**Introduction**

Predicting future regional trends is one of the biggest challenges currently facing climate scientists. Presently, climate models are reasonably adept at capturing large-scale historical features in response to prescribed observed radiative forcing. For example, global average surface temperature trends, are well simulated by general circulation models (GCMs) (Kirtman et al. 2013). However, it is knowledge of trends at a regional scale that is crucial to adapting to the effects of climate change. Accurate regional climate trend prediction would enable policy makers and local officials to better organize efforts to adapt to the negative effects of climate change. Moreover, the average person is likely to have a greater interest in regional climate trends, how climate change will affect where he or she lives, and a lesser interest in global trends.

Unfortunately, GCMs perform poorly at sub-continental and smaller scales. At these scales, observed trends typically fall in the tails of the distribution of modeled trends (Bhend and Whetton 2012; Knutson et al. 2013; van Oldenborgh et al. 2013; Kirtman et al 2013). In fact, the recent IPCC report asserts that the Coupled Model Intercomparison Project Phase 5 (CMIP5) ensemble cannot be used to make reliable regional forecasts, and that uncertainty can be larger than the model spread. It is true that natural variability and model spread is larger at smaller scales, but these differences do not fully account for the discrepancies between model prediction and observations at regional scales. One phenomenon that negatively impacts GCM accuracy is too-high climate sensitivity (Kumar et al. 2013). Meaning that the temperature increases too rapidly in response to increased CO­2. As a result of excessive climate sensitivity GCMs tend to project positive temperature trends across the globe, and therefore only seemingly accurately predict positive regional temperature trends.

The main goal of this study is to assess the predictability of regional climate trends. One widely used definition of a regional climate trend is the area averaged trend over a particular region. A major motivator behind this study is assessing regional forecast skill so individuals planning for the future have a better idea of how much trust they can put into climate forecasts for where they live. For this application an area averaged climate trend is not necessarily useful. Therefore, in this study when asking if models capture regional trends, we’re asking if models are able to reproduce the spatial patterns of climate trends over a particular area. Is it possible for contemporary climate models to accurately forecast the spatial patterns of regional trend? To answer this question regional climate trends generated by the phase-1 North American Multi-Model Ensemble (NMME) were compared with observations. These comparisons will be made using a novel Monte Carlo approach.

The NMME is an experimental seasonal forecasting system that utilizes coupled models from various modeling centers and is ideal for analyzing inherent predictability at the regional scale. Each component model provides an approximately 30-year hindcast that can be used to diagnose the predictability of the ensemble. Additionally, it has been shown that the multi-model ensemble approach generates better predictions than any single model (Palmer er al. 2004; Hagedorn et al. 2005; Doblas-Reyes et al 2005). What makes the NMME most suited for this undertaking is that, unlike forecasts from other model ensembles such as CMIP5, NMME forecasts have short lead times: from 1 to 9 months. If reliable regional trends cannot be predicted at lead times that are only months removed from observations, the current generation of climate models may be incapable of reliably predicting regional climate trends. Alternatively, an inability to recreate regional trends at extremely short lead times may indicate that there is little to no inherent predictability in climate trends over small areas.

There are past studies that evaluate the predictability of regional climate trends but they either focus on one region (van Oldenborgh et al. 2009) or a set of hand-picked regions (Bhend and Whetton 2012). Studies show that certain regions have more predictive skill than others at decadal time scales (Meehl et al. 2014) Therefore, using a few specific regions and ignoring the rest of the globe may not be an accurate way to determine how a model predicts regional trends in general. A Monte Carlo Approach will be taken to assess the predictability of regional trends in an attempt to eliminate this partiality.

Simply put, a Monte Carlo simulation is the computer generation of random objects (Krouse et al 2014). In this case, the Monte Carlo Method (MCM) will be used to generate thousands of randomly located regions with a set dimension size for which regional trends can be calculated. The correlation between the model derived climate trend and observed climate trend will be calculated for each of the regions randomly generated by the MCM. The analysis of the MCM results should reveal how each model is able to forecast trends over a range of regional dimension sizes.

The MCM is utilized in three ways in this study. Not only is it used to evaluate regional predictability by comparing the NMME to observations, but also it is also used to estimate an individual model’s maximum predictability via the “perfect model assumption” (Becker et al 2014) and as a measure to compare against in to lend to or detract from the confidence in a particular regional forecast.

The homogeneous test builds on the idea of using a Monte Carlo simulation to assess how models perform at regional scales without having to analyze a relatively small number or specified regions. In this case the MCM is used to correlate randomly selected regions across ensemble members from the same model. How well a model is able to reproduce results from one ensemble member run to another for a given regional dimension is a measure of potential predictability.

Confirming or adding confidence to forecasted climate trends in a specific region is an application of this MCM that could be applied to countless smaller scale long-term projects that are dependent on climate change. Currently, there is so much uncertainty in climate prediction on a regional scale that information from GCMs is essentially useless. One example of a current project that could benefit from this analysis is the Comprehensive Everglades Restoration Plan (CERP). Climate change is a significant threat to the everglades, which as a wetland environment is extremely sensitive to changes in precipitation patterns. GCMS project that South Florida will experience a -10% to +10% change in precipitation for 2060 (National Research Council 2014). This projection is not helpful and contributes to the fact that climate change is not largely considered by CERP in its planning process.

If a GCM ensemble produced precipitation trends over Florida have a higher correlation with observations compared to random regions of the same dimension produced by an MCM, then project developers may have more confidence in GCM projections and in integrating climate change into plans for the future. We’ll test how precipitation trends over Florida provided by the 133 NMME members stack-up against the MCM in this study.

**Data and Methods**

1. *North American Multi-Model Ensemble*

The NMME is an experimental seasonal forecasting system that uses coupled models from NOAA/NCEP, NOAA/GFDL, IRI, NCAR, NASA, and Canada’s CMC modeling centers (Kirtman et al. 2014). This ensemble was constructed to address the absence of a forecast between the typical 10-day weather forecast and a full season. The prediction systems that comprise the NMME were readily available and independently developed prior to the development of the NMME. These products were assembled to utilize the multi-model approach to resolving forecast uncertainty.

The models included in the NMME are listed in Table 1 along with additional information including the center at which each model was produced and the ensemble size. The CanCM3, CanCM4, GFDL, GFDL-FLOR, CFSv1, CFSv2, ECHAM4-a, ECHAM4-f, CCSM3, and CCSM4 models, a total of 133 ensemble members, were utilized to assess regional climate predictability. A requirement of the NMME is that the models include lead times out to at least 9 months, however this study will only included analysis of 1-month, 3-month, and 6-month lead times. In phase-1 of the NMME project each model was required to output forecasts for sea surface temperature (SST), 2-meter temperature (T2M), and precipitation rate (PREC).

Every participating model generated an approximately 30-year hindcast for each of the 1 to 9 month lead times in order to validate model runs with observations. In this study trends were calculated over the period of January 1982 to December 2009 (28 years) for each of the models. Least squares regression was used to determine the trend at each grid point (Räisänen 2007).

1. *Observations*

The CPC Merged Analysis of Precipitation (CMAP) was used as the verification data set for precipitation rate. CPC-CMAP combines satellite precipitation data and raingauge observations on a 1° latitude by 1° longitude grid. Global monthly data is available from January 1979 onwards.

SST was verified against a NOAA produced optimum interpolation (OI) analysis (Reynolds et al. 2002). This dataset interpolates SST using satellite data and on-site measurements from ships and buoys. Like the NMME output, the SST OI analysis is on a 1.0° latitude by 1.0° longitude grid.

1. *Trend Maps*

Trend maps were calculated for the ensemble mean (EM) of each individual climate model that comprises the NMME for a 1-month, 3-month, and 6-month lead-time as well as for observations. Each ensemble member forecast can be represented by *F(s,y,m,l),* where *s* is the forecast grid point, *y* is the year, m is the month, and *l* is the lead-time. The EM for a model is:

The trend maps for each model EM were then averaged to obtain the trend maps for the mean of the entire NMME for 1-month, 3-month, and 6-month lead times. The difference between averaging the trends of all of the 99 ensemble members that make up the NMME and averaging the trends of each model EM is negligible (Kumar et al. 2014).

Before calculating the trend the mean-bias was removed from the hindcasts and observations. These anomalies were computed by subtracting the climatological means. Using the formulas from Becker et al 2014 for each model:

Where {} denotes the mean over the 1982-2009 time period. The anomalies of the observations are defined as,

Where is the local climatology over the 1982-2009 time period.

1. *Monte Carlo Method*

A novel approach was taken to evaluate the predictability of regional trends. A Monte Carlo scheme was employed to randomly compare thousands of regions on the globe instead of focusing on only one, or an assortment of hand picked regions, as in previous studies. The MCM was used to generate correlation values between the map of the 28-year, point-by-point trend of SST and precipitation rate and the corresponding observed trend for various sized regions. The regional 2-D correlation between the forecast and observation trend maps is calculated as:

Where is the global mean of the forecast values, and is the mean of the observed values over the 2-D region.

First, a dimension size is selected for the regions that will be randomly isolated. Once the region dimension has been assigned, 10,000 sections are selected by randomizing the coordinate of the north-eastern corner. This number of iterations was chosen to be large enough to ensure that the entirety of the global map would be encapsulated in the MCM yet small enough that the MCM would minimize computer run time. The spatial correlation between the predicted and observed trend was calculated for all 10,000 regions.

Some restrictions were placed on the randomly selected regions. Only regions that fell entirely between -60.0° and 60.0° latitude were included in the MCM because of the unreliability of observations from the Polar Regions. Additionally, the MCM regions were required to have square dimensions. All 10,000 iterations of the MCM were repeated for each trend map for a maximum dimension of 120.0° x 120.0° and minimum dimension of 5.0° x 5.0° and every 5.0° increment in between.

This MCM was applied to the data in three fashions. Firstly, the MCM was used to assess the regional predictability of model produced climate trends. In this case the MCM was simply applied to a trend map of observations and a trend map derived from the NMME. Secondly, the MCM was utilized to compare individual ensemble runs within a model to serve as a “perfect model assumption.”

Secondly, the “perfect model assumption”, or homogeneous experiment, was first carried out by randomly choosing one ensemble member from a model to be the “truth.” This “truth” ensemble member was essentially treated as observations. Every other ensemble member from the model was compared to the “truth” using the MCM. This procedure was repeated for every individual model for SST and precipitation for 1-month, 3-month, and 6-month lead times.

Thirdly, the MCM was slightly modified to assess how trends in one particular region compare to trends over all regions of that same size. Instead of using a single trend map, the first step of the MCM was to randomly select one of the 133 ensemble members from the NMME. After the ensemble member was chosen the MCM carried on, as previously described, randomly selecting a region from the globe and evaluating correlation with observations. Because a second random element was added to the MCM the number of iterations had to be increased from 10,000 to 50,000 to ensure that repetitions of the MCM do not yield different results. The correlation values from this modified MCM were compared to the correlation values for trend patterns over one specific region was generated by each of the 133 ensemble members in the NMME and the observations.

**Results**

1. *Global Trends*

Given the short lead-times of the NMME model there is little surprise that the global predicted trend maps of SST and precipitation in fig. 1 closely resemble the trend maps of the observed SST and precipitation trends. The 1-month and 3-month lead time NMME SST trend maps have a cooling signature in the eastern Pacific, which is also present in the observed trend map. With a 6-month lead-time the cooling SST trend is mostly lost in the eastern Pacific, and the SST warming trend expands to cover most of the globe.

The global precipitation trend map with a 1-month lead-time at first glance appears to be a very good reproduction of the trend from the observational estimate. However, upon closer inspection the spatial precipitation pattern in the tropical Pacific is well captured while trends in the Indian Ocean and extra-tropics are lost. Additionally, one can see that the precipitation signal in the Pacific that dominates the observations has deteriorated by the 6-month lead-time. The observed trend in the Pacific resembles a La Niña signal along the equator while the 6-month trend more resembles El Niño.

1. *Regional Trend Over North America*

At first glance, the NMME produced global trend maps of SST and precipitation (Fig. 2) trends appear to be a close match to the global observations, but what if the maps are scaled down? Here the precipitation trend maps have been zoomed into the North American continent. Immediately it is apparent that at this regional scale the predicted trend maps do not correspond as closely with the observed trends. Interestingly, like the NMME and observed global trend maps, the regional 1-month lead precipitation trend over North America has a pattern indicative of La Niña. There is anomalously high precipitation in the northern half of North America and low precipitation in the southern portion of North America. This pattern persists to a 3-month lead-time.

In general, the modeled precipitation trend is a fair match to the observed precipitation trend over the North American continent given the correlation between the two is .43. However, this is only one particular region of one given size. The MCM will give a quantitative measure of how the NMME is able to recreate regional trends across the entire globe for a range of region dimensions.

1. Regional Monte Carlo

Figure 3 exhibits the spread of the spatial correlation between the NMME and the observed trend for both SST and precipitation. Contours represent the percentage of the 10,000 iterations of the MCM with a specific correlation value between the NMME average and the observations. For example a correlation density of .4 corresponding to a region size of 60° and a correlation value of .5 indicates that 40% of the 10,000 Monte Carlo iterations for a 60°x60° region have a .5 correlation with observations. The solid black line is the mean correlation for a particular region size, and the dotted black lines are one standard deviation from the mean.

The spread of the correlation between the NMME 1-month lead SST trend and the observed SST trend for a 120° square region is from approximately 0.7 to 0.9 and begins to increase at a region size of about 75° square. The spread of the correlation for the NMME 1-month precipitation rate trend doesn’t drastically increase until the region sized has been decreased to less than a 60° square region. However, the correlation spread for the precipitation trends initially has a larger spread and a smaller correlation value, about 0.3 to 0.7, than the correlation spread of the SST trends.

Given a 1 month lead time the SST forecast skill is quite good at the regional scale. The mean correlation value remained steady as the region size diminished, never falling below .7. The spread of the correlation values grew as the region size decreased, but one standard deviation resides well above the zero correlation mark down to a 5°x5° dimension size. The multi-model ensemble does not perform as well in the case of precipitation. Once the region size has been diminished to approximately 20° a correlation value of zero falls within one standard deviation of the mean. This cutoff is used to designate the region size at which there is no forecast skill. In other words, at a 1-month lead time the NMME average has no forecast skill for PREC for regions that are 20° or smaller. Given that the models were re-initialized with observations every month, these results corroborate the assertion that contemporary GCMs cannot be used to make regional climate trend predictions for domain sizes 20° and smaller.

For the 3-month lead-time the initial SST correlation spread, 0.4 to 0.7, begins to grow as soon as the region size decreases, but the average correlation value remains fairly consistent. Unlike the 1-month lead case, a zero correlation value falls within one standard deviation at a region size of about 15°. The mean correlation value for precipitation decreases at a faster rate than the mean for SST. At a 3-month lead there is no forecast skill for precipitation at any regions smaller than 40°.

