Sub-seasonal to seasonal predictability of weather statistics using NMME
Final Report

1. General Information

Project Title: Sub-seasonal to seasonal predictability of weather statistics using NMME
PI/co-PI names and institutions: Simon Mason, International Research Institute for Climate and Society, The Earth Institute of Columbia University
Report Year: Progress Report, May 1, 2016 – April 30, 2017
Grant #: NA15OAR4310076

2. Main goals of the project, as outlined in the funded proposal

The overarching goal of this project is to provide guidance on the creation of high-impact weather event forecast products by determining how sub-seasonal to seasonal predictive capability, using currently available operational models, of such events varies as a function of lead-time, averaging time and spatial area. One of the components of the project, more specifically, aims to use a hidden Markov model (HMM) toolkit based on generalized linear models and non-homogeneous HMMs, in order to (1) assess the seasonal predictability of varying degrees of spatio-temporal aggregation of daily rainfall statistics, and (2) to inform the development of spatio-temporal downscaling methods for NMME output.

- *Objective 1*: Quantify how spatial and temporal averaging affects predictive skill at increasing lead-times
- *Objective 2*: Diagnose the ability of the NMME to predict realistic sequences of weather, and thereby identify possible systematic errors that may limit the models' ability to predict high-impact weather events

3. Results and accomplishments

Introduction

When predicting extreme rainfall event frequencies at seasonal timescales, events are counted not over the season, but over space also. Just as the temporal aggregation is based on the assumption that it is not possible to predict the exact timing of weather events within a season, so also it is assumed that the exact tracks of storms are unpredictable. Similarly, just as the season needs to be long enough to minimize the noisiness of individual weather events, but short enough to represent a period with a consistent climate forcing, so also the area needs to be of an optimal size. The area should be large enough to minimize the noisiness of weather

sampling, and small enough to represent an area with a consistent climate forcing. However, the regions cannot be defined purely in terms of size if the latter criterion is to be upheld: the areas should not be arbitrarily defined by an optimal grid size, but should rather be defined as homogeneous areas in which extreme rainfall event frequencies co-vary in time. Therefore, a key preliminary step before developing seasonal forecasts of extreme rainfall frequencies is to identify appropriate homogeneous regions. An appropriate technique for identifying such regions is cluster analysis, but the details of how to cluster for extreme rainfall frequencies are unclear.

Cluster analysis effectively partitions sample data into groups such that objects in the same cluster are very similar, by minimizing the distance from each data point to its assigned cluster. It can be used to rigorously quantify spatial distinction of subregions, for variables such as rainfall and temperature, representing potential improvement over assuming a predefined box or domain over which characteristics are expected but not demonstrated to be similar. Here, as part of an effort to improve the seasonal prediction of extreme rainfall in the US, we analyzed daily rainfall using cluster analysis over the winter December-March (DJFM) season.

Data and methods

The gauge-based National Oceanic and Atmospheric Administration (NOAA) Climate Prediction Center (CPC) Unified Precipitation Analyses, CONUS v1p0 was used (Chen et al. 2008). The product has daily temporal resolution from 1948 to the present ("retro" and "real-time" periods were concatenated), and 0.25-degree spatial resolution over the entire CONUS region (129.875W to 55.125W, 20.125N to 49.875N)[1]. For this analysis, the data were truncated to the end of 2016, representing 69 years in total. The DJFM season is the focus of this investigation, representing a total number of days of 8349 (121 days per season for 69 years, excluding leap year days).

1. *Cluster analysis*

Two primary approaches to clustering were evaluated initially, k-means and hierarchical. Due to the sample size of the data the former results were found to be much more coherent. Therefore the experiments and results discussed here utilized the k-means approach, employing squared Euclidean distances, for which each centroid is the mean of the points in that cluster. The k-means++ algorithm uses a heuristic to find centroid seeds for k-means clustering. According to Arthur and Vassilvitskii (2007), k-means++ improves the running time of Lloyd's algorithm [2], and the quality of the final solution. In all experiments performed here the number of clusters was chosen a priori, beginning with three and progressing to ten in increments of one.

[1] http://iridl.ldeo.columbia.edu/SOURCES/.NOAA/.NCEP/.CPC/.UNIFIED_PRCP/.GAUGE_BASED/.CONUS/.v1p0/.RETRO/rain/

This analysis was specifically concerned with clustering of extreme daily events of rainfall, above a respective threshold. Therefore, the raw data (pooled daily rainfall over all years at each point in space) were first converted into respective quantiles, effectively 'standardizing' them in space, and allowing them to then be ranked in descending order of daily rainfall extremity at each gridpoint. Four experiments were then performed using different quantile indices, and in each case all values below the respective threshold were replaced with a zero.

1. A binary approach was first employed, in which all values above the .9 quantile threshold (the most extreme ten percent of all days, 835 of 8349) were replaced with a 1.
2. In the next experiment all values above the same .9 threshold were replaced with their respective quantile closest to an increment of .01 quantile. For example, the most extreme one percent of daily rainfall events at any given point would now be represented by a .99, and the next most extreme one percent would be represented by .98, etc. This approach retains some information about the relative severities of the extreme of each event, unlike the binary approach, which only retains information about whether an event is extreme or not.
3. Because the .9 threshold for daily rainfall extremes represented essentially a one in every ten days type of event, the next experiment employed a more restrictive threshold for extreme values. In this case, all values above the .99 threshold were replaced with their respective quantile closest to an increment of .001 quantile. For example, the most extreme one tenth of one percent of daily rainfall events at any given point would now be represented by a .999, and the next most extreme one tenth of one percent would be represented by .998, etc.
4. Last, the binary approach was again employed as in the first experiment, after which the numbers of extreme events in each year were then summed at each gridpoint, as well as the number of extreme 'events,' defined as periods of consecutive days exceeding the threshold in each year. This represents a much smaller sample size in time (69 years rather than 8349 days) but is more useful in clarifying the relationship between seasonal forcings (sea-surface temperatures, for example) and the increased likelihood of extreme daily rainfall events during a given season.

2. Generalized linear model with a Bayesian hidden Markov model

Hidden Markov Models (HMMs) have been widely used to simulate daily rainfall occurrences and amounts across multiple weather stations, based on rainfall observations and exogenous meteorological variables. The HMM fits a stochastic model to observed daily rainfall data by introducing a small discrete set of rainfall states, not directly observable, hidden states. The Markov chain allows transition between the hidden states each day. A nonhomogeneous HMM (NHMM) was introduced to condition the state transition probabilities by one or more

exogenous climate variables and the transition probabilities are not homogeneous in time. In order to overcome a limitation of NHMM (the characteristics of rainfall states cannot change as the climate changes), Holsclaw et al., (2016) combined a generalized linear model (GLM) with HMM. The GLM approach was suggested by Kenabatho et al. (2012) and Ambrosino et al.(2014) to allow state dependent model parameters to be modulated directly by exogenous climate variables. In this study, the recently developed Bayesian Non-Homogenous Markov and Mixture Models R-software package (https://CRAN.R-project.org/package=NHMM) introduced in Holsclaw et al. (2017) was used to implement GLM-HMM to test the potential for spatio-temporal downscaling of NMME output. More details on the GLM-HMM can be found in Holsclaw et al., (2016) and Holsclaw et al. (2017).

