8.3 STATISTICAL ANALYSIS OF MESOSCALE CONVECTIVE SYSTEM MOUNTAIN INITIATION LOCATIONS

Elisabeth F. Callen* and D. F. Tucker
University of Kansas, Lawrence, Kansas

1. INTRODUCTION

There are roughly three categories of convective weather systems: single cell systems (including supercells), multiple cell systems, and mesoscale convective systems (MCSs). MCSs produce large amounts of precipitation, making this category the focus for this analysis. A further step into this analysis is looking at the components needed for orographic MCS initiation. There is a connection established between mountain convection and eventual MCS initiation since Rocky Mountain convection tends to lead to numerous MCSs, but the cause of the connection is still under investigation (Tucker and Crook, 1998). According to Banta and Schaaf (1986), Tripoli and Cotton (1988), and Tucker and Crook (2005), convective initiation occurs in preferred locations in mountainous regions. There is a possibility of super-preferred locations to convective initiation, but this is also still under investigation.

Banta and Schaaf (1986) determined some preferred locations by examining satellite images and tracing the thunderstorms, not necessarily MCSs, back to the initiation locations. Tripoli and Cotton (1988) determined areas for MCS initiation in the Rocky Mountains by examining diurnal flow regimes interacting with topography. Tucker and Crook (2005) determined preferred locations by examining the outflow from mountain convection.

According to Tucker and Li (2009), their data base of thunderstorms in the south central United States area was one percent MCSs. An analysis of a subset of this one percent will be performed. The studied MCSs originate west of 104 W (mountain initiation) in the Arkansas-Red River basin in the warm season (April-September) in the years 1996 to 2006. In this paper, one preferred location will be analyzed after a cluster analysis is performed.

Once the data was selected, a Multiple Linear Regression (MLR) and Principal Component Analysis (PCA) were run on the entire cluster for each of the 6 hours leading up to initiation, the initiation hour, and the 3 hours following initiation.

2. DATA AND METHODS

The original data used in this analysis was first used by Tucker and Li (2009). This data is multi-sensor precipitation data, as described in Young et al (2000). It is a combination of radar, satellite, and rain gage data. This combination provides the best estimate of precipitation over other known methods. This original data was then delineated into systems. Cells from one hour had to be connected to cells in the next hour for the same system to continue through the next hour. Criteria were set for single cell systems, multiple cell systems, and MCSs. For a system in this dataset to be considered an MCS, it had to have a footprint size of 21 cells or greater and a lifetime of at least 6 hours. The cells, with an approximate size of 5 kilometers by 5 kilometers, were from the original multisensor precipitation data from Tucker and Li (2009). A cell is added to the footprint size if precipitation is detected in the cell within a system. Initiation hour location was the location used in determining the variable values. The initiation location was the location where the first cells connected to the MCS appeared, not where the system met MCS criteria. Once a cluster analysis was performed on the approximately 1,700 MCSs originating west of 104 W from the original data set, the cluster used contained the most members out of all the clusters (154 members) and is considered to be the most commonly occurring pattern for this data set. Tucker and Li (2009), originally, looked at the occurrence of all types of systems in the entire Arkansas-Red River basin in the years 1996 to 2006 in the warm season.

The MCSs originating west of 104 W during the warm season were used in this analysis. These MCSs were extracted from the original data using MATLAB. The coordinates were transformed from Hydrologic Rainfall Analysis Project (HRAP) coordinates (ARM: Climate Research Facility) into latitude and longitude (Polar Stereographic coordinate system). The initiation times were also changed from local time to Coordinated Universal Time (UTC) time. This was done to make it less of a challenge to search for the correct initiation data and to place all the MCSs in a well-known reference frame.

A cluster analysis was run on the subset in SPSS using between-groups linkage and Euclidean distance. The cluster analysis was performed in hopes that the individual clusters were centered on the mountain peaks in the general area that the MCSs formed. The end result of the analysis was to find the initiation characteristics of each cluster, possibly concluding that each cluster had different initiation conditions. For this paper, only one cluster from that cluster analysis will be examined. This cluster was centered on Elk Mountain in New Mexico at approximately 35˚ 46’ N and 105˚ 33’ W, as can be seen in Figure 1. Figure 2 gives the distribution of the start years, start months, start days, and start hours of all the MCSs included in the Elk Mountain cluster. The years 2004 and 2006 had the highest occurrence of MCSs in the Elk Mountain vicinity. The highest occurrence of MCSs occurred in July and August.

* *Corresponding author address:* Elisabeth F. Callen, Univ. of Kansas, Dept. of Geography, Lawrence, KS 66045; email: bcallen@ku.edu

Approximately half of the MCSs occurred at a start hour of 0000 UTC. The next highest initiation hour was 0100 UTC, then 2300 UTC in the Elk Mountain vicinity. This highlights the fact that most of the first cells of the eventual MCSs have late afternoon to early evening mountain initiation. Figure 3 contains a histogram of the day of year the MCS initiations occurred. The mean day of year for MCS initiation is 204.36. This is July 22 for leap years and July 23 for the other years. Days 215 to 220 are when the MCS initiations occurred most often at 15 out of 154 times. Days 205 to 210 were the next most likely time frame for MCSs to initiate at 13 out of 154 times. Days 220 to 225 were the next most likely time frame, after days 205 to 210, for MCSs to initiate at 12 out of 154 times. All of these days are in middle to late July, showing that the best time of year for MCS mountain initiation is in the summer in July.

Surface data, upper air data, and North American Regional Reanalysis (NARR) data was gathered for each MCS in the cluster. The surface data and upper air data were gathered from Iowa State's online database and the NARR data was gathered from NOMADS. The surface data and upper air data were downloaded for the day of initiation and examined at the hour of initiation. NARR data was downloaded 2 model runs before and 1 model run after initiation of the MCS. If initiation occurred in the hour a NARR model run was performed, then that model run was also downloaded. To find the data in the files for each MCS, GEneral Meteorology PAcKage (GEMPAK) (DesJardins and Petersen, 1986) was used on the surface data and on the upper air data. On the NARR data, Integrated Data Viewer (IDV) (Murray, 2003) was used to find the variable values. The U component of the wind runs west-east. The V component of the wind runs south-north.