By a 6-month lead time a zero correlation value falls within one standard deviation from the mean at all region sizes for SST and precipitation. Essentially there is no regional forecast skill for either SST or precipitation at any region dimension at a 6-month lead.

The regional prediction ability of individual models from the NMME was also assessed for the same lead times. A few of the models managed to perform as well as the NMME average for SST, but none of the models noticeably beat the multi-model ensemble at predicting precipitation trends. Conversely, it is clear which models were outperformed by the multi-model ensemble. The correlation spread diagrams of a couple of the “best” and “worst” models compared to the observation datasets are included in Figs. 4, 5 and 6.

At one 1-month and 3 month lead times CanCM3 (Fig. 4) and CanCM4 have the same forecast skill for SST trends as the NMME. The individual models have a marginally larger spread in the correlation; however, like the NMME average, both models exhibit a one standard deviation spread that exists well above the zero correlation designation. NCEPv2 performs nearly as well as CanCM3 and CanCM4 at a 1-month lead but interestingly becomes one of the worst models at a 3-month lead times and longer with no SST forecast skill at regions small than 90° and no predictability for precipitation whatsoever (Fig. 5 and 6).

Examining the performance of the individual models at a 6-month lead time it is apparent that two models, GFDL-CM2.1 and GFDL-FLOR, stood out. In Fig. 5 you can see that GFDL-FLOR has a non-zero quantity of regional predictability for SST for regions larger than 70°. GFDL-CM2.1 had very similar results to GFDL-FLOR, the major difference being that GFDL-CM2.1 displayed slightly lower correlation values at the largest dimension sizes. None of the other models had a one standard deviation spread completely above zero correlation. Surprisingly, more models (CanCM3, CanCM4, GFDL-CM2.1, and IRI-ECHAM4f h) had one standard deviation fall above zero correlation in the case of precipitation than in the case of SST, but only for the larger region dimensions (>80°).

The CCSM3 model (Fig. 4) is the “bad” model from the ensemble, which is not surprising. CCSM3 solely has ocean initialization while the other models attempt to have realistic atmospheres (Becker 2014). Also the CCSM3 model does not include radiative forcing. There is no SST predictability at any region sizes and, a zero value correlation falls within one standard deviation for SST at all region dimensions smaller than 80°.

There are a couple of common traits that appear in the precipitation results of the MCM for all of the individual models. Most of the precipitation plots exhibit a “banding” where there is a higher concentration of correlation values. The banded pattern, an example of which is shown in figure 6, persists throughout all lead-times and appears in each of the models. Originally it was believed that this was most likely a product of the difference in the skill for predicting precipitation over land vs. over water. However, upon further experimentation, running the Monte Carlo with a land mask and with an ocean mask, the banding pattern persisted. The MCM was also repeated to single out the tropics and extra-tropics (-23°N to 23°N) and -45°N to 45°N. In every case the banded precipitation pattern appeared in the results. Therefore, it seems that this recurring pattern in the precipitation correlation is due to various regions spread across the globe that are better forecasted than others, and not necessarily as systematic as land vs. ocean or tropics vs. extra-tropics.

Another pattern that appears in the precipitation correlation, but in this case for the 6-month lead-time only, is a very small spread of negative correlation values for larger regions. CCSM3, ECHAM4-a, and NCEP2 (Fig. 4) exhibit this pattern; therefore, the NMME mean for a 6-month lead also has this pattern, although to a lesser degree. The overconfidence of the models combined with the negative correlation indicates that the models systematically produce regional precipitation trends that are spatially out of phase with the observational estimates.

**Confirming Forecast Skill at Specific Regions**

The driving idea behind using a MCM was to be able to analyze a model or ensemble’s regional forecast skill without selecting specific regions on which to base an analysis. However, the MCM can be used to add confidence to a forecast over a specific region. If a particular region has a higher correlation with observations, persistent across multiple models, than the Monte Carlo spread of correlations for a region of the same size may lend more confidence to the ensemble’s predictions for that area. A slightly modified MCM was used to explore this notion for three regions of varying dimension, the entire North American Continent (60°x60°), the southeastern United States (20°x20°), and the state of Florida (8°x8°).

In this case, the MCM randomly selected one of the 133 ensemble members and then randomly selected a region to correlate to observations. Fig. 7A the 50,000 correlation values for a 60° region produced by the MCM are plotted in the red histogram. The blue histogram depicts the 133 correlation values between the actual 60°x60° regions that contain North America from each individual NMME member ad observations. The 1-month lead-time results indicate that North America is a region where the forecast skill is quite good. The blue histogram peaks at a higher correlation value than the red and the spread is quite small. This is not surprising given that one can see in Fig.2 that the NMME 1-month lead-time precipitation trend had a fairly good likeness to the observed precipitation trend. Also noteworthy is that the second peak in the blue histogram around .1 correlation is composed of values from only one model, CFSv2.

At a 3-month and 6-month lead-time the MCM is nearly a Gaussian distribution around zero. The blue histogram of correlation values for North America from each ensemble member is slightly more positively skewed for a 3-month lead-time than the MCM, but not convincingly enough to add confidence to the North American regional forecast. And interestingly, at a 6-month lead North America appears to be an area with particularly poor forecast skill.

Not only was it suspected that North America would prove to have above average precipitation forecast skill due to the trend maps in Fig. 2, the North American region is fairly large and given the results of the MCM in Fig. 3 it’s known that the NMME yields higher correlation values and smaller spreads of correlation values for larger regions. What happens if we look at a smaller area? In Fig. 7B it appears that the southeastern United States is not a particularly strong region for precipitation trend forecast skill. The correlations between the individual NMME members and observations peaks at about .2 while regions of that dimension (20°) peak at a correlation of about .5. Both histograms are evenly distributed across zero correlation for a 3-month and 6-month lead.

Finally, how does a very small region compare to the results of the MCM? The precipitation trend maps in Fig. 8 cover an 8°x8° region that contains the entire state of Florida. The precipitation trend pattern is fairly persistent for all three lead-times. In all three cases the precipitation trend is negative throughout the entire region. The 1-month lead-time trend pattern is similar to the observations, both maps exhibit a drying trend along the Atlantic coast and over the Gulf of Mexico just south of the panhandle.

Fig. 9 shows how Florida correlates to observations in each of the ensemble members compared to all 8° regions. The 1-month lead histogram of Florida correlation values from the 133 ensemble members looks to be quite spread out, however, like in the case of the North American region, the second peak at -.1 correlation is comprised of values only from the CFSv2 model. Knowing that the second peak in the set of Florida correlations is due to one model, the 1-month lead-time plot does add some confidence to the precipitation forecast skill for the state of Florida. The blue histogram has a distinct peak at a correlation of .5, while the MCM results for an 8° region are more evenly distributed across the entire range of correlation values.

Like in the cases of the two larger regions, the Florida region does not appear seem to be a particularly well forecasted region compared to the MCM results. The Florida correlations are also spread across all correlation values. This was somewhat surprising given that all of the maps in Fig. 8 had entirely negative precipitation trends. Unlike the analysis on the two larger regions a further investigation was made into the Florida region. Given that Florida has distinct wet and dry seasons and the trends were originally calculated across all seasons, the precipitation trends were recalculated for the wet and dry seasons and the MCM was reapplied.

At a 1-month lead-time there is a noticeable difference between the correlation values for Florida and observations from the 133 ensemble members. Precipitation trends for Florida over the dry season have correlation values that are skewed negative. In this case Florida does not appear to be a region with confident forecast skill. The 1-month lead precipitation trend correlations over the wet season are evenly distributed around zero. These results are counterintuitive, typically forecast skill for the southeastern US is greater for the Boreal winter (Florida dry season) than for the Boreal summer (wet season) (Kirtman et al 2014). Relatively greater predictability in the summer months could be a result of the westward expansion of the North Atlantic Subtropical High over Florida in the summer (Li et al. 2012). Westward expansion of high pressure would stymie hard to predict, small-scale convection. Interestingly, neither the wet or dry season forecasts for a 1-month lead are an improvement over the trend calculated over all seasons. The dry and wet season Florida correlation values have an even distribution over all correlation values at the two-longer lead times, which is the same as the correlation values for Florida precipitation trends over all seasons.

**Homogenous testing**

This Monte Carlo method can also be applied with the “perfect model assumption” to measure an individual model’s confidence in its ability to predict regional trends. In this case one model ensemble member is designated as the “truth” and takes the place of observations. Then the Monte Carlo experiment is repeated for each of the remaining ensemble members compared to the “true” ensemble member. Quantifying a model’s ability to reproduce its own results, gives us an idea of a model’s predictability for a particular variable, which can add confidence to the model’s forecasts.

The correlation plots for these homogeneous experiments are similar to previous figures however there are more individual correlation values plotted. For example, the CanCM3 model has 10 ensemble members; therefore, the MCM must be used 9 times in order to compare every other ensemble member with the “true” ensemble member. The results of all 9 MCM runs are plotted on the same figure. Consequently, there are 90,000 correlation values for each dimension size rather than 10,000 correlation values.

Each of the individual models is able to almost perfectly recreate regional SST trends down to the smallest region sizes. Fig. 11 shows the results of the homogeneous experiment for the canCM3 model. Not only does the CanCM3 model have nearly perfect homogeneous SST predictability, but it is also one of the best models at forecasting observations. Therefore, in this case the homogeneous experiment adds confidence to the model’s forecasts. The homogeneous MCM results were also included for the NCAR-CCSM3 model (Fig. 11), which happens to be the model with the highest homogeneous predictability for SST and precipitation. This illustrates the fact that a model with great regional predictability may not provide an accurate regional forecast.

All of the models performed similarly at self-predicting SST trends. Mean correlation between all the other ensemble members and the “truth” is greater than .8 for all individual models, at all region sizes, and all lead times out to 6 months. However, this is not the case for precipitation trends. With only a 1-month lead-time the individual models typically have an average correlation value of .8 for the largest region sizes, but average correlation values dramatically decrease and the spreads of correlations increase as the region size shrinks. Therefore, not only do the models fail to accurately represent observed precipitation trends, the have very little homogenous predictability as well. This result is not surprising, as it is well known that forecast skill for precipitation is considerably lower than SST (Kirtman et al. 2014).

**Conclusions and Summary**

Traditionally, one would determine a model’s regional forecast skill by evaluating correlation patterns on a global map, as in the first portion of the previous section. However, when looking at global maps it is difficult to discern at what scales the models maintain predictability and what is the cutoff at which the models’ trends are no longer trustworthy. The MCM does not single out particular regions that are more predictable; it gives an idea of which individual models or model ensembles are more accurate at regional scales in general and at what scales models lose all predictability. However, the findings of the MCM can be utilized to determine if a region is well predicted compared to all other regions of that dimension.

Applying the MCM on the NMME average SST and precipitation trends from 1981-2010 reveals that there is very little forecast skill at regional scales when the models are run with a 1-month lead time, and there is no forecast skill when the models are run with longer lead times. Even though there is some forecast skill at a 1-month lead time, the results still point to model shortcomings. Given that the models are being fed observations at the beginning of every month, discrepancies between the models and observations must be a result of model failure and not natural variability. Also there are significant model deficiencies at these scales to develop such low mean correlation and large correlation spread at smaller regional scales after merely one month. These findings are consistent with the assertion that GCMS are unreliable at regional scales.

The MCM can also be used along with a “perfect model assumption” to perform homogeneous tests on the individual models that comprise the NMME to quantify model predictability at various spatial scales. All of the models were consistent from one ensemble run to the next in regards to SST. However, this was not the case for precipitation trends. Even with monthly re-initializations with observations the models are unable to confidently recreate their own precipitation trends.

Using the MCM as a measure with which to assess the trustworthiness of a regional trend forecast was also explored. Precipitation trends over the entirety of North America, the southeastern United States, and the state of Florida were compared to MCM results for 60°, 20°, and 8° dimensions respectively. At a glance, precipitation trends for North America are a fair likeness to observations, therefore North America is expected to be an area with good forecast skill. Compared to the MCM, North American trends are well modeled given a 1-month lead time; however, with a longer, 6-month lead time North American precipitation trends appear to be more poorly correlated with observations than most 60° regions according to the MCM.

One specific motivation behind this application of the MCM was to see if it would lend confidence to forecasted precipitation trends for the planning of Everglades restoration. Currently, GCMs project a -10% to +10% change in precipitation by 2060 for south Florida. Unfortunately, the 133 NMME precipitation trend patterns for Florida correlated with observations compared the MCM results for an 8° region dimension do not impart significant confidence in the GCM projections. Florida has a higher correlation with observations than the MCM for a 1-month lead time, but performs roughly equal to the MCM with 3-month a 6-month lead times. Despite Florida appearing to be a well modeled region compared to other 8° regions at a 1-month lead-time these results do not instill confidence for forecasts made with longer lead-times. A well-forecasted area with a 1-month lead-time could be a one of the poorer forecasted regions at a longer lead-time as in the case of the North American region in this study.

**References**

Adler, R. F., and Coauthors, 2003: The version-2 global precipitation climatology project (GPCP) monthly precipitation analysis (1979-present). *J Hydrometeor* **4,** 1147-1167.

Becker, Emily, Huug van den Dool, Qin Zhang, 2014: Predictability and Forecast Skill in NMME. *J. Climate*, **27**, 5891–5906. doi: http://dx.doi.org/10.1175/JCLI-D-13-00597.1

Bhend J and Whetton P 2012 Consistency of simulated and observed regional changes in temperature, sea level pressure and precipitation *Clim. Change* doi:10.1007/s10584-012-0691-2

Delworth, T. L., and Coauthors, 2006: Gfdl's cm2 global coupled climate models. part i: formulation and simulation characteristics. *J. Climate*, **19**, 643–674.

doi: <http://dx.doi.org/10.1175/JCLI3629.1>

DeWitt, D. G., 2005: Retrospective forecasts of interannual sea sur- face temperature anomalies from 1982 to present using a directly coupled atmosphere–ocean general circulation model. Mon. Wea. Rev., 133, 2972–2995, doi:10.1175/MWR3016.1.

Doblas-Reyes, F. J., R. Hagedorn, and T. N. Palmer, 2005: The rationale behind the success of multi-model ensembles in seasonal forecasting-II. Calibration and combination. *Tellus,* **57A,** 234-252, doi:10.1111/j.1600-0870.2005.00104.x.

Hagedorn, R., F. J. Doblas-Reyes, and T. N. Palmer, 2005: The rationale behind the success of multi-model ensembles in seasonal forecasting-I. Basic concept. *Tellus,* **57A,** 219-233, doi:10.1111/j.1600-0870.2005.00103.x.

