

*Central Region Technical Attachment
Number 09-01
August 2009*

**A Statistical Analysis Approach for Inferring Preferred
Wind Direction for Enhanced Snowfall Amounts at
Eastern Kentucky COOP Sites**

Gary S. Votaw
National Weather Service
Jackson, Kentucky

1. Introduction.

Spatial distribution of precipitation and quantifying precipitation are forecast problems with many root causes. Among those causes are: 1) atmospheric conditions are often difficult to model at a resolution that sufficiently captures the processes causing the variability; 2) precipitation cannot yet be directly modeled; 3) critical detail in atmospheric spatial variability is at a finer scale than the resolution of the remote sensors and upper-air network; and 4) topographic influences lead to a wide range of modeling problems and cause a wide range of conditions over very short distances.

These forecast problems are magnified when the precipitation is in the form of snow since water-snow ratios magnify not only snowfall rates but also the distributive differences on a spatial plane. Snowfall amounts often vary greatly across short distances, partly due to a tight temperature gradient often inducing a sharp rain-snow line, partly due to convection and banding, but also partly due to topographic variation. Since among these variables topography is the only one invariant, this influence might be visible in long-term data.

Could it be determined which wind directions over eastern Kentucky enhance snowfall rates and which have a suppressive effect? Might topographic influences be noticeable in the

snowfall amounts received? Statistics has found little application in meteorological analysis other than regression, but this study was arranged to compare data involving non-parametric categories of wind direction. How well does statistical analysis answer these questions?

2. Study Area.

The county warning area (CWA, Figure 1) for the Weather Forecast Office (WFO) at Jackson, Kentucky spans 33 counties in eastern Kentucky between the Ohio River and the higher part of the southern Appalachian Mountains. Topography in the CWA encompasses an area considered part of the Bluegrass in the extreme north to the highest terrain in the state at the southwest Virginia border. Elevations range from approximately 550 ft above MSL in northern Martin County to 4145 ft above MSL (Black Mtn.) in Harlan County. In between, steep slopes are common with 500-ft deep valleys while ridge crests often lay less than one mile either side. This terrain is exceptionally complex but generally rises toward the southeast. On the northeast side of the CWA lay the Ohio River, Big Sandy River, and Tug Fork of the Big Sandy River which allow for another significant slope, rising towards the south and southwest.

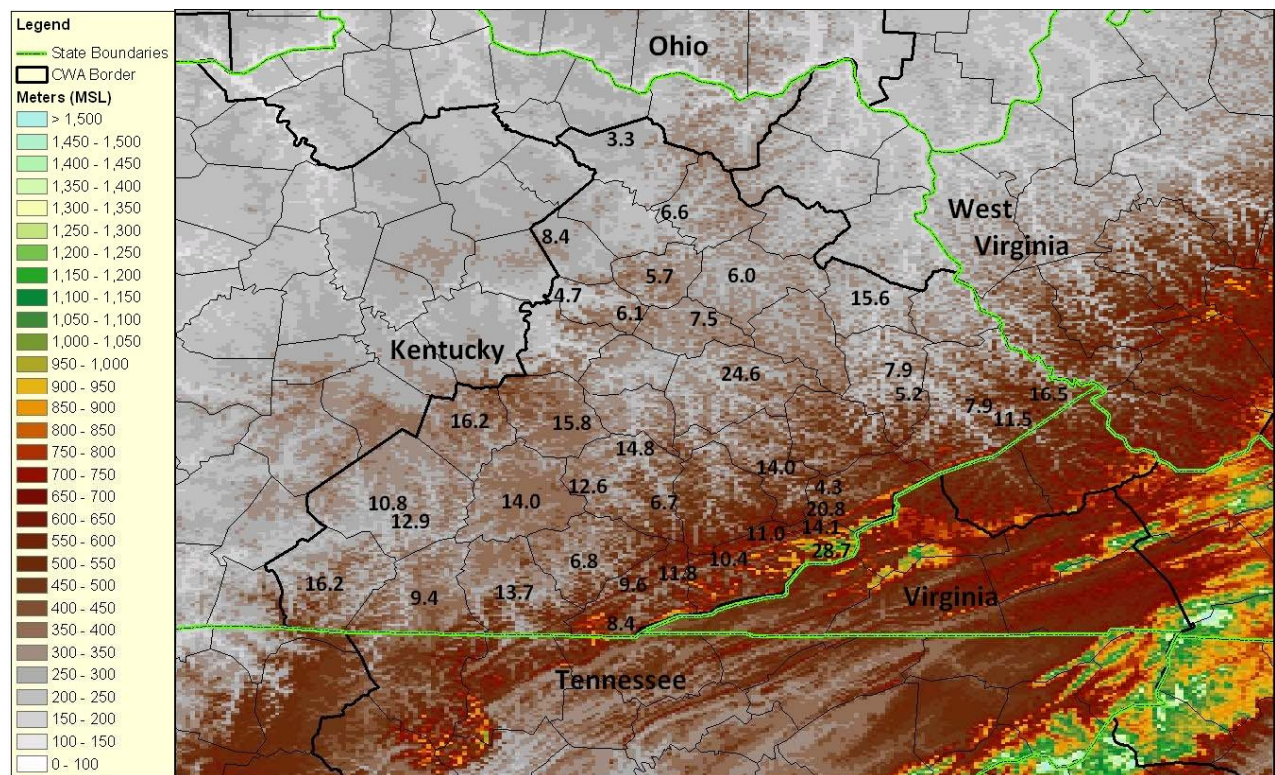


Figure 1. Topographic map with outlined state (green), counties (gray), outlines of county warning areas (black), and centered on WFO Jackson's CWA, and average annual snowfall (inches) for selected COOP sites, from Midwest Regional Climate Center (1971-2000).