The variables were chosen for study because of the possibility that these variables are significant to MCS initiation. For each MCS, the following variables were recorded for use in the MLR and PCA:

Surface data (recorded from the closest station reporting all the variables):

- Surface potential temperature in Celsius (STHC)
- Surface mixing ratio in g/kg (SMXR)
- Surface saturated mixing ratio in g/kg (SMXS)
- U component of the wind in m/s (UWND)
- V component of the wind in m/s (VWND)

Upper air data (recorded from the closest station reporting all the variables):

- Lifting Condensation Level (LCL) [mb]
- Lifted Index (LI)

NARR model data (recorded at the initiation location, for each model run):

- 1000-500 hPa thickness (thickness) [gpm]
- Precipitable water (PW) [mm]
- Convective Available Potential Energy @ surface (CAPE) [J/kg]
- Convective Inhibition @ surface (CIN) [J/kg]
- Storm relative helicity (SRH) [m^2/s^2]
- Geopotential height @ 600, 500, 300, 200 mb (GH600, GH500, GH300, GH200) [gpm]
- Specific humidity @ 850, 800, 600, 500, 300, 200 mb (SH850, SH800, SH600, SH500, SH300, SH200) [kg/kg]
- U component of the wind @ 600, 500, 300, 200 mb (UC600, UC500, UC300, UC200) [m/s]
- V component of the wind @ 600, 500, 300, 200 mb (VC600, VC500, VC300, VC200) [m/s]
- Temperature @ 600, 500, 300, 200 mb (T600, T500, T300, T200) [˚C]

Once all the variables were gathered for each MCS, MLR and PCA were run for each hour in SPSS. The MLR and PCA runs resulted in equations for the 1 through 6 hours prior to initiation, the initiation hour, and the 1 through 3 hours after initiation.

For the MLR, the dependent variable used was the footprint size (FP). Since initiation was known to occur in that location for that MCS, FP was used to determine the characteristics needed to produce an MCS of that size. It was reasonable to assume that the characteristics needed to produce that FP would also be needed to produce that MCS. The type of MLR run was a stepwise MLR. Each variable was taken into account separately and the best fit variable was added into the MLR equation one at a time. An entry value of 0.15 and a removal value of 0.20 were chosen for each MLR run. These values were chosen because the R square value increased significantly from using an entry value of 0.10 to an entry value of 0.15. For a variable to be entered into the MLR, the error associated with that variable had to be less than the entry value. If there were multiple variables with an error less than the entry value, then the variable with the smallest associated error was the one included for that run in the MLR. A variable was only removed from the MLR if the error associated with that variable became larger than the removal value. One issue that could arise with MLR is that the fit of the independent variables may not be a linear fit to the dependent variable. MLR was used here because it made it easier to compare from hour to hour what is important to the formation of an MCS with that size of a FP. Also with MLR, there could be multicollinearity issues. If the independent variables are not truly independent of each other then multicollinearity is likely to occur. This is an issue with MLR and is corrected when using PCA.

PCA was performed to correct any possible multicollinearity issues seen when performing MLR and also to have an equation that included all the variables instead of a select few. PCA has components that form a linear equation but the fit to the original data is better since for each component, each variable is used. Since the variables were standardized, the eigenvalues greater than 1 were kept to account for most of the variance. The eigenvalues were used to determine the loadings of each variable. How much a variable loads into a component depends on how much of the variance of that variable is accounted for in the PCA for that component. Another option would have been to look at the scree plot that is outputted to determine the best place to stop to account for most of the variance, but the eigenvalue greater than 1 was used instead. The eigenvalue value is more efficient to set in SPSS than looking at the individual scree plots. To compare the different PCAs, the variables used most often in the PCA were compared.

3. MULTIPLE LINEAR REGRESSION

3.1 *MLR Results*

Once the MLRs were run for the 6 hours prior to initiation, at initiation, and 3 hours after initiation, an analysis and comparison of the output was performed. The most important data output for each model run were the variables included in the final run of each model and the resulting R square value. The complete list of the data is included in Table 1. Only the approximate equations for each ending model run are given since this is an analysis and comparison of the variables used in each equation. Each model run equation was equal to FP. When looking at the original data, FP varied widely from MCS to MCS. FP was only chosen as the dependent variable because it is included as part of the criteria needed for an MCS to occur in the data set first used by Tucker and Li (2009).

While no two MLR runs were exactly the same, there were some similarities. The most used variable was UC500 (7 out of 10 MLR runs). Since UC500 is included in many of the MLR runs, it can be assumed that, if looking at the runs individually, the value of UC500 helps MCS initiation throughout most of the time frame studied. With a component of the wind being included so often in the MLR, the inclusion of a proxy for wind shear becomes a possibility, especially if another wind component with the opposite sign is used. The U component of the wind at 500 mb was used more often than the V component of the wind at 500 mb (4 out of 10). Just because it seems the U component of the wind is important to MCS initiation, it does not necessarily mean that the V component is also important to MCS initiation. This does seem to be the case for the MCSs forming off Elk Mountain in New Mexico. This could be the result of the surrounding topography. According to Tucker and Crook (2005), storm initiation off a mountain feature is sensitive to the wind direction. The reason the wind components are not included in the MLR equations more frequently is that the given FP size is not sensitive to wind direction. There is a possibility that FP size is sensitive to wind direction, but it is not that noticeable in the MLRs since it is sensitive to multiple wind directions instead of a single direction. MCS initiation may be sensitive to wind direction, but to have an MCS of a certain FP size, wind shear is not as important. The only other variable used at 500 mb was GH500 (1 out of 10 runs). This shows that the geopotential height at 500 mb is not as essential to the MCS initiation as the other variables.