**IPCC**, 2014: Climate Change 2014: Impacts, Adaptation, and Vulnerability. Part A: Global and Sectoral Aspects. Contribution of Working Group II to the Fifth Assessment Report of the Intergovernmental Panel on Climate Change [Field, C.B., V.R. Barros, D.J. Dokken, K.J. Mach, M.D. Mastrandrea, T.E. Bilir, M. Chatterjee, K.L. Ebi, Y.O. Estrada, R.C. Genova, B. Girma, E.S. Kissel, A.N. Levy, S. MacCracken, P.R. Mastrandrea, and L.L. White (eds.)]. Cambridge University Press, Cambridge, United Kingdom and New York, NY, USA, 1132 pp.

Kirtman, B. P., and D. Min, 2009: Multimodel ensemble ENSO prediction with CCSM and CFS. Mon. Wea. Rev., 137, 2908– 2930, doi:10.1175/2009MWR2672.1.

Kirtman, B. P., and Coauthors, 2014: The North American Multi-Model Ensemble (NMME): Phase-1 seasonal interannual prediction, phase-2 toward developing intra-seasonal prediction, phase-2 toward developing intra-seasonal prediction. *Bull. Amer. Meteor. Soc,* **95,** 585-601, doi:10.1175/BAMS-D-12-00050.1.

Kirtman, B., S.B. Power, J.A. Adedoyin, G.J. Boer, R. Bojariu, I. Camilloni, F.J. Doblas-Reyes, A.M. Fiore, M. Kimoto, G.A. Meehl, M. Prather, A. Sarr, C. Schär, R. Sutton, G.J. van Oldenborgh, G. Vecchi and H.J. Wang, 2013: Near-term Climate Change: Projections and Predictability. In: *Climate Change 2013: The Physical Science Basis. Contribution of Working Group I to the Fifth Assessment Report of the Intergovernmental Panel on Climate Change* [Stocker, T.F., D. Qin, G.-K. Plattner, M. Tignor, S.K. Allen, J. Boschung, A. Nauels, Y. Xia, V. Bex and P.M. Midgley (eds.)]. Cambridge University Press, Cambridge, United Kingdom and New York, NY, USA.

Kroese, D. P., Brereton, T., Taimre, T., & Botev, Z. I. (2014). Why the Monte Carlo method is so important today. *Wiley Interdisciplinary Reviews: Computational Statistics*, *6*(6), 386-392.

Kumar, S., Merwade, V., Kinter III, J. L., & Niyogi, D. (2013). Evaluation of temperature and precipitation trends and long-term persistence in CMIP5 twentieth-century climate simulations. *Journal of Climate*, *26*(12), 4168-4185.

Li, W., Li, L., Ting, M., & Liu, Y. (2012). Intensification of Northern Hemisphere subtropical highs in a warming climate. *Nature Geoscience*, *5*(11), 830-834.

Masson D, Knutti R (2011) Spatial-scale dependence of climate model performance in the CMIP3 ensemble. J Climate 24(11):2680-2692

Meehl, G. A., Goddard, L., Boer, G., Burgman, R., Branstator, G., Cassou, C., ... & Yeager, S. (2014). Decadal climate prediction: an update from the trenches. *Bulletin of the American Meteorological Society*, *95*(2), 243-267.

Merryfield, W. J., and Coauthors, 2013: The Canadian Seasonal to Interannual Prediction System. Part I: Models and ini- tialization. Mon. Wea. Rev., 141, 2910–2945, doi:10.1175/ MWR-D-12-00216.1.

National Research Council. Progress Toward Restoring the Everglades: The Fifth Biennial Review, 2014. Washington, DC: The National Academies Press, 2014.

Palmer, T. N., and Coauthors, 2004: Development of a European Multimodel Ensemble System for Seasonal-to-Interannual Prediction (DEMETER). *Bull. Amer. Metero. Soc,* **85,** 853-872, doi:10.1175/BAMS-85-6-853.

Saha, S., and Coauthors, 2006: The NCEP Climate Forecast Sys- tem. J. Climate, 19, 3483–3517, doi:10.1175/JCLI3812.1.

Suranjana Saha, Shrinivas Moorthi, Xingren Wu, Jiande Wang, Sudhir Nadiga, Patrick Tripp, David Behringer, Yu-Tai Hou, Hui-ya Chuang, Mark Iredell, Michael Ek, Jesse Meng, Rongqian Yang, Malaquías Peña Mendez, Huug van den Dool, Qin Zhang, Wanqiu Wang, Mingyue Chen, and Emily Becker, 2014: The NCEP Climate Forecast System Version 2. *J. Climate*, **27**, 2185–2208.

Räisänen J 2007 How reliable are climate models? *Tellus* A **59** 2-29

Reynolds, R. W., N. A. Rayner, T. M. Smith, D. C. Stokes, and W. Wang, 2002: An improved in situ and satellite SST analysis for climate. *J. Climate,* **15,** 1609-1625, doi:10.1175/1520-0442(2002)015<1609:AIISAS>2.0.CO;2.

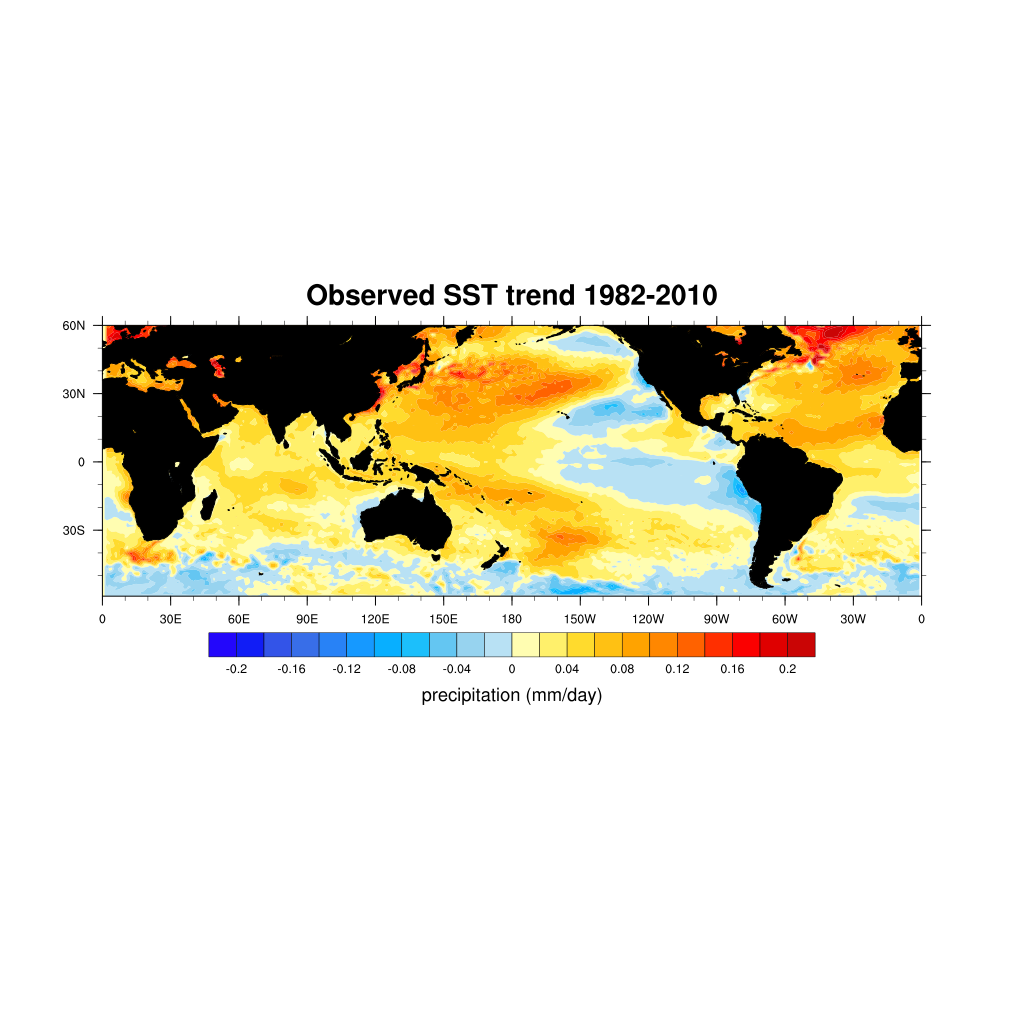
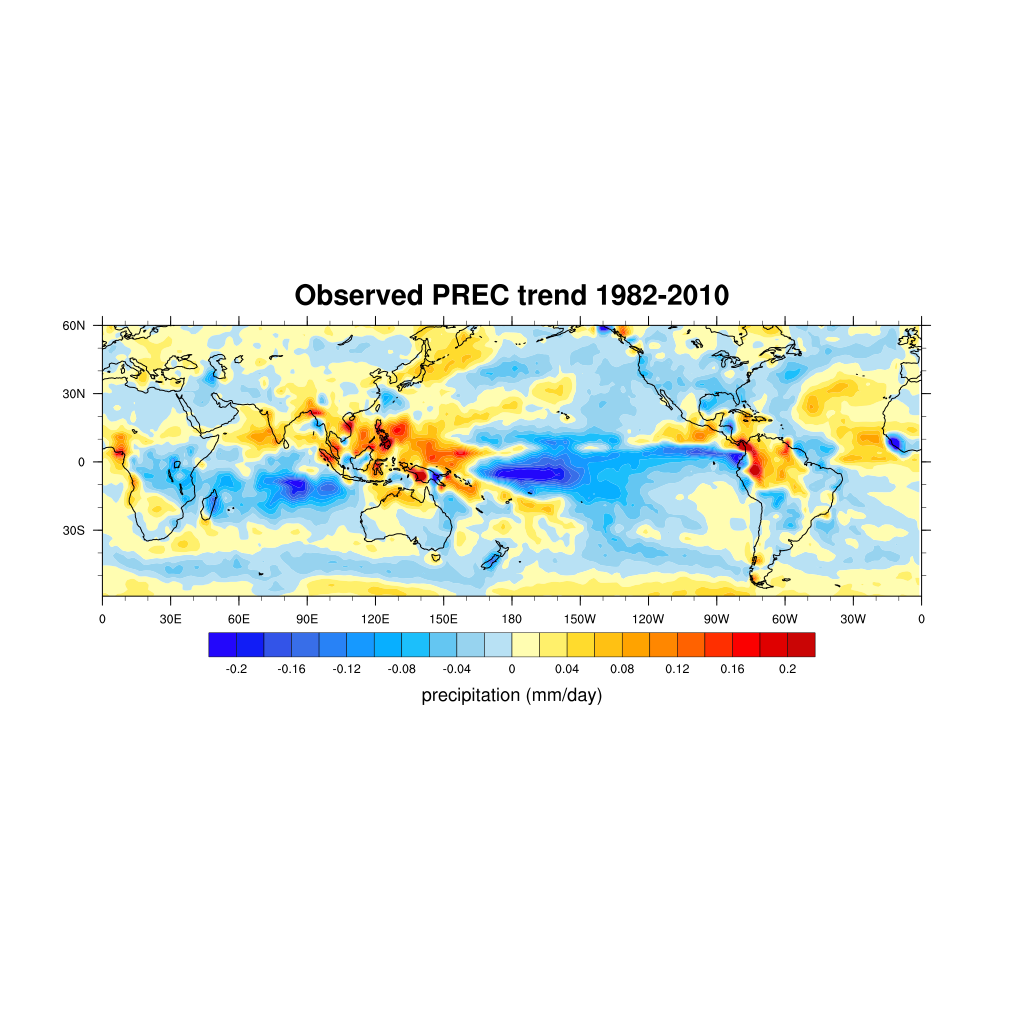
Vecchi, G.A, and Coauthors, 2014: On the Seasonal Forecasting of Regional Tropical Cyclone Activity. *J. Climate*, **27**, 7994–8016. doi: <http://dx.doi.org/10.1175/JCLI-D-14-00158.1>

Yokohata T, Annan J, Collins M, Jackson C, Tobis M, Webb M and Hargreaves J 2012 Reliability of mutli-model and structurally different single-model ensembles *Clim. Dyn.* **39** 599-616

