

Spatial Distributions of Precipitation Events from Regional Climate Models

N. Lenssen

September 2, 2010

1 Scientific Reason

The Institute of Mathematics Applied to Geosciences (IMAGe) and the National Center for Atmospheric Research (NCAR) has an active research program in the statistical analysis of large numerical simulation of the climate over North America. What is unique about these simulations is their high resolution (50 km) which approaches the scale needed to assess local effects of climate change. The goal of this research is to quantify the validity of these models when compared to observed data. This project uses my past work on precipitation using flexible statistical analysis software (R package) and observational data.

In this project, a statistic is developed that summarizes different climate models' abilities to localize precipitation events. A successful analysis method will enable developers and users of the models to determine how well a model is localizing storms. The study uses five regional climate models from the North American Regional Climate Change Assessment Program (NARCCAP) to test this statistical approach. Investigating the area that has received a significant amount of precipitation for each time-point has proved an effective way of summarizing the extent of the clustering of rainfall. Curves are created for each time point and then analyzed through scaling and singular value decomposition (SVD) also known as principle components analysis. Comparison of the basis functions generated by the SVD show differences in how different models localize precipitation events.

2 Data

Model output from the five regional climate models was provided by the North American Climate Change Assessment Program(NARCCAP). The goal of this program is to use four atmosphere-ocean general circulation models (AOGCMs) to drive six regional climate models (RCMs) in very high resolution. Information of this program as well as the original data in netCDF format can be found on their website (<http://www.narccap.ucar.edu/>). All of the models have been re-gridded to the same grid to facilitate comparison. The exact dataset used is an R data file in three hour increments that has been compiled by Doug Nychka of NCAR (for information on the conversion of the data from the native netCDF to Rbin contact Dr. Nychka at nychka@ucar.edu). The RCMs used in this study are the CRCM (Quebec, Ouranos), ECPC (Scripps), MM5 (Iowa State/PNNL), RegCM3 (UC Santa Cruz ICTP), and WRF(NCAR/PNNL). The models contain thirty years of data divided into in three hour increments with

rainfall rate reported in $\frac{mm}{hr}$. The region used in this experiment is bounded by the corners $(-104.0, 32.0)$, $(-104.0, 45.5)$, $(-89.5, 45.5)$, and $(-89.5, 32.0)$ and comprise approximately 800 grid points at a spacing of 50 km.

3 Methods

3.1 Proportion Plots

The basic method of analysis used is a plot of the number of grid cells that exceed a sequence of values. Because the grid cells are of equal area, this curve is also the fraction of area where precipitation exceeds a given amount. This was computed by constructing a histogram of the precipitation across grid cells and then forming the cumulative sum of the counts. The shape of this curve compactly summarizes the distribution of rainfall across the region of interest for one time-point. This curve can be described by a function

$$\hat{z}(x) = \frac{1}{N} \{ \#P_{\mu,\nu} : P_{\mu,\nu} > x \} \quad (1)$$

where $P_{\mu,\nu}$ is the rainfall image indexed at pixel location (μ, ν) and N is the total number of grid cells.

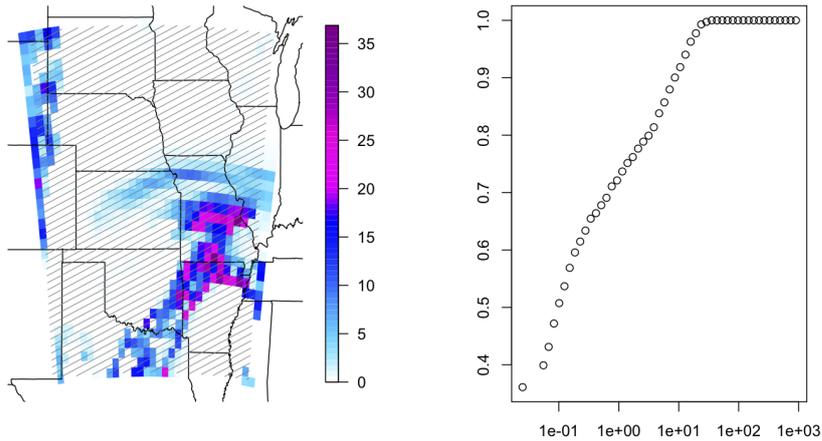


Figure 1: An example of the proportion summary for one three hour time period from the CRCM model. The left plot is the 2-d image and the right plot is the proportion curve. Notice the first point is well above zero indicating a fraction of non-zero precipitation cells.