The U component of the wind at 600 mb (UC600) was used in 4 of the 10 runs and the V component of the wind at 600 mb (VC600) was used in only 1 run. At 600 mb, the surrounding topography would play a bigger role in the movement of the winds than at 500 mb. The winds at 600 mb are much more strongly influenced by the elevation changes than at 500 mb. Since UC600 and VC600 are used less often than UC500, it can be inferred that the winds at 600 mb are not as critical to MCS initiation as the U component of the wind at 500 mb. Also, the V component of the wind at 200 mb is of the same importance as the U component at 600 mb. VC200 was used as a variable in 4 out of the 10 separate runs. The winds at 200 mb are not influenced by topography but are rather influenced by the jet stream. This shows that the possible inflow of air from the south can be crucial to initiation. This inflow of air has to occur at 200 mb according to these model runs for it to be of use.

A possibility is that the winds at different heights were used as a proxy for wind shear. When the stepwise MLRs were created, SPSS selected from those available instead of creating wind shear variables. SPSS used opposite sign wind component variables to account for the wind shear needed in MCS initiation. Looking at Table 1, the signs (positive or negative) are opposite for two levels of wind components. In multiple model runs, UC500 is positive and UC600 is negative. This could be used as the proxy for wind shear. Wind shear is definitely vital to MCS initiation. Moderate shear is needed for MCSs to initiate. Too much shear and the storms that would form would be supercells rather than MCSs. The U component wind shear is much more prominent than V component wind shear. This gives the conclusion that wind shear in the U component is more central to MCS initiation than in the V component. Sign changes between V component levels only occur rarely. These changes are not as consistent as can be seen with the U components. The V component levels used change where the U component levels were consistently at 500 mb and 600 mb.

The surface winds were not as valuable in these model runs. The U component of the wind was only used in 2 runs and the V component of the wind was only used in 1 run. These surface winds were not taken at the initiation location. They were recorded from the nearest reporting station that had a complete data set. There was not a reporting surface station close to Elk Mountain so the exact surface conditions are not known and are assumed to be similar to the reporting surface station. This is a source for error and if this could be corrected then there will be a change in the variables used for the MLR runs. Surface conditions should play an important role in the initiation of an MCS. For the MLRs, the surface conditions are not as critical as other variables. The SMXR, used in 1 run, and the STHC, used in 3 runs, also came from the same reporting surface station as the wind components. There would also be an error associated with the SMXR and STHC since the values were not taken at the initiation location. These variables would also become more useful if values were taken from the initiation location. This could also show that the mountain-valley winds (surface winds in the area around Elk Mountain, NM) in that location are not as important as other variables to MCS initiation. The only way to determine if this was the case is to get the variable values located in the surface and upper air files from the initiation location.

The data recorded from the upper air files was also used in a few model runs. The LCL values and LI values were both used in 3 out of 10 model runs. The closest reporting station for most of the MCSs was Albuquerque, New Mexico (ABQ) which is close but still a distance away from Elk Mountain. For 1 MCS, El Paso, Texas (EPZ) was used for upper air data. For 2 MCSs, Amarillo, Texas (AMA) was used for upper air data. AMA is not located in the mountains but to record upper air data for those two MCSs, it was the closest reporting station. These three MCSs have the potential to skew the data since the less representative data is at EPZ and AMA rather than at ABQ. Since the upper air data could not be taken at the initiation location, this is also a source of error. Another source of error is that the soundings were only taken every 12 hours at ABQ. Many of the MCSs formed at 2300 UTC, 11 hours from the last, closest sounding. A lot can change in the atmosphere in 11 hours, and the soundings taken at 1200 UTC are not representative of the atmosphere 11 hours later. For each model run LCL was included in, it was always a positive value. Since LCL is also always positive, this gave a positive value to the equation. LI was different. Two times LI was used, it was a positive value, and the 1 time LI was used, it was a negative value. It is hard to tell what significance different LI values will give since LI can be positive or negative.

The T300 and T200 variables were included in 4 out of 10 model runs. This is surprising because as the data was being recorded there was not much of a change seen in the T300 and T200 variables from case to case. The changes in T300 and T200 must have been considered significant enough and the error associated with those variables small enough to be included in the MLR runs. If the depth of convection is large, spanning up to 300 mb and 200 mb, then the temperatures at 300 mb and 200 mb becomes important. The temperature at that pressure is very cold and only ice particles should exist (-30 ˚C to -50 ˚C). If the convection reaches that high, then the amount of ice present at 200 mb and 300 mb affects MCS formation. The ice falls from low pressures into higher pressures, affecting the ice-to-liquid water content. This occurs when the convection is deep enough to reach high into the atmosphere. This deep convection generates ice particles high in the atmosphere that fall, affecting the ice-to-liquid water content. The ice in connection with outflow was discussed as a possibility in affecting storm initiation in Tucker and Crook (1998). When T200 and T300 occurred in the same equation, opposite signs were employed. T200 most often had a positive sign, meaning it would give a negative value in the equation (temperatures recorded in Celsius). T300 most often had a negative sign, meaning it would give a positive value in the equation. The one time apiece that T200 and T300 were not included in the same model run, opposite signs of what was most often used occurred. T200 had a negative sign, giving a positive value for the equation and T300 had a positive sign, giving a negative value for the equation.

The rest of the variables were included between 1 and 3 times. This is also somewhat surprising since CAPE and CIN would be a factor in the initiation. Lower than expected values of CAPE were recorded giving rise to the idea that MCS initiation does not need a lot of CAPE. However, CAPE was included in the 4 hours prior to initiation model run. This model run had the highest R square value. This gives the conclusion that while CAPE is not important at other hours prior to initiation, CAPE does become important 4 hours prior to initiation. CIN values were also relatively small, so little energy would be needed to overcome it. When CAPE and CIN were included in the MLR equations, these variables were given as positive values. Since CAPE was already a positive number, CAPE will give a positive value in the equation. CIN was a negative number, and would therefore give a negative value to the equation.