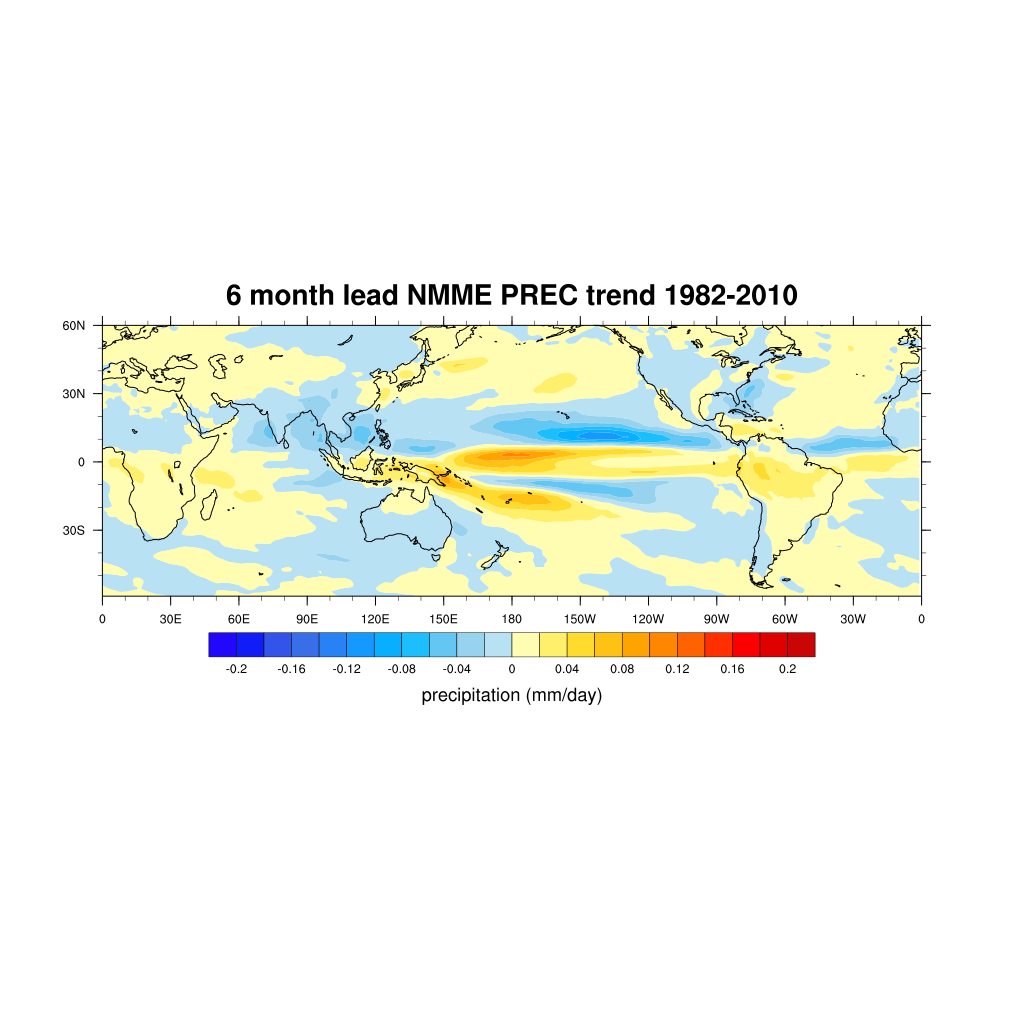
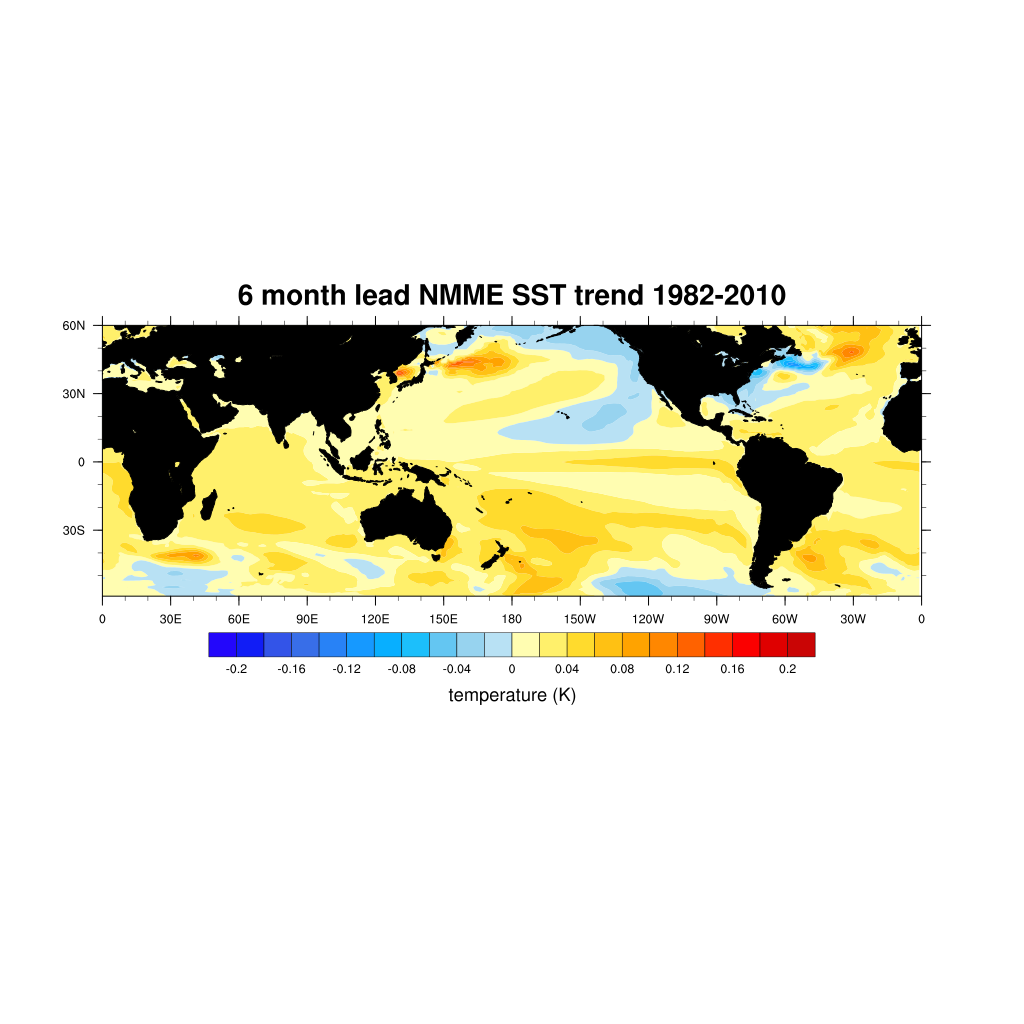
Zhang, S., M. J. Harrison, A. Rosati, and A. Wittenberg, 2007: System design and evaluation of coupled ensemble data as- similation for global oceanic climate studies. Mon. Wea. Rev., 135, 3541–3564, doi:10.1175/MWR3466.1

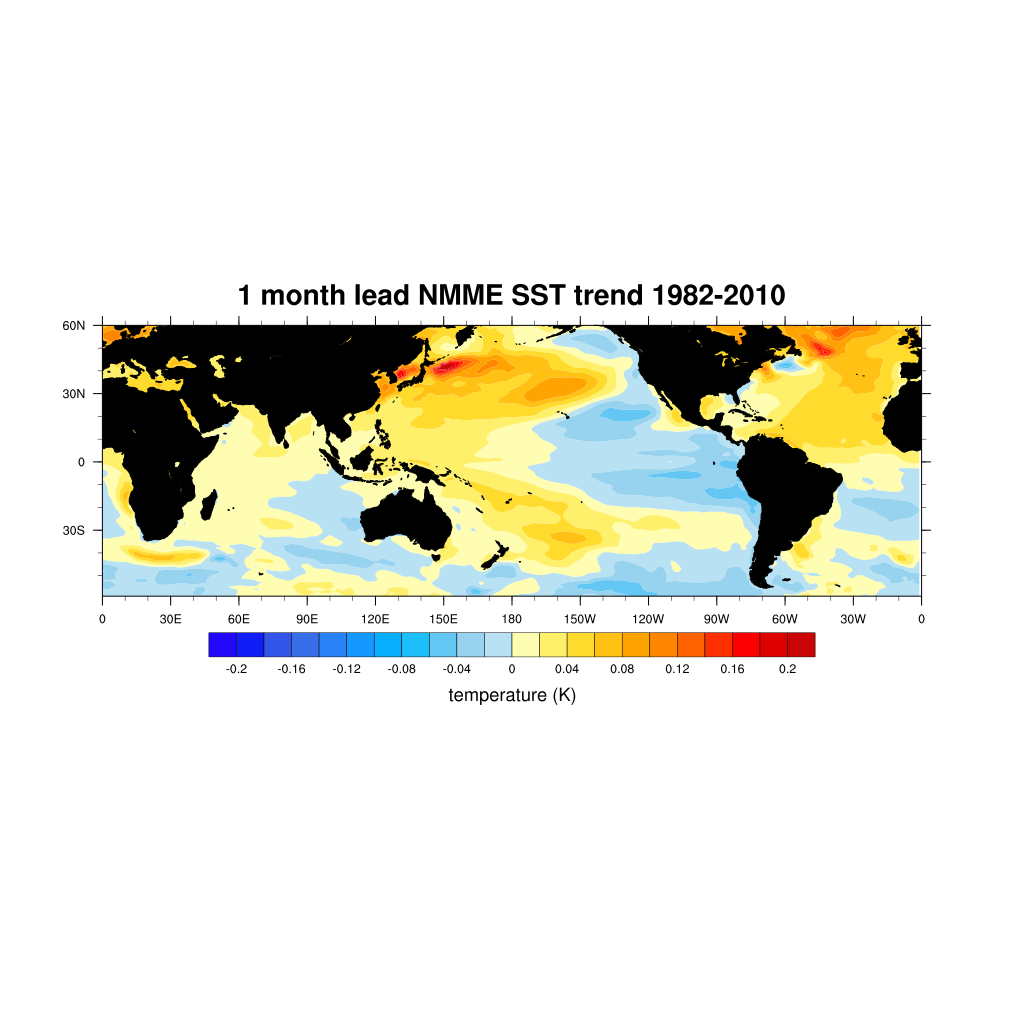
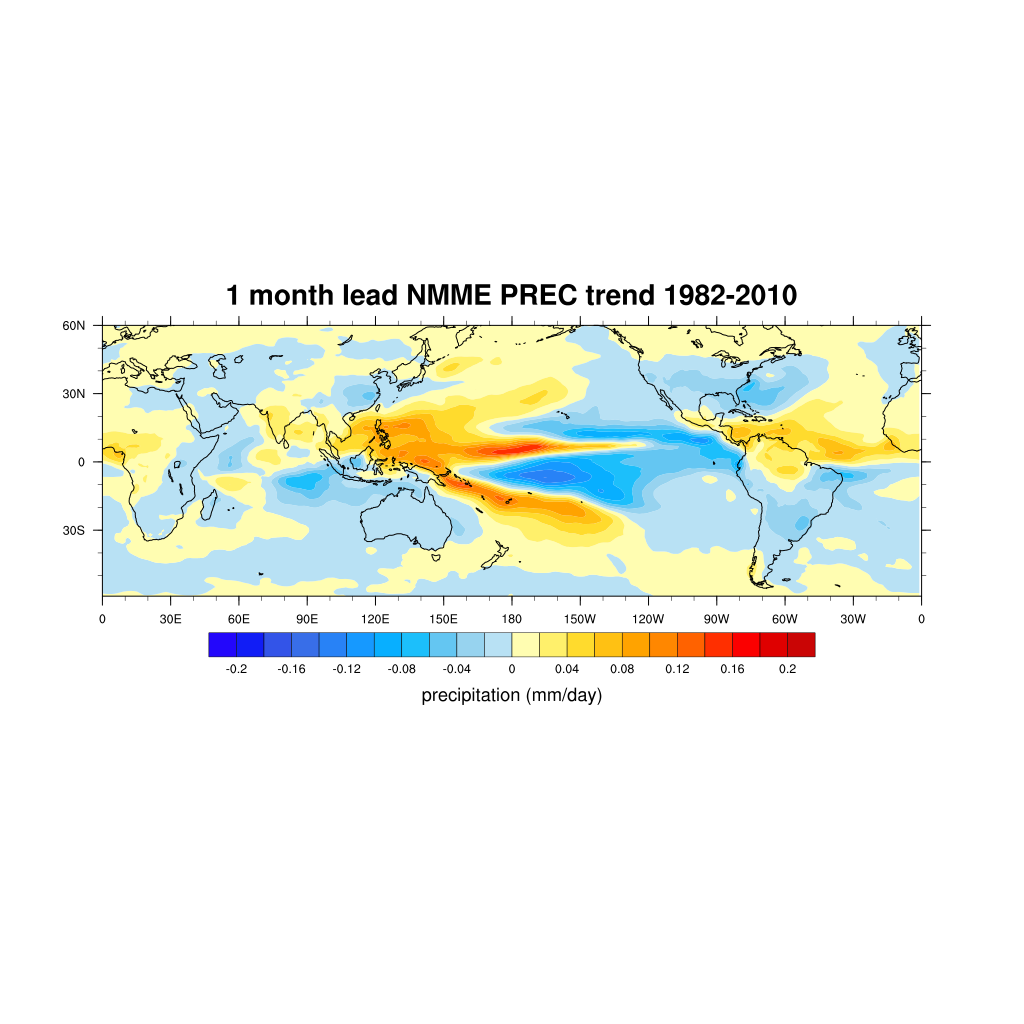
|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **Model** | **Organization** | **Ensemble Size** | **Lead Times** | **Reference** |
| CFSv1 | NCEP | 15 | 0-8 months | Saha et al. 2006 |
| CFSv2 | NCEP | 24 | 0-9 months | Saha et al. 2010 |
| GFDL CM2.1 | GFDL | 10 | 0-11 months | Delworth 2006 |
| GFDL CM2.5 (FLOR) | GFDL | 24 | 0-11 months | Vecchi et al. 2014 |
| ECHAM4-a | IRI | 12 | 0-7 months | DeWitt 2005 |
| ECHAM4-f | IRI | 12 | 0-7 months | DeWitt 2005 |
| CanCM3 | CMC | 10 | 0-11 months | Merryfield et al. 2013 |
| CanCM4 | CMC | 10 | 0-11 months | Merryfield et al. 2013 |
| CCSM3 | NCAR | 6 | 0-11 months | Kirtman and Min 2009 |
| CCSM4 | NCAR | 10 | 0-11 months | Kirtman et al. 2014 |

Table 1: Models used from the North American Multi-Model Ensemble project – Phase 1