When the proportion function is evaluated over an image with 50 x values, the result is a vector of length 50 containing the fraction of grid cells above each x value. These 50 values are then formatted as a row of a data matrix, \mathbf{Z} , which is simple way to organize the results across all time points. Specifically for a given model experiment, let $\mathbf{Z}_{i,j}$ be the j^{th} value of the proportion curve for the i^{th} time period. Moreover, we must consider a set of matrices, $\mathbf{Z}^1, \dots, \mathbf{Z}^5$ that are the proportion curve summaries for each of the five model experiments. These are stacked

together into a single large data matrix, \mathbf{Z} , so that linear algebra techniques can be applied to analyze the complete set of experiments.

$$\mathbf{Z} = \begin{bmatrix} \mathbf{Z}^1 \\ \mathbf{Z}^2 \\ \vdots \\ \mathbf{Z}^5 \end{bmatrix} \quad (2)$$

3.2 Empirical Orthogonal Functions

The singular value/principle component analysis used here is based on the concept of expanding the proportion curve in a sequence of basis functions. A single proportion curve $z(x)$ for a particular image can be represented as

$$z(x) = \sum_{j=1}^J \alpha_j \delta_j(x) \quad (3)$$

where α_j are coefficients and $\delta_j(x)$ are the basis functions. Typically the basis functions will be fixed for the whole profile and the coefficients vary to produce differently shaped curves. The goal is to find a small number of basis functions, in our case substantially smaller than 50, and approximate the proportion curves with these limited set. Since the basis functions are kept the same for every profile the differences among proportion curves can be analyzed by the differences of a small number of coefficients.

The singular value decomposition of a matrix can be used to achieve these goals in the case when a discrete set of values are recored for each curve. The main step is to decompose the full data matrix, \mathbf{Z} as a product of three special matrices.

$$\mathbf{Z} = \mathbf{U} \mathbf{D} \mathbf{V}^T \quad (4)$$

For an original \mathbf{Z} matrix with dimensions $i \times J$, \mathbf{V} is an $J \times J$ matrix whose columns contain the discrete basis functions that combine to replicate the original curves. These are analogous to the δ basis functions in the continuous representation. The \mathbf{D} matrix is a $J \times J$ diagonal matrix that contains the weights for each of the basis functions and the \mathbf{U} matrix is $i \times J$ and contains the coefficients for each time period of data. To create the desired approximations of the dataset, \mathbf{D} values after the first n can be set to zero and the result will just be the combination of the first n basis functions.

$$\begin{bmatrix} D_{1,1} & 0 & 0 & 0 & 0 & 0 \\ 0 & D_{2,2} & 0 & 0 & 0 & 0 \\ 0 & 0 & \ddots & 0 & 0 & 0 \\ 0 & 0 & 0 & D_{n,n} & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 \end{bmatrix} \quad (5)$$

This truncation of \mathbf{D} effectively limits the number of columns of \mathbf{V} used to represent the data. Often a small number of columns is effective in representing the entire data matrix and in this case, we will see that three columns are adequate to capture the variation in the proportion

curves among different time periods and different models. Thus, interest lies in the basis functions V and their respective weights, D . Investigation of these basis functions reveal the nature of the variation between time-points.

4 Results

To make sure that all of the precipitation values were comparable across the 5 models, the model outputs were standardized by a the global mean of all the 2-d fields. Proportion curves were calculated for each time-point and the matrix containing these curves was reduced by a singular value decomposition. Relative weights of the basis functions were calculated by dividing each of the D matrix values by the cumulative sum of the D values. The weights of the first three functions were 0.904, 0.0799, and 0.00948. These three accounted for 99.4 % of all variance between the curves. Plots of the first basis function (the first column of DV^T) and the effects of the D -scaled second and third were generated to see how much variance could be demonstrated.