Some of the variables recorded were not used in any of the MLR runs. These variables are SMXS, SH800, SH500, VC300, and T500. It would have been expected that SMXS was important since it is the surface saturated mixing ratio but this seems to have not been the case. There is also an error associated with SMXS since it was recorded from the closest reporting surface station like the other surface variables. If the variable had been recorded at the initiation location, then the outcome of how frequently this variable was used could have been different. The rest of the variables not included had the potential to be of use but either varied too widely to be associated with a certain FP size or did not vary enough.

The fit to the data changed dramatically from model run to model run. The lowest R square value was reported at 2 hours prior to initiation at 0.266. The highest R square value was reported at 4 hours prior to initiation at 0.862. The model run at 4 hours prior to initiation gives the best mix of ingredients needed for MCS initiation. This implies that the best time for forecasting MCS initiation at a particular location, according to MLR, is 4 hours prior to the first cells initiating. This model run contained multicollinearity issues, but that is to be expected. The variables were expected to be interrelated to some degree. If MCS initiation occurs with given characteristics, then those characteristics are expected to be related to each other. The highest R square value, not containing multicollinearity issues, was reported at 2 hours after initiation at 0.746.

The higher entry value was used (0.15) in order to increase the R square value. The higher R square value was the result for all but 1 model run. The two-hours-prior to initiation equation did not change from using an entry value of 0.10 to an entry value of 0.15. This shows that the variables at that time contained too much error to be of any use. This model run also contained the lowest R square value. A higher entry value could be used, but then there is the issue of allowing that much error to be included in the model run.

The model runs with the higher R square values are deemed more important since it is a better fit to the data. The number of times a variable is used is important, as well as how the variable was used and the combination of the variables used was also important. When a variable is used more often, it indicates that the value of the variable is important to MCS initiation and creating the associated FP size. For example, the combination of UC500 and UC600 is given as a proxy for wind shear. Positive values of CAPE are important to MCS formation rather than CAPE negatively affecting the MCS formation. Also temperatures at different levels can become important when convection reaches high into the atmosphere, as was discussed earlier.

3.2 *MLR Conclusions*

Some of the variables that are considered to be important to initiation were included in the model runs, while others that have been considered important in previous papers were not included. There were a few noted similarities between the model runs. The most important variable was UC500. Several other variables were used often: UC600, VC200, T300, T200, and VC500. The change between UC500 and UC600 is worth noting again, since wind shear is important to MCS initiation. Examining the data and realizing the signs of the variables were also important since the signs of the variables will affect the value of FP. These variables could be considered major supporters of MCS initiation. No surface data or upper air data were considered to be truly important. There is some error associated with the surface data and upper air data and, if that were to be corrected, it could change the importance associated with those variables.

Also the MLRs run after initiation may not have been needed. These were run to see if there were variables that affected the FP size after initiation had already occurred. This did prove to be the case, as can be seen in Table 1. Looking just at the variables needed for initiation, the MLRs after initiation were pointless since, by that time, initiation has already occurred.

An issue to consider with MLR is that the fit to the data may not be linear. This linear fit was chosen because the similarities between the clusters were much more obvious. Since the R square values are relatively low, this could be an indication that the fit to the data may not be linear. If the fit to the data is not linear, then MLR is not a valid regression model to run on the data.

One issue seen with these MLRs is the question of multicollinearity discussed earlier. Most of the model runs had no multicollinearity issues, which was expected. The variables were expected to be independent enough from each other that the variables were not considered similar, but 2 of the model runs had multicollinearity issues. The model runs 4 hours prior to initiation and 3 hours after initiation had multicollinearity issues. With those model runs, more so than the other runs, multiple variables were removed after being used earlier. With multiple independent variables being removed, this could be the cause of the multicollinearity. These issues are also seen in the models that have some of best fits to the dependent variable and contain the most independent variables. But since these model runs are not the only runs with good fits to the data, it is difficult to accept that this is the reason for the multicollinearity issues seen.

4. PRINCIPAL COMPONENT ANALYSIS

4.1 *PCA Results*

PCA was performed to overcome the multicollinearity issues seen in MLR. Two model runs in the MLRs had these issues. An assumption of PCA is that the variables are independent of each other, the same assumption seen in MLR. In the MLRs, there were two runs with multicollinearity issues, meaning the variables were not considered to be truly independent. However, PCA was still performed under the assumption that the variables are independent. For PCA, there is no dependent variable; therefore, FP is not included in the PCA model runs since its value is not known until after the MCS has dissipated and a dependent variable is not needed. The same variables and values are used for the PCAs that were used for the MLRs.

The PCAs were run for the 6 hours prior to initiation, the initiation hour, and the 3 hours after initiation. A comparison and analysis of the output is done. A note should be made that a comparison of the PCA model runs can be difficult to do since all the variables are included in every model run. The comparison will be made between the variables that have most (90 percent and above) of the associated variance accounted for in the PCAs. It is done this way to see which variables are the most important from hour to hour. The complete list of the data is included at the end of this paper in Table 2.

There was a smaller range of variables used the most in the PCAs than all of the variables in the MLRs. There were multiple variables that were never in the most used list. These variables were STHC, SMXR, UWND, VWND, LCL, and LI from the surface data and upper air data. From the NARR data, the variables not on the most used list were CIN, SRH, SH300, SH200, UC600, UC200, VC600, VC500, VC200, and T200. This is unlike the MLRs since the MLRs included most of the wind components in at least one equation. For the PCAs, most of the wind components were not included in the most used. Again, the issue with the surface data and upper air data is that it is not at the initiation location. If those variables had been recorded at the initiation location, the most used list of variables would have changed.