This downscaling model requires a predictor "input" for each CPC rainfall pixel location, for every day of the DJFM 1981-2015 period, in order to train it. We constructed a large-scale "predictor" variable for predicting the daily characteristics of these sequences via an HMM-GLM (Holsclaw et al. 2016). Before turning to the NMME as a predictor, we first use reanalysis precipitation to test the methodology and assess weather-within-climate S2S predictability, taken from the NASA - Modern-Era Retrospective analysis for Research and Applications, Version 2 (MERRA-2) dataset (obtained via IRI Data Library). By constructing various spatio-temporal averages (or "coarse graining") of this MERRA2 predictor, we are then in a position to assess scale-dependence of seasonal predictability of weather statistics.

Spatial smoothing of the MERRA2 is accomplished using eigenfunctions of the Laplace operator which form an orthogonal basis set that can be ordered by length scale (DelSole et al., 2017). Fig. 1 shows an example of the MERRA2 maximum precipitation field from January 1, 1981 (left), together with the same field smoothed using the leading 10 spatial eigenfunctions, and a 3-pentad moving average.

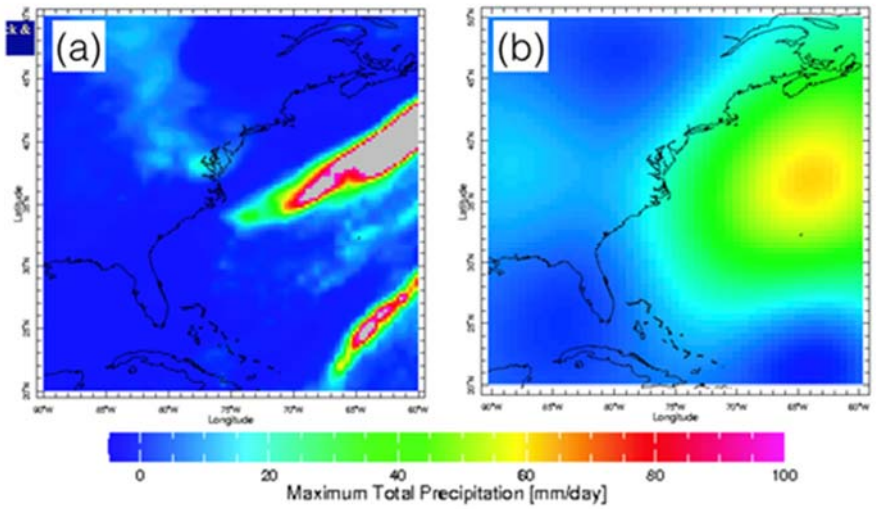


Fig. 1 MERRA2 Maximum Total Precipitation on Jan 1, 1981. (a) Raw MERRA2 estimate, and (b) spatio-temporally smoothed field.

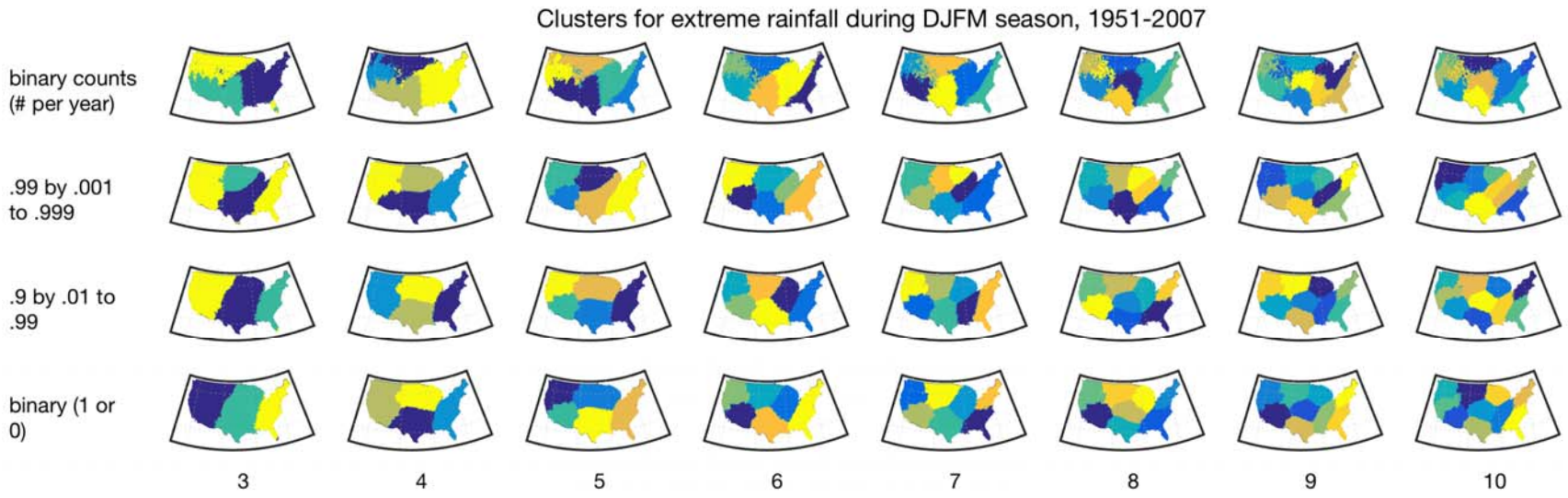


Fig. 2: Clustering of winter (DJFM) daily extreme rainfall (NOAA NCEP CPC Unified Precipitation Analyses, CONUS v1p0).

We divided the rainfall data sets for training (1981 DJFM to 2010 DJFM, 30 years) and separate prediction and validation period (2011 -2014, 4 years) in order to assess the quality of model in terms of reproducing the daily rainfall sequences. The first set of data (1981-2010) was used to train the model and calibrate model parameters and the second set of data (2011-2014) was used to evaluate the model.

Results and discussion

The clustering results are shown in Figure 2, with case one on the bottom and case four on the top row. There are some key similarities and differences between the four experiments outlined above. In case one (bottom row), the clustering using the binary approach reveals a very strong similarity to case two, which uses quantile differentiation above the threshold. This implies that for the 10% most extreme rainfall days very little additional information is provided by using the quantiles above the threshold.

In case three, which uses only the 1% most extreme rainfall days there are some differences in the hierarchy as seen by the similarity between clusters four through six across the first three cases, but the divergence in clustering over the midwestern US for clusters above six. The Pacific and Atlantic coasts are consistent across all three cases.

In case four (top row) the clustering was performed based on counts of the number of events occurring during the same year rather than all days pooled. Here there is some divergence from the clustering in the previous three cases. The western half of the US appears to converge with the other cases as the number of clusters increases to nine, but the Atlantic coast in particular remains distinct from the other cases until the number of clusters reaches ten. The spatial coherence in case four is less in this case due to the reduced sample size in time, more notably in the north and west of the US. Clustering results using extreme 'events,' defined as yearly counts using periods of consecutive days of extreme rainfall, were nearly the same as these results and are therefore not shown here.

In each row, as the number of clusters increases, branching of the hierarchical tree can be seen, as one cluster splits and becomes two, or two become three, etc. In the binary case (bottom row) for instance, the Atlantic coast cluster remains the same for up to six clusters, before splitting into a northeastern cluster and a southeastern cluster for clusters seven through ten. This also is seen in the second and third cases. In case four, as mentioned above, the Atlantic coast has a different cluster shape than in the other cases, but as with the others this cluster remains the same for clusters five through nine before splitting into two clusters. For these reasons, the Atlantic coast, or Eastern US region represents a good candidate for further analysis with respect to seasonal to subseasonal forecast improvement. Based on these results, we examined the predictability of extreme rainfall events in orange region shown in Figure 3.

