NUREG/CR-2358

Subregional Variability in Missouri Tornado Statistics

Covering the Period
April 1977 - June 1979

Prepared by S. W. Jamison, G. L. Darkow

Department of Atmospheric Science
University of Missouri

Prepared for
U.S. Nuclear Regulatory Commission

Available from

GPO Sales Program
Division of Technical Information and Document Control
U. S. Nuclear Regulatory Commission
Washington, D. C. 20555

Printed copy price: $4.50

and

National Technical Information Service
Springfield, Virginia 22161

NUREG/CR-2358
RB, R6

Subregional Variability in Missouri Tornado Statistics

Covering the Period
April 1977 - June 1979

Manuscript Completed: Septembe' 1981
Date Published: November 1981

Prepared by
S. W. Jamison, G. L. Darkow

Department of Atmospheric Science
University of Missouri
Columbia, MO 65211

Prepared for
Division of Health, Siting and Waste Management
Office of Nuclear Regulatory Research
U.S. Nuclear Regulatory Commission
Washington, D.C. 20555
NRC FIN B6196
Under Contract No. NRC-04-77-055

ABSTRACT

The reality of subregional variability in tornado occurrence density as evidenced in the county to county variability in Missouri is examined. Reported tornadoes for the period from 1916 through 1975 were used. Demographic and geographic factors known to impact on tornado reporting efficiencies and accuracies are related to county tornado report densities by step wise multiple linear regression techniques. The analysis suggests that over 75 percent of the county to county apparent variability in reported tornado densities in Missouri is explainable in terms of variability in population density, other related demographic variables and regional scale geographic factors. The remaining 25% of unexplained variance appears quite randomly distributed and may be due to random noise in the initial report data or to physical and geographic factors not included in the study. The analysis suggests that during the 1916-1975 period actual tornado occurrences in Missouri exceeded reported occurrences by at least 65%. Most of these reporting deficiencies are manifest in the reports prior to 1954, the start of press clipping service era. Although the population induced bias in reported tornado densities during the post 1954 reporting period were not examined in this study they are likely still present to a lesser extent and should be taken into account in interpreting subregional scale variability in reported tornado densities.

TABLE OF CONTENTS

LIST OF TABLES

Table		Page

LIST OF FIGURES

Acknowledgements

This research was conducted under Contract No. NRC-04-77-055 with the United States Nuclear Regulatory Commission with Dr. Robert F. Abbey, Jr. as technical monitor. The authors wish to thank Dr. Abbey for his helpful suggestions, guidance, and encouragement during the course of this research.

CHAPTER 1

INTRODUCTION

Tornadoes are among the most destructive of natural phenomena. The average number of deaths per year (1953-1972) ascribed to tornadoes was 113, and estimated average annual property damage for the same time period was $75 million. While other weather phenomena may exceed these figures (see Table 1), the tornado's high energy density, widespread distribution, erratic behavior, and potential for destructiveness has earned it a high place in the nation's "most feared disasters" list.

"A severe tornado event leaves a community momentarily stunned and disorganized and draws a response of the relative magnitude demanded in war." (Moore, 1958)

Statement of Problem

There is an obvious need to assess both absolute and relative risk from natural hazards such as tornadoes. Not only is there concern for the tornado risk at critical facilities such as nuclear power plants or nuclear materials handling facilities, but interest has also developed in the risk to less critical facilities, where protection of function and occupant protection are important. Hospitals,

Table 1

United States Losses Attributed to Some Weather Phenomena

Type of Storm	Average Annual Deaths	Average Annual Property