Regression Analysis for Testing Tornado Modeling Techniques

Mary Sue Passe-Smith, Lecturer
Department of Geography, University of Central Arkansas

ABSTRACT

ArcGIS 9.3's new regression analysis tools were to be used to explore whether terrain features, especially valley orientation, of areas that experience more tornadoes can be successfully modeled for predictive purposes. Overcoming difficulty in operationalizing directionality is a prime task to be tackled. Past research has shown ***visual*** tendencies, and analysis of mean direction confirms, but incorporating this information in to a model has thus far been difficult. With the new regression tools, it is hoped that the effect of valley orientation versus storm travel direction can be effectively modeled. The regression tools, further, will allow for far more effective testing methods than in the past. The study area will be the high-tornado state of Arkansas and low-tornado portions of southern Missouri for good comparison of factors which encourage or inhibit tornado formation.

Note: A small correction on the abstract; after testing for a week without success in the state of Arkansas as a whole, the comparison with Missouri was dropped. Findings did not warrant further investigation.

Introduction

As discussed in Passe-Smith (2005, 2006, 2008), there is some evidence that local topography is thought to enhance or even initiate tornadogenesis. Bosart *et al.* (2004) notes that storms crossing a river valleys perpendicular to their movement experience strong increases in shear within the parent supercell thunderstorm; Peckham *et al.* (2004) discuss horizontal convective rolls which form along upslopes in the Texas panhandle as related to enhanced convection; these rolls were likely a mechanism for initiating convection on May 3, 1999 (the deadly Oklahoma City outbreak). Others have cited changes in land cover or vegetation type as instigating factors (cf. Raddatz and Cummine, 2003; Weaver and Avissar, 2001; Esau and Lyons, 2002). Encouraging findings related to land cover were found in 2005 in a study in Oklahoma and Arkansas, but little further was found in 2006 and 2008 other than a tendency for tornadoes to seemingly cross stream valleys perpendicularly, a finding that was repeated in 2008. The goal, as stated in the Abstract, was to definitively operationalize stream valley orientation to test whether this, or other lucrative past findings, can be tested using regression analysis to model tornadogenesis.

Methodology

I obtained a National Elevation Dataset for the entire state of Arkansas with which to begin my analysis. In order to attempt to remove any population bias in tornado reporting, rather than divide the state up into statistical units such as census tracts, block groups, etc. I used Hawth's Tools to generate a grid of 20x20 kilometer cells. My dependent variable was tornado touchdowns—whether any topographic or land use/land cover influence would affect the counts within each of these grid cells. Tornado touchdown points were obtained for 1950-2008 from the Storm Prediction Center in tabular form, including latitude/longitude fields (NAD83 GCS), which were displayed and stored as points. These were spatially joined to the grid cells so that counts could be obtained.

In order to work with raster data within each cell, many, many zonal statistics were calculated for each polygon. Figure 1 shows the process for the derivation of just one attribute. Land use/land cover (LULC) data was separated into "is forest," "is urban," etc. Boolean layers so that sums could be obtained (i.e., the sum of all the cells equal to 1 would define how much of each grid cell was forest, urban, etc.). This was also done with aspect, one way I hoped to delineate directionality or facing of valley walls: all northeast-facing cells were separated out using the raster calculator and SQL statements such as that shown in Figure 1. Again, the sums were added to the 20-kilomter grid cells using zonal statistics (joined to the spatial layer, then the needed field—in this case sum--was named and calculated permanently using the field calculator before removing the join).