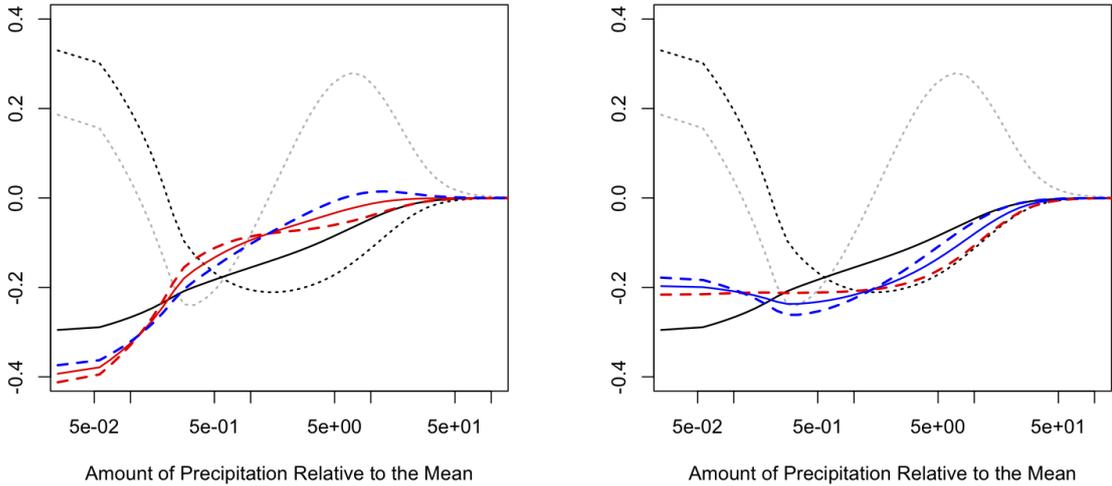


Figure 2: The addition and subtraction of the second and third basis functions to the first. The solid black line is the first function, the dotted black line is the second, and the dotted grey line is the third. On the left, the second function is added to the first and then \pm the third. The right shows the second function subtracted from the first and then \pm the third.

Figure 2 provides a visual representation of how the first three basis functions work together to modify shape of the approximated proportion curve. It also makes it apparent how much variation could potentially exist between two curves and where in the curve that variation would be demonstrated.

In addition to analyzing the weighted basis functions in the DV^T matrix, the values of the

coefficients found in the U matrix can quickly reveal aspects of the data. To determine how much the functions actually are varying and if this variation is different between the models, plots between the first and second and second and third basis function coefficients were generated. Contours were then taken from these scatterplots at enclosing regions that contain approximately 25%, 50%, and 75% of the points. To relate these contours plots back to the different shapes implied in the original proportion curves, points near the contour lines were pulled off and were used as the coefficients, or a row of U matrix, for a reconstruction with the previously calculated DV^T . For figures 2–7, black is CRCM, red is ECPC, green is MM5, blue is RCM, and cyan is WRF.

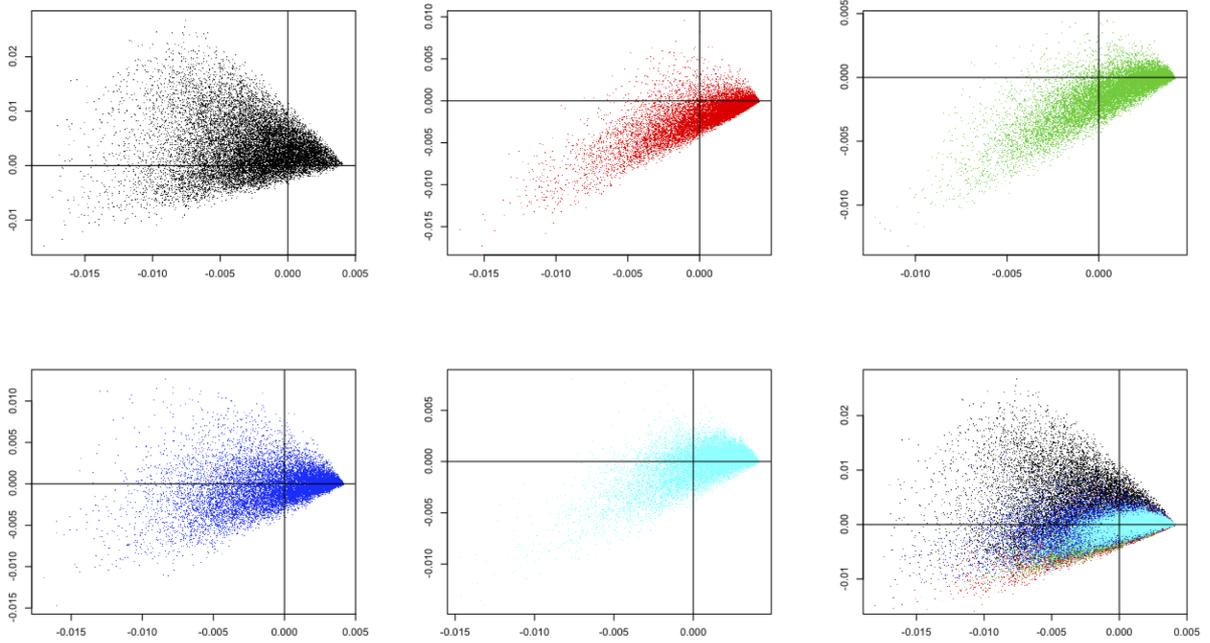


Figure 3: Plots with the first basis function coefficients on the x-axis and the second on the y-axis for all the models separately then overlaid.