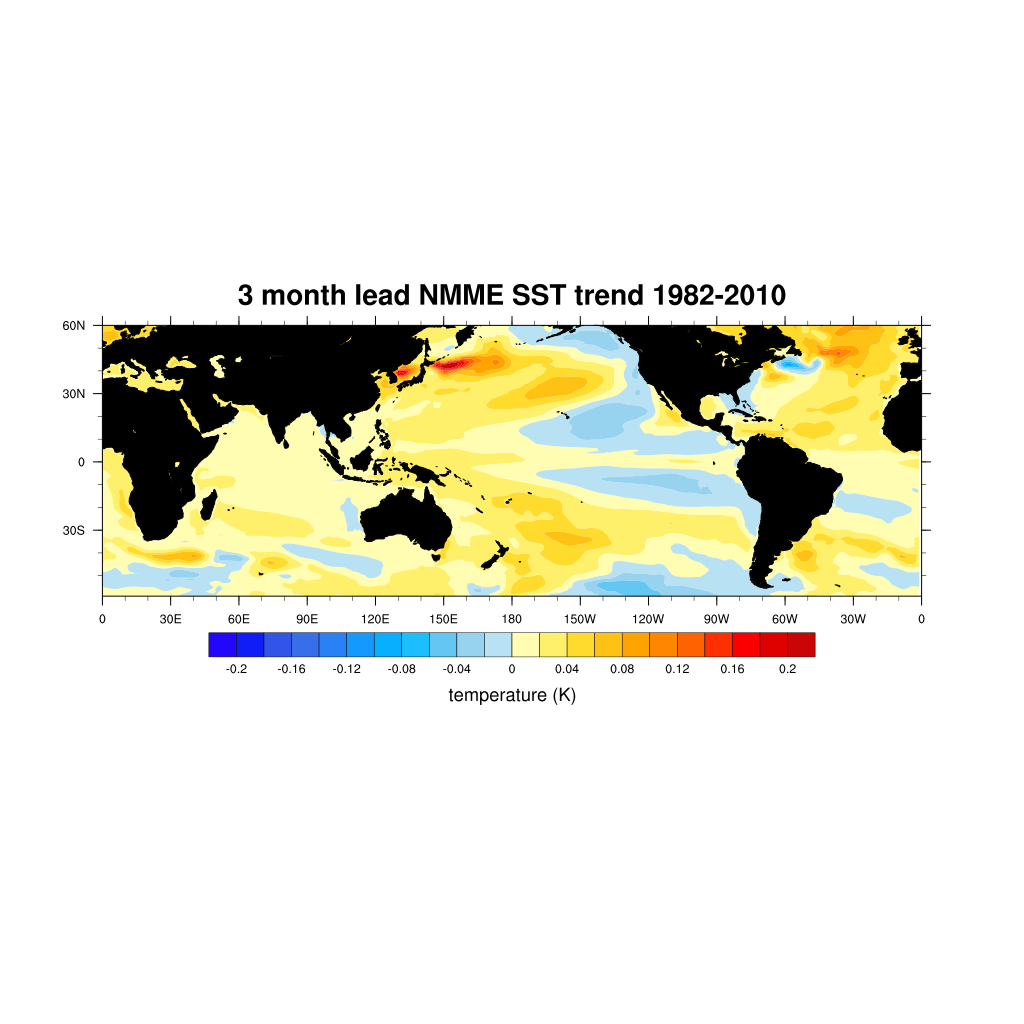
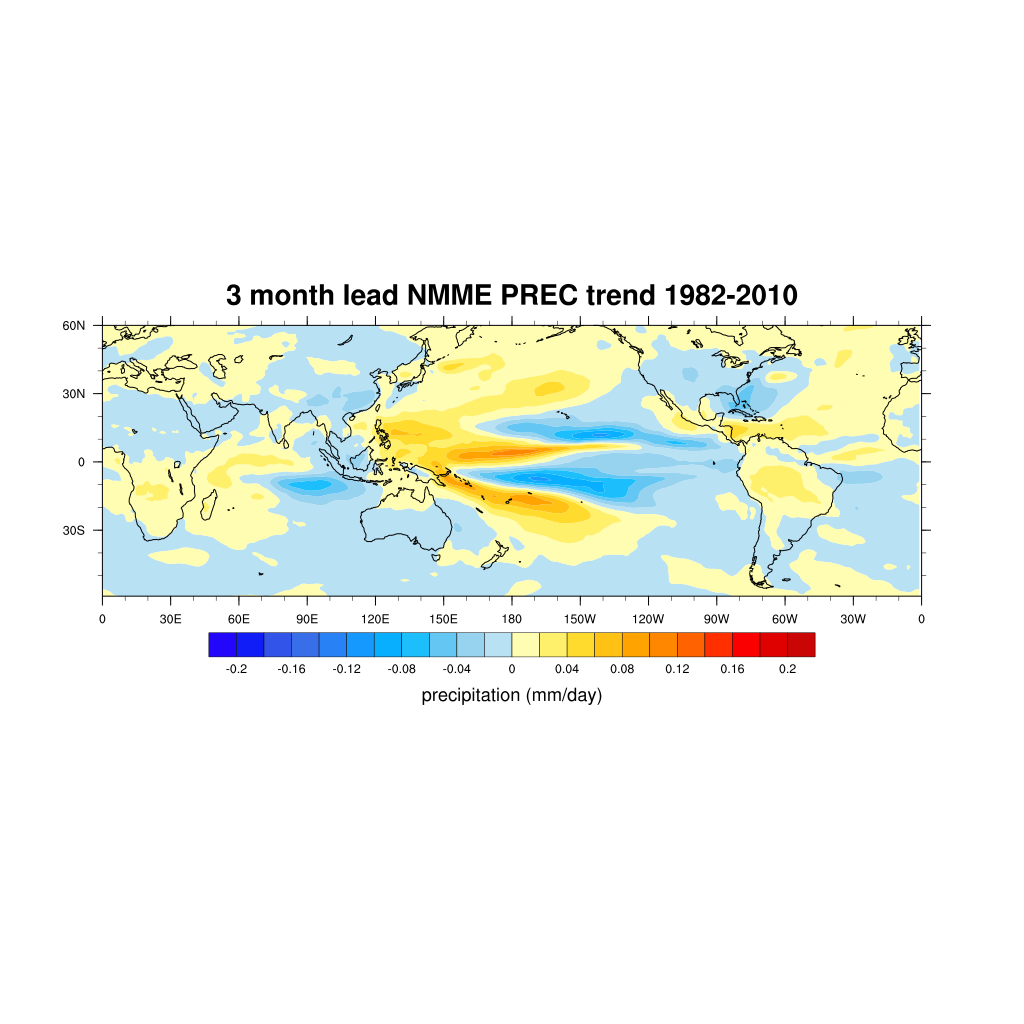
****

a)

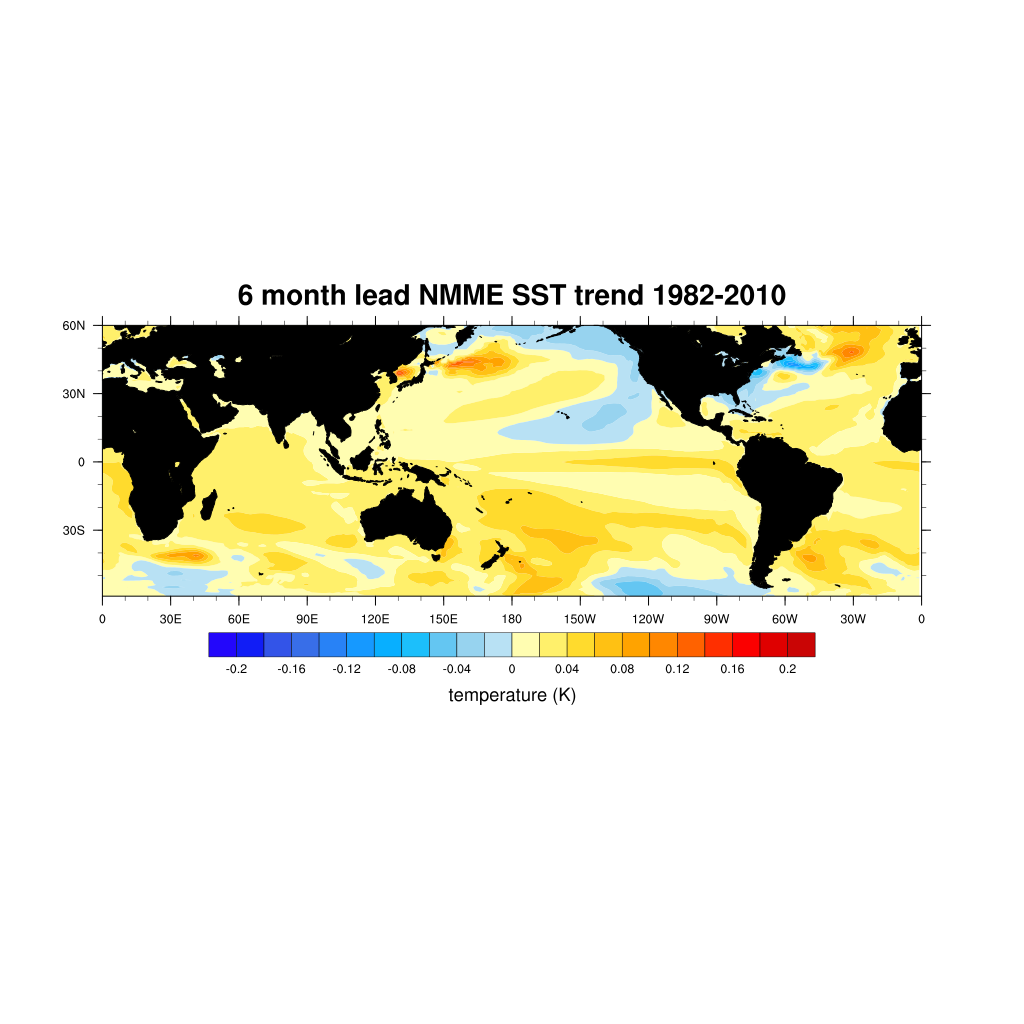
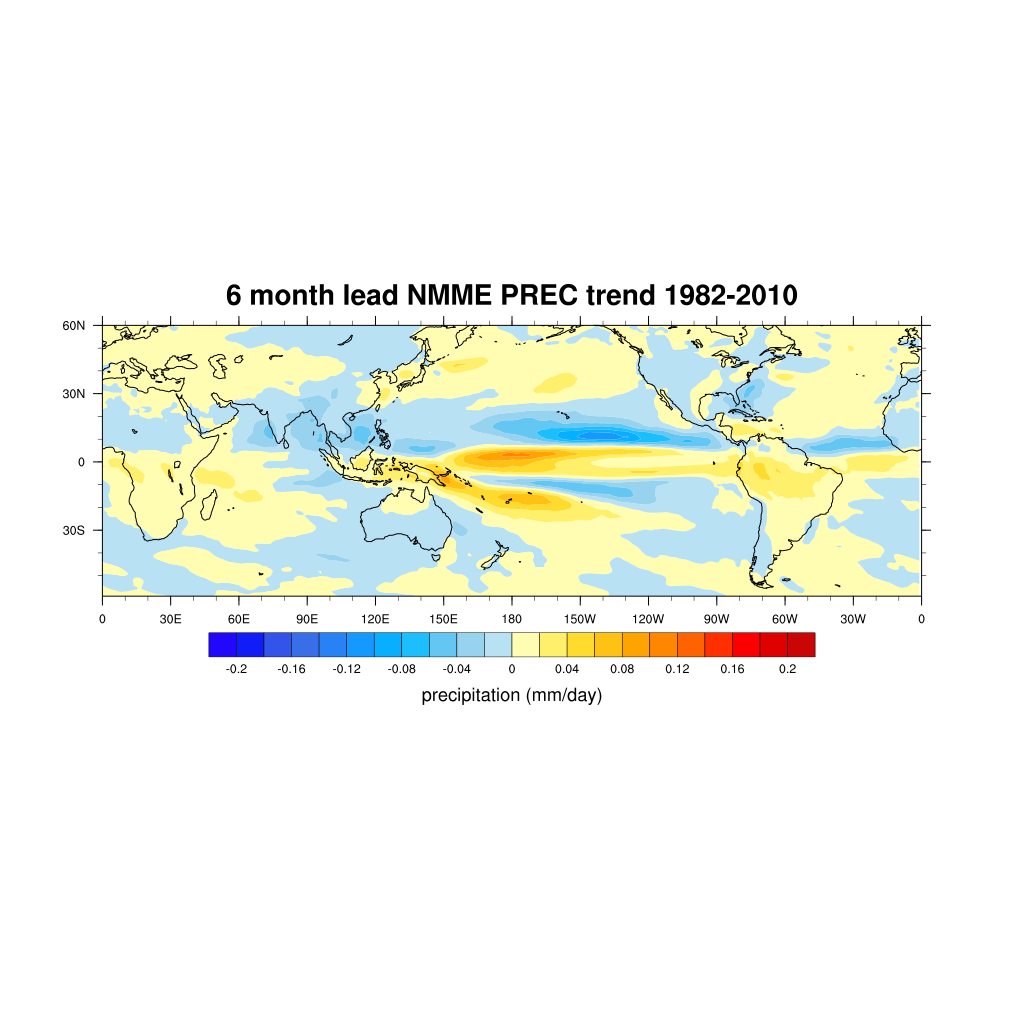
****

****

b)

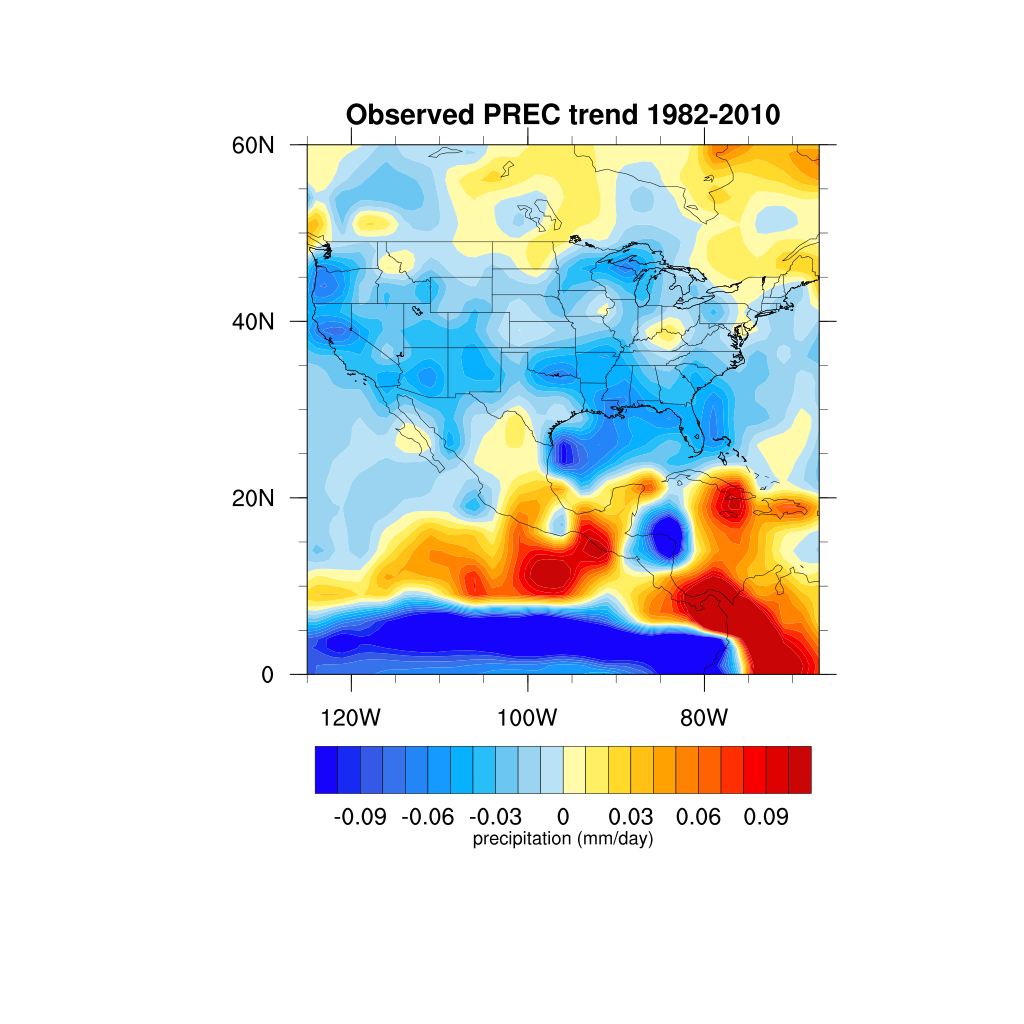
****

c)

****

d)

Figure 1: Global linear trend in SST and PREC for a) observations, b) NMME 1-month lead time ensemble mean, c) NMME 3-month lead time ensemble mean, d) NMME 6-month lead time ensemble mean.

