Analyzing Statistical Models of Hourly Precipitation Events

Jennifer Esker
Oklahoma Weather Center Research Experience for Undergraduates
NOAA/National Severe Storms Laboratory, Norman, Oklahoma
Southern Illinois University Edwardsville, Edwardsville, Illinois

Mentors

Harold Brooks
NOAA/National Severe Storms Laboratory, Norman, Oklahoma

Michael Baldwin
Cooperative Institute for Mesoscale Meteorological Studies, University of Oklahoma and
NOAA/National Severe Storms Laboratory, Norman, Oklahoma

1 August 2003

Corresponding author address: 711 E. Broadway, Steeleville, IL 62288
Email: jesker@siue.edu

Abstract

Understanding the national precipitation distribution can be useful in many fields of study, but finding those patterns is not easy. Overwhelming amounts of data create roadblocks for detailed analysis, but constructing statistical models can reduce the mount of data needed. This study applied gamma distributions to a year's worth of processed hourly precipitation data to examine the national precipitation. The set represents all precipitation events of the contiguous United States as elliptical objects and produced a precipitation regime classification based on the gamma parameters assigned to the precipitation within the object. Starting with a general model of the national precipitation the analysis continues to categorize the data by location, season and precipitation regime to produce detailed relationships. Examining plots of the gamma parameters also provides insights into the variability of these categories and additionally confirms that these models present an accurate representation of annual precipitation.

I. Introduction

Understanding annual precipitation distributions is useful for all types of fields—forecasting, public planning, farmers, flood areas, etc.—but doing this proposes many difficult problems. The amount of data for every hourly precipitation event is overwhelming, so it is necessary to develop methods of describing the data and still producing useful information. This study presents the results of a gamma distribution model. Using a processed set of the National Center for Environmental Prediction Stage IV data, patterns are identified and analyzed in different regions of the nation during different times of the year, with different precipitation regimes and then employ other plots to further describe the variability of these trends.

II. Data

The data set comprises the hourly precipitation from 2002 from the 48 contiguous states in the form of precipitation "objects" from the NCEP Stage IV data processed by Baldwin (2003), hereafter B03. The Stage IV data, a combination of radar images and rain gauge measurements, is separated into objects of rain and for each object B03 interprets the implied precipitation into two parameters, α and β, creates a regime classification, and simplifies all other information into a small set of parameters (Baldwin 2003).

To construct this set, B03 modeled each Stage IV precipitation object as an ellipse. The amount of rain at points within each object is patterned with a gamma distribution given by:

$$f(x; \alpha, \beta) = (x/\beta)^{\alpha - 1} [\exp(-x / \beta)] [\beta \Gamma(\alpha)^{-1}, x >= 0, \alpha, \beta > 0$$

where x is an amount of rain, α is the shape parameter, and β is the scale parameter (Wilks 1995). A gamma distribution promotes itself as a reasonable choice for precipitation because it is nonnegative, positively skewed, and compares well with other distributions used for precipitation (Wilks 1989). In order to use the gamma distribution, B03 used the method of moments where the mean and variance of the object were employed to produce the α and β parameters. With these values, the distribution produces a curve that with larger values of β creates a heavier tail with increasing probability of extreme rain events.

Once the precipitation is modeled, the objected is labeled with parameters describing the location and size. B03 created a 4 km by 4 km grid over the contiguous United States and each 4 km by 4 km box is equivalent to 1 point. The program calculated the number of points included in an object and output it along with the date and time, an x-y coordinate for the grid, a latitude/longitude coordinate, and several other parameters that describe the ellipse.

The α, β and the classification made by the program were the key parameters for analyzing the patterns of precipitation while the other parameters created boundaries for sorting. B03 used the object's points to analyze the objects by size in three categories: small, medium and large. Reanalyzing those groups generated better boundaries between the sizes. The regression lines in Figure 1 represent the parts of the curve where the points of an object have a significant relationship. Examining only the third and fourth groups includes objects with greater than about 700 points, and even though this consisted of about 10% of the total data possible, it also provided 92% of the total precipitation for the United States and allowed for a more manageable set of points.

III. Methodology

For a complete analysis, the data was split into categories that could represent the precipitation better than one national plot. First, seven regions were constructed. The Southeast region was given the most importance due to the annual extreme amounts of rainfall (Brooks and Stensrud 1999), but then all other regions were created along major latitude/longitudinal lines. Figure 2 displays the regions with boundaries and in Table 1 presents the areas of each region. Only results from the Southeast (SE), North Central (NC), Central Plains (P) and West Coast (W) will be shown here because they display the most distinct differences over the course of the year.

Next, models had to be created to display the most possible information and this led to the reapplication of the gamma distribution. The shape and scale parameters from each object were placed into the distribution with a specific threshold. This produced the fraction of the object's implied precipitation for thresholds equal to and above that specified amount. This fraction is then directed to the points that constitute the object and the result is the number of points within an object that produce over the threshold amount of precipitation. This number is summed for all objects and plotted with the threshold producing a curve that describes the probability of large precipitation events. Figure 3 is the curve for the national precipitation. Each point on the curve is a threshold that was actually analyzed and the vertical lines mark particular thresholds. The dark line marks the 1mm threshold where 17.8% of the points can be found with precipitation totaling 1mm or above. On the other hand, the lighter line marks the 100mm threshold where about 1 in every 5 million points will have at least 100mm of precipitation.

IV. Results

The national plot in Figure 3 is the overall representation of the contiguous United States precipitation. The figure provides an original idea of the distribution of precipitation amounts, but the regional composition of this curve is better explained by Figure 4. Here the national plot acts like an average of these regions. As expected, SE has a large probability of extreme events, but P, also above the national average, jets out to have the largest probability of precipitation events above 120mm. On the other extreme, NC dips below the national curve at larger thresholds and W is drastically below the other curves, the first indication that one general model of national precipitation is not accurate.

By continuing to categorize the data into seasonal outlooks displayed in Figure 5, information about the behavior of the regional curves becomes visible. The extreme regions, SE and W, keep their own similar patterns throughout the year. SE varies more in the probabilities of lower thresholds than on the tail, but in the summer still hits its highest probabilities, while W stays in the lower thresholds with higher probabilities and makes one significant increase in the summer. Both NC and P make regular fluctuations during the year. The NC curve increases and decreases in probability and thresholds as temperatures rise and cool through the seasons. P has a different style. Large thresholds stay highly probable through the warmest seasons, especially in the spring and fall, but still make a big drop in the winter displaying vulnerability to seasonality at times when both regions have less moisture in the air.

Thus far, each breakdown of the data provided more information about the national precipitation, but these tell us nothing about the parameters, α and β. If α and β are scattered to different extremes, then these curves could be miss representing these categories. Plotting the parameters in Figure 6 verifies the variability of the seasonal patterns seen previously. Each plot has a unique distribution, but the patterns from the curves in Figure 5 are distinctly recognizable. SE has a slightly high probability in the summer than in the winter, but the two are very similar. Also, referring back to Table 1, there is a large difference between the number of objects in the summer and winter. P is very similar to SE. Both plots exemplify the same domain and range, but the P plot displays a bigger distinction between the two seasons as shown in the previous figure. NC follows very closely behind P only with lower values while W has a defined plot that is in direct relation to the small differences between W winters and W summers. The previous curves show that the two seasons have different probabilities for fairly low thresholds. The idea verifies the curve while also displaying a relationship that might be described by more specific plots.