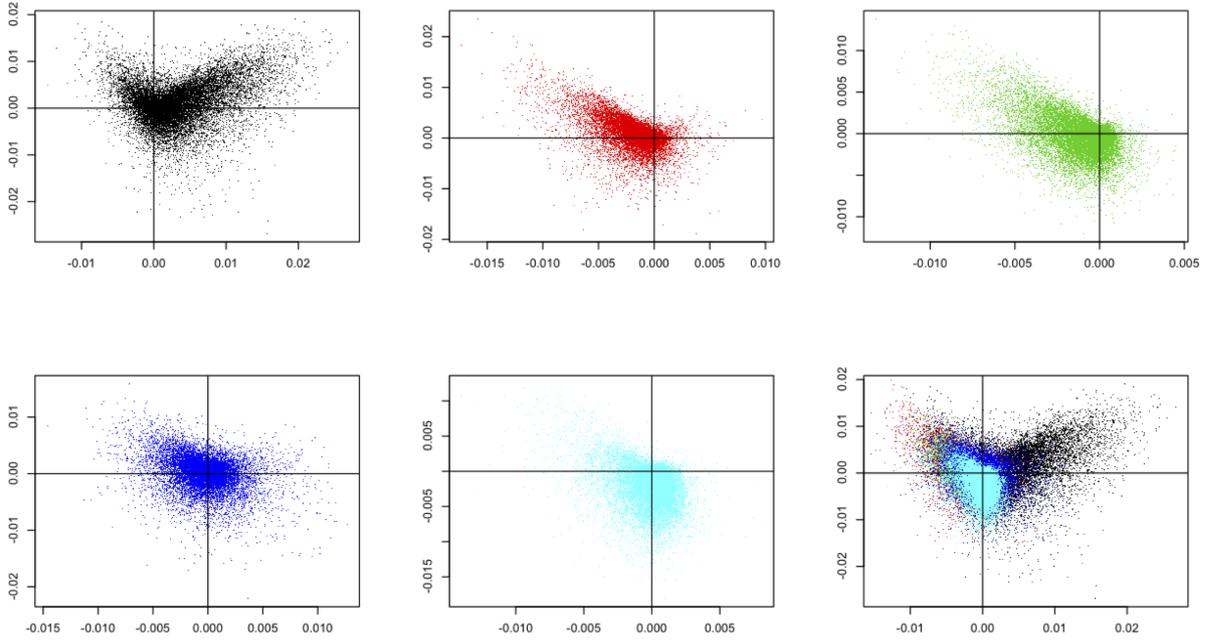


Figure 4: Plots with the second basis function coefficients on the x-axis and the third on the y-axis for all the models separately then overlaid.

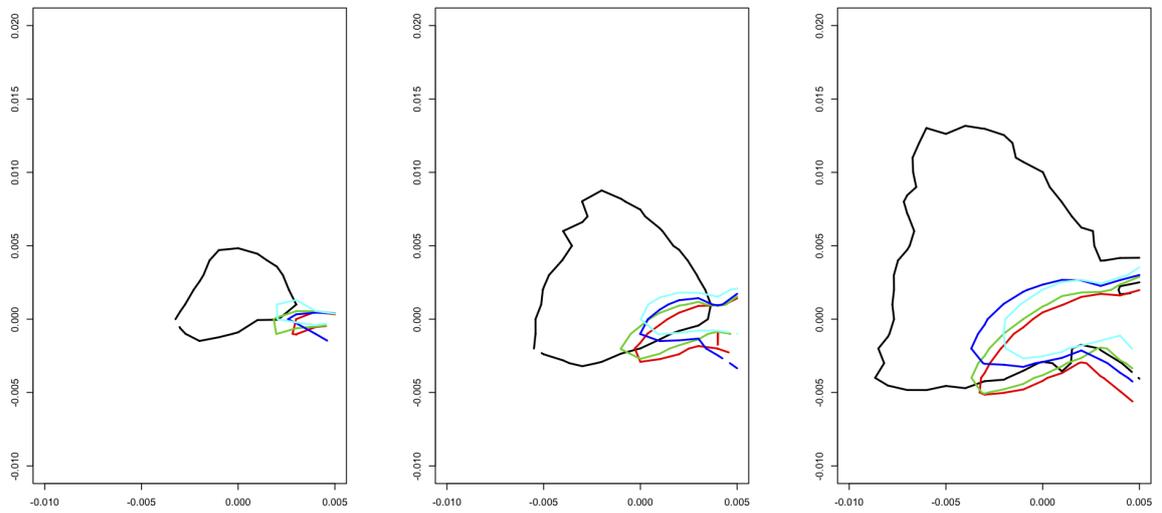


Figure 5: 25%, 50%, and 75% contours [left to right] from the scatterplots in figure 3