Eastern Kentucky has topography that ranges from gentle hills to mountainous terrain. It seems likely that topography significantly influences precipitation patterns at many locations across this area. But the subject of how the boundary layer behaves across such a diverse topographic landscape is much beyond the scope of this study. No assessment can be made here regarding how the boundary layer changes near a recording station since it will not parallel the ground surface.

3. Data and Methodology.

This study uses data from 58 cooperative network (COOP) stations that are located in the CWA along with airport locations at Jackson and London. Daily snowfall as recorded in the Applied Climate Information System (ACIS, n.d.) and accessed in the WFO Jackson office was

recorded for each of the stations for the 1948-2008 period except that trace amounts were treated as zeros. If any one of the stations recorded measurable snowfall for a day, then all stations were recorded on that same day (herein called a “snow day”). The COOP stations normally reported snowfall daily at 0700 EST/EDT and their reports represent snow that fell in the previous 24 hours. The two airports reported snowfall for the calendar day. This resulted in 2,076 snow days during the 61-year period.

Upper-air data were recorded from the Earth Systems Research Laboratory (ESRL, n. d.) web site for each snow day. Due to the very sparse vector field on the reanalysis maps (none in Kentucky), general interpolation between vectors was required to derive a useful wind direction for use in this study and thus a 16-point compass was used to identify wind directions. This also precluded using surface wind data from the ESRL data source since the existing topography is vastly more complex than can be represented by the vectors available. One wind direction and speed was recorded for each snow day from the web site for each of the following millibar levels: 925, 850, 700, and 500. The location of the recorded upper-air winds was at the center of the CWA (near Jackson). To compare data from the COOP sites, the upper-air data were recorded at the 0000 UTC (1900 EST) on the report day. This time was the mid-point during the snow period being reported. For the London and Jackson airports, since they reported data based on a calendar day, the relevant times of recorded wind data were for 1200 UTC (0700 EST) and also 0000 UTC (1900 EST) on the day the snow fell.

These data represent but an instant during each 24-hour period. A wind field displayed with vectors as coarse as this and involving such a limited temporal aspect would have limited value for analysis involving one or just a few meteorological events. In fact there are so little data

collected for each event that traditional analysis is impossible. But in a data set with a very large number of days, statistical analysis is expected to be able to identify significant long-term relationships.

Data for each station were analyzed individually. Days with missing data were removed. Wind direction, wind speed and snowfall total were the only variables input for analysis, and the objective was to determine which wind directions were significant for enhancement of snowfall at each station. To analyze this large data set including a non-numeric parameter (wind direction) an analysis of covariance (ANCOVA) was chosen as the statistical test to use (Mertler and Vannatta 2005). Wind speed was used as the covariate which would enhance the test's ability to determine differences. These data were sorted into groups according to wind direction and pressure level but the ANCOVA would analyze the differences between only two wind directional groups at a time. To determine whether a wind direction of interest significantly enhanced snowfall at the station, data associated with that wind direction were paired with corresponding data representing exactly the opposite direction (the two groups in each ANCOVA test). The ANCOVA then analyzed whether the wind direction in question was associated with higher reported snowfall than was the case when the wind was from the opposite direction. For example, in data at the Jackson airport, all daily snowfall reports associated with an 850-mb wind direction from the east-northeast were paired with data on days when 850-mb wind directions were west-southwest, creating a subset along this directional radial.

Each coupled data pair was then analyzed by statistical software by using the analysis of covariance (ANCOVA) method. The dependent (predicted) variable was daily snowfall. While

wind direction was the only independent (predictor) variable used, wind speed was used as the covariant.

Each ANCOVA was conducted to determine the effect of wind direction on snowfall amounts while controlling for wind speed. The statistical software that was used (Winks SDA 6.0.5, n.d.) examined each of the paired groups (such as the group for northeast and southwest) for group effect. It then examines the linear relationship between the dependent variable (snowfall amount) and the covariate (wind speed). However since these data sets never involve more than two group directions at a time there were never enough degrees of freedom for this part of the analysis to work (three are needed). The covariate still helps the program determine significance between groups of wind direction better than an analysis of variance without it. But, an analysis that would involve significance in the singular relationship between snowfall amounts and wind speed without consideration of wind direction could not be done in this study. A representative hypothesis for these ANCOVA is:

Hypothesis for a group effect:

H₀: The group means between paired wind directions (snowfall amounts, as adjusted by the wind speed) are equal.

H_a: The group means between paired wind directions are not equal.

The null hypothesis (H_0), if found to be true, represents those cases when the means of two groups of paired wind directions were not found to be different at more than a 95% probability ($p < 0.05$) whereas the alternative hypothesis (H_a) are those cases when the paired directions were tested to be different at least to that level. This is a two-tailed test, meaning that the values one group of snowfall amounts might be found to be either significantly lower or higher relative to the other group it is compared to.

Initial data screening led to elimination of all snow days when snowfall was missing and reassigning trace amounts to zero. An example of ANCOVA results for the COOP station at Monticello 3NE (see Table 1) for the northeast-southwest wind direction couplet at 925 mb follows.