The next categorization of the data was made by precipitation regime. Figure 7 is a cumulative distribution plot of the convective objects throughout the year for each region and Figure 8 displays curves for the stratiform objects. Compared to the curves in Figure 5, the plots in Figure 7 are almost the exact curve except shifted toward higher probabilities, with the exception of W making the most change in curvature and most noticeable shift. This is an expected happening when comparing the figure with Table 1. The totals for SE, NC and P are dominated by convective objects while W precipitation falls mostly in stratiform. These numbers quantify the changes in curves from Figure 5 to Figures 7 and 8. The SE, NC and P curve movement happened at the lower thresholds while the convective precipitation numbers increased at large thresholds.

The convective plots also produced the more interesting regional curve comparisons. Figure 7 reveals the linear jabs from the P curve over the SE curve. When plotted as convective objects, the spring curve SE intersection decreases from 50 mm to 30 mm and the fall intersection occurs at 20 mm instead 70 mm. In conjunction with Figure 8, Table 1 reflects the increase of convective objects with the increase of temperature throughout the year for SE, NC and P and reveals the convective maximum for W occurs in the winter.

The stratiform plots are very reflective of uniform precipitation. The plots all have the same curvature and they all stay within the same thresholds, but each region goes through a seasonal stratiform pattern. The SE curve remains the same for most of the year except for a slight shift into higher probabilities from winter to spring followed by a slight decrease from spring to summer. NC stratiform precipitation becomes more frequent from winter to spring and remains the same through the summer and decreases drastically for the fall and winter. P increase probability from winter to spring and spring to summer where it remains the same though fall and decreases into winter. W has the highest probability of stratiform precipitation in the winter and then it decreases for the spring and slightly increases to summer and through fall.

The final figures display the variability between P and NC with different regimes. Figure 9 is an $\alpha-\beta$ plot of the NC winter and summer stratiform curves discussed previously and plotted in Figure 8. The threshold cutoff of the stratiform classification is obvious as well as the distinction between the seasons, the summer plot exhibits the moisture available during that season. The seasonal convective differences in the winter and summer of P are displayed in Figure 10 are much larger. The extreme threshold values reached in the summer extend pass the limitations of the stratiform plot, but the winter is not as defined stratiform winter. The figures reiterate the convective drive of the seasons and analyzing with the numbers from Table 1 quantifies and emphasizes one more time, the steady formation of stratiform precipitation throughout the year and the decrease of convective objects in the winter.

V. Conclusion

The results of this analysis show explicit trends of national precipitation distributions. Breaking the data into groups based on size and location and examining this data with the threshold curves showed the lower probability for large precipitation events in W. Then, this information is expanded by looking at the regions through time. Figure 5 displays the strong seasonality of NC and P, and the α-β plots further illustrate this difference and reiterate the patterns of the W and SE. The precipitation regime plots explained these common patterns. While W is annually dominated by stratiform precipitation and SE is annually dominated by convective precipitation, NC and P shifts are based on the convective decrease due to the lack of moisture in the winter. Once again, the specific α-β plots detail the variability of these regions.

These findings confirmed the gamma models as representations of national precipitation and provided leads for future possibilities. While more in depth measurements are always possible and can bring more patterns of national precipitation to the forefront, interannual variability is also an interesting direction for the analysis. Finally, understanding these trends can aid those with an interest or a need for understanding precipitation and the results from the models can be instrumented as input for other precipitation models or be based as truth for testing numerical weather prediction output.

Acknowledgments. I would like to thank the National Science Foundation for sponsoring this REU program with Grant no. 0097651, as well as everyone at NSSL, OU, NWS and SPC who make this program possible. I would also like to especially thank Harold Brooks for his guidance this summer and Michael Baldwin for the use of the data set and his assistance with this project.

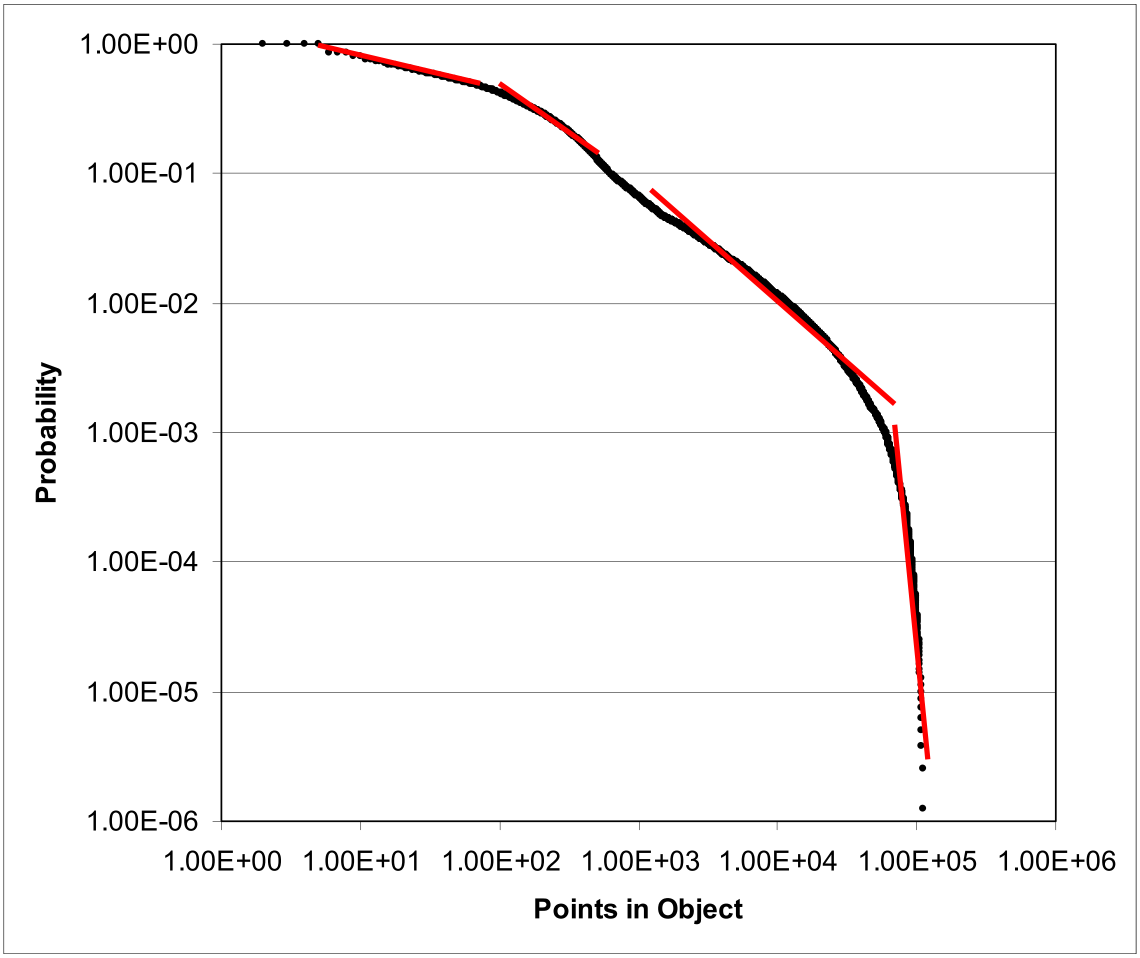
References

Baldwin, Michael E, 2003: *Automated Classification of Rainfall Systems Using Statistical Characterization*, PhD. Thesis, University of Oklahoma, 28, 64, 84, 101-133.