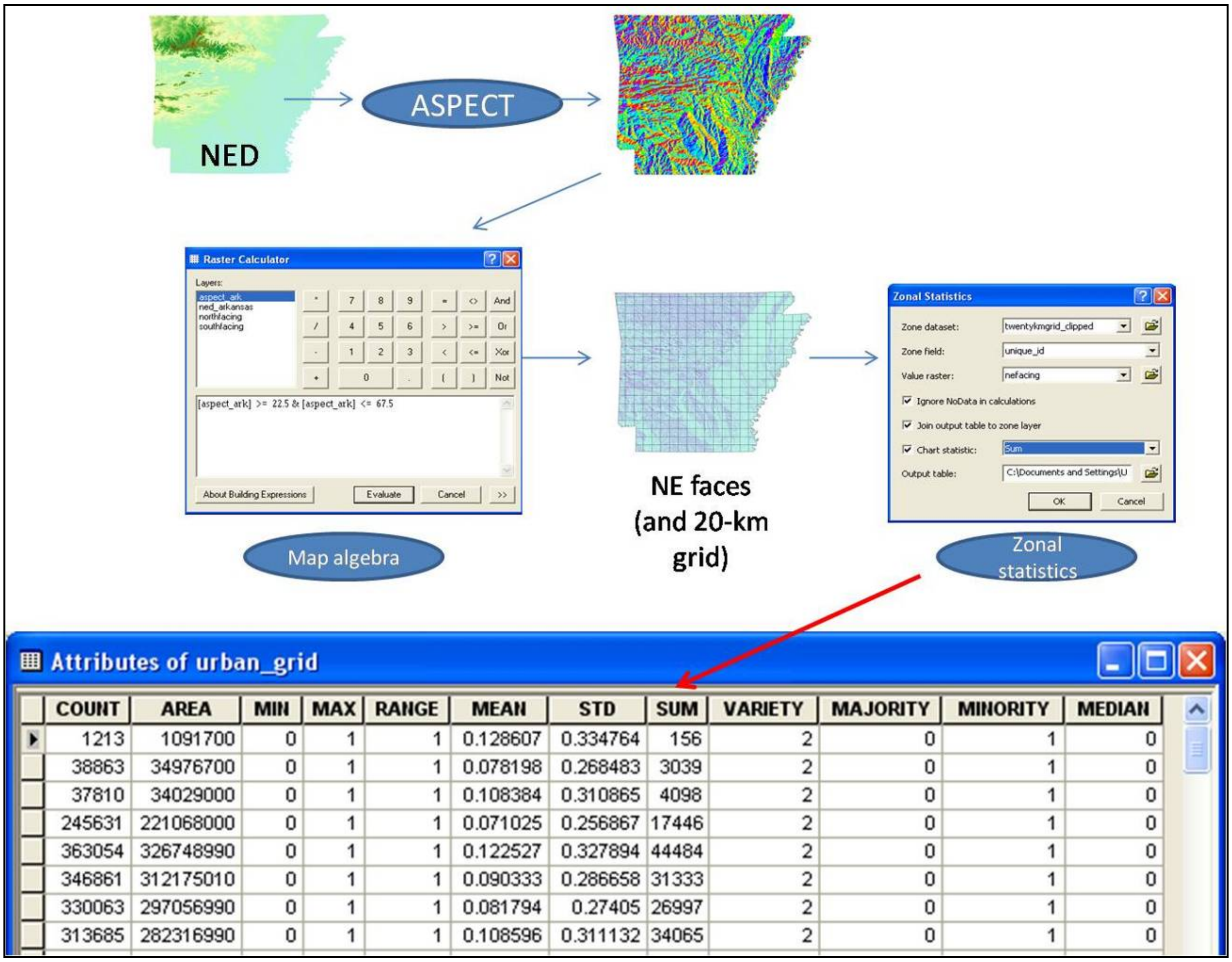

Figure 1: Extraction and use of northeast-facing cells within 20-kilometer grid polygons. Derived value "Sum" is all the cells within each grid that are equal to one, meaning they face northeast. Most raster data was obtained in this manner.

Spatial Analyst was used to calculate the slope for each region. To create 'roughness' neighborhood analyst was used to create a layer based on standard deviation from the mean slope—those areas with high standard deviations are much rougher than are surrounding areas. I extracted cells with slopes over 10 degrees, with the idea of creating fairly steep NE and SW-facing walls. Due to either the small size of my NED's cells (30x30 meters) (later remedied by making them ten times larger and still having no luck) or a possible lack of veracity of the NED as a representation of the earth's surface or some other problem inherent in using aspects in this manner, I was not able to construct a layer that showed primarily northeast/southwest facing 'walls' of valleys, no matter how steep—it simply reflected that tornadoes seemed fairly randomly distributed across all aspects. This was borne out by doing a Chi

Square test of the distribution of tornado touchdowns and the aspects underlying them. While there were slightly more touchdowns on south faces than would be expected given the proportion of the state which is flat, and slightly fewer on some other faces, the numbers are unremarkable and the Chi Square statistics (p = .059) is not significant.