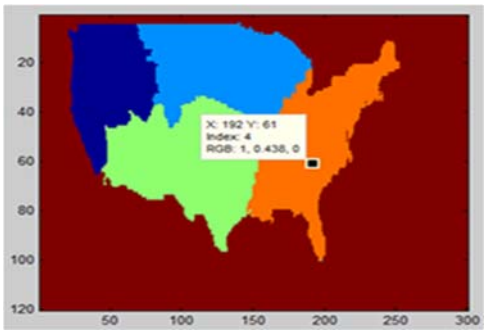


Fig. 3 Location of the Eastern US study domain (orange).

To fit the GLM-HMM we tried different numbers of hidden states (K= 1, 2,…8) and then selected the value of K that minimized the Bayes Information Criteria (BIC) and Akaike's Information Criterion (AIC) which are approximate log-likelihood to assess the model fit. The initial analysis showed that the GLM-HMM has a minimum BIC (and AIC) with K = 8 (Fig. 4). The 8-state HMM was selected as the model with the smallest number of rainfall states that allowed a robust estimation of the model parameters.

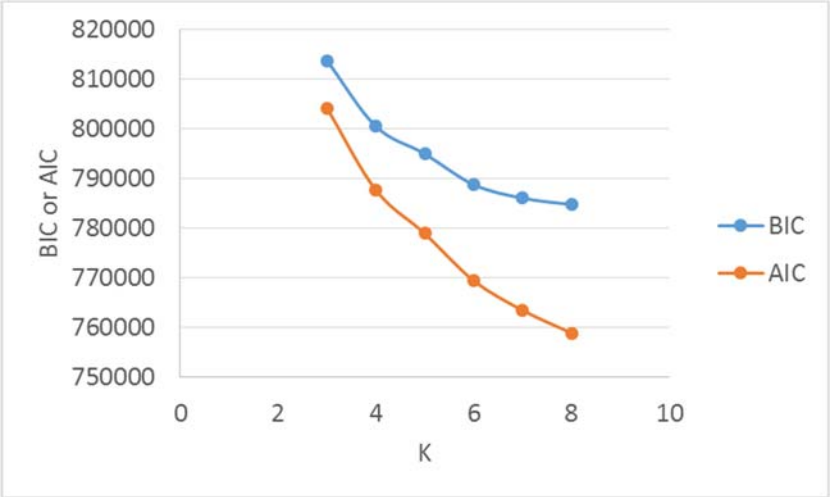
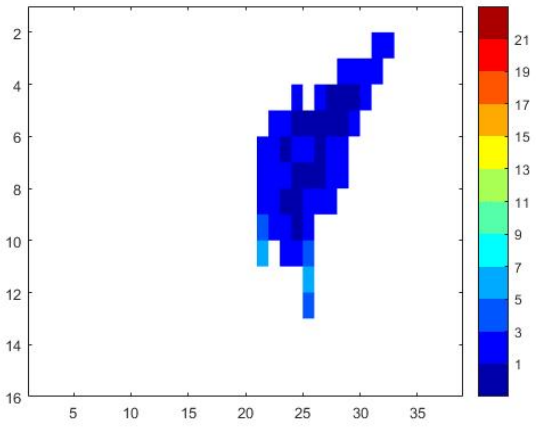
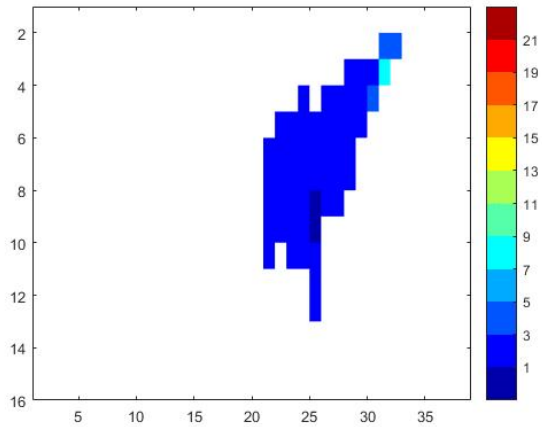
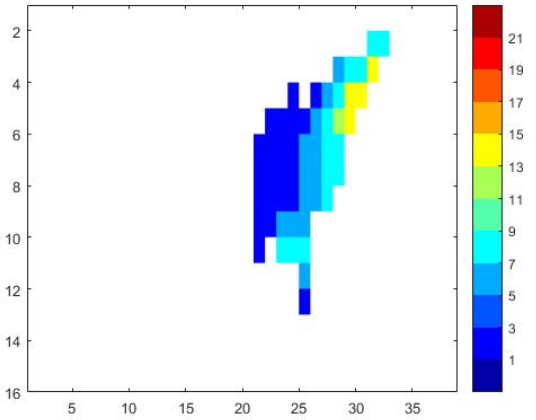
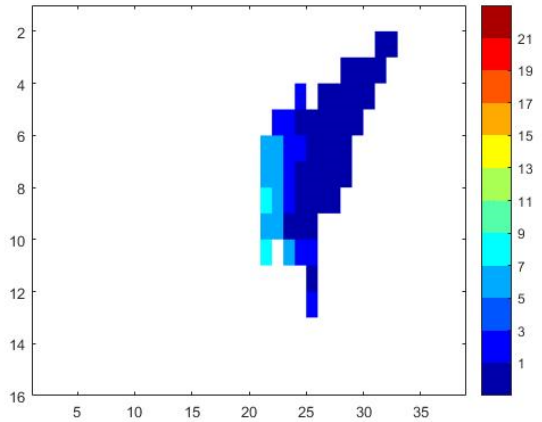


Fig. 4 Number of states in HMM model versus BIC and AIC

Figure 5 illustrates examples of the 8 model states, in terms of the rainfall intensity at each of the 56 pixels, averaged over the days when daily rainfall is greater than 0.1 mm day^{-1}, assigned to the 8 states, during the DJFM 1981-2010 training period. Rainfall intensity is relatively low in State 1,2 and 3. In State 7, high rainfall intensities are concentrated over the Gulf Coast, while in State 8 they are located over the Mississippi Valley. The other states exhibit spatial rainfall maxima in other parts of the domain. Figure 6 shows rainfall occurrence probabilities for the same period. State 1 represents dry days or days with less probable rainfall (rainfall probabilities less than 0.5) while State 6 represents days with high probability of

rainfall, except Florida. There is general correspondence between high rainfall probability and high mean intensity for wetter states (K=6,7,8) and for a dry state (K=1).

<table>
<tr><td>K=1
  </td><td>K=2
  </td></tr>
<tr><td>K=3
  </td><td>K=4
  </td></tr>
<tr><td>K=5</td><td>K=6</td></tr>
</table>

