Average	0.35
Region 3	0.45	Region 62	0.25	Region 3	0.33	Region 62	0.09
		Region 63	0.50			Region 63	0.36
		Average	0.38			Average	0.23