Table 1: Tornado Touchdowns by Aspect (Chi test)

Face	Cells	Proportion of Total Area	Actual Tornadoes	Expected Tornadoes
Flat	572	0	0	0.51
N	146894	9.5%	109	131.86
NE	168946	10.9%	130	151.65
E	209462	13.5%	185	188.02
SE	238819	15.4%	219	214.37
S	253141	16.3%	262	227.23
SW	215861	13.9%	210	193.77
W	187171	12%	169	168.01
NW	133203	8.6%	111	119.57

Chi p statistic: .0588

In the end, I kept the aspect attributes to test them using regression. I derived a total of 13 attributes within each 20-kilometer grid cell covering the state of Arkansas, as listed below in Table 2. Each attribute was tested as an independent variable using both Ordinary Least Squares (OLS) regression and Geographically Weighted Regression (GWR); all were checked for spatial autocorrelation (and all displayed it to varying extents which will be explained below); many of the residuals were mapped using Hotspot mapping, which showed clusters of areas of extremely high tornado occurrence, or local tornado alleys. It is these I have primarily sought to explain.

Table 2: Attributes Tested with Regression		
Field	Description	Explanation
SW_Facing	Sum of all 30x30 cells facing SW	Theory is that valleys that cross normal storm paths in perpendicular manner increase shear; most storms travel SW to NE
NE_Facing	Sum of all 30x30 cells facing NE	See above; other side of same valley (in theory)
Mean_Roughness	See text for method of creation	Roughness thought to inhibit storm formation by virtue of fact that local torando minima occur in Ozarks and Ouachitas.
Slope_gt_10	All cells with slope over 10 degrees	Test if slope aids/hinders touchdown and help delineate valley walls
NE_and_SW	All cells facing NE or SW in a grid cell	Goal to capture valleys with NW-SE orientation (walls facing NE/SW) which is perpendicular to storm movement (may enhance shear)
NE_SW_OVER10	All cells facing NE/SW and over 10 degree slope	Goal to capture steeper valleys with NW-SE orientation, as above
Is_Urban	All urban land use cells	Control for population
Is_Forest	All forest land cover cells	Suspected forest inhibits tornadogenesis (Passe-Smith, 2005)
Is_Crop	All agricultural land cover cells	Suspect crop land encourages tornadogenesis due to increased soil moisture during tornado season (Passe-Smith, 2005)
Mean_Elev	Mean elevation in 20-km grid cell	Test if lower elevation more likely to induce touchdown (Ozarks, Ouachitas local minima; discussed widely)
Count	Spatial join count of all tornado touchdowns (N = 1488) 1950-2008	Dependent variable
Dist_120m	Distance to 120 meter contour	Chosen as the contour that separates low/uplands across entire state—hope to capture

In light of the failure to use aspect to aid in the determination of stream valley orientation, I explored other methods. One such method is to insert COGO attributes, including direction, into linear data using ArcCatalog. This I did to my streams layer, obtained from the National Atlas and clipped to the state. I then edited, selected, and populated them in ArcMap using the COGO toolbar; however, the direction field was peopled only in a small number of cases due to the convoluted nature of the streams; attempts to simplify them resulted in only about one quarter of the streams acquiring COGO direction attributes. This was not acceptable for my analysis. Scripts were found that ostensibly created an azimuth, but only for small portions of streams. To this day, I continue to seek the method, but it will not be presented in this paper.

FINDINGS

Let me begin by stating definitively and emphatically that population is by far the best predictor of the pattern of tornadoes in Arkansas. No matter how the state is divided—by statistical boundaries such as census tracts or block groups, or 20-kilometer grid cells—the spatial autocorrelation of residuals, when mapped, shows nearly every urban area in Arkansas (see Figures 2 and 3). Even a random sample of tornadoes was clustered about population centers. This is not to say that there are not patterns that exist outside of population, but that population must be controlled for before any topographic variable can be assessed. When using population to predict tornado touchdown occurrence, using the number of urban cells per 20-km grid and Ordinary Least Squares (OLS) regression, the resulting R^2 is .24. That population is such a good predictor is reflected in Figure 3, showing mapped residuals while testing another independent variable altogether, clearly depicting areas of Arkansas where there are far fewer tornadoes than would be expected—in every urban area in the state (Figure 2).