A test for group effects yields the following results:

 $F(1, 208) = 18.6663 \quad p = < 0.001$

This F-test (with a resulting ratio) determines if the null hypothesis is true in which case the two groups of data would have a continuous probability distribution, normally distributed and with the same standard deviation. The numbers in parentheses represent the degrees of freedom, those variables that are allowed to vary freely (The Animated Software Company, n.d.) In this study, the first value of degrees of freedom is always one since there are only two groups of snowfall at a time being compared ($k-1$). The second number represents the degrees of freedom for the number of days of snowfall recorded at the station being tested. The resulting F-ratio is equal to the variance (mean squares between snowfall groups) divided by the variance (mean squares) expected due to chance, or error (Mertler and Vannatta 2005). Finally, the p -value is the probability of falsely rejecting the null hypothesis, $< 0.1\%$ in this case.

Winks SDA software also runs a post hoc comparison test (Scheffe') after each ANCOVA to further determine significance in the mean differences of the two groups. This serves as a check and verification for the ANCOVA results. After the northeast-southwest subset at Monticello 3NE had computed means for the two snowfall groups and was adjusted according to the covariate (wind speed) the Scheffe' test then compared the means of the groups with each other and

against a computed a critical value. If the difference in the means of the two groups exceeded the critical Scheffe' value, then those means were determined to be statistically different.

The general formula for the Scheffe' critical value is:

$$S_{\text{critical}} = \sqrt{(k-1)F_{.05(1), k-1, N-k}}$$

where k = # groups or levels of the independent variable,

F is a value from the F-distribution table for $k - 1$ and $N - k$ degrees of freedom

N is the sample size per group.

(Zar, 1999).

For this Monticello 3NE data the Critical S was determined to be 1.972 while the computed S value for the data was 4.32. The Scheffe' test thus determined that the group means were different at the $p < 0.05$ level.

Both the ANCOVA and Scheffe' tests for group effects for this subset of Monticello 3NE were significant; we may conclude that there is a significant difference in the group means. The analysis concludes that to a greater than 95% probability, the difference in snowfall recorded between the two group means is significant. But since the only variable to determine differences is wind direction then it may be inferred that conditions associated with wind flow in these directions are the cause of the difference in snowfall amounts.

Figure 2 is a graphical display of the output of the ANCOVA for Monticello 3NE for the 925-mb winds from the northeast and southwest directions. Linear regression lines are drawn through the data after the ANCOVA adjusted the data according to the covariate, wind speed. In most instances, if slopes of the regression lines through the snowfall data separated into the two groups were less different than this example, the software would determine that there is no difference between the groups of snowfall. There would then be insufficient reason to conclude

that wind direction plays a significant role in such cases for enhancing or suppressing snowfall amounts.

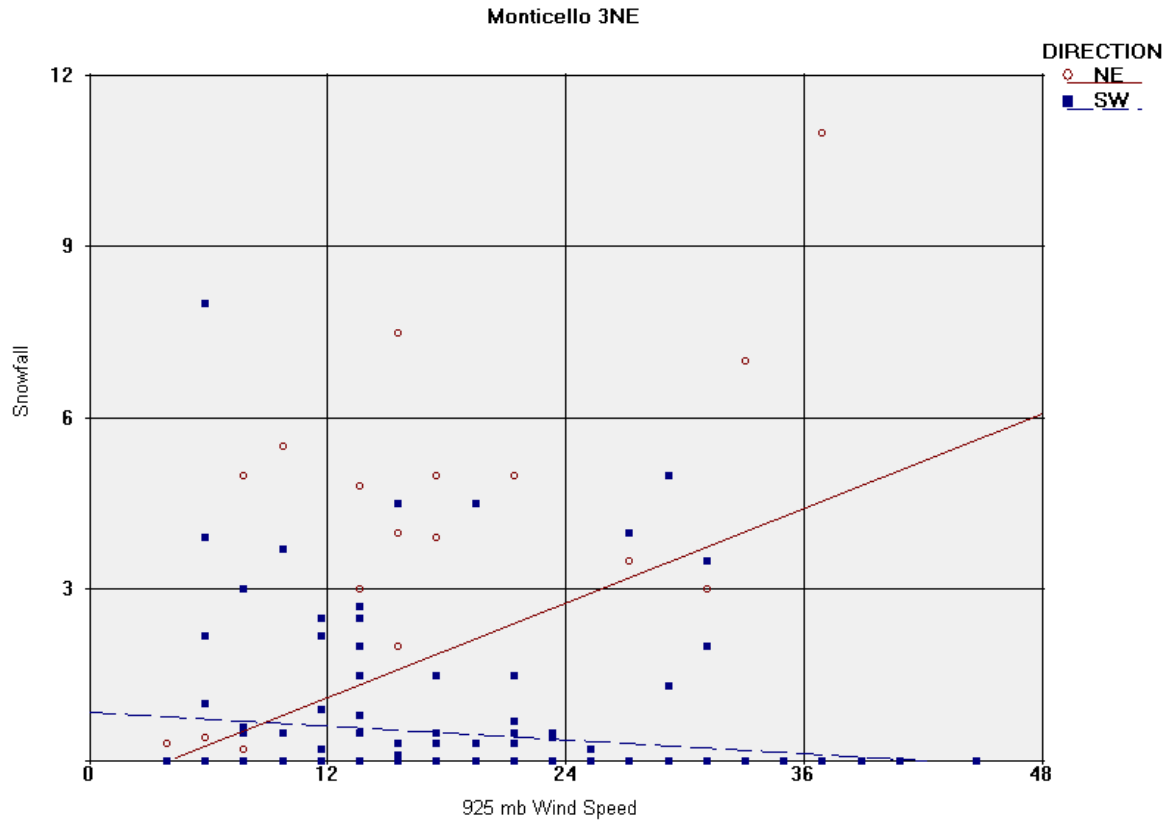


Figure 2. ANCOVA results for analysis of snowfall in the northeast and southwest groups. The linear regression lines in the figure are drawn through snowfall amounts after having been adjusted by the ANCOVA.