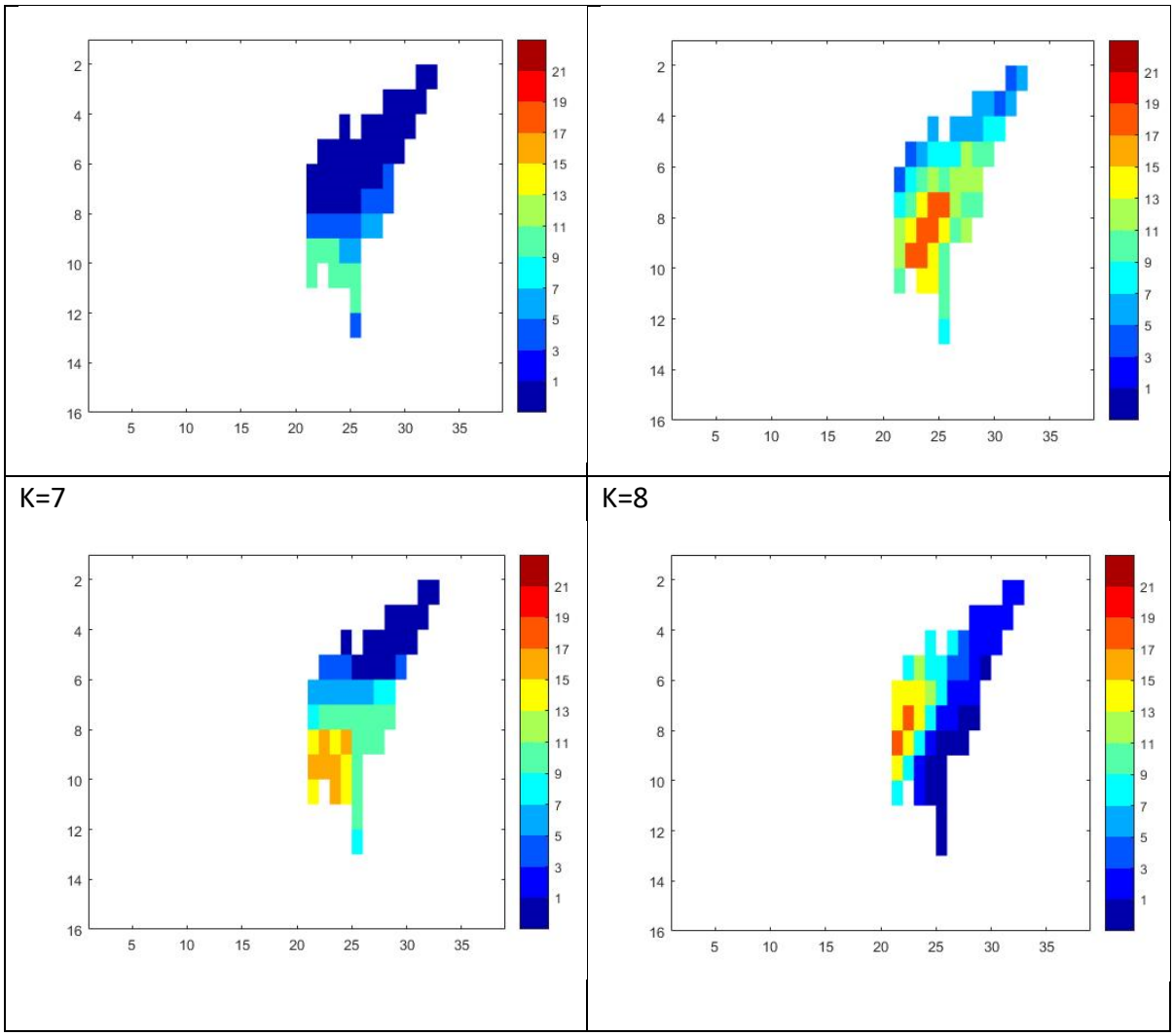
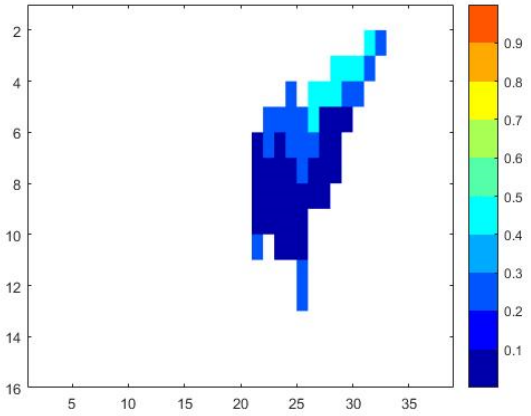
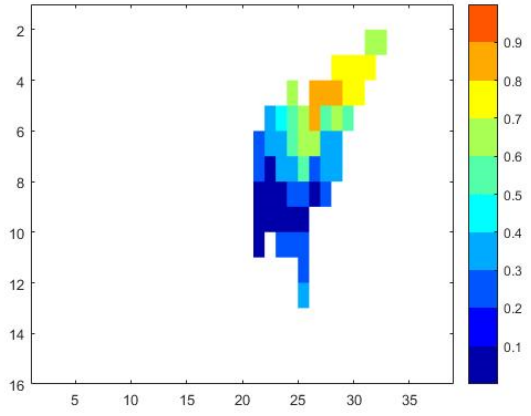