This will necessarily be a short discussion—simply put, there was not one independent variable that explained much of anything at all outside of population using OLS. Although all contributed some

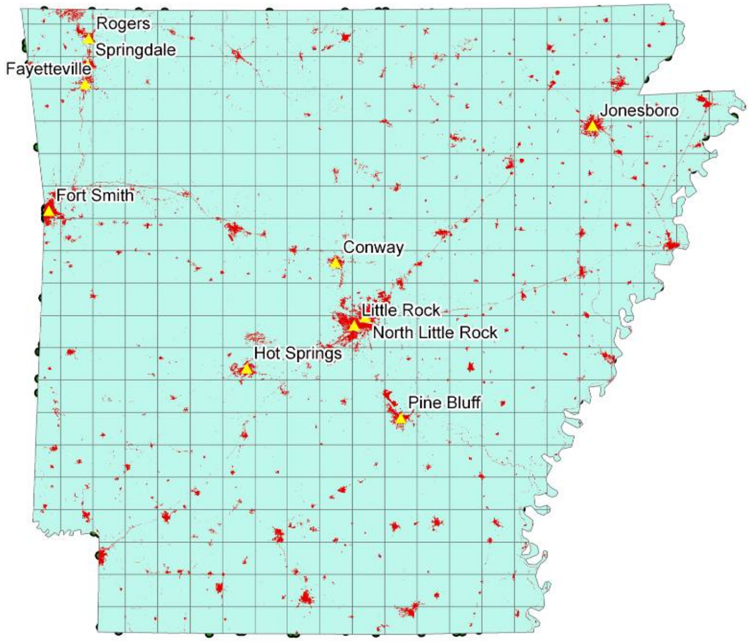

Figure 2: Urban areas; labeled cities have populations of over 30,000 and are reflected in Figure 3, the mapped residuals of an OLS regression equation where mean roughness was the independent variable chosen to explain tornado touchdown count.

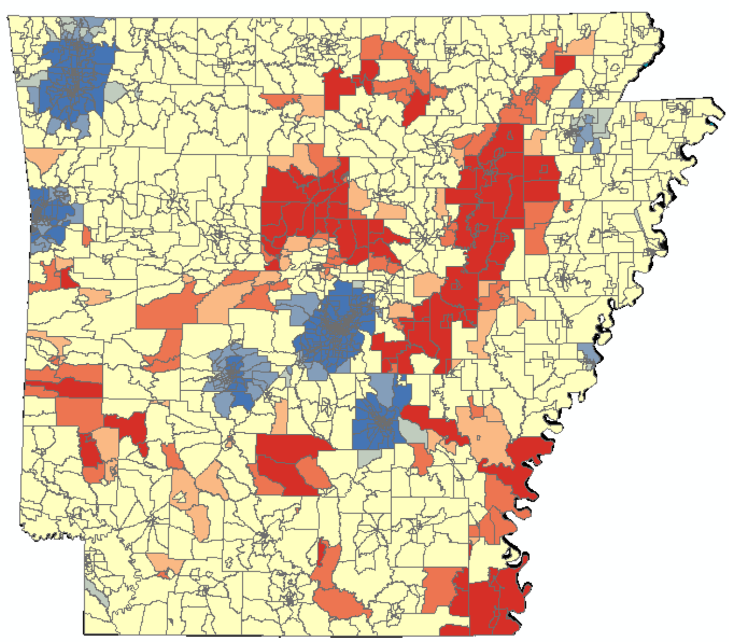

Figure 3: Example of hotspot (GiZ score) map of residuals showing clusters at major urban areas in Arkansas, as shown in Figure 2; fewer than expected tornadoes occur in these areas, meaning population is strongly influencing the distribution of tornadoes.

small amount and when used as the independent variable using GWR displayed regionally high R^2s, when controlled for by including population as an independent variable, the R^2 changed little if at all. When using OLS, with population removed, R^2 also sank to nothing (examples such as .003, etc. abound). Although as an instructor I tell my students that no findings are legitimate findings, I am amazed that nothing explained anything. Using GWR, several of the variables rendered R2's of .65 or even higher in small areas of the state, but they were scattered regionally, and exploration made no coherent sense, nor did it seemingly explain the hotspots discussed below.

The best GWR I could build—one with an overall R2 of .59—was one in which population was used as a control, and cropland was my primary independent variable. However, with an outcome that varies across space, it would be impossible to apply any result as a model to other places when the coefficients, intercepts, etc are specific to Arkansas. When using either forest or crop as my explanatory variable (with or without population as a control) in OLS, where I can actually get a regression equation to apply to other areas, my results seem to wash out. It has become apparent that I first perhaps need to better understand the GWR tool before utilizing it, as in the past I have worked solely with least squares regression and my goal was to use it in this work.