****

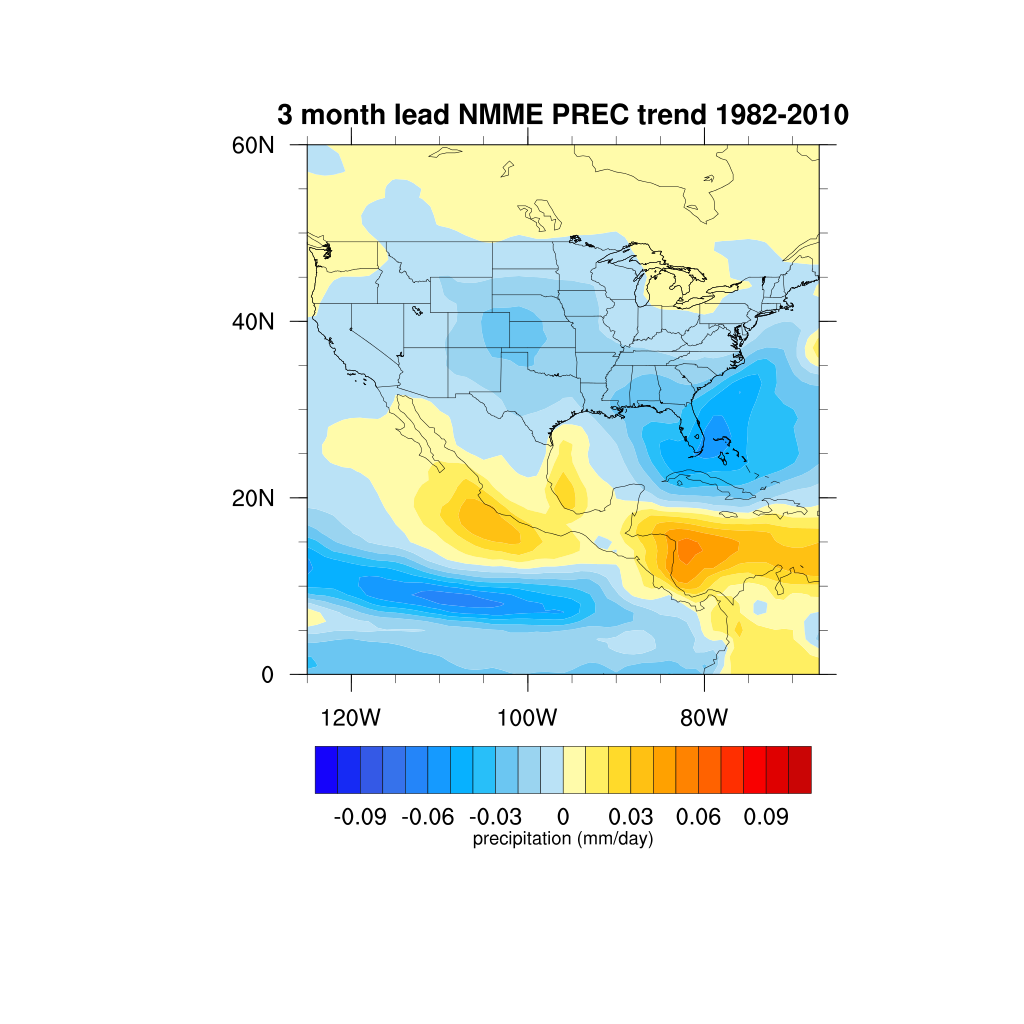
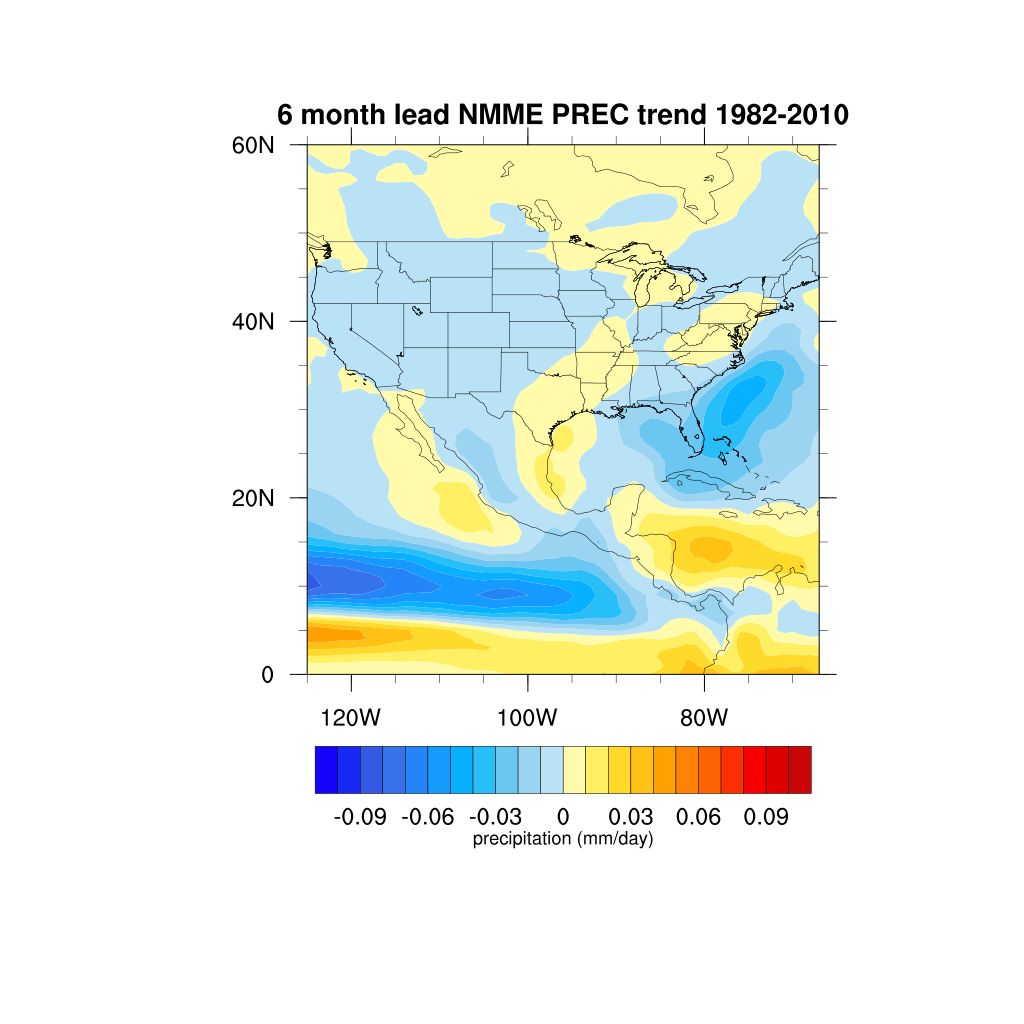
****

Figure 2: Linear precipitation trend over North America from 1982-2010 for observations and NMME 1-month, 3-month, and 6-month lead-time ensemble means.

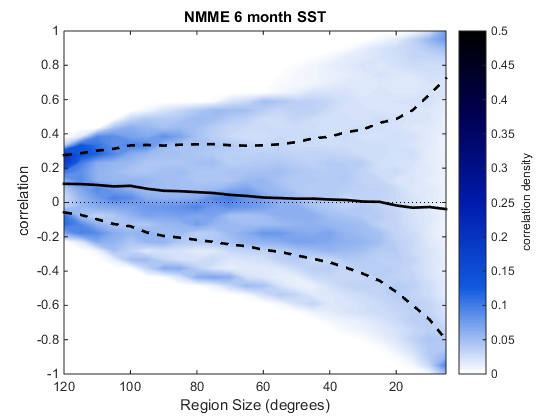
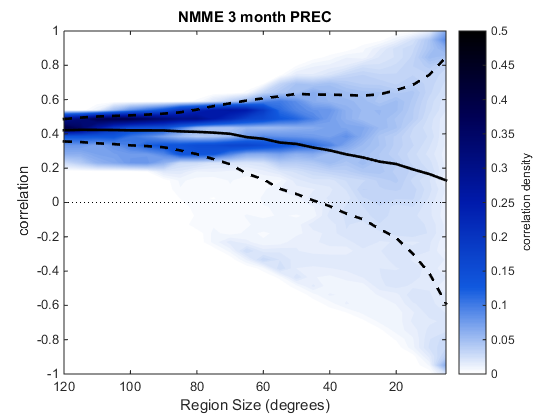
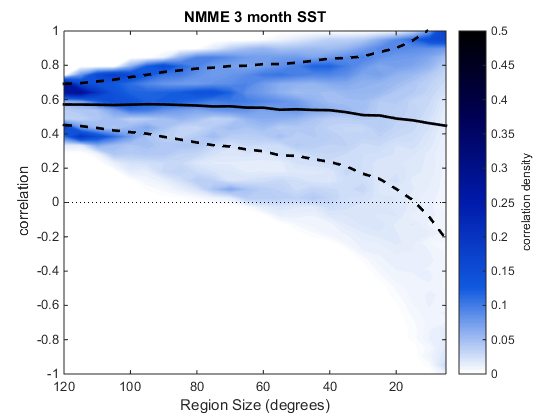
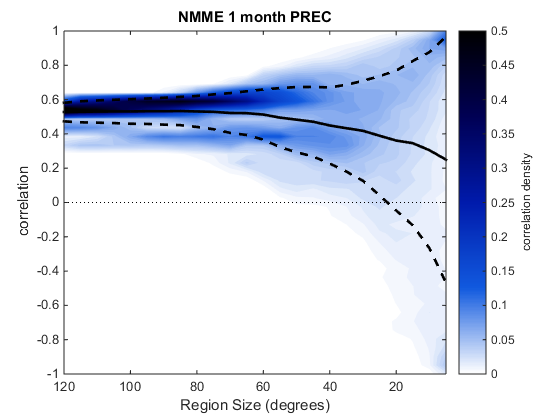
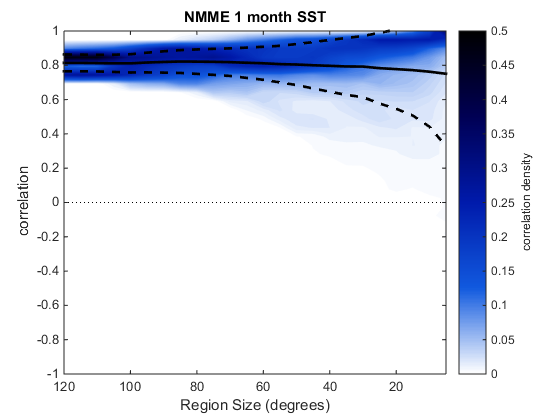
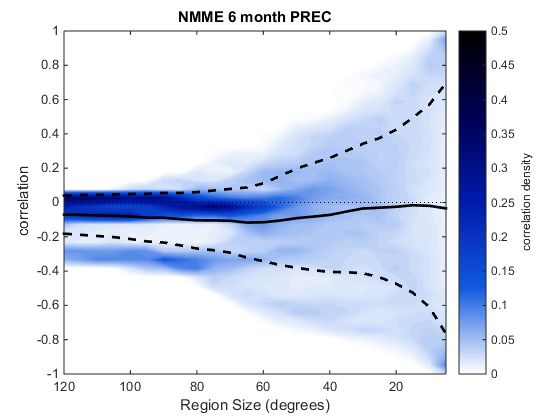


Figure 3: Contours of correlation values for various regional dimensions generated by the Monte Carlo Method (MCM) for the NMME mean for 1-month, 3-month, and 6-month lead times. The solid black line represents the mean correlation value for a given dimension and the dotted black lines are one standard deviation from the mean.

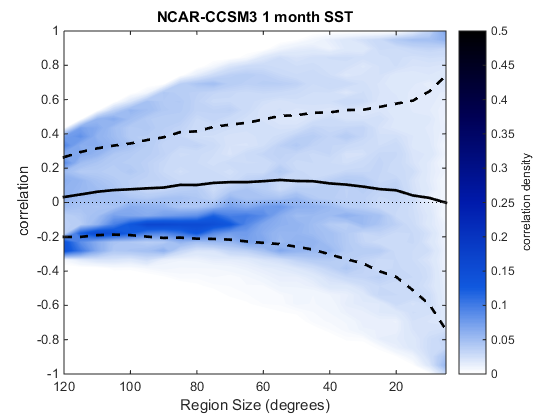
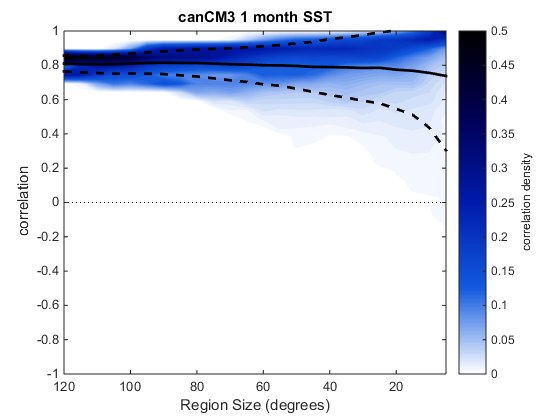
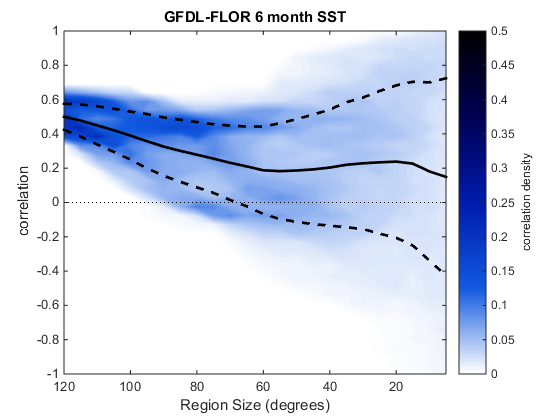


Figure 4: MCM correlation values for the best (canCM3) and worst (CCSM3) performing models for SST at a 1-month lead time.

****

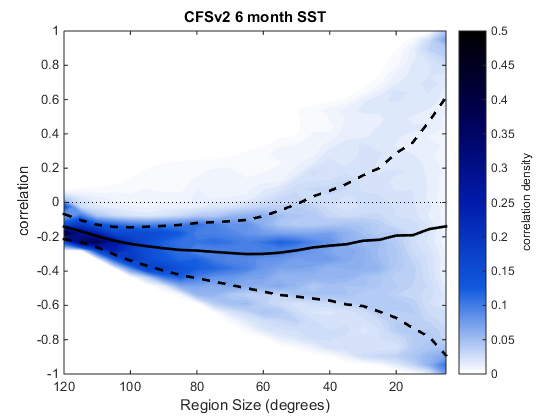


Figure 5: MCM correlation values for the best (GFDL-FLOR) and worst (CFSv2) performing models for SST at a 6-month lead-time.