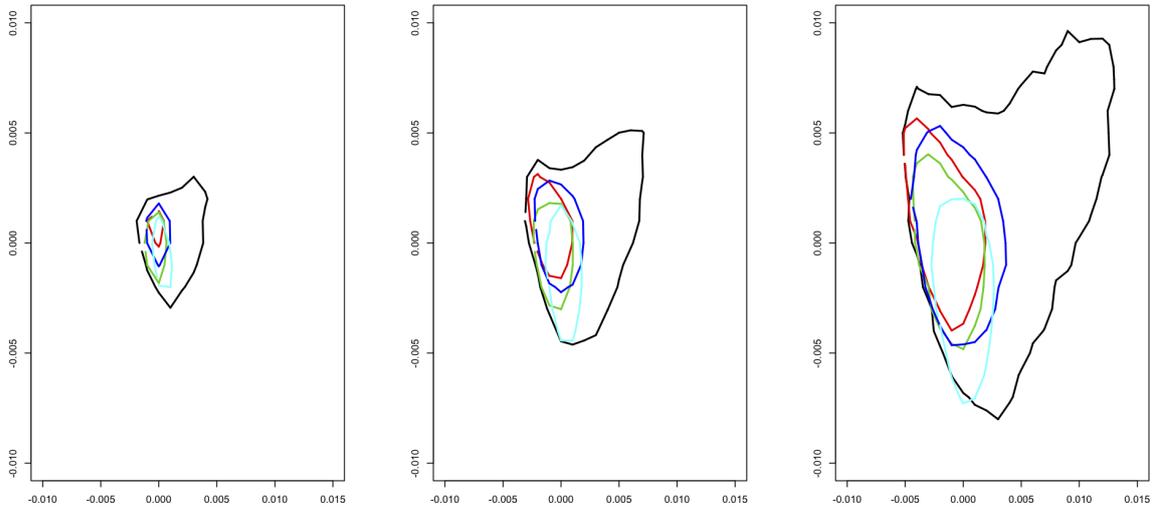


Figure 6: 25%, 50%, and 75% contours [left to right] from the scatterplots in figure 4

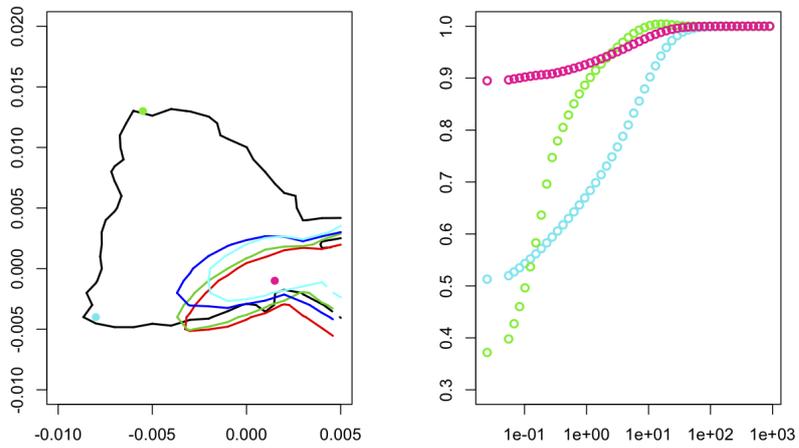


Figure 7: Three points selected from the plot in figure 5 and their proportion plots when multiplied with the DV^T matrix

5 Conclusions

By looking at figures 2–7, it is apparent that there is a large amount of variation in the proportion curves both within and between the models. Figure 2 demonstrates just how much variation there can possibly be by plotting the extreme additions and subtractions of the first three basis functions. The best demonstrations of variation occur in figures 5 and 6, by showing how much the CRCM model’s proportion curves vary from the other four. This analysis suggests that the CRCM model has a significant amount of distinctly different proportion curves from the other models. This difference appears to be in the amount of the variability. On closer investigation of the CRCM coefficients in figure 7, it is explicitly shown how the curves vary and how the CRCM model has more curves with larger precipitation events.

5.1 Future Work

One way this analysis could be strengthened is by adding in historical rainfall data from NOAA’s NEXRAD doppler system. This raw data could serve as the control by providing proportion curves from actual precipitation events.