Brooks, Harold E. and David J. Stensrud, 1999: Climatology of Heavy Rain Events in the United States from Hourly Precipitation Observations, *Monthly Weather Review*, **128**, 1194-1201.

Wilks, Daniel S, 1993: Comparison of Three-Parameter Probability Distribution for Representing Annual Extreme and Partial Duration Precipitation Series, *Water Resources Research*, **29**, 3543-3549.

Wilks, Daniel S., 1995: *Statisical Methods in the Atmospheric Sciences*, Academic Press, 86 pp.


1.00E+00
1.00E-01
1.00E-02
1.00E-03
1.00E-04
1.00E-05
1.00E-06
Probability
1.00E+00
1.00E+01
1.00E+02
1.00E+03
1.00E+04
1.00E+05
1.00E+06
Points in Object

	SE	NC	P	W
Region Area	895376.35	713673.2	1580111	1336588
OBJECTS				
Winter	2020	609	747	4105
conv	1145	118	142	412
strat	875	491	605	3693
Spring	2398	1342	1619	3931
conv	1552	532	806	306
strat	846	810	813	3625
Summer	6979	2028	3079	1609
conv	5616	1464	2385	325
strat	1363	564	694	1284
Fall	4707	1474	1896	1813
conv	3374	619	895	243
strat	1333	855	1001	1570
TOTAL	16104	5453	7341	11458
Total-conv	11687	2733	4228	1286
Total-strat	4417	2720	3113	10172
Total/km^2	0.01798573	0.007641	0.004646	0.008573
conv/km^2	0.01305261	0.003829	0.002676	0.000962
strat/km^2	0.00493312	0.003811	0.00197	0.00761
POINTS				
Winter	19487795	1659677	2470901	23431240
conv	14710247	265145	658994	5826625
strat	4777548	1394532	1811907	17604615
Spring	13383128	6606853	4905427	14185295
conv	11815657	3542410	2908898	1124676
strat	1567471	3064443	1996529	13060619
Summer	35191072	10347788	11305880	3825621
conv	33043331	8964306	10008079	873132
strat	2147741	1383482	1297801	2952489
Fall	34107896	5616604	8682288	8397490
conv	30392715	3138389	5735166	1334057
strat	3715181	2478215	2947122	7063433
TOTAL	102169891	24230922	27364496	49839646
Total-conv	89961950	15910250	19311137	9158490
Total-strat	12207941	8320672	8053359	40681156
Total/km^2	114.10832	33.95241	17.31809	37.28873
conv/km^2	100.473896	22.29347	12.22138	6.852144
strat/km^2	13.6344242	11.65894	5.096706	30.43658

Table 1. A tally of the specific number of objects and points analyzed along with the area for each region.

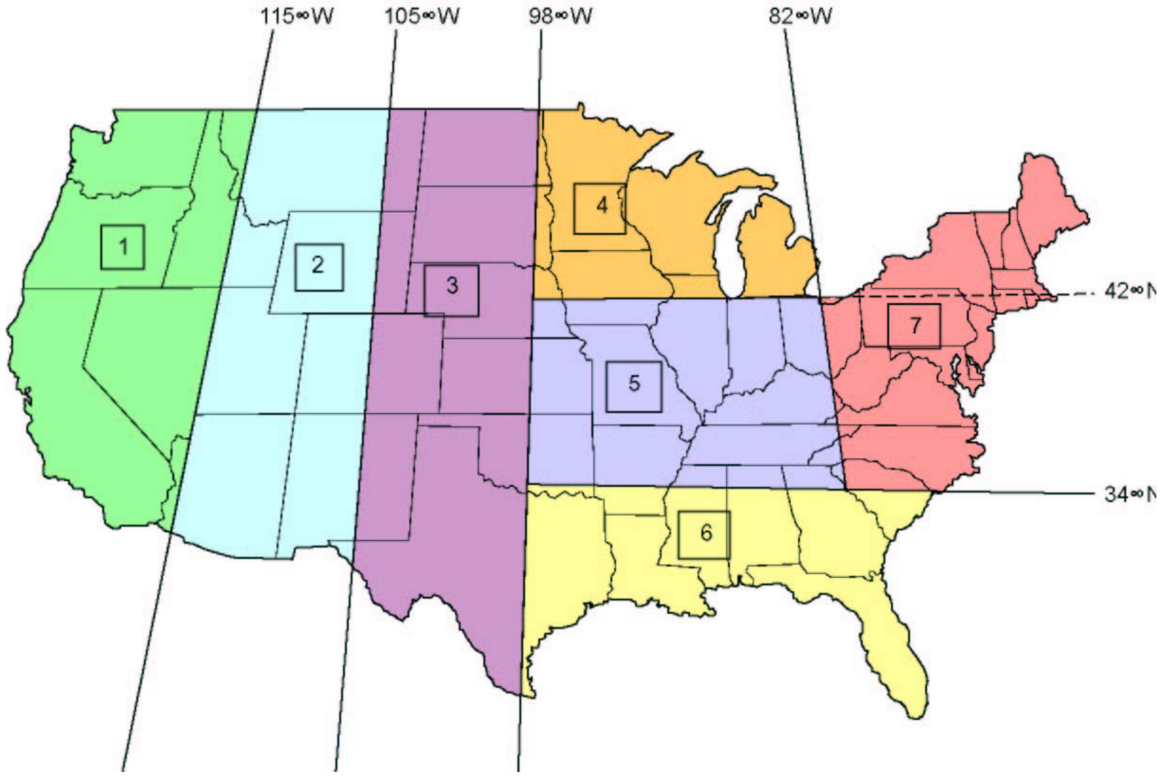


Fig. 2. The regional divisions of the data. 1—West (W); 2—West Central (WC); 3—Plains (P); 4—North Central (NC); 5—Midwest; 6—Southeast (SE); 7—Northeast.