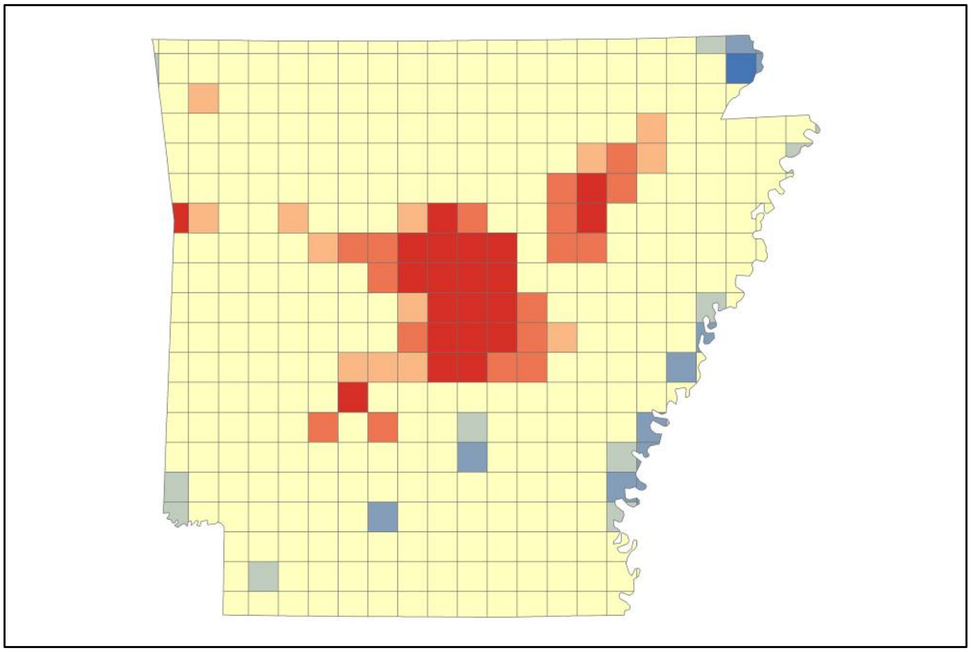

Figure 4: Clustered hotspot region showing area of extremely high tornadic activity that is not well explained by any variable.

There were a few areas that repeatedly showed up as hotspots when mapping residuals for many of the variables; one such region is shown in Figure 4. The largest area replicates the southern end of a country-wide high-tornado-density area, as detailed in Passe-Smith (2008) and occurs at and near the triple junction of the flat "delta" region of Arkansas (the Mississippi Embayment physiographic region), the Arkansas River Valley, and the Ozark upland. Attempts to capture these high numbers in my regression equations eluded me. The 120-meter contour appeared to be the perfect dividing line between the Ozark (and Ouachita) uplands and the valley/delta lowland, and so it was selected as the source of a distance surface to test if proximity to this boundary had explanatory value. It did not really add any explanatory power. Yet visually the fact remains, as is shown in a 3D rendering of Arkansas in Figure 5—tornadoes are *not* distributed across the highest parts of Arkansas, and *are* clustered (with an extremely high Global Moran's I of 0.17, a significance level of .01) in the area just north of the eastern extent of the Arkansas river valley and along the northeast-southwest trending uplands. This is not to say tornadoes do not occur elsewhere; they do, but clusters of high and low areas are evident and

appear to be related to topography. Yet nothing that can currently be modeled using the data available—and the techniques this author is familiar with—is showing any relationship to topography. The reasons for a lack of tornado touchdowns may vary across space in such a way that it cannot be modeled. It may be completely related to meteorological features or the fact that land cover varies so much across time, both in the 50-year period covered and within the course of a year, that it cannot be modeled in a static manner. The alternative explanation is simply that humans are also distributed in the same places as are tornadoes—they do not reside where the terrain is extreme; in a related issue, tornadoes may not be reported in areas of rough terrain and mountain/forested regions due simply to inability to see in these places. Because population itself explains roughly ¼ of the variability in tornado touchdowns, something is certainly missing, or there is no pattern that can be explained at all.