Fig. 5 Average daily rainfall intensity (mm) for the eight-state model

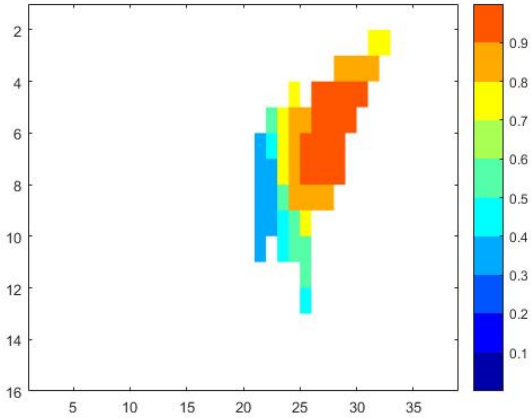
K=1



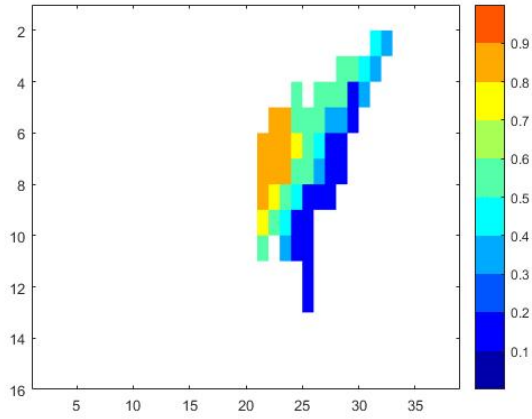
K=2



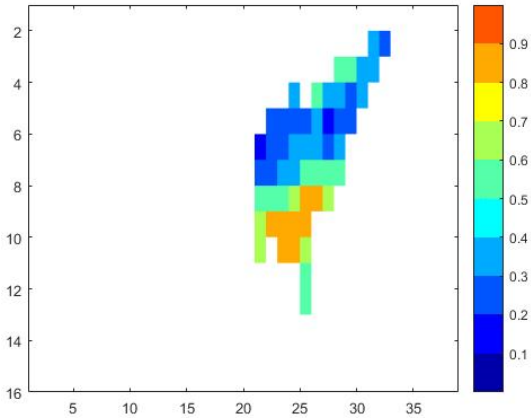
K=3



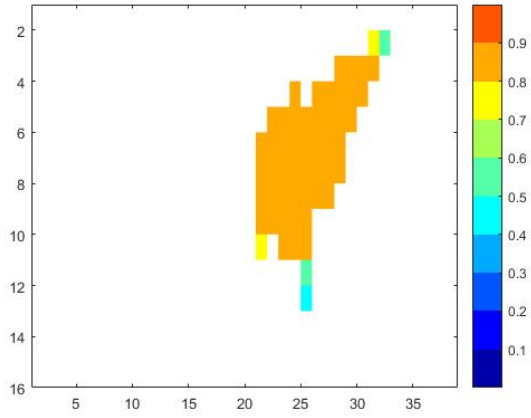
K=4



K=5



K=6



K=7

K=8

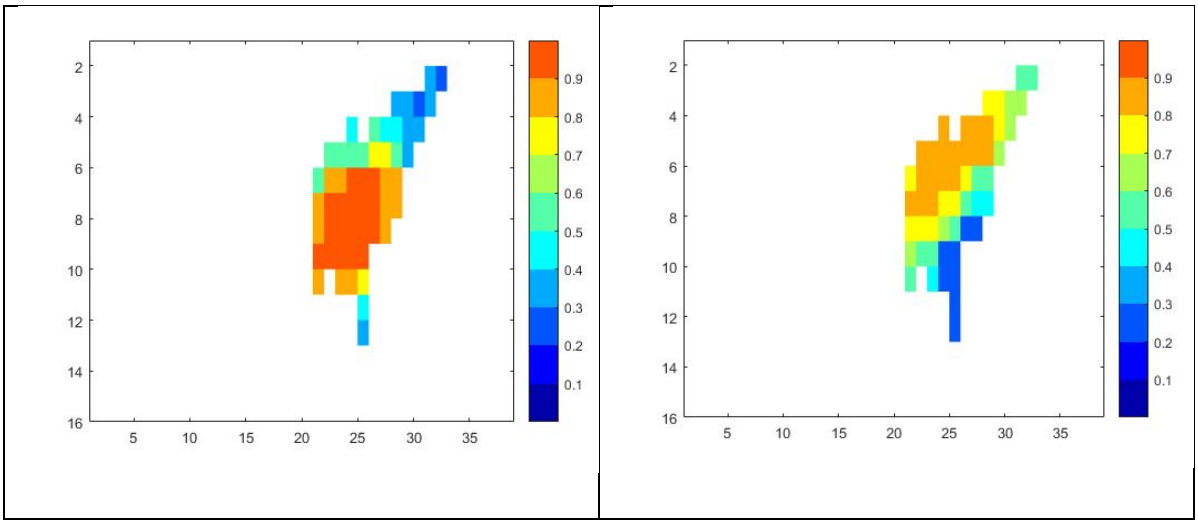


Fig. 6 Occurrence probabilities for the eight-state model

Figure 7 illustrates the most likely daily sequence of the hidden states over the analysis period 1981-2010 and seasonality. The sequence is noisy and average seasonality is not easy to discern. This is partly because of a function of the total number of states (8 states) and illustrates the high complexity of the region's sub-seasonal rainfall variability.

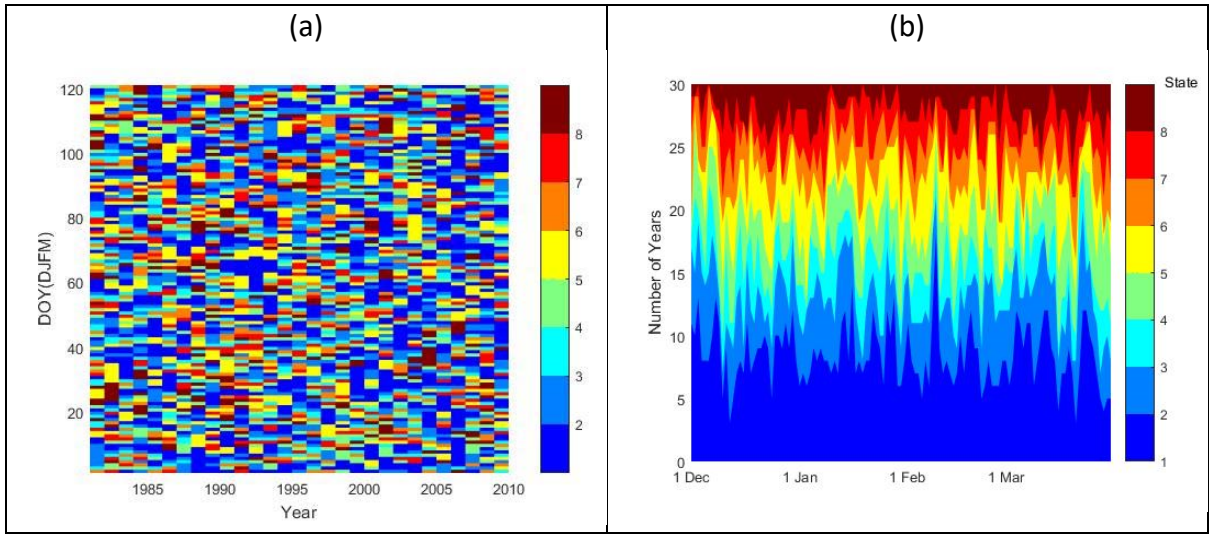


Fig. 7 Viterbi sequence of most likely states in DJFM from 1981-2010 (a) and seasonality for the target area

The Bayesian approach of the GLM-HMM generates samples of the unknown parameters and the hidden states from their posterior distributions conditioned on the data. Figure 8 shows the average estimated coefficients for each gridpoint ("station") and their 95% probability intervals for the exogenous variable (i.e. spatially smoothed MERRA2 data). The average values and 95% intervals were computed using 2000 parameter samples (iterations). At 38 out of total 54 gridpoints, the probability intervals do not include zero, which means the β coefficients associated with exogenous variable are statistically significant from a Bayesian perspective.

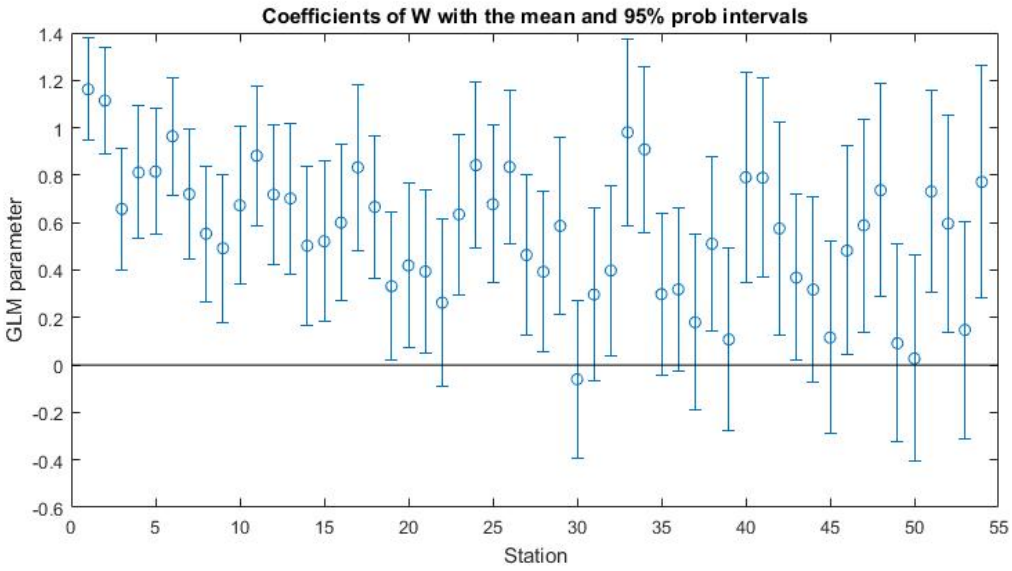


Fig. 8 Coefficients for the GLM exogenous variable (spatially smoothed MERRA2 data) at each gridpoint ("station").

The simulated daily rainfall was validated by comparing with observed rainfall in terms of reproducing the seasonality. Figure 9 shows the observed rainfall amount, obtained by averaging rainfall amount over all gridpoints (spatial average) for each calendar day of the training period (1981-2010) in black line. The mean average simulated rainfall amount per day and 95% confidence bands were created based on 500 30-yr datasets simulated spatially (over grids) and temporally (over each calendar day) and are shown in grey lines in Figure 9. The model captures the climatological sub-seasonal variability of observed rainfall in the target region in general. There is no distinct seasonality found in the observed rainfall probably because of spatial heterogeneity in seasonality across the target analysis region (Eastern US).