Two of the variables, GH300 and T500, were included in all of the PCAs. Most of the variance of these two variables were accounted for in every model run. This demonstrates that these two variables play a very prominent role in the formation of these PCAs and MCS initiation. This is unexpected since these variables did not play a major role in the MLRs. T500 could be considered a proxy for stability since it is part of the LI equation. If T500 is a proxy for stability, CAPE and LI should be of importance in the PCAs as well. Using T500 in every equation gives a proxy for stability instead of using the stability variables recorded. A temperature variable closer to the ground was expected to be of more importance (T600 or STHC) than T500. In 8 out of 10 model runs, T600 was used; however, STHC was never included on the most used list. Also T300 was included in 8 out of 10 of the model runs. This is surprising since the data changed very little and was consistently almost the same temperature for each MCS. T200 also did not vary much, but it was not included on the most used list.

All of the geopotential height variables gathered from the NARR data were included in the most used list. In only 2 of the 10 model runs, GH600 was used. This demonstrates that the significance of GH600 is much less than the other geopotential height variables. According to PCA, GH300 is an important variable for MCS initiation as already discussed. Included in 9 out of the 10 model runs were GH200 and GH500; therefore, this shows that these variables are also important and continue to be important to most of the model runs. Also the thickness, taken from the NARR data as well, was used in 8 out of 10 runs. This is not surprising since thickness and GH500 were usually of similar value. It seems that temperature values and geopotential height values are the most important to MCS initiation according to PCA.

Other variables were included in the most used listed but not as often for some, as the ones mentioned above. Included 8 out of 10 times were PW and SH800, indicating importance. PW and SH800 are expected since water is needed in the atmosphere for the MCS to form and rain out precipitation. Specific humidity at 800 mb should be close to a surface value. Since this should be the case, the values for SH800 and SMXR/SMXS should be relatively close if these variables were in the same units. One is listed in kg/kg and the other is listed in g/kg. SMXR is never included in the most used list and SMXS is only included once in the most used list. Since the values between the variables should be similar this is a surprise.

The rest of the variables are included 6 times at most, for example, SH850 is included 6 out of 10 times. This variable should be closer to the surface value than SH800. Since SH850 is included more often than the surface values, the conclusion is made that the values of SMXR/SMXS and SH850 are very different. Other specific humidity variables included were SH600 and SH500 at once each. The importance of the specific humidity decreases as the atmospheric pressure decreases.

CAPE was only included 2 out of 10 times in the most used list. CAPE was a variable that varied widely from MCS to MCS. PCA could not account for most of CAPE's variance in most cases because of the widely fluctuating values. CAPE should be an important variable to MCS initiation, which can be seen more clearly in the MLRs, but that is not the conclusion drawn by the PCA model runs.

While the wind components were used often in the MLRs, those variables were rarely included on the most used list for the PCAs. The only wind variables used were UC500 (4 model runs), UC300 (1 model run), and VC300 (2 model runs). According to the PCAs, the wind at 300 mb is the most important and has the most variance accounted for of any of the wind component variables. The U component of the wind seems to be of more importance than the V component since more of the U components' variance is accounted for than that of the V components.

The loadings of the individual variables in the PCA components varied widely. In MLR, a proxy for wind shear was noticed. In the loadings for the same two variables, throughout all the model runs, the loading signs (positive or negative) were usually the same. So there is not a proxy for wind shear present using UC500 and UC600 in PCA. For the most part, all of the loadings for the wind component variables were the same sign, indicating there was not a proxy for wind shear present in the PCA. If the signs of the wind components had been opposite at different levels, then a possible proxy for wind shear could have been present. There was, however, a possible wind shear proxy between the surface winds and upper levels winds. The surface U component usually loaded into the PCA positively, while the upper air U components usually loaded into the PCA negatively. So the wind shear is over a greater depth of the atmosphere in the PCA than in the MLR. The same could not be said for the surface V component and the upper air V component. Both usually loaded negatively and would not give a proxy for wind shear.

For the upper air temperatures (T600, T500, T300, and T200), approximately half the time the variables were loaded positively and half the time negatively. The same came be same for the surface temperature, STHC, as well. For every first component, the upper air temperature variables are all loaded positively, with the exception of one variable in one of the first components. This means that if the value of the temperature used is negative, then that part of the component is a negative addition.

The upper air specific humidities, for the first component, all loaded positively. Since the values of specific humidities are always positive, then the addition of these variables to the component is always positive as well. The other components vary on how positively or negatively loaded the variables are. Overall, most of the specific humidity variables in most of the components are loaded positively, but there are occasions when the loading is negative.

For the most part, CAPE and CIN were not loaded highly. Approximately half of the loadings were positive loadings and the other half were negative. A good portion of the CAPE and CIN loadings were very small. This shows that these variables are not as important to MCS initiation since the loadings are so small. For the first component, PW is consistently loaded highly. For the rest of the components, PW is not loaded highly. PW should be important and related to the amount of precipitation that will fall over the lifetime of the MCS.

The variables loaded consistently high are the ones that have most of the variance accounted for in the components. These variables usually loaded the highest in the first component, then slowing decreased in loading in the components after that. This really is no surprise. The most used variables were expected to load highly.

In general, PCA contained fewer variables on its most used list than MLR included in its entirety. PCA is more accurate since every variable is used for each component. It is more difficult with PCA to compare model run to model run, but it gives a more accurate representation of the data than MLR. This more accurate representation also gives a better depiction of the variables required for MCS initiation.

4.2 *PCA Conclusions*

The loadings of the individual variables can give an idea of how these variables affect MCS initiation with the PCA. There seems to be a proxy for wind shear, mainly the U components, between the surface winds and the upper level winds. There is not much of a proxy for wind shear seen in the V components. Patterns are more difficult to recognize in PCA because of how PCA works. Looking for shear in the components is difficult since the signs of the winds change from component to component. Wind shear was much more obvious is the MLRs. The specific combinations given for each component of each PCA gives the best fit to the data. The combinations are important but difficult to attain given the nature of PCA.

An issue with PCA is that it is difficult to compare the model runs to each other since each variable is used in every model run. However, what can be compared is the amount of variance that is extracted from each variable for each PCA. After separating the variables into most used (90% or more of the variance is accounted for) and least used, trends are noticed. Certain variables are used multiple times in the PCAs. There are some variables that are never in the most used list.