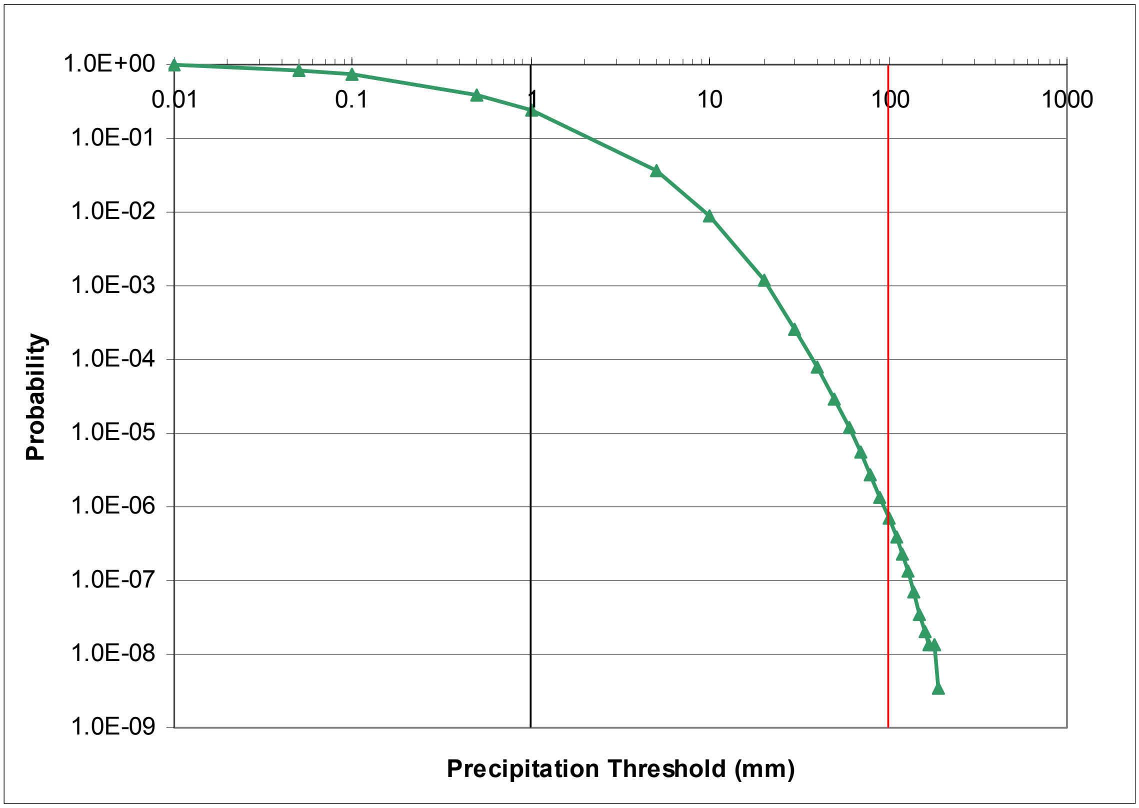


Fig. 3 The national hourly precipitation displayed as a function of a threshold of rain verses the probability that a point in an object has over that threshold. Each triangle on the curve represents an analyzed threshold amount. The vertical lines provide specific examples of the information given by the curve. The darker line displays 17.8% of the points have more than 1mm of precipitation, while the lighter line expresses that 1 in every 5 million points have at least 100 mm of precipitation.

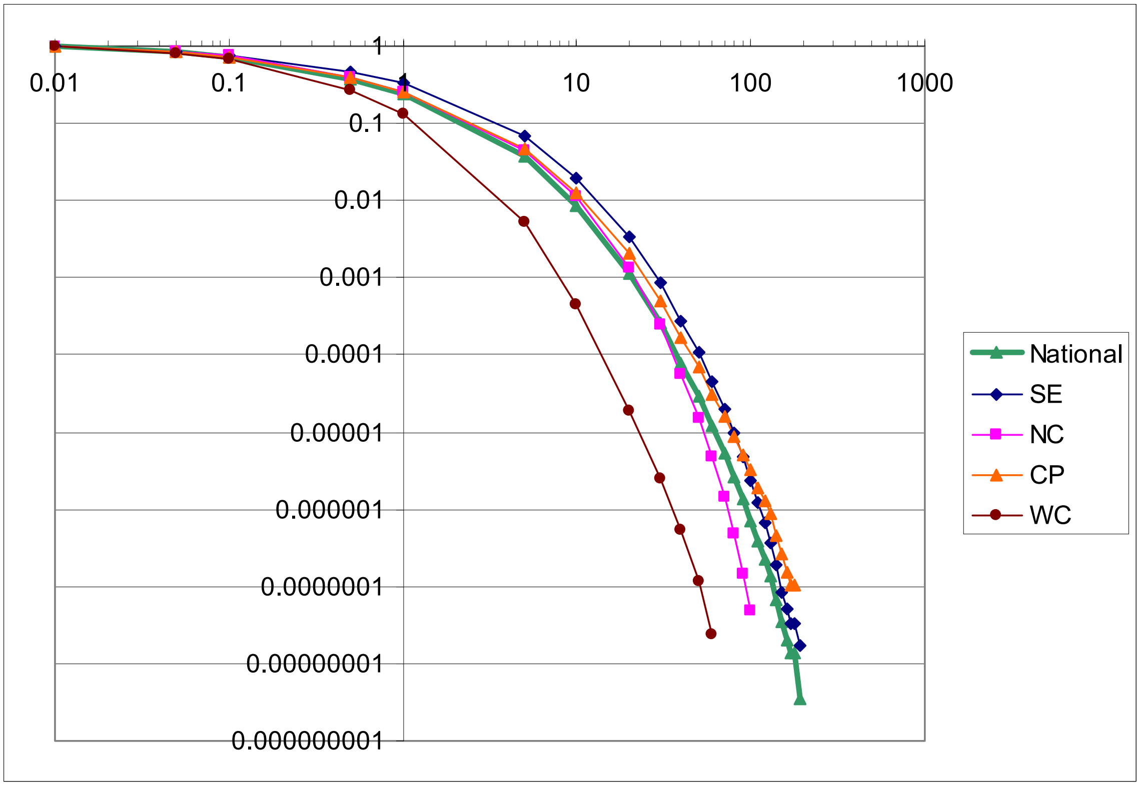
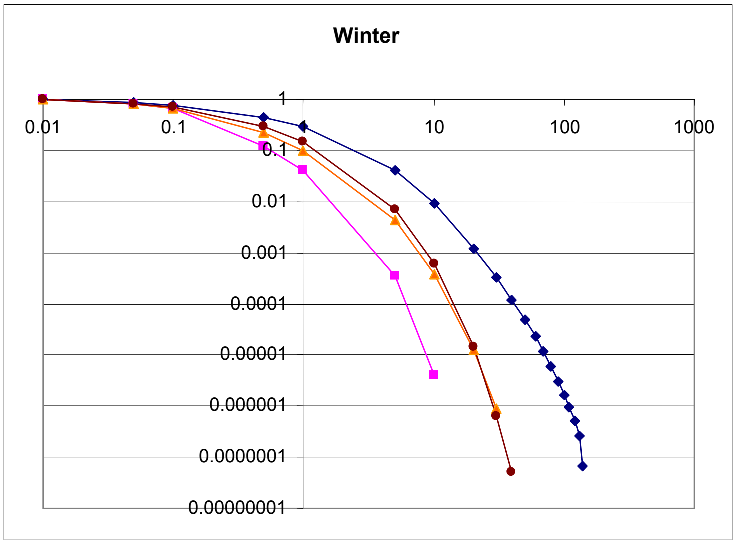
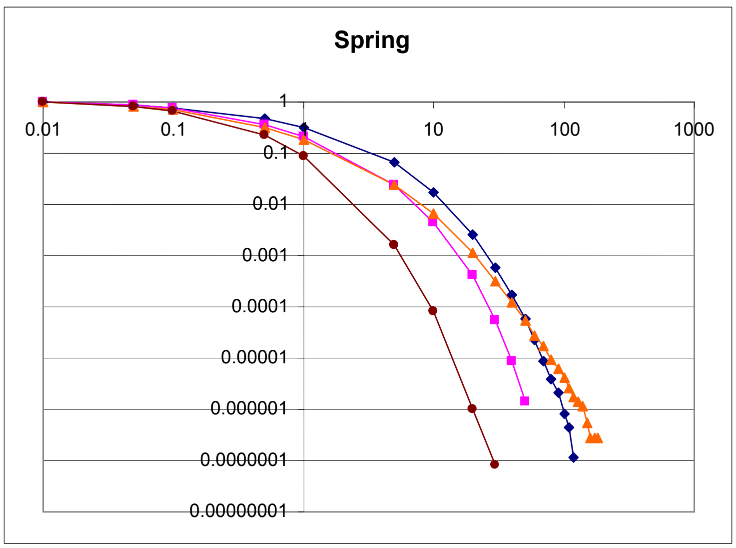
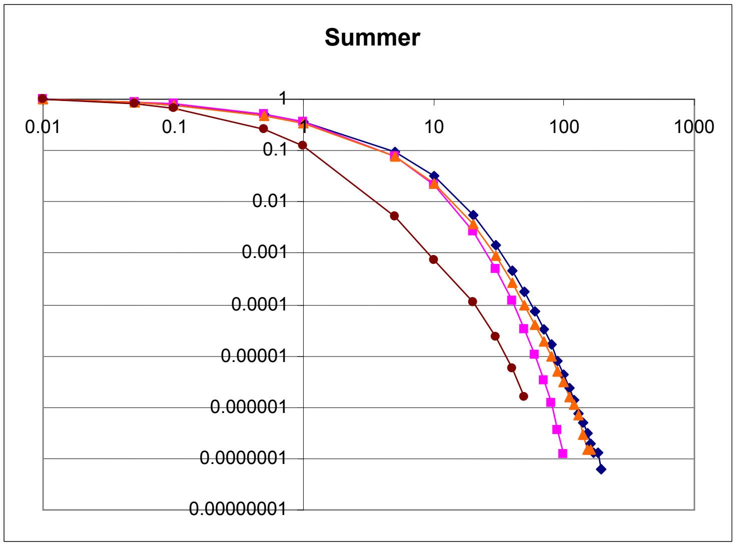