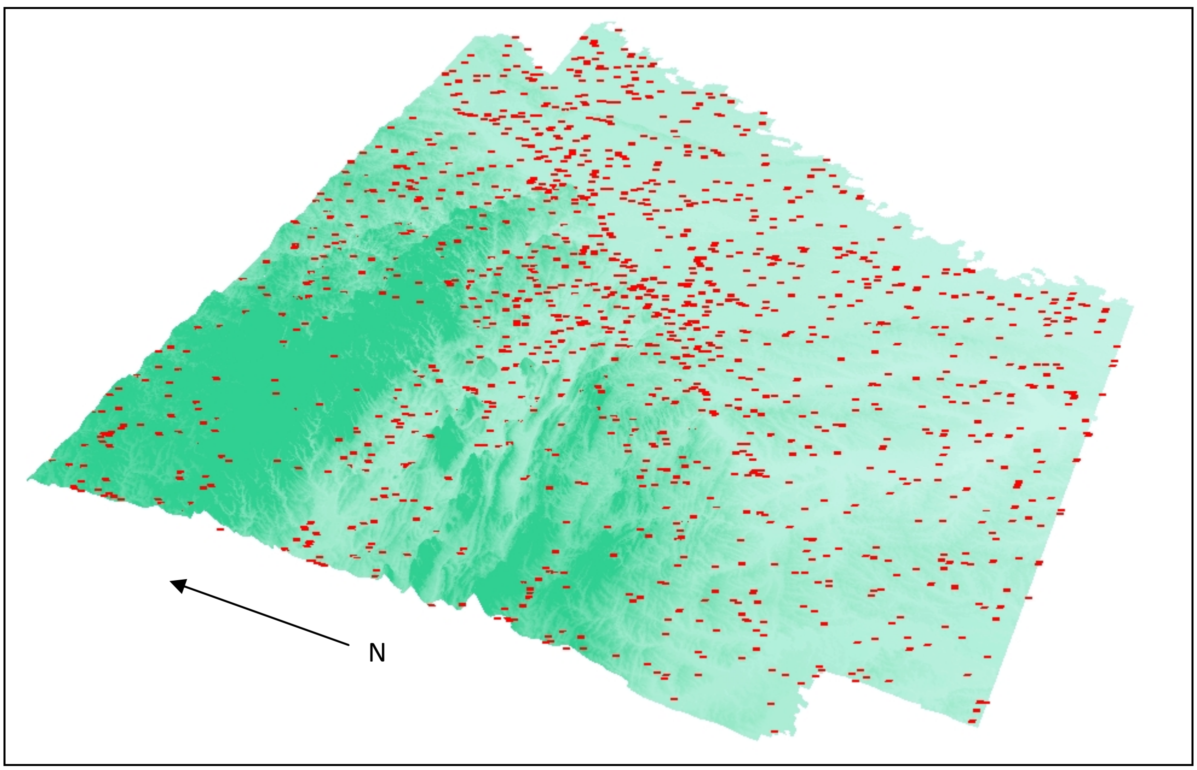

Figure 5: Arkansas relief and tornado touchdowns, 1950-2008.

There is, of course, something causing the local hotspots in Arkansas. One portion of the local tornado alley in northern Conway, Faulkner, and southern Van Buren counties (shown as 1 in Figure 6), overlaid upon a Landsat TM (bands 432) taken in the Spring of 1999 (ten passes from April-June, or during the height of tornado season; Figure 7) show an interesting feature: a wedge of clouds directly over this hotspot area. There is some feature on the earth that contributed to this local cloud patch, likely related to an uplift such as Cadron Ridge. Both Areas 2 and 3 clearly lie on the juxtaposition between the upland forested areas (red) and the not-yet-verdant croplands of the Mississippi Embayment, which is nearly flat. It stands to reason that, in the absence of a semi-permanent feature such as the West Texas dryline that often enhances storm development in the region, any feature that could enhance shear or lift could enhance tornadogenesis. Using relatively simple GIS models to attempt