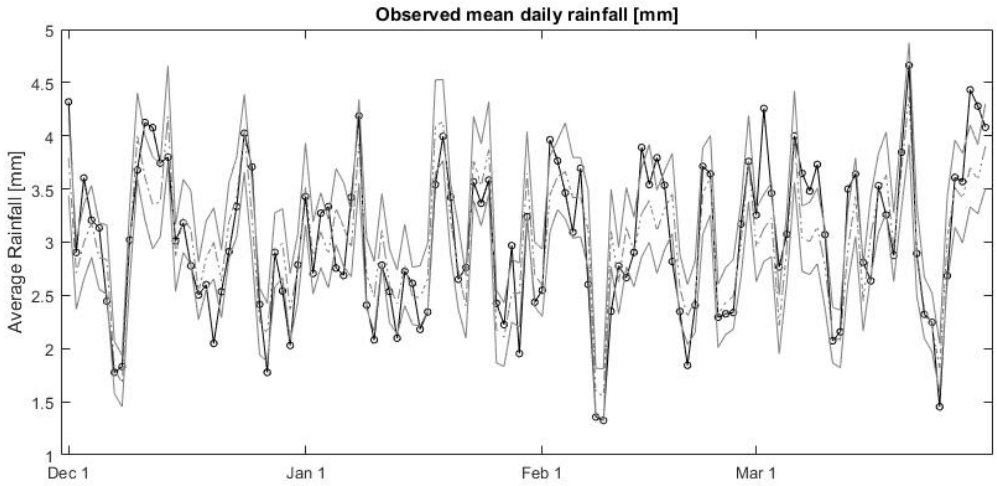


Fig. 9 Climatological sub-seasonal variation of spatially-averaged observed (black line with o-marker) and simulated (grey dotted line) with 95% confidence bands.

The observed rainfall distribution (log frequencies of rainfall histogram) was compared to 500 simulated rainfall replicates of 30 years of data generated via conditional simulation from the model in Figure 10. 95% confidence intervals were also plotted on the histogram. Figure 10 (a) and (b) represent a drier grid (grid # 44 which covers borders between Tennessee and Mississippi) and a wetter grid (grid #12 which covers western New York). In general, the model captures the rainfall distribution of both wet and dry grids. There is a slight bias for wet grid points, where the model underestimates the number of days with high rainfall.

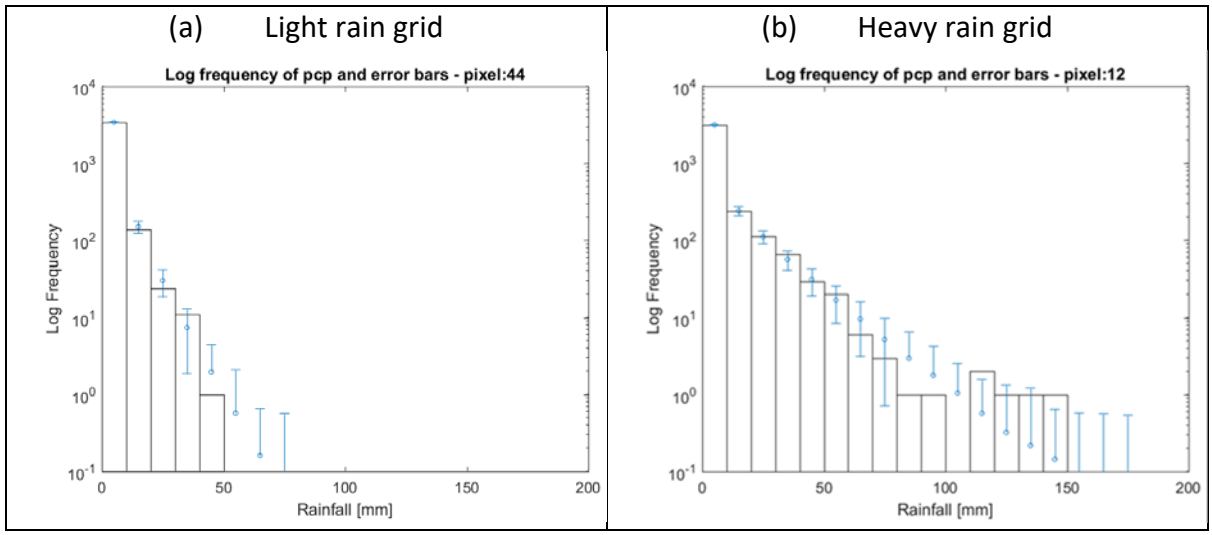


Fig. 10 Histograms of log frequency of observed rainfall and error bars from 500 simulated replicates from the model.

The model's ability to capture extreme rainfall events in terms of daily amounts was explored by fitting a generalized extreme value (GEV) model to the annual maximum rainfall at each grid point. The GEV model fitting was done using the Generalized Extreme Value Distribution function of MATLAB. The GEV fitting produces the return levels of 50-yr return period events. Note that the model we implemented was not designed to model extreme events only relying on the tail behavior of the gamma distribution to simulate the extremes. Fig. 10 shows box plots for each grid (pixel in X-axis) which was created from 500 simulations from the model. Note that on each box, the central mark indicates the median, and the bottom and top edges of the box indicate the 25th and 75th percentiles, respectively. The whiskers extend to the most extreme data points not considered outliers, and the outliers are plotted individually using the red '+' symbol. In Figure 11 observed values were indicted in blue line. The result shows that the model can capture the extremes well for most of the grids in general.

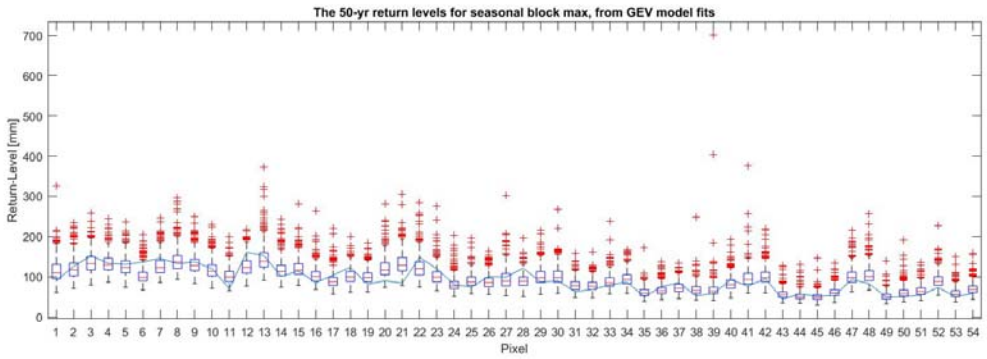


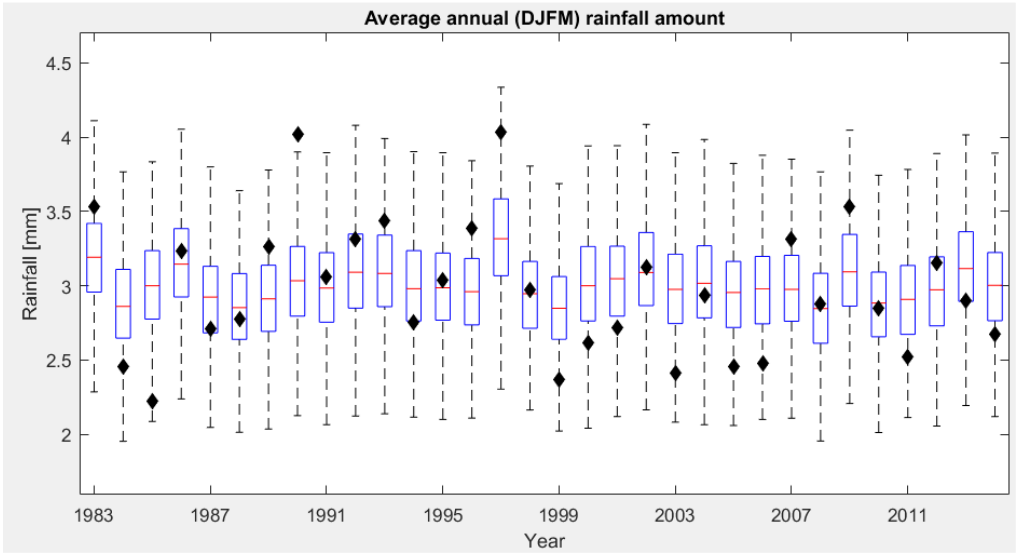
Fig.11 The 50-yr return levels for annual maximum rainfall.