4. Results.

Table 1 is a section of a spreadsheet showing several of the largest data sets at the 925-mb pressure level. Monticello 3NE was the largest set with 1,977 snow days while the size of the data sets (the number of snow days during 1948-2008) decreases with each station toward the right. The values in the spreadsheet are mean snowfalls (unadjusted by ANCOVA) associated with days when these wind directions occurred (according to the single wind vector recorded for that

day). The numbers are typically low since all of the stations' snowfall reports were recorded even when only one station in the CWA received a measurable amount. Shaded in green and blue are data that the ANCOVA found significantly higher or lower than data associated with its opposite direction.

925 mb	Monticello 3NE (1,977)	Paintsville 1E (1,924)	Manchester 4W (1,738)	Farmers 2S (1,711)	Somerset 2N (1,669)	Mt. Vernon (1,666)	West Liberty 3NW (1,5,69)	Baxter (1,505)	Jeremiah 1S (1,437)	Hazard Wtr Wrks (1,428)	Flemingsburg 2N (1,387)	Williamsburg (1,377)	London Apt (1,346)	Mt. Sterling (1,318)	Skyline 1SE (1,304)	WFO Jackson (1,122)
N	0.4	0.8	0.5	0.4	0.1	0.4	0.4	0.4	0.3	0.6	0.1	0.4	0.7	0.4	0.9	0.7
NNE	1.0	0.5	0.6	0.6	0.8	1.0	0.8	0.6	0.7	0.9	0.3	0.3	1.1	1.0	1.1	2.1
NE	1.6	1.7	1.4	1.6	0.8	1.5	1.3	1.1	1.1	0.8	0.9	1.2	1.1	0.9	1.1	0.7
ENE	0.4	0.6	0.5	0.6	1.1	0.3	0.1	0.3	0.0	0.1	0.3	0.8	0.5	0.8	0.3	0.6
E	1.1	1.3	0.7	0.7	0.4	1.1	0.4	0.3	0.3	1.0	0.3	0.7	0.8	0.5	1.3	0.1
ESE	1.1	0.7	0.4	0.9	1.0	1.9	0.6	0.8	1.0	0.4	0.0	0.6	0.6	0.5	1.0	0.4
SE	0.6	0.5	0.4	0.5	0.3	0.6	0.4	0.1	0.3	0.2	0.5	0.1	1.2	0.6	0.3	0.3
SSE	0.4	0.8	0.3	0.1	0.4	0.5	0.3	0.3	0.3	0.0	0.4	0.6	1.2	0.9	0.4	0.4
S	0.3	0.3	0.2	0.1	0.5	0.3	0.2	0.1	0.1	0.2	0.1	0.3	0.5	0.3	0.2	0.4
SSW	0.5	0.4	0.4	0.3	0.5	0.6	0.2	0.3	0.5	0.1	0.4	0.2	0.4	0.4	0.4	0.3
SW	0.4	0.3	0.3	0.3	0.4	0.5	0.1	0.3	0.3	0.1	0.2	0.2	0.5	0.4	0.2	0.3
WSW	0.3	0.3	0.2	0.4	0.3	0.4	0.2	0.2	0.2	0.2	0.5	0.3	0.5	0.4	0.2	0.2
W	0.3	0.2	0.2	0.3	0.2	0.3	0.2	0.2	0.2	0.1	0.2	0.2	0.3	0.2	0.3	0.4
WNW	0.3	0.3	0.3	0.3	0.1	0.3	0.2	0.3	0.2	0.2	0.3	0.2	0.4	0.2	0.5	0.5
NW	0.4	0.5	0.4	0.4	0.1	0.3	0.3	0.4	0.2	0.4	0.3	0.3	0.4	0.2	0.6	0.6
NNW	0.5	0.4	0.6	0.3	0.2	0.4	0.2	0.6	0.4	0.4	0.3	0.3	0.6	0.0	0.6	0.4

Table 1. Raw snowfall means (1948-2008), unadjusted by ANCOVA, partitioned according to 925-mb wind direction with significant higher values shaded in green and their opposite directions in blue.

One of the most striking aspects of these results is how prevalent the northeast and east directions were determined to be different than their opposite directions, southwest and west.

For 28 of the 31 largest data sets, the northeast-southwest couplet was significant. While the

east-west couplet was the second most dominant couplet, only 17 of those 31 were significant. The 29 COOP sites that had the smallest data sets ($< \sim 850$ snow days) were usually too small for the program to determine significance.

It is curious that at 925 mb, east-northeast (and west-southwest) would only be significant for only 3 of the 60 locations analyzed since the northeast-southwest and east-west radials were significant so often. Would the same processes not be at work for the east-northeast to west-southwest radial? The east-northeast to west-southwest data set is typically the smallest data set with fewer than 30 snow days included. The only difference noted in the data for this direction was in the vertical wind profile. In the data sets involving every other direction, the low-level flow was relatively deep with the 925-mb wind direction usually extending through 850 mb or else slowly veering with increasing height. But the east-northeast flows that occurred were mostly shallow and highly sheared. It was not known why this might make a difference resulting in less snowfall when this direction is occurring.

Figure 3 is the topographic map with 925-mb wind vectors showing the directions along a radial that are associated with the largest differences in snowfall between polar directions (from the spreadsheet that Table 1 is a part of). Fourteen stations were not used for this map due either to unreliability of the data or of the analysis. Comparing these vectors with verifying wind directions might explain why some locations receive greater snowfall amounts while other events do not produce as much where it might have been expected.