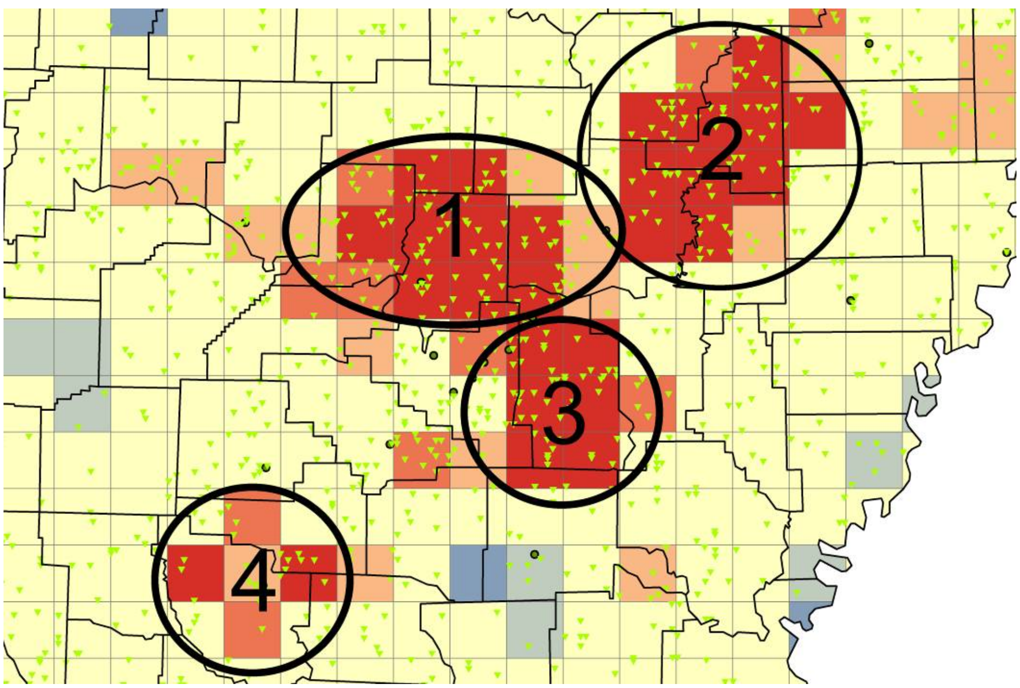

Figure 6: The four major tornado hotspots: Central/northern Conway County, most of Faulkner County, eastern White County, and Southern Van Buren county (1); Lonoke County (3); Jackson, Northern Woodruss, and Eastern White Counties (2); and Clark/Hot Spring Counties (4)

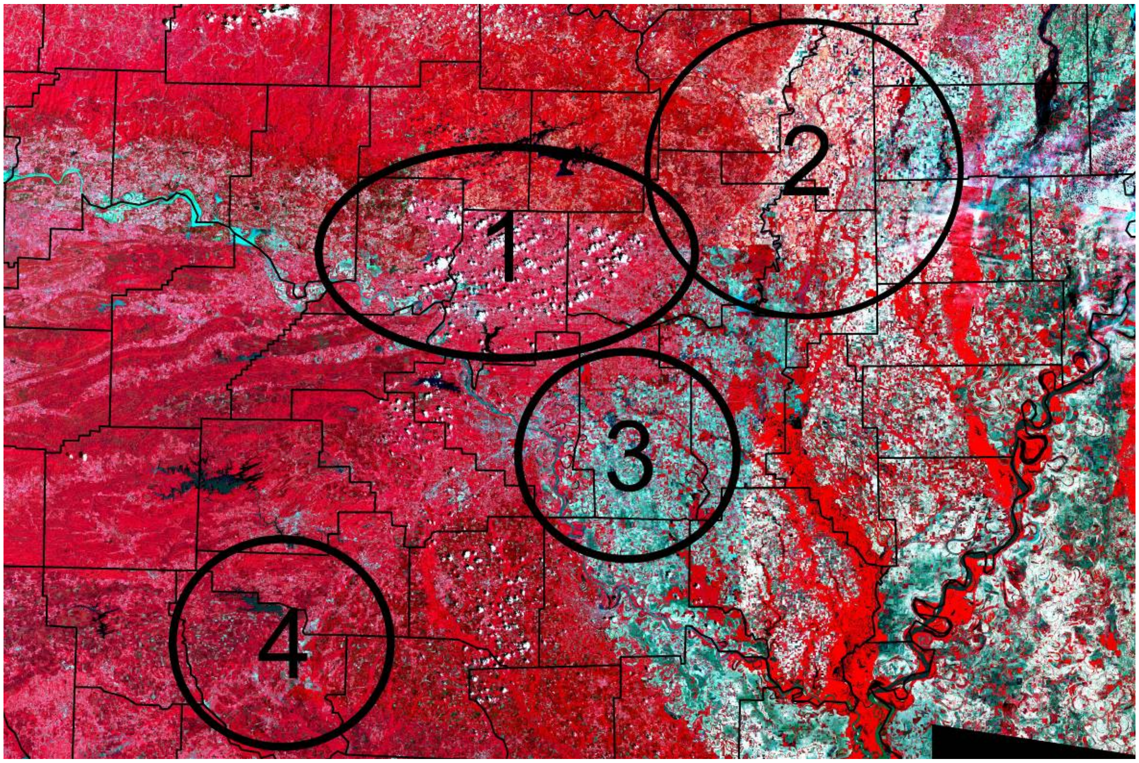

Figure 7: Landsat photo taken in Spring 1999. Note cloud shield over high-tornado region 1.