The interannual predictive skills of the model in response to the exogenous variable were evaluated using three metrics of interest: 1) mean daily rainfall (averaged over DJFM), 2) the count of rainfall days over 50mm, and 3) the count of dry days (with zero rainfall). The evaluation was carried out using cross-validation by withholding 4-yr blocks of data consecutively as shown below. The model is fit to the other 30 years of data and the simulated results were validated with the held out 4 years of data.

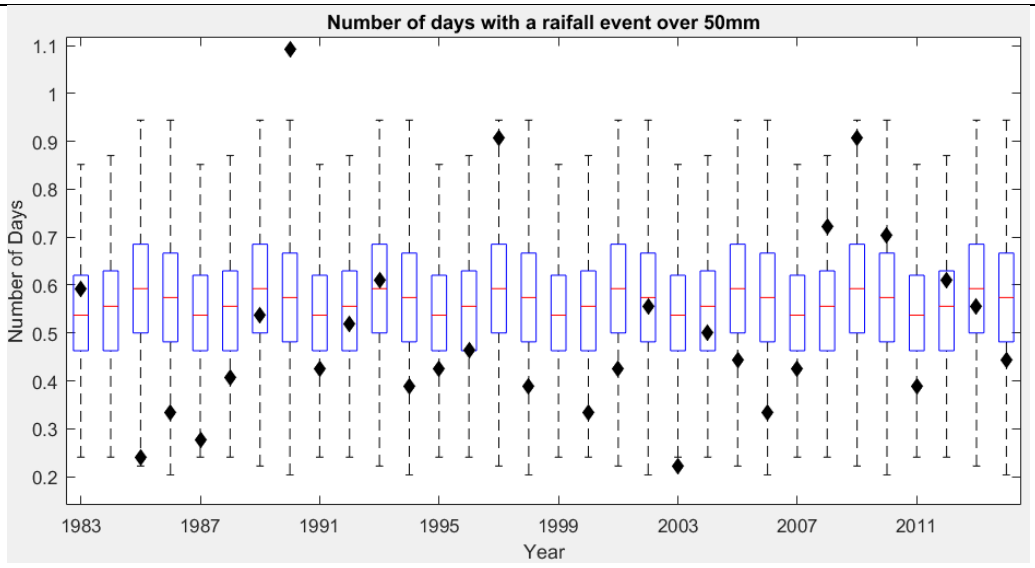
	Training (30yrs)	Prediction (validation) – 4 yrs
1	1981 - 2010	2011 -2014 (Dec. 2011- Mar. 2015)
2	1981 - 2006 & 2011- 2014	2007-2010
3	1981 - 2002 & 2007- 2014	2003-2006
4	1981 - 1998 & 2003- 2014	1999-2002
5	1981 - 1994 & 1999- 2014	1995-1998
6	1981 - 1990 & 1995- 2014	1991-1994
7	1981 - 1986 & 1991- 2014	1987-1990
8	1981 - 1982 & 1987- 2014	1983-1986

Figure 12 shows the results of the conditional predictions by comparing with observed rainfall statistics with diamond-shape markers. The various evaluation metrics were computed by spatially averaging the simulated rainfall across all grids in the study area. Observed rainfall has high interannual variability for all metrics (mean daily rainfall, count of rainfall days over 50mm and the count of dry days), while the model simulations remain relatively constant in their distribution. For most years before 2002, simulated rainfall highly overestimated the number of dry days. The smaller variability of simulated rainfall may be partly attributed to the spatial average over a large area (entire Eastern US).

(a)



(b)



(c)

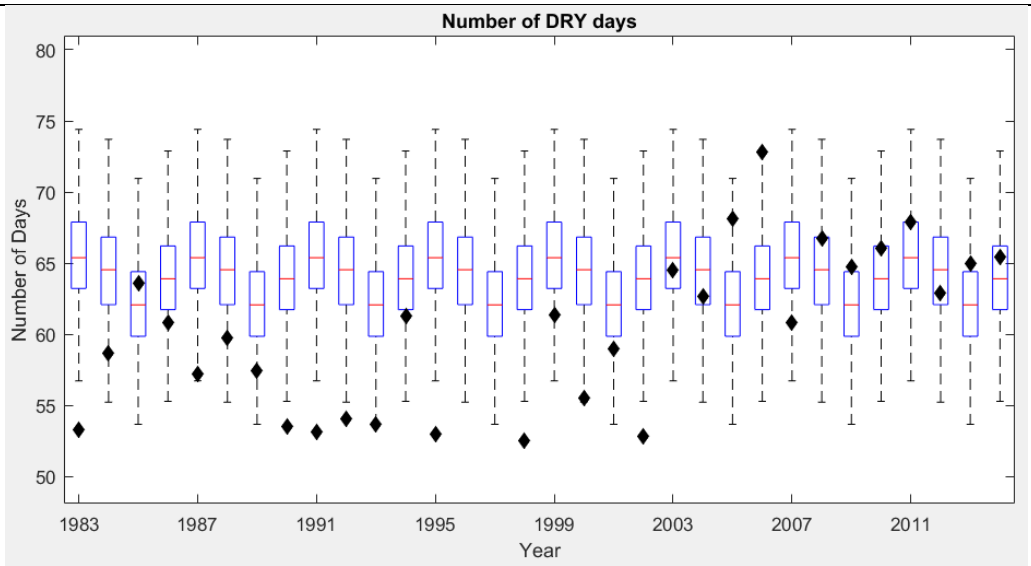


Fig.12 Interannual predictive skill of model: (a) Mean daily rainfall (averaged over DJFM and over all grid points), (b) counts of rainfall days over 50mm and (c) counts of dry days (with zero rainfall).

Figure 13 shows the results from the same simulations in terms of the root-mean-square error (RMSE) for the same metrics of interest (i.e., mean daily rainfall averaged over DJFM, count of rainfall days over 50mm and count of dry days), for each grid and for a particular year. For the mean daily rainfall and number of rainy days over 50mm shown in Figure 13 (a) and (b), northern areas with higher pixel numbers have relatively lower RMSE than the southern part with lower pixel numbers. This indicates that the model did not perform well in wetter southern areas compared to the relatively dry northern areas. For the number of dry days shown in Figure 13 (a) and (c), in the southern areas, the years before 1990 have relatively higher RMSE than the years after 2000. This corresponds to the results of Fig. 12(c). Further investigation is required to find out the reasons for the different RMSE distribution.