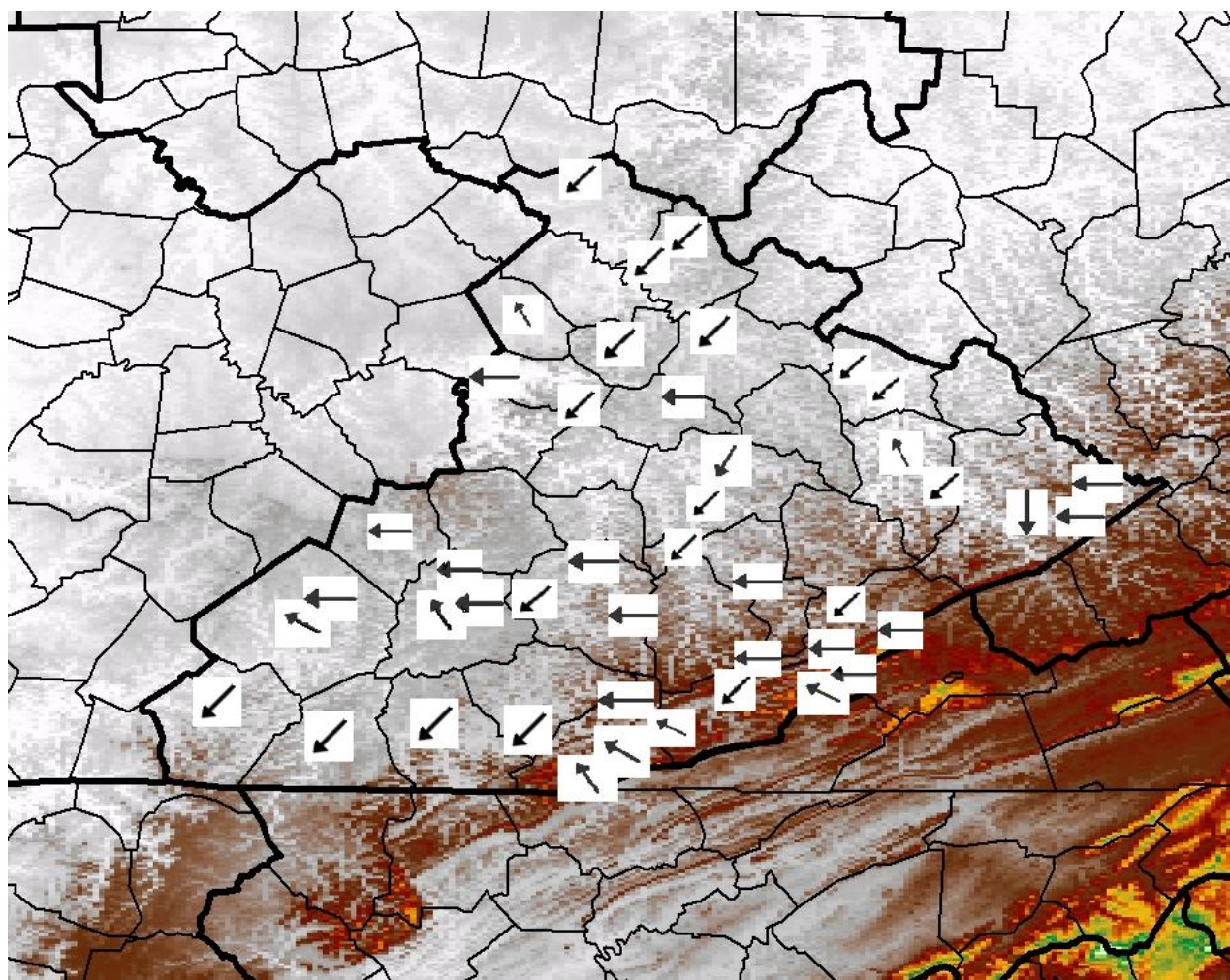


Figure 3. Topographic map with outlined counties and county warning area. Arrows represent the directions that have historically maximized snowfall at the COOP sites indicated.

It is clear from these maps that if one considers the maximum widespread lift that could occur by simply topographically lifting an air mass over this CWA, then it should occur in northwest or north-northwest flow. Topographically lifted air would certainly produce more snowfall given all other factors being equal. But in this study directions down the slope (especially from east or east-southeast) show higher mean snows than those from the northwest. The likely interpretation is that deep upslope flow from the northwest typically does not produce as much snow per event as southeast flow (downslope) which has compensating better dynamic lift, warm advection, and moisture afforded in these conditions.

5. Upslope Flow.

Much research has recently been done on the southern Appalachian Mountains regarding snowfall during and due to northwest flow (e.g. Perry 2006; Perry & Konrad 2006; Holloway 2007; Perry et al. 2007) which broadly is upslope for this mountain chain. Much of the effort has focused farther east and southeast of WFO Jackson's CWA but (Holloway 2007) showed solid evidence for northwest flow to often tap moisture from the Great Lakes for increasing snowfall amounts between the lakes and the Appalachian crest, the trajectory of moist flow being one of the key forecast issues. Perry (2006) defined northwest flow snow events as those events when 850-mb winds were between 270 and 360 degrees at the hour of the greatest snow extent. The vectors from the ESRL reanalysis maps do not allow for the level of directional precision used in that study and upslope flow for the WFO Jackson CWA also includes northeast. The focus of this study thus changed the definition of upslope as including everything from (clockwise) west through northeast. While 850 mb is the level used by most forecast applications estimating Appalachian snowfall, greater significance was found in the ANCOVA results in this study by using 925 mb thus was the primary level used.