Fig. 4. The same as Figure 3, but shown with regional plots. The heavy line with triangles still represents the national precipitation, the line with circles represents the West (W), the line with squares represents the North Central (NC), the line with diamonds represents the Southeast and the smaller line with triangles represents the Plains (P).


Winter
1
0.1
0.01
0.001
0.0001
0.00001
0.000001
0.0000001
0.00000001
0.01
0.1
10
100
1000


Spring
1
0.1
0.01
0.001
0.0001
0.00001
0.000001
0.0000001
0.00000001
0.01
0.1
10
100
1000



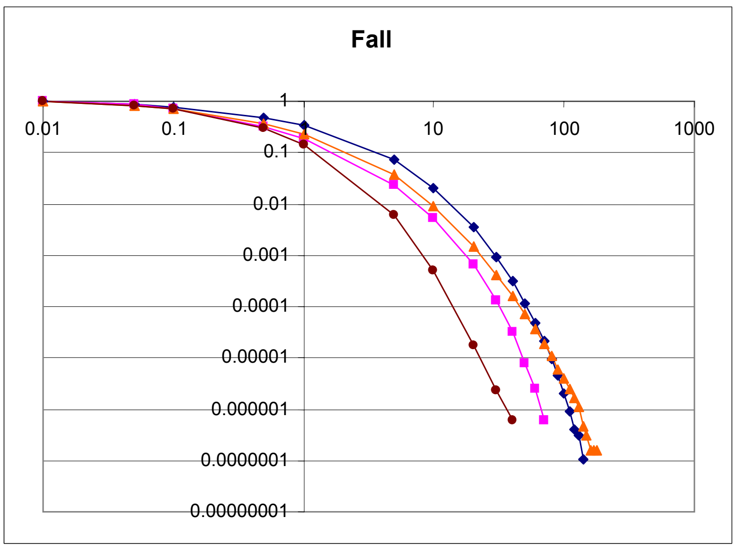
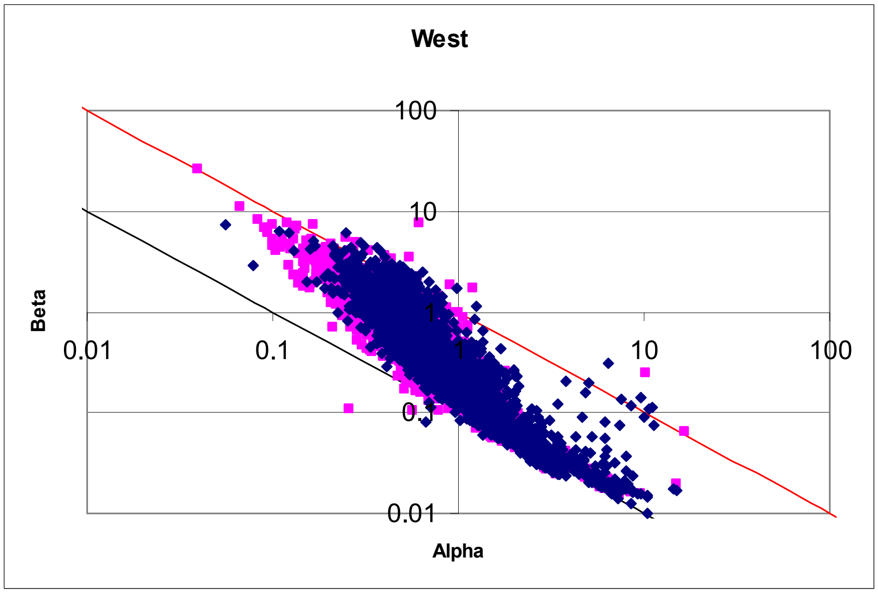
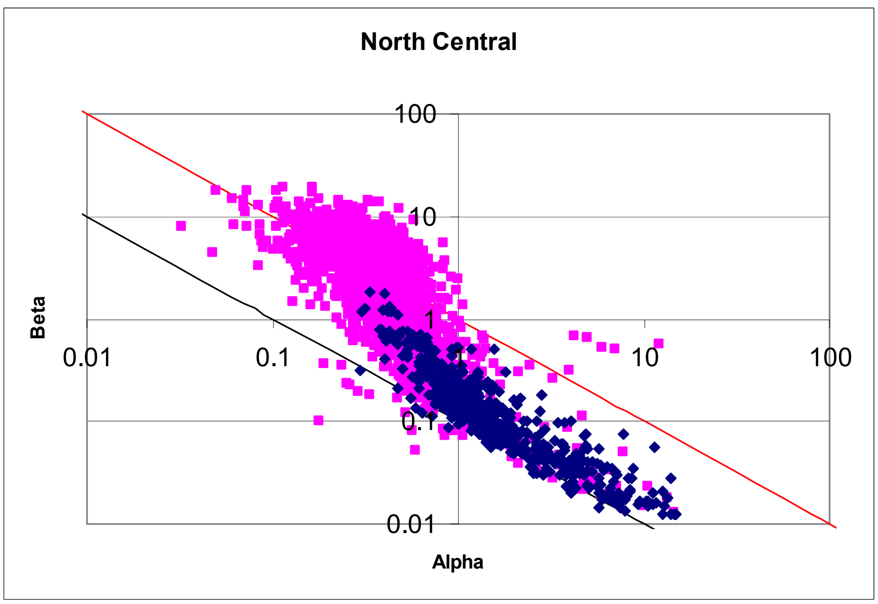


Fig. 5. Seasonal depictions of the region curves in Figure 4 without the national curve. The curve descriptions remain the same.