prediction is in this case hampered by the inability to adequately operationalize the orientation and even the true existence of boundaries between dissimilar surface features at this time. It may well be time to put to rest my quest: there might be, as others have suggested, absolutely nothing that across-the-board can be used as a predictor for tornado hotspots on the local scale. Harold Brooks, research meteorologist, NOAA National Severe Storms Laboratory, Norman, Oklahoma, is quoted in *USA Today* as stating that topography has been looked at, but "little conclusive evidence can be drawn… population density…often [is] involved with topography to make it difficult to separate effects in the observations. There appear to be areas that are tornado minima, such as the Ozark Mountains, but it is hard to draw definite conclusions." At this time, I must agree; any comments and suggestions are welcome, and further research and use of the available tools will be ongoing, but at this time I believe Mr. Brooks has hit the nail on the head.

References

Bosart, L., K. LaPenta, A. Seimon, and M. Dickinson. 2004. Terrain-Influenced tornadogenesis in the Northeastern United States: An examination of the 29 May 1995 Great Barrington, Massachusetts, tornado. *Preprints, 22nd Conference on Severe Local Storms,* American Meteorological Society, 4-8 October, 2004, Hyannis, MA. http://ams.confex.com/ams/pdfpapers/81734.pdf, last accessed 5/22/2009.

Brooks, Harold, quoted in *Experts answer tornado questions, USA Today Online* http://www.usatoday.com/weather/wtwistqa.htm (last accessed 5/20/2009). Special tools used

Esau, I.N. and Lyons, T.J. 2002. Effect of sharp vegetation boundary on the convective atmospheric boundary layer. Agricultural and Forest Meteorology 114: 3-13.

Passe-Smith, M.S*. 2008.* Effect of Topography on Weak and Moderate Tornadoes. *Proceedings of the 2008 ESRI International Users Conference.* Online publication, http://proceedings.esri.com/library/userconf/proc08/papers/papers/pap_1396.pdf.

---------. 2006. Exploring Local Tornado Alleys for Predictive Environmental Parameters. *Proceedings of the 2006 ESRI International Users Conference.* Online publication, http://proceedings.esri.com/library/userconf/proc06/papers/papers/pap_1339.pdf

---------. 2005. Test-driving a tornado likelihood model. *Proceedings of the 2005 ESRI International Users Conference.* Online publication, http://proceedings.esri.com/library/userconf/proc05/papers/pap1185.pdf.

Peckham, S.E. and R.B. Wilhelmson, L.J. Wicker, and C.L. Ziegler. 2004. Numerical simulation of the interaction between the dryline and horizontal convective rolls. *Monthly Weather Review,* 132, No. 7, pp. 1792-1812.

Raddatz, R.L. and Cummine, J.D. 2003. Interannual variability of moisture flux from the prairie agro-ecosystem: Impact of crop phenology on the seasonal pattern of tornado days. Boundary-Layer Meteorology 106:283-95.

Weaver, C.P. and R. Avissar. 2001. Atmospheric disturbance caused by human modification of the landscape. *Bulletin of the American Meteorological Society* 82, No. 2, pp. 269-77.

Data sources (all last accessed in May, 2009)

Streams: http://nationalatlas.gov/atlasftp.html?openChapters=chpwater (national atlas)

USGS National Map Seamless Server (http://seamless.usgs.gov/): National Elevation Dataset for Arkansas.

State of Arkansas GeoStor GIS data server (http://www.geostor.arkansas.gov/Portal/index.jsp) : Spring 2004 image of land cover, Spring 1999 Landsat TM Image.

Tornadoes: Storm Prediction Center archived tornado data,
http://www.spc.noaa.gov/wcm/index.html#data

ESRI software/extensions/tools used

ArcInfo Desktop 9.3 (SP1)

Spatial Analyst toolbar (Slope, Aspect, Neighborhood analysis, zonal statistics, raster calculator, reclassify)

ArcToolbox (Projections and Transformations, Raster mosaic, Spatial Statistics: GWR, OLS Regression, Hot Spot Analysis, Spatial Autocorrelation (Moran's I))

Hawth's Tools, available online at http://www.spatialecology.com/htools/tooldesc.php (last visited May 20, 2009)