In spite of the prevailing significance for a greater chance of producing higher snowfall amounts from low-level winds from the east or northeast as identified in this study, more total snow occurs during northwest flow. This, too, is upslope for most of WFO Jackson's CWA and one must consider if, how, and where snow amounts to forecast should be adjusted due to topographic effect. Can the topographic effect even be isolated?

Figure 4 shows the percentage of total snow that has occurred at the COOP sites and airports while in upslope flow (925-mb winds, west through northeast). While the annual

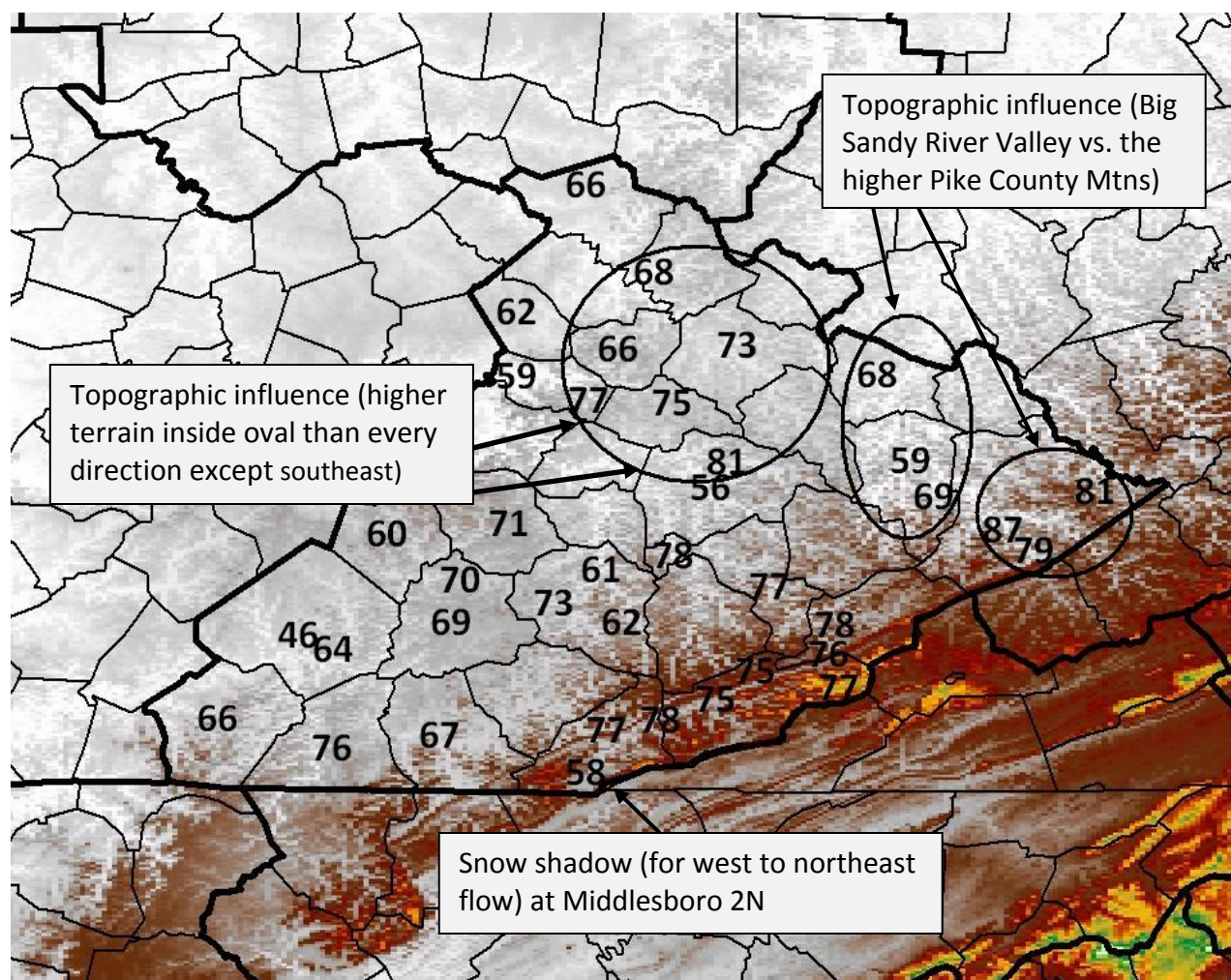


Figure 4. Percentage of total snowfall reported at selected COOP stations when the recorded 925-mb flow was west through northeast (1948-2008).

snowfall (Figure 1) as reported by the sites in the Bluegrass Region (Mt. Sterling, Flemingsburg 2N, Farmers 2S, and Clay City 1WNW) only ranges from 3 to 8 inches, the percentage of the totals from these directions ranges from 59 to 68 percent. These percentages are lower than the values in the 70s farther southeast or south where the terrain is higher. WFO Jackson, at 81 percent, is the only truly ridge-top location among the sites included on this map. If we compare that value to the 56 percent at Quicksand, only 5 miles to the southwest of the WFO and in the North Fork Kentucky River valley, this highlights the importance of upslope (and a generally higher location) for the ridge tops. Though Quicksand has not had enough years of reliable data

to compute a normal snowfall amount, their annual average in this study, plus some years from the 1920s and 1930s, suggests that their annual normal snow is around 14 inches or slightly more than half the normal for the WFO location.