West
100
10
1
0.1
0.01
0.01
0.1
10
100
Beta
Alpha


North Central
100
10
1
0.1
0.01
0.01
0.1
10
100
Beta
Alpha

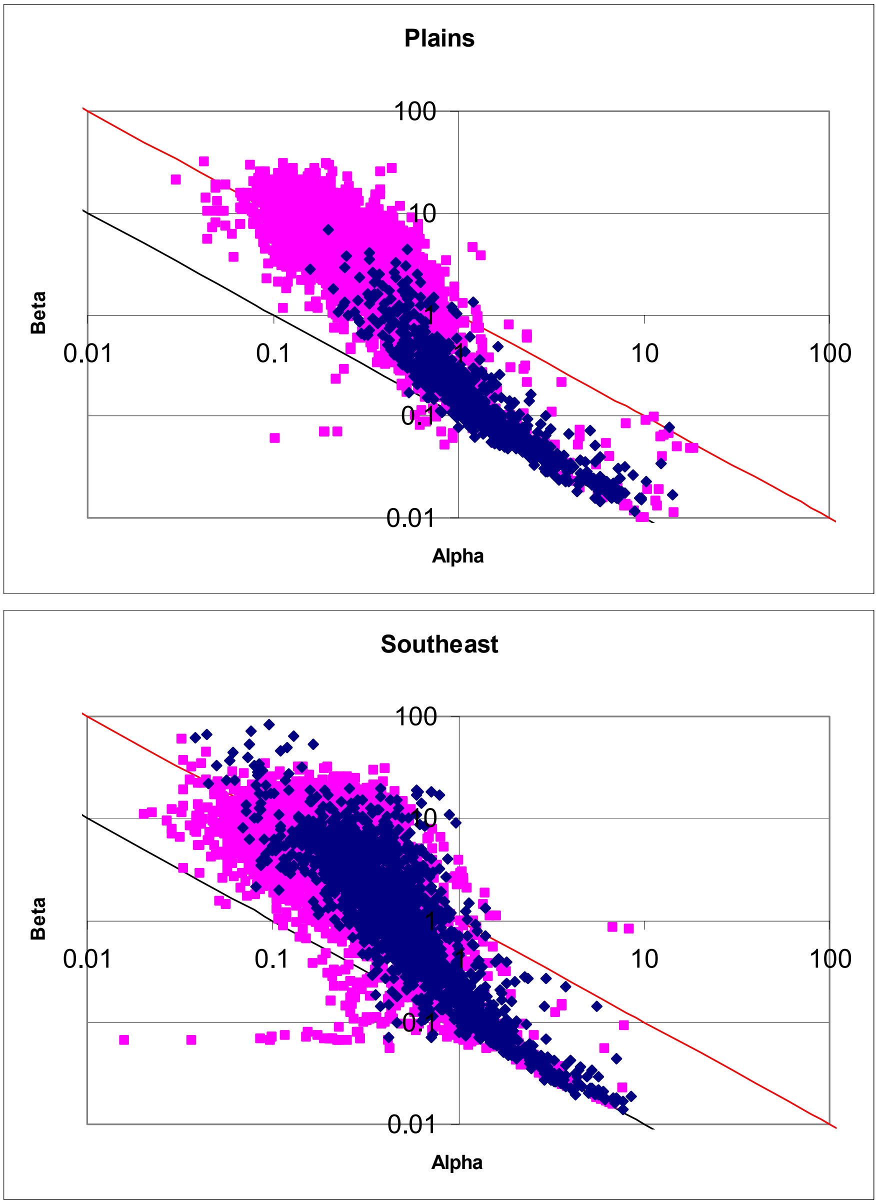
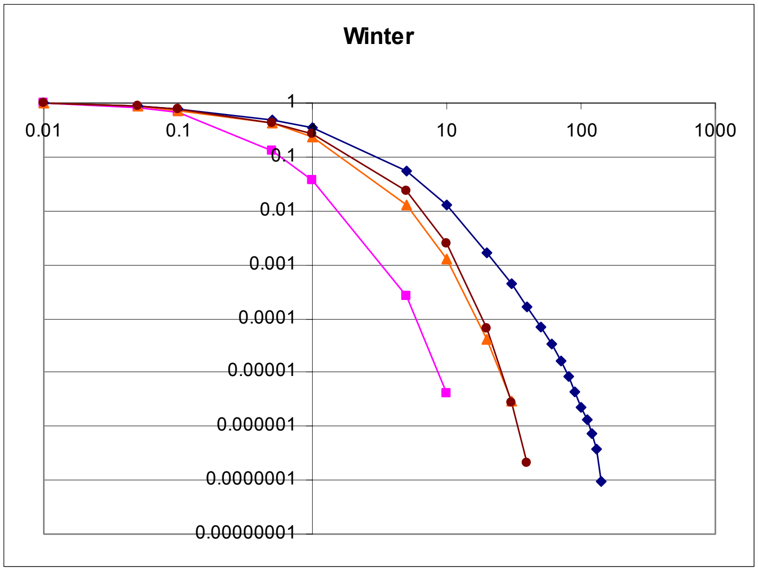
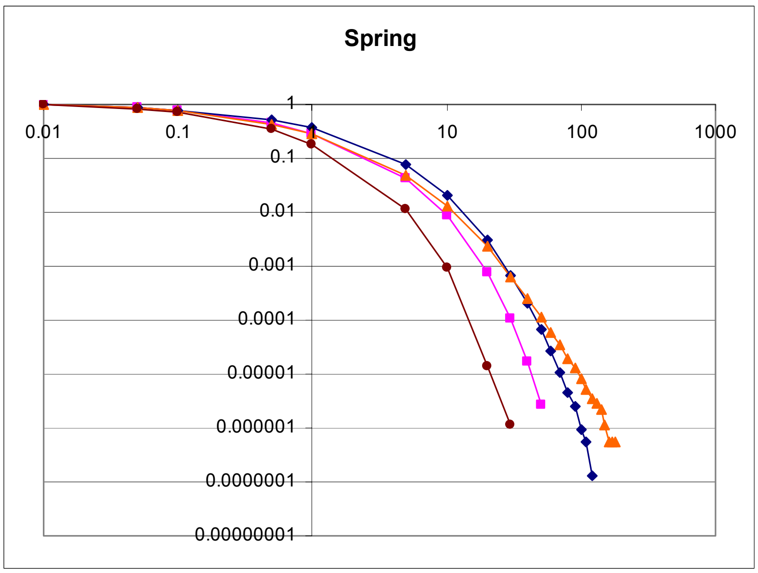
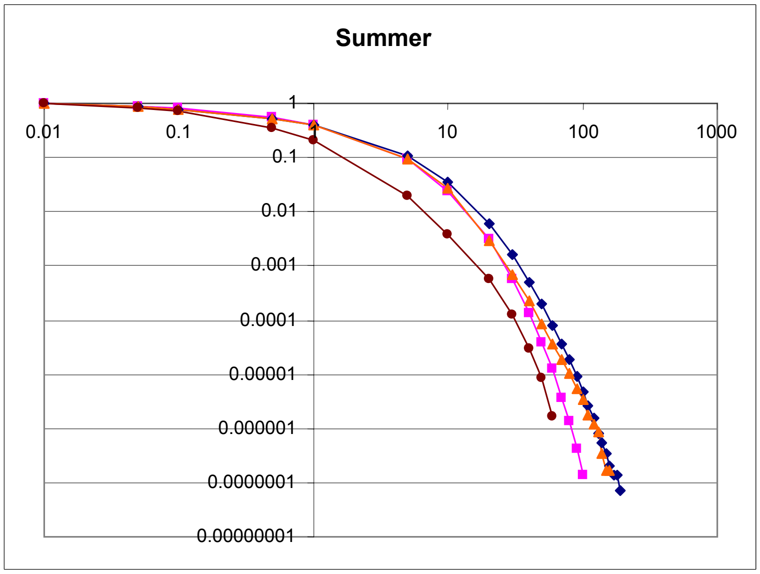