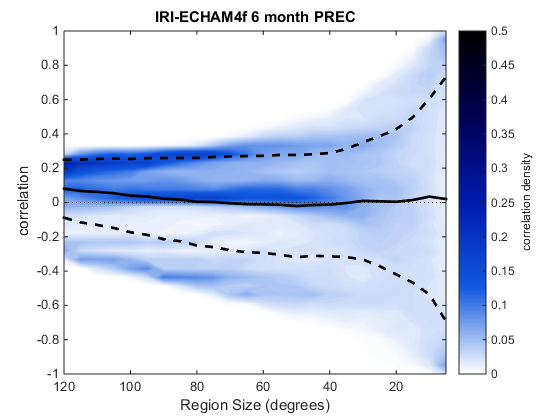
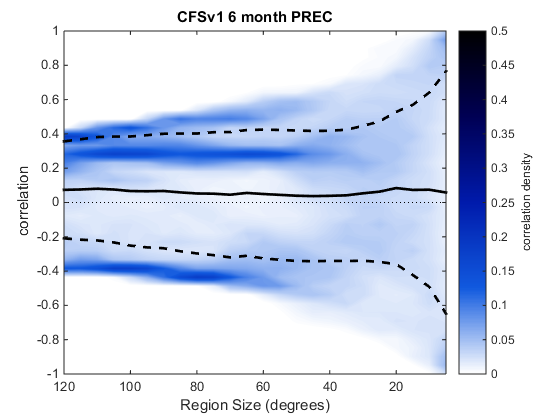


Figure 6: Example of the banding pattern (IRI-ECHAM4f) found in the correlation plots of precipitation at all lead times, and example of negative precipitation correlation pattern (NCEP1) generated by some models with a 6-month lead time.

A B

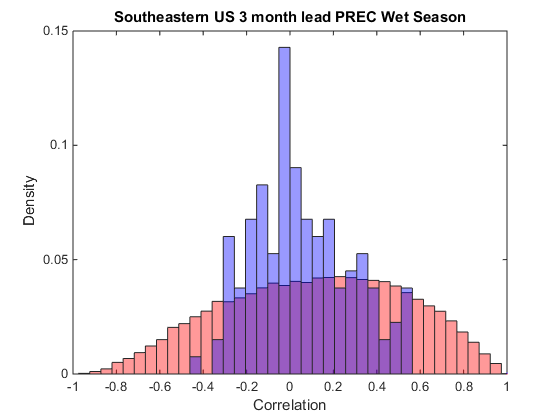
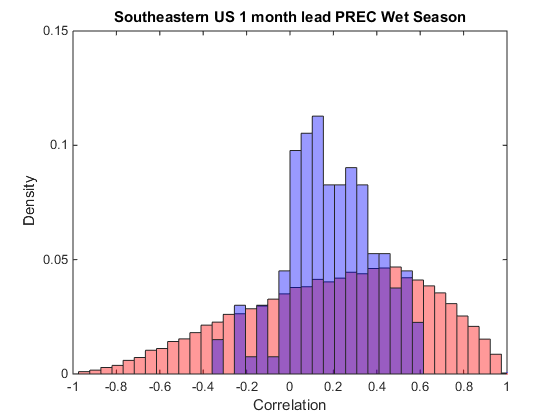
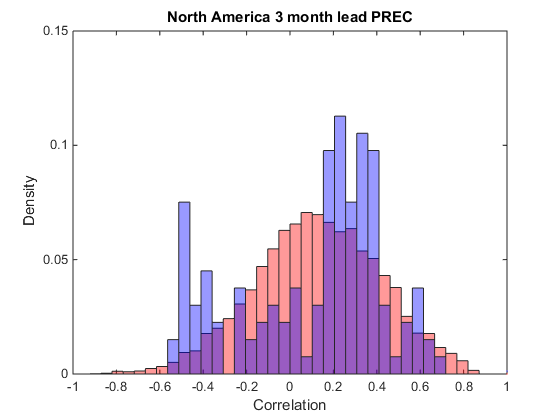
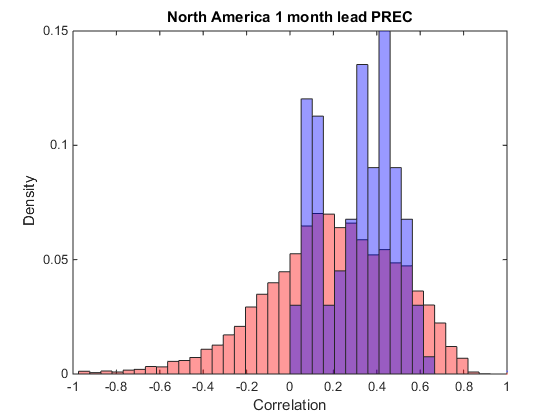


Figure 7: A) Normalized histogram of the correlation between the North America region in all 133 individual NMME members and observations in blue. Normalized histogram of correlation between 10,000 randomly selected North America sized (60°x60°) regions and observations in red. B) as A) for the South Eastern US and South Easten US sized region (20°x20°).

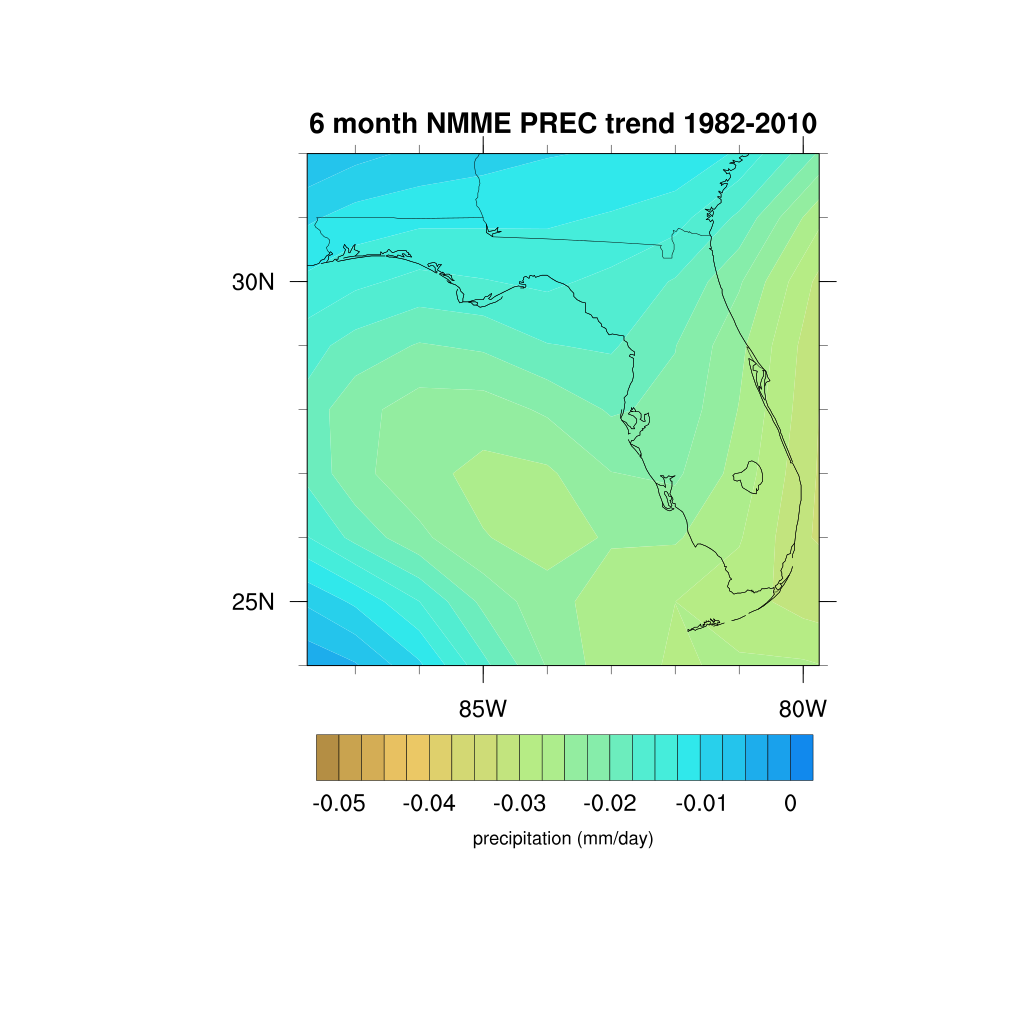
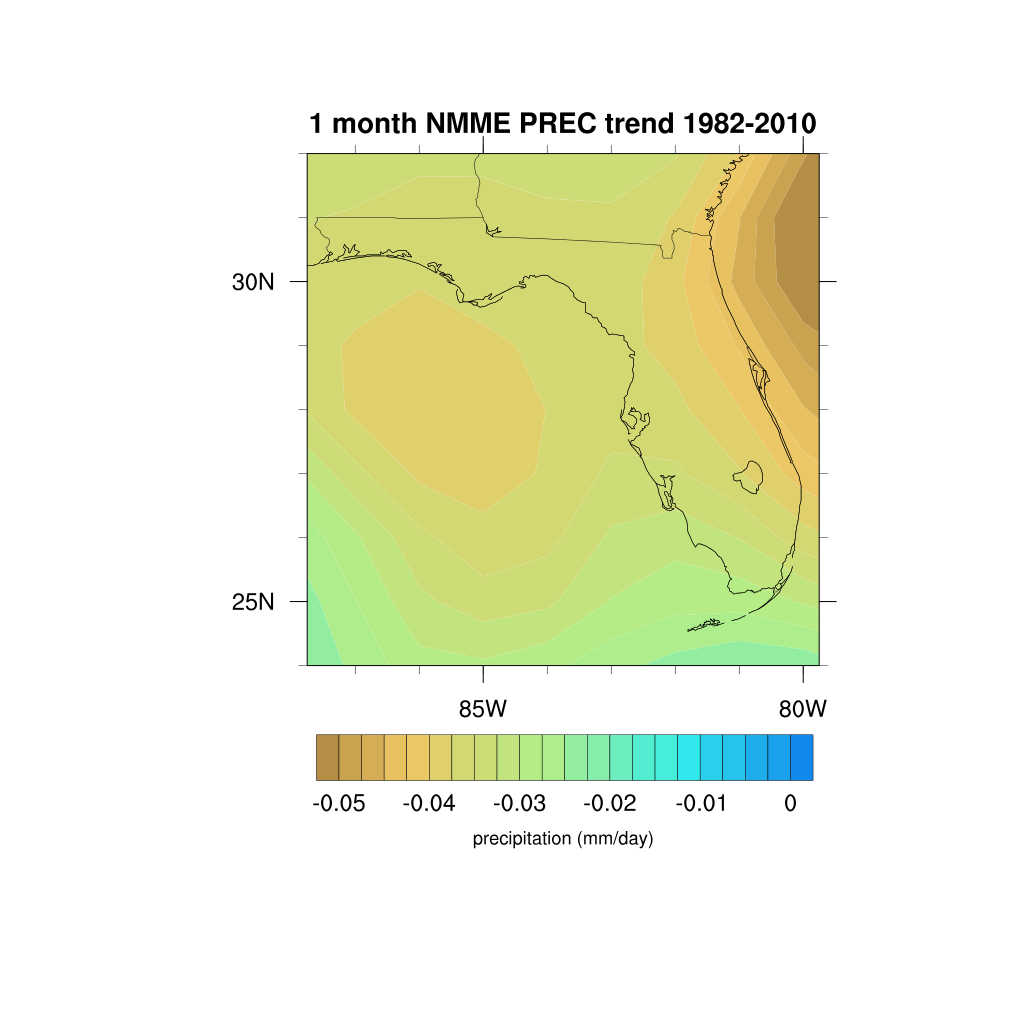
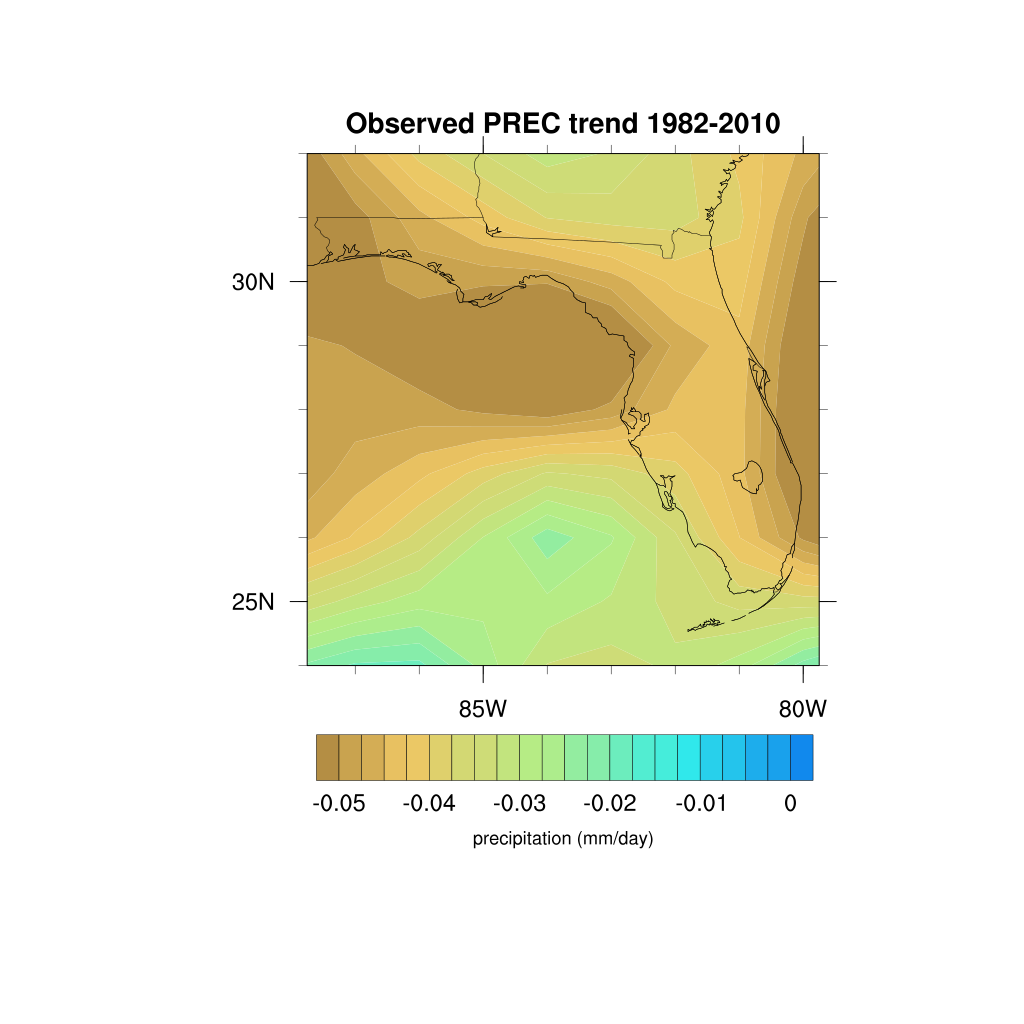


Figure 8: Linear precipitation trend over Florida from 1982-2010 for observations and NMME 1-month, 3-month, and 6-month lead time ensemble means.