The Big Sandy River valley has similar percentages (Figure 4) as does the Bluegrass. These stations are only about 200 feet lower in elevation relative to those at the southeast corner of Kentucky, very likely an insufficient difference in elevation change to conclude that an upslope flow should produce this much more of their percentage of snowfall. However, the terrain immediately surrounding the stations at the southeast corner of Kentucky is almost 1,000 feet higher than terrain in the Big Sandy River valley and a general upslope flow inclusive from the west through northeast would enable synoptic-scale flow to lift moist air and produce more snow when from these directions. Note also, the conspicuously low value at Middlesboro 2N which represents a snow shadow relative to upslope conditions at most other locations. This station is on the lee slopes of higher terrain north of Middlesboro and is near the lee side (while in northwest flow) of Pine Mountain.

It is difficult to assign cause of higher snow amounts by direction though it seems very likely that upstream slope for a location and whether it is a ridge or valley location are the primary causes in much of the extreme differences between nearby stations. Figure 5 shows the percentage of events at WFO Jackson sorted according to wind direction at 925 mb and compared directly to the percentage of total snow recorded at the station, sorted according to wind direction at the same level. While northwest is the most common direction for events to occur, this direction shows an even higher percentage of the total snow that fell, implying that more snow usually falls per event than when the wind direction is from most other directions.

Still, the wind direction that shows the most inordinate amount of greater snow is north-northeast.

While these data cannot identify conclusively what direction is the most significantly upslope flow for a site since there may be other dynamic and synoptic causes of lift more common for various directions, it may be said that these directions are climatologically more likely to produce more snow than others.

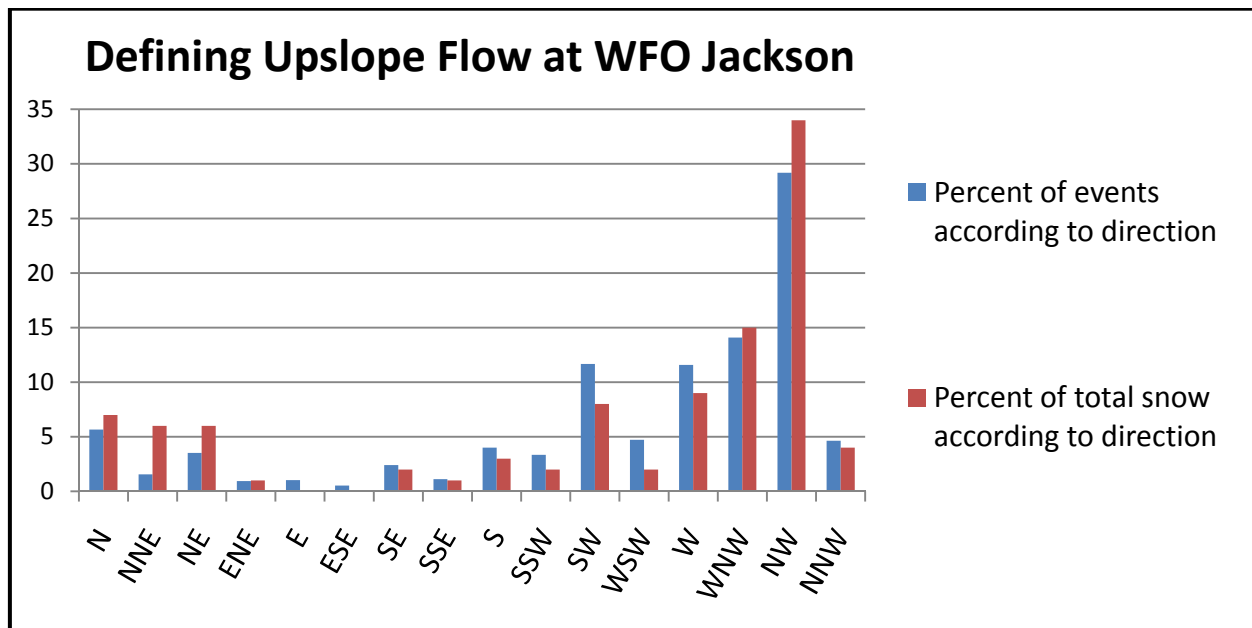


Figure 5. Comparing the percentage of snow days with the percentage of total snow at WFO Jackson, distributed by wind direction (1981-2008). A positive ratio from total snow to events implies that direction produces a higher than average snowfall.

A different perspective is represented at Middlesboro 2N (Figure 6). This COOP site is near the southeast slopes of Pine Mountain and Log Mountains which are just to its west through north. The 925-mb wind directions that produce the most snowfall there are northeast through south-southeast. In fact, in only 2 percent of snow days at Middlesboro 2N a northeast flow results in at least 10 percent of total recorded snow occurred on those days.