Fig. 6. Gamma parameters for regional plots of summer and winter in the Figure 5. The summer objects are the lighter squares plotted behind dark diamond winter points. α is plotted on the x axis and β is potted on the y axis. The guide lines are set at specific means. The dark line is set at .1 and the light line is set at 1.


Winter
1
0.1
0.01
0.001
0.0001
0.00001
0.000001
0.0000001
0.00000001
0.01
0.1
10
100
1000


Spring
1
0.1
0.01
0.001
0.0001
0.00001
0.000001
0.0000001
0.00000001
0.01
0.1
10
100
1000



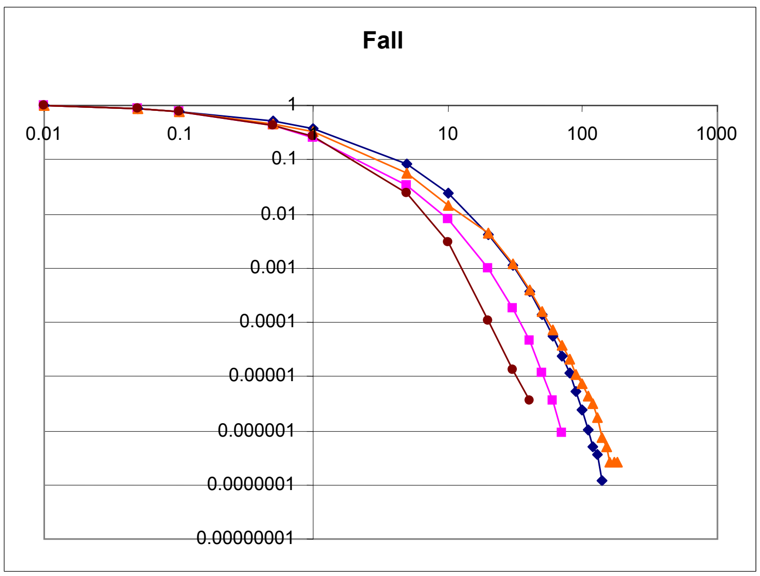
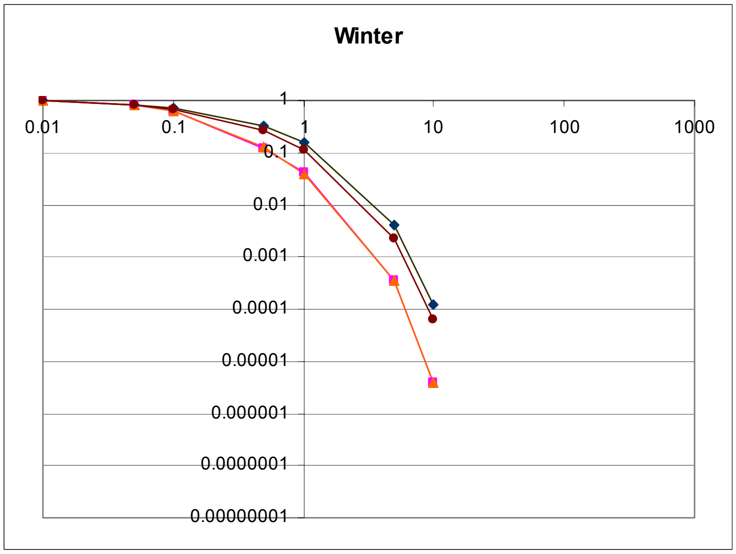
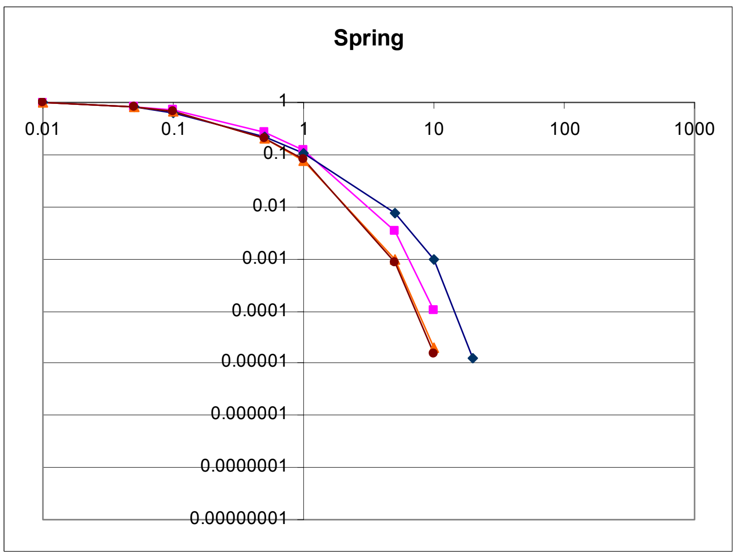


Fig. 7. The same plot as Figure 5 for only the convective objects of each region. The curves still represent the same regions.



Winter
1
0.01
0.1
1
10
100
1000
0.1
0.01
0.001
0.0001
0.00001
0.000001
0.0000001
0.00000001



Spring
1
0.01
0.1
1
10
100
1000
0.1
0.01
0.001
0.0001
0.00001
0.000001
0.0000001
0.00000001