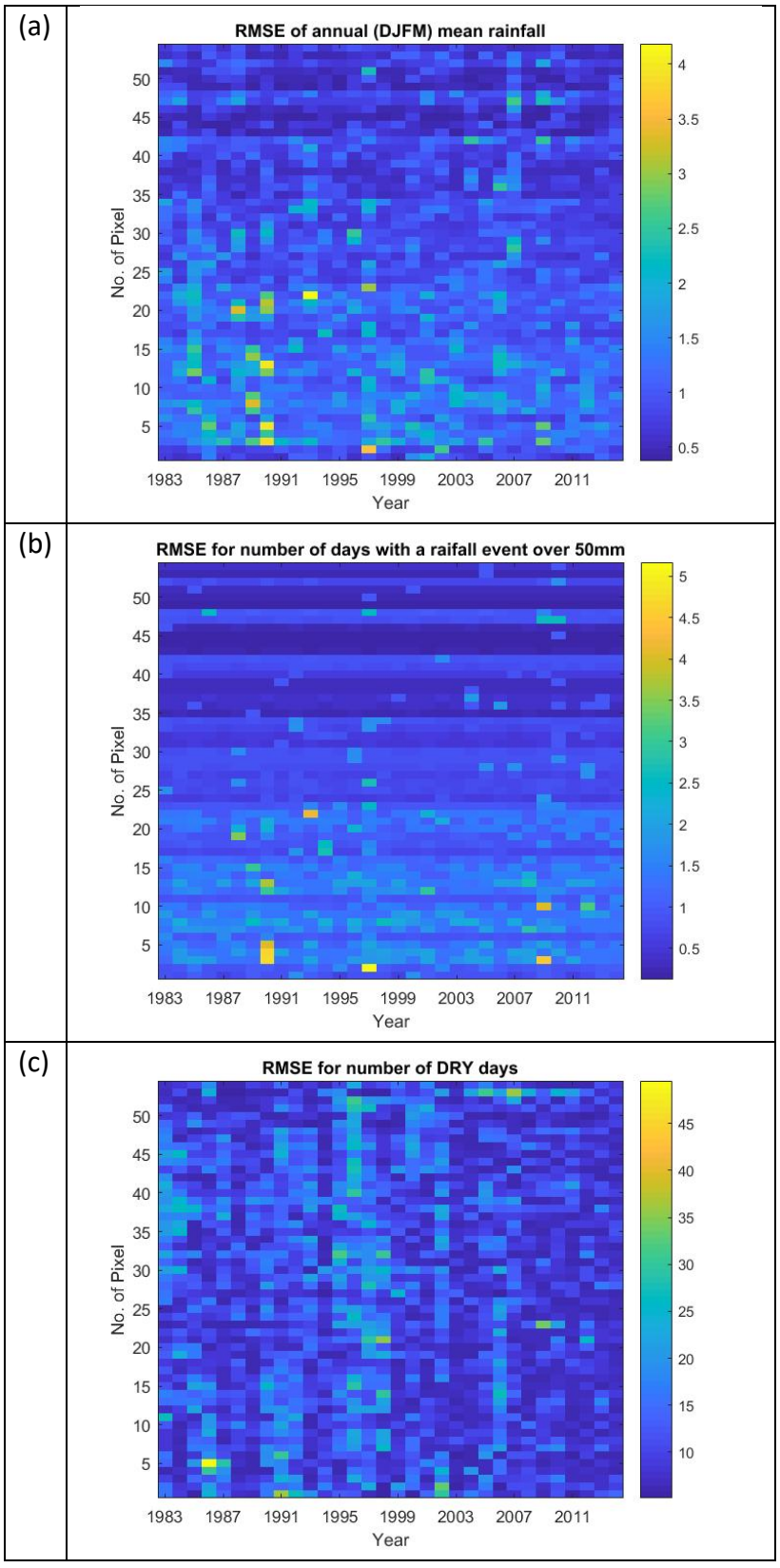


Fig.12 RMSE values per year and per grid: (a) Mean daily rainfall (averaged over DJFM), (b) count of rainfall days over 50mm and (c) count of dry days.

Conclusion

The project aims to explore the potential of using the North American Multi-Model Ensemble (NMME) database of seasonal forecasts to investigate the predictability of sub-seasonal high-impact weather events, and to develop an NMME-based downscaling system for their prediction. Various ways of delineating homogeneous extreme rainfall regions were considered, and an appropriate methodology using cluster analysis was identified. A spatio-temporal downscaling methodology was proposed to apply a hidden Markov model (HMM) toolkit based on generalized linear models and non-homogeneous HMMs. A merit of the GLM approach is to allow exogeneous variables to directly influence the distributuional characteristics of rainfall over time. The performance of the GLM-HMM was examined using the observed rainfall data, CPC Unified 0.25-degree gridded daily precipitation product and an exogeneous climate variable, NASA MERRA2 reanalysis dataset, for a target season (DJFM) in Eastern US. In general, the model can reproduce seasonality and daily rainfall distribution well. Extreme rainfall events with return levels of 50-yr return period were also well-captured. In terms of reproducing interannual variabilities, the model showed small skill for the number of dry days while moderate skill for mean daily rainfall and count of rainfall days over 50mm. Future study will include analysis with various spatio-temporal aggregates of exogenous inputs (e.g., currently pentad average to monthly average) in order to assess the predictability of them.

References

Ambrosino, C., Chandler, R.E. and Todd, M.C., 2014. Rainfall-derived growing season characteristics for agricultural impact assessments in South Africa. Theoretical and applied climatology, 115(3-4): 411-426.

Arthur, D., and Vassilvitskii, S., 2007: "K-means++: The Advantages of Careful Seeding." *SODA '07: Proceedings of the Eighteenth Annual ACM-SIAM Symposium on Discrete Algorithms*. 2007, pp. 1027–1035.

Chen, M. et al., 2008. Assessing objective techniques for gauge-based analyses of global daily precipitation. Journal of Geophysical Research: Atmospheres, 113(D4).

DelSole, T., Trenary, L., Tippett, M.K. and Pegion, K., 2017. Predictability of Week-3–4 Average Temperature and Precipitation over the Contiguous United States. Journal of Climate, 30(10): 3499-3512.

Holsclaw, T., Greene, A.M., Robertson, A.W. and Smyth, P., 2016. A Bayesian hidden Markov model of daily precipitation over South and East Asia. Journal of Hydrometeorology, 17(1): 3-25.

Holsclaw, T., Greene, A.M., Robertson, A.W. and Smyth, P., 2017. Bayesian nonhomogeneous Markov models via Pólya-Gamma data augmentation with applications to rainfall modeling. The Annals of Applied Statistics, 11(1): 393-426.

Kenabatho, P., McIntyre, N., Chandler, R. and Wheater, H., 2012. Stochastic simulation of rainfall in the semi-arid Limpopo basin, Botswana. International Journal of Climatology, 32(7): 1113-1127.

Lloyd, S. P., 1982: Least Squares Quantization in PCM. *IEEE Transactions on Information Theory*. 28, 129–137.

4. Highlights of Accomplishments

- An appropriate regionalization procedure for occurrence of extreme weather events has been developed.
- A stochastic weather generator for daily rainfall has been constructed using large-scale predictor variables from reanalysis data.

5. Transitions to Applications

The regionalization procedure of extreme weather events was used in collaborative work with the Caribbean Institute of Meteorology and Hydrology (CIMH) to predict heatwaves at the May 2017 Caribbean Climate Outlook Forum, Kingstown, St Vincent. The testing and implementation of this method was funded by CIMH in a sub-contract to IRI: *Seasonal Heatwave Predictability in the Caribbean*.

The HMM methodology developed here has been applied in the International Research and Applications Project (IRAP) to generate stochastic sequences of daily weather in seasonal forecasts for input to rice crop models for Bihar, India.

6. Publications from the Project

A paper on the use of cluster analysis for identifying homogeneous regions of climate extremes is in preparation.

7. PI Contact Information

Simon Mason
International Research Institute for Climate and Society
The Earth Institute of University of Columbia
PO Box 8000
Palisades
NY 10968
Email: simon@iri.columbia.edu
Telephone: 845 680 4514