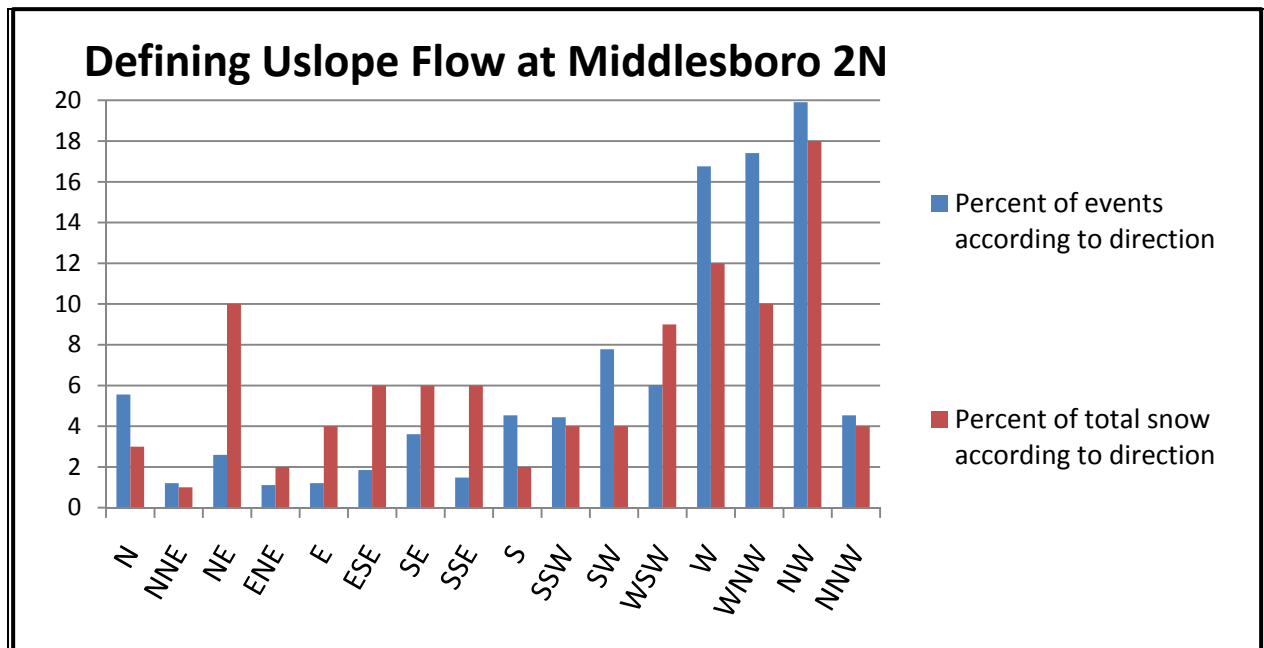


Figure 6. Same as figure 5 for Middlesboro 2N (1948-1991 and 2007-2008).

6. Conclusions.

The maximum for snow production from the northeast at the 925- and 850-mb levels is a direction that benefits from some topographic lift and upslope from the Ohio and Tug Fork River valleys that form the northeast border of Kentucky. At the same time flow from this direction stands a better chance of drawing from excellent moist flux and maximum ascent in the trough (trough of warm air aloft), as shown by Penner (1955), Galloway (1958), and Martin ((1998), of a system passing by just to the south or southeast.

ANCOVA, and its partnering Scheffe' test, was found to be an effective method of analyzing wind directions for a large data set of snowfall and associated winds aloft, though analyses that accessed fewer than 850 snow days at a station were often questionable. However, a simple histogram that compares the number of snow days with the total snow fallen, such as Figures 5 and 6, is probably more useful for identifying what is significant upslope at each COOP site.

Through ANCOVA, a dominant wind profile was found to be a prime condition for production of snow throughout the CWA:

500 mb SSE to S
700 mb E to S
850 mb NE to SE
925 mb NE to ESE

These data can be viewed as another check against other data to have higher confidence in boosting (or reducing) forecast snowfall amounts. This method effectively identified which wind directions are associated with higher or lower snowfall amounts, was done at localized spots, and provided insight to why a COOP site might report more (or less) snow than a forecaster expects.

It is unlikely that the ANCOVA method identified very many of the directions involved in topographically enhanced snowfall. That signature appeared to be overwhelmed by the overall synoptic and dynamic signatures. However, producing a histogram of each observation site, as done in Figures 5 and 6 probably identifies enhancement from upslope better since it shows which directions usually produce more snow per event.

7. References.

ACIS, n.d. Applied Climate Information System (ACIS). NOAA Regional Climate Centers.

ESRL, n.d. Six-Hourly NCEP/NCAR Reanalysis Data Composites.

[Available at: <http://www.cdc.noaa.gov/data/composites/hour.>]

Galloway, J. L., 1958. The three-front model: Its philosophy, nature, construction and use.

Weather, **13**, 3-10.

Holloway, B. S., 2007. *The role of the Great Lakes in northwest flow snowfall events in the southern Appalachian Mountains*. (M.S. thesis, North Carolina State University).

Martin, J. E., 1998. The structure and evolution of a continental winter cyclone. Part I: Frontal structure and the occlusion process. *Mon. Wea. Rev.*, **126**, 303-328.

Mertler, C. A., and Vannatta, R. A. 2005. *Advanced and Multivariate Statistical Methods*. Pyrczak Publishing.

Penner, C. M., 1955. A three-front model for synoptic analyses. *Quart. J. Roy. Meteor. Soc.*, **81**, 89-91.

Perry, L.B., 2006. *Synoptic climatology of northwest flow snowfall in the southern Appalachians* (Doctoral dissertation, University of North Carolina, Chapel Hill, 2006).

Perry, L.B., and C.E. Konrad, 2006. Relationships between NW flow snowfall and topography in the Southern Appalachians, USA. *Clim. Res.*, **32**, 35-47.

Perry, L.B., C.E. Konrad, and T.W. Schmidlin, 2007. Antecedent upstream air trajectories associated with northwest flow snowfall in the southern Appalachians. *Wea. Forecasting*, **22**, 334-352.

The Animated Software Company, n.d. Degrees of freedom. Internet Glossary of Statistical Terms. [Available at: <http://www.animatedsoftware.com/statglos/sgdegree.htm.>]

Winks SDA, n.d. Winks Statistics Software, www.texassoft.com.

Zar, J. H., 1999. Biostatistical Analysis (4th ed.) New Jersey, Prentice Hall.