Other variables are used in either every/almost every PCA model run. These variables are more prominent and are believed to be a major supporter of MCS initiation. The conclusion is that while some variables are important throughout the time frame examined, certain variables are only important at certain times in the time frame. The changes and values in these certain variables could mean the difference between MCS initiation or not.

5. MLR VERSUS PCA

There were more differences between the MLR and PCA than there were similarities. For example, the most used variable in MLR, UC500, was not in the most used list for any of the PCAs. So UC500 held a much more prominent position in the MLR runs than in the PCAs. Also one of the most used variables in the PCAs, T500, was never included in any of the MLRs. These differences are due to how each analysis is set up and run.

MLR creates issues that are corrected by the PCA, as discussed earlier. MLR uses a wider range of variables than is seen in the PCA's most used. Overall, PCA uses all the variables where MLR considers each variable for the model but does not include every variable.

The proxy for wind shear was much more noticeable in the MLR than in the PCA. The wind shear in MLR was seen between UC500 and UC600. The wind shear in PCA was seen between the U component surface winds and the U component upper air winds.

Temperatures were also important in MLR and PCA. The upper level temperatures affecting the MCS initiation show that the depth of convection reaches high into the atmosphere and generates ice particles which fall into the lower atmosphere. This affects the outflow and creates new cells, furthering the development of the MCS.

A proxy for stability was used in the PCAs. There is not a proxy for stability used in the MLRs. This proxy was T500. It is considered a proxy for stability since T500 is included in the LI equation. Stability did not have a very prominent role in the MLRs but it did have a prominent role in the PCAs.

Certain variables appear much more often in the PCA's most used list than those same variables appear in the MLR's list of variables used. For example, PCA includes the geopotential height variables more often than MLR does in the model runs. The wind components are more prominent in the MLR model runs, while the geopotential heights and temperatures are more prominent in the PCA model runs.

6. SUMMARY

Depending on the type of analysis used, the supporters for MCS initiation change. Also the variables needed for MCS initiation change depending on the time frame before and after initiation. These analyses show that the conditions needed to produce an MCS are numerous. The conditions needed to produce the MCSs change depending on the time frame before and after initiation of the first storm.

MCS formation is favored in the Elk Mountain, New Mexico, region when the upper atmospheric temperatures are low (-30 ˚C to -50 ˚C). For the most part, T200 gives a negative addition to the equations and T300 depends on its placement and loadings into the equations. Wind shear needs to be present. In most of the model runs, proxies for wind shear were used by using components of the wind of opposite signs in different levels. The levels used for this wind shear proxy depended on the type of analysis being performed.

CAPE and CIN were not as important to MCS as was expected. In the MLRs, these variables were rarely used. CAPE added to the equation, while CIN subtracted from the equation. In the PCAs, it was similar, except the loadings were positive or negative with both, never giving an exact showing of how the variables were used. The rest of the variables used in the MLRs and PCAs were either not used in the final equations (MLR) or presented such a wide range of information that nothing definitive could be gleamed from the outputted information.

PCA was a better fit to the data since it took into account each variable on each component. MLR did not include every variable and there is no proof that the fit to the data to create that FP size is linear. While the wind shear proxies were very noticeable in the MLR runs, the proxies were also evident in the loadings given in the PCAs. These wind shears are important to the creation of MCSs. The wind shear needs to be just right for an MCS to form. If the wind shear is too strong, a different type of system will form and it will not become an MCS. Also upper air temperature loadings were relatively high for each PCA. MLR only used a fraction of the possible temperature variables. These temperature variables help show that the depth of the convection in the atmosphere spans high into atmosphere, as discussed earlier.

One improvement in the analysis that could be made is to choose a different type of analysis (possibly logistic regression) rather than MLR. After that analysis is run a comparison between PCA and the new analysis could occur. It is possible that the results given in the new analysis and the PCA would be more similar than the compared results between MLR and PCA. Another possible improvement could be to include variables that were not considered here and include even earlier model runs than just the two previous ones. Other variables include possible wind shears over different depths of the atmosphere for both the U component of the wind and the V component of the wind. Another improvement could be to look at the exact wind direction for each component rather than just looking at the wind speed in the U and V components.

REFERENCES

ARM: Climate Research Facility, cited 2011: ARM XDC Datastreams. [http://www.arm.gov/xdc/xds/abrfc].

Banta, R.M., and C.B. Schaaf, 1986: Thunderstorm genesis zones in the Colorado Rocky Mountains as determined by traceback of geosynchronous satellite images. *Mon. Wea. Rev.,* **115,** 463-476.

DesJardins, M., and R. Petersen, 1986: GEMPAK: Applications in Research and Education. Preprints, *Second International Conference on Interactive Information and Processing Systems for Meteorology, Oceanography, and Hydrology,* Miami, FL, American Meteorological Society.

Murray, D., 2003: The Integrated Data Viewer – a web-enabled application for scientific analysis and visualization. *19th International Conference on Interactive Information Processing Systems*, Long Beach, CA, American Meteorological Society.

Tripoli, G.J., and W.R. Cotton, 1988: Numerical study of an observed orogenic mesoscale convective system. Part I: simulated genesis and comparison with observations. *Mon. Wea. Rev.,* **117,** 273-304.

Tucker, D.F., and N.A. Crook, 1998: The generation of a mesoscale convective system from mountain convection. *Mon. Wea. Rev.,* **127,** 1259-1273.

-------, D.F., and N.A. Crook, 2005: Flow over heated terrain. Part II: generation of convective precipitation. *Mon. Wea. Rev.*, **133,** 2565-2582.

-------, D.F., and X. Li, 2009: Characteristics of warm season precipitating storms in the Arkansas-Red River basin. *J. Geophys. Res.*, **114,** D13108, doi:10.1029/2008JD011093.

United States Geological Survey, 1983: *Santa Fe, New Mexico, 30 x 60 Minute Quadrangle*. United States Department of the Interior.