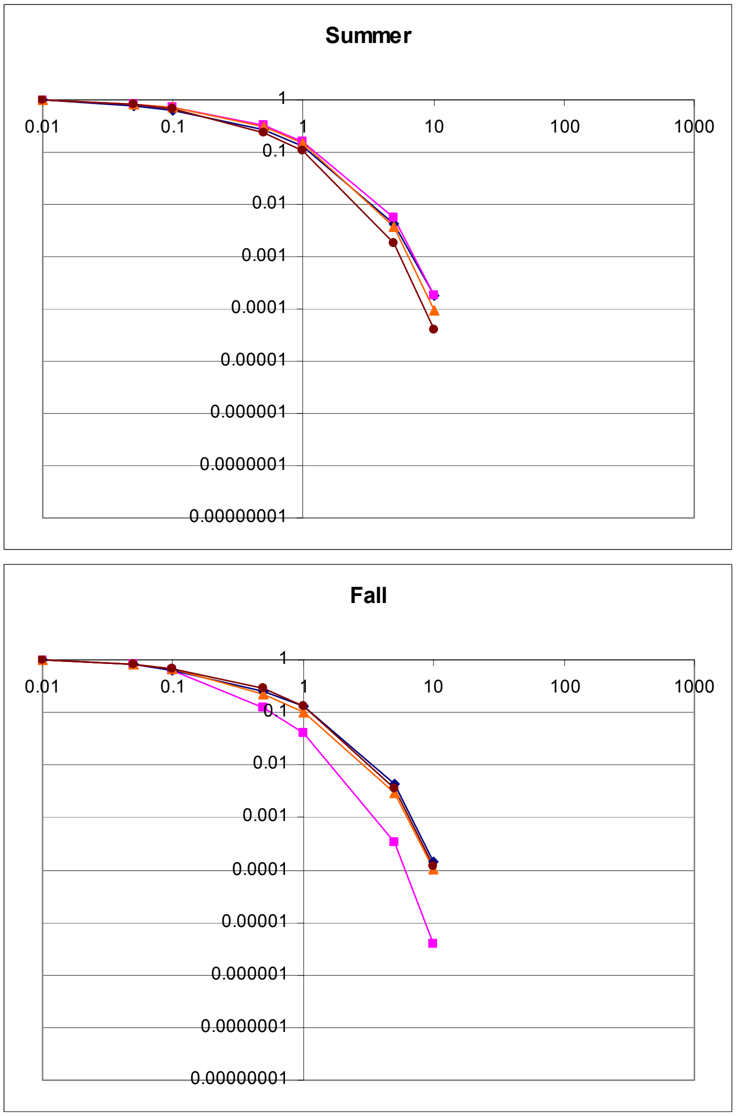


Fig. 8. The same as Figure 5 for only the stratiform objects. The curves still represent the same regions.

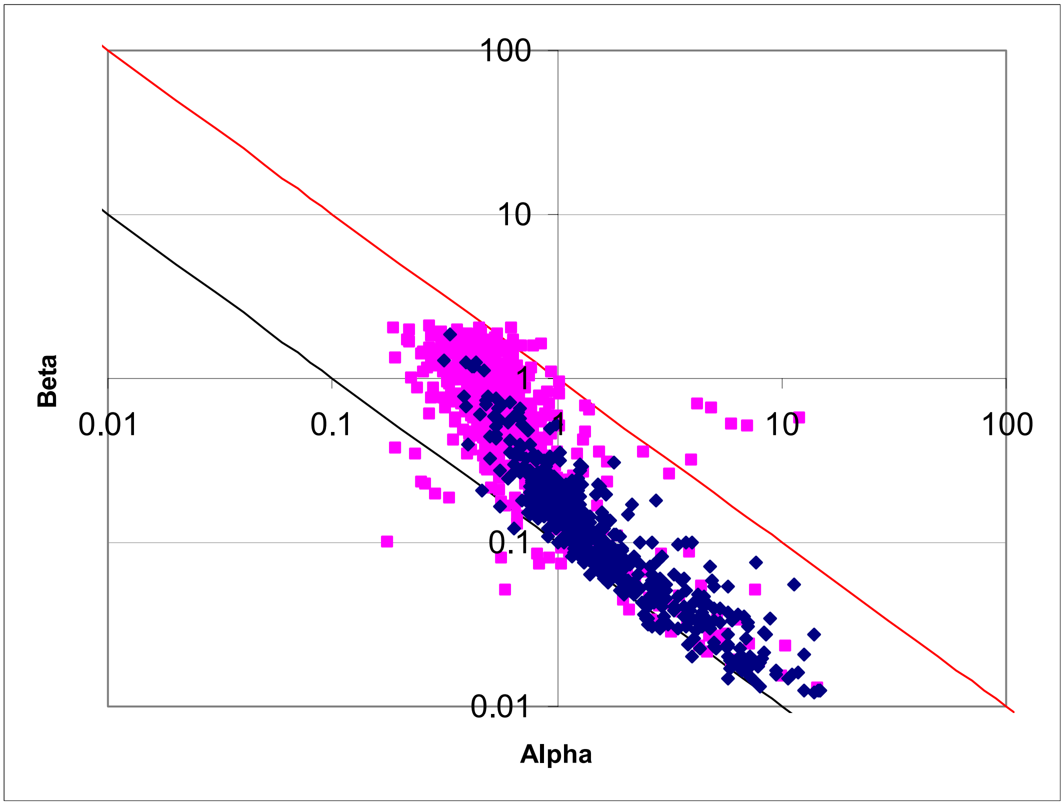


Fig. 9. NC gamma parameters for the summer and winter plots in Figure 8. Same plot type as Figure 6.

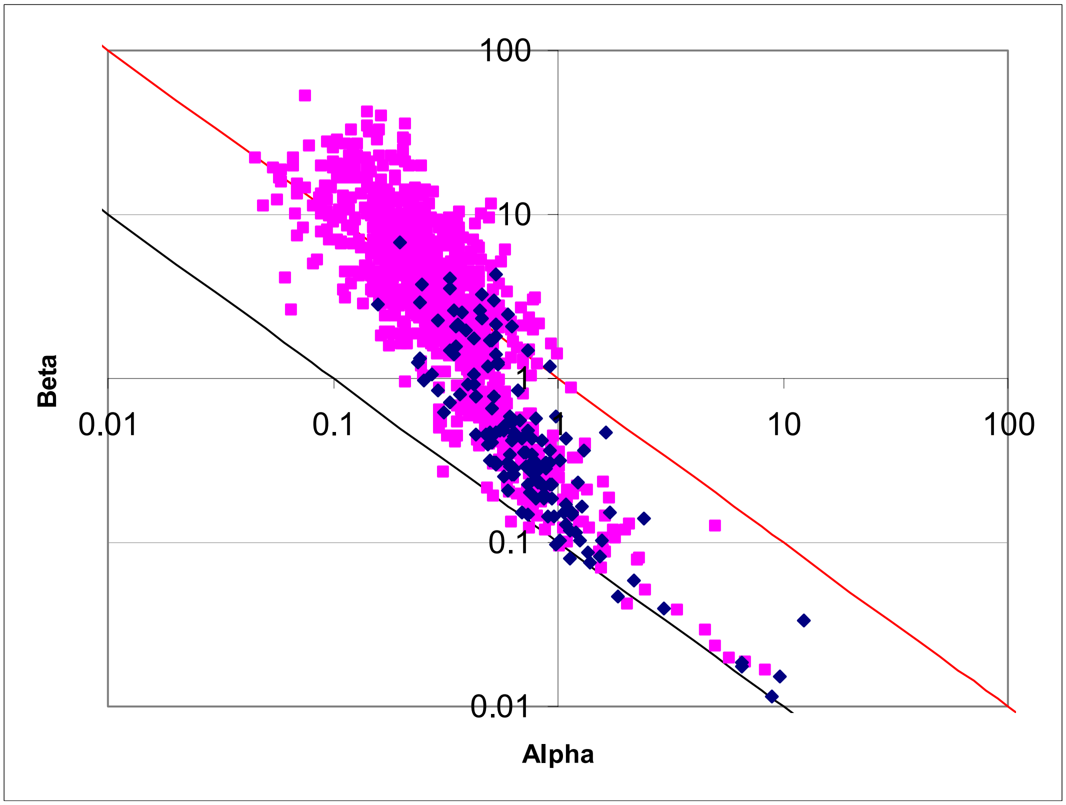


Fig. 10. P gamma parameters for the summer and winter plots in Figure 7. Same plot type as Figure 6.