A B

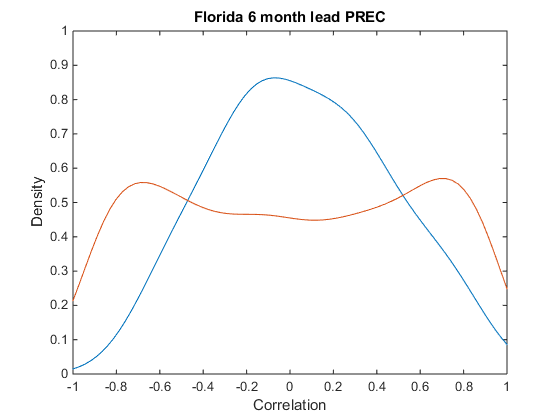
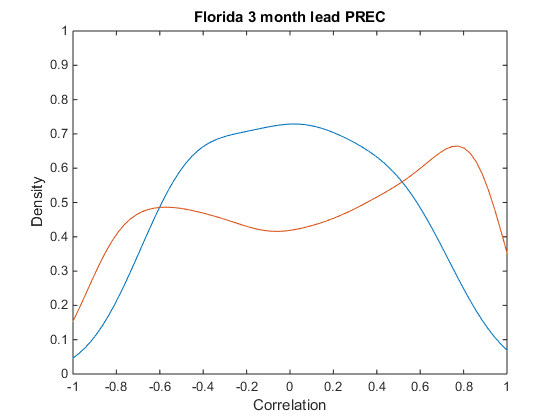
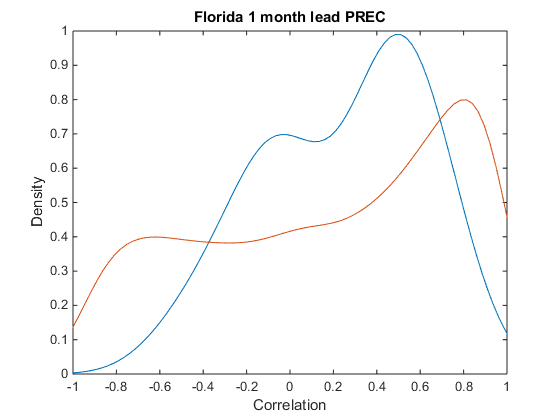
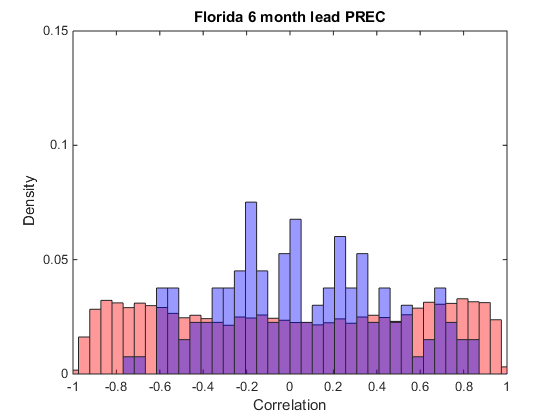
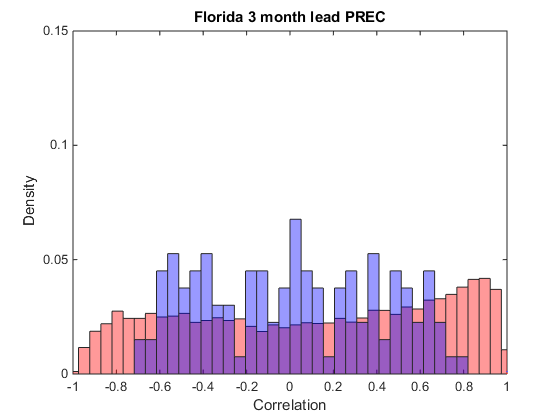
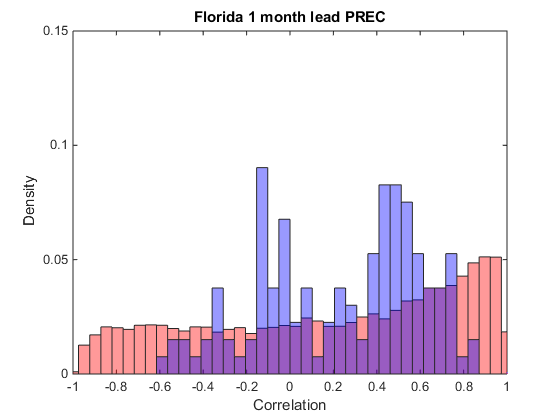


Figure 9: Column A) Normalized histogram of the correlation between the Florida region in all 133 individual NMME members and observations in blue. Normalized histogram of correlation between 10,000 randomly selected Florida sized (8°x8°) regions and observations in red. Column B) PDF of the correlation between the Florida region in all 133 individual NMME members and observations in blue. PDF of correlation between 10,000 randomly selected Florida sized (8°x8°) regions and observations in red.

A B

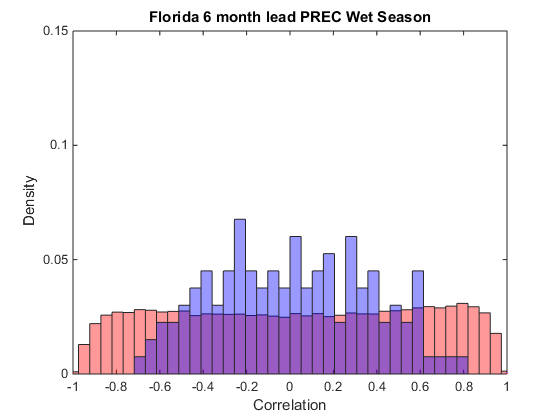
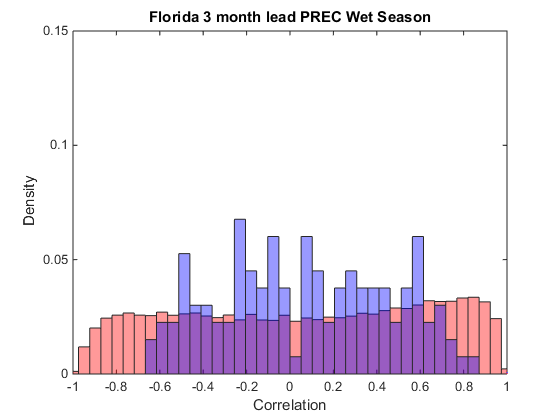
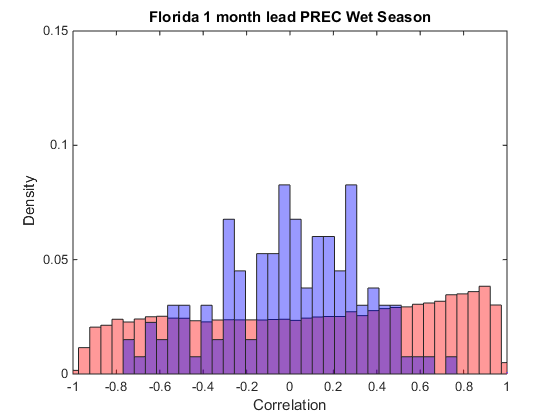
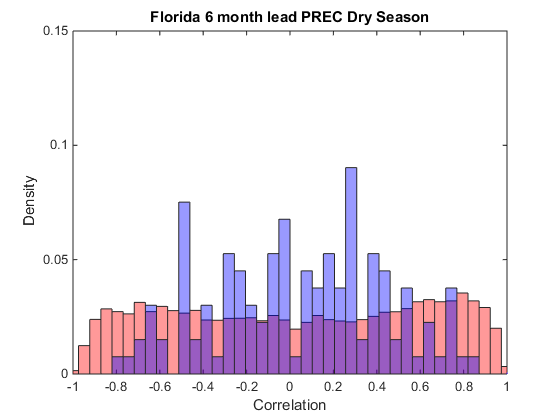
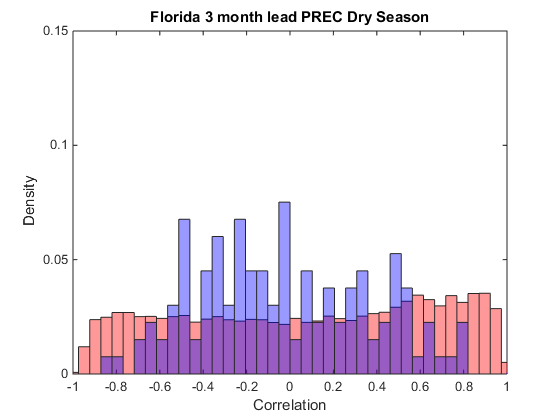
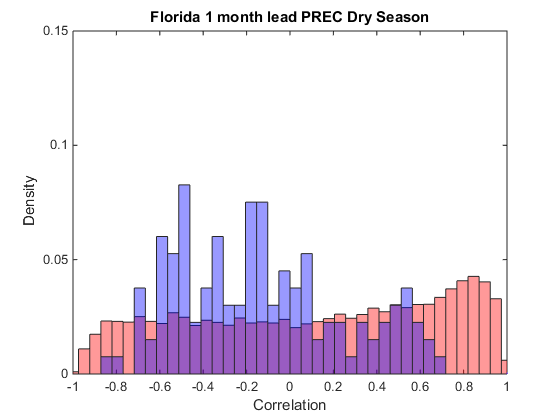
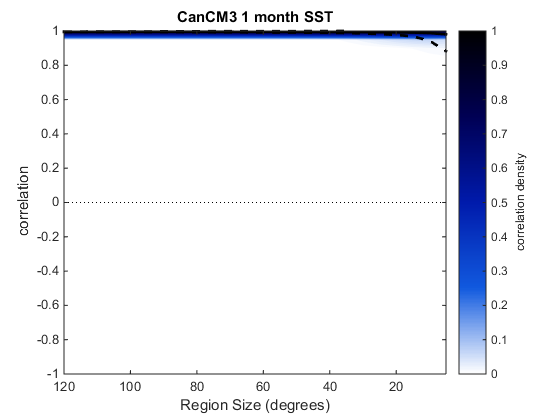
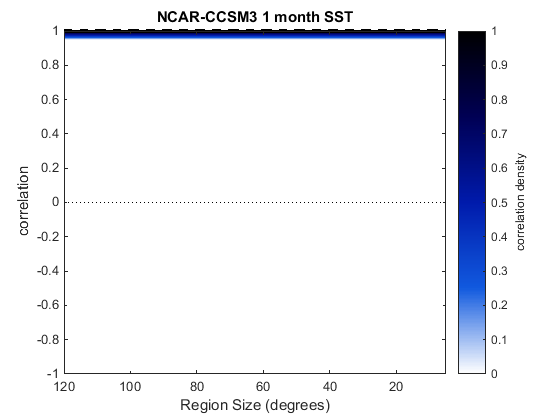
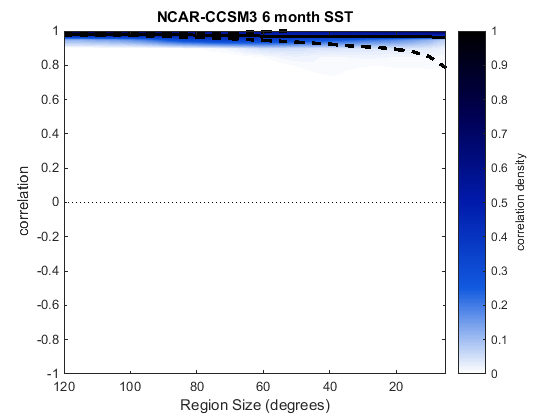
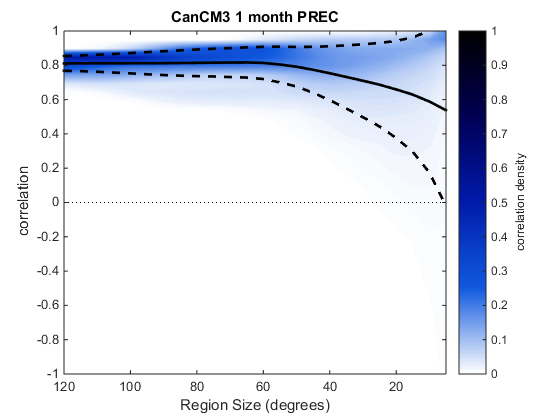
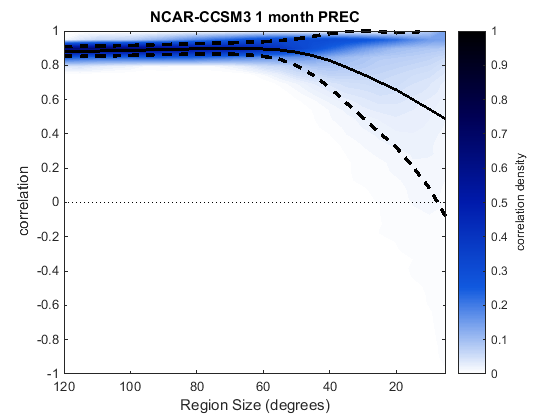
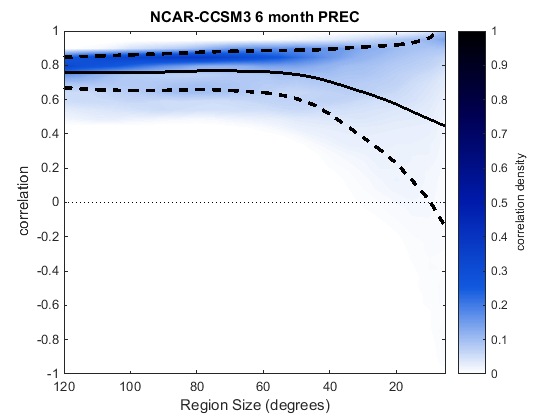


Figure 10: A) Normalized histogram of the correlation between the Florida region in all 133 individual NMME members and observations in blue for the dry season (DJF). Normalized histogram of correlation between 10,000 randomly selected Florida sized (8°x8°) regions and observations in red for the dry season. B) as in A) for the wet season (JJAS).



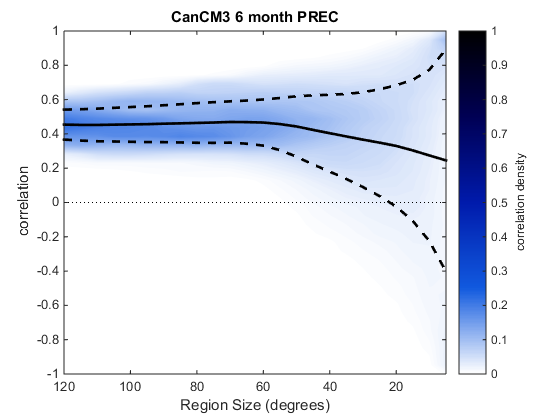


Figure 11: Example of homogeneous MCM for canCM3 1-month and 6-month lead times and for NCAR-CCSM3 1-month and 6-month lead times.