Young, C.B., A.A. Bradley, W.F. Krajewski, and A. Kruger, 2000: Evaluating NEXRAD multisensory precipitation estimates for operational hydrologic forecasting. *J. Hydrometeor.*, **1,** 241-254.

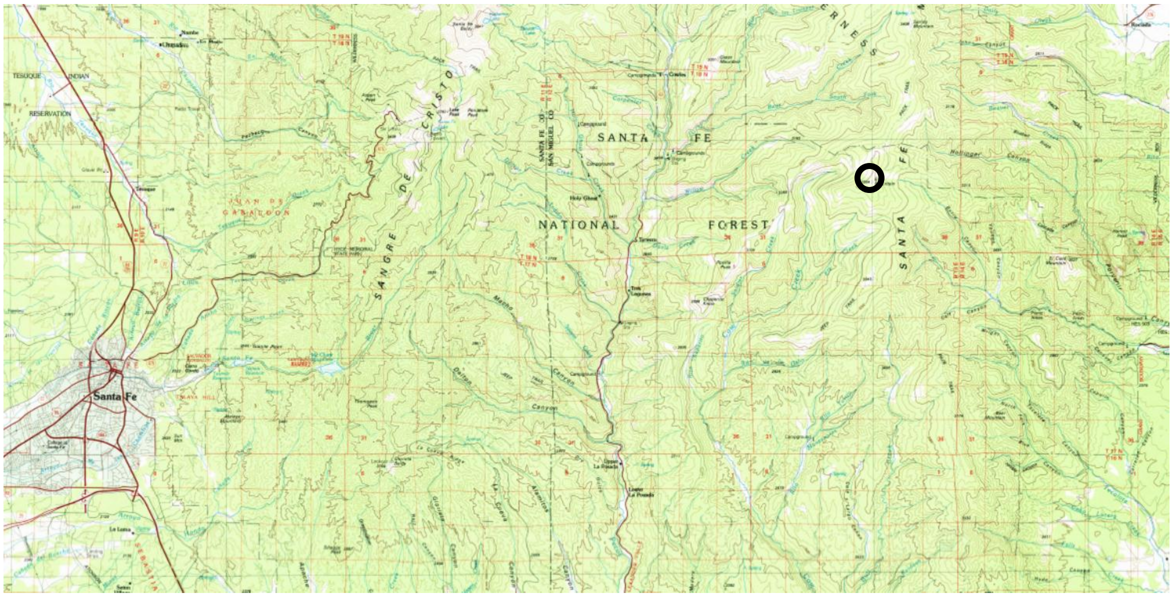
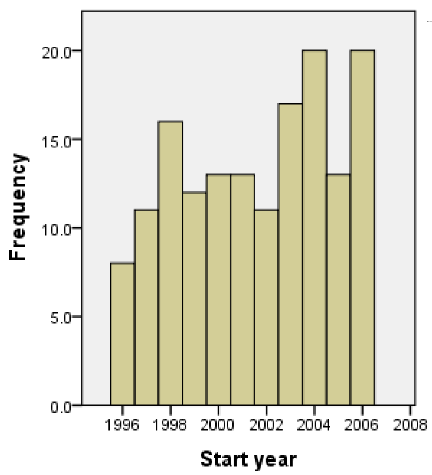
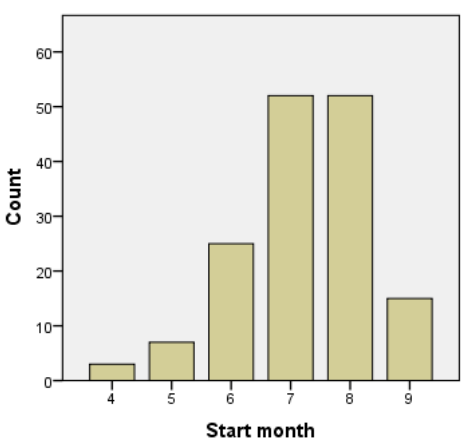


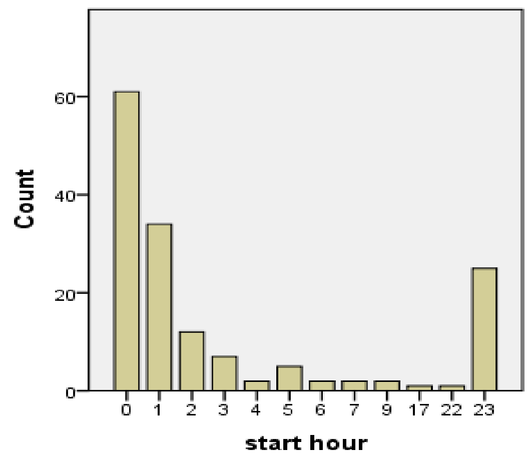
Figure 1: Portion of the Santa Fe, New Mexico, 30 x 60 Minute Quadrangle (United States Geological Survey, 1983). The black circle highlights the location of Elk Mountain. Santa Fe, New Mexico is included in the map to give a point of reference.



(a)



(b)



(c)

Figure 2: Frequency was used on the y-axis for the occasions where the frequency was low overall. Count was on the x-axis for the occasions where the frequency was high overall. (a) A histogram detailing the distribution of the start years MCS initiation occurred. (b) A histogram detailing the distribution of the start months MCS initiation occurred. (c) A histogram detailing the distribution of the start hours MCS initiation occurred.

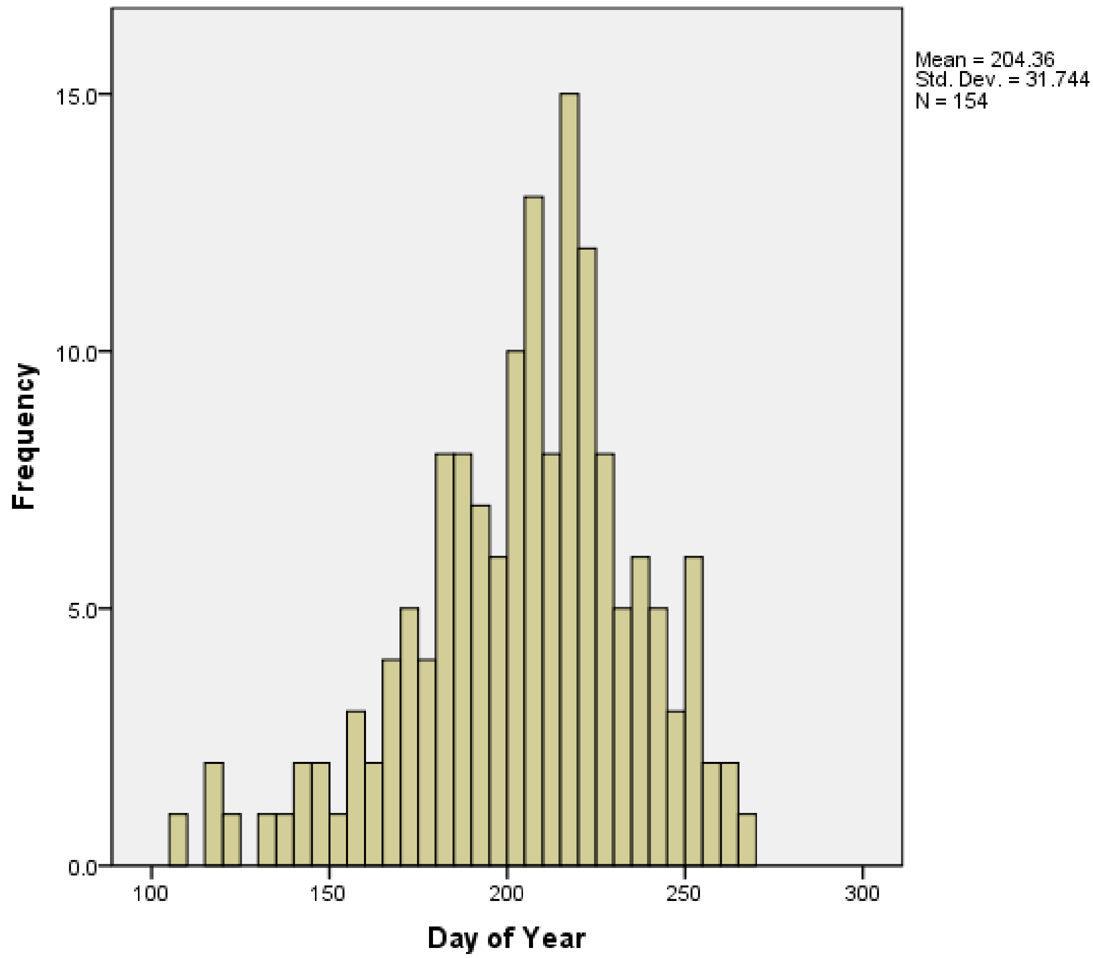


Figure 3: A histogram of the day of year the MCS initiations occur.

Table 1: Results of the MLRs run on the Elk Mountain, NM cluster. Model run gives the hour the approximate equation is used and the R square value is the resulting fit of the approximate equation to the given data.

Model Run	Approximate Equation	R square
6 hours prior	FP ≈ – GH200 – UC600 + UC500 + SH200	0.458
5 hours prior	FP ≈ VC200 + LCL + CIN	0.268
4 hours prior	FP ≈ CAPE + LI + SH300 – UWND – SH500 + T300 – GH200 – STHC + GH500 – SRH	0.862
3 hours prior	FP ≈ T200 + VC500 – T300 + SH500 – UC600 + UC500 – UC300 – SMXR + VC200	0.453
2 hours prior	FP ≈ VC200 + LCL	0.266
1 hour prior	FP ≈ – UWND –GH600 + SH600 – T200 + UC500	0.577
Initiation	FP ≈ T200 + SH200 – T300 + VC500 – UC200 + SH850 + UC500 – UC600	0.389
1 hour after	FP ≈ VC200 – UC200 + LCL + UC500 – VC500	0.385
2 hours after	FP ≈ UC500 + LI – VC500 + CAPE + SH200 – STHC + UC300 + VC600	0.746
3 hours after	FP ≈ T200 – STHC – UC600 + UC500 + Thickness – T300 – T600 + SRH – LI – VWND + GH200 – GH300	0.753

Table 2: Results of the PCAs run on the Elk Mountain, NM cluster. For each model run, the most used variables (over 90% of variance accounted for) are listed along with the number of components with eigenvalue greater than one and the resulting accounted for variance.

Model Run	Most Used Variables	# of components with λ > 1	Accounted for Variance
6 hours prior	Thickness, GH500, GH300, T600, T500, T 300	9	81.102%
5 hours prior	SMXS, Thickness, PW, GH600, GH500, GH300, GH200, SH850, SH800, T600, T500, T300	7	82.133%
4 hours prior	Thickness, PW, CAPE, GH500, GH300, GH200, SH800, SH500, UC300, T600, T500, T300	9	86.754%
3 hours prior	Thickness, PW, CAPE, GH500, GH300, GH200, SH800, UC500, VC300, T600, T500, T300	9	82.463%
2 hours prior	Thickness, PW, GH500, GH300, GH200, SH850, SH800, UC500, T600, T500, T300	7	82.574%
1 hour prior	PW, GH600, GH500, GH300, GH200, SH850, SH800, SH600, UC500, T600, T500, T300	9	86.023%
Initiation	PW, GH300, GH200, SH850, SH800, VC300, T500	8	80.203%
1 hour after	Thickness, PW, GH500, GH300, GH200, SH850, SH800, T600, T500, T300	7	82.884%
2 hours after	Thickness, PW, GH500, GH300, GH200, SH850, SH800, UC500, T600, T500, T300	8	82.998%
3 hours after	Thickness, GH500, GH300, GH200, T500	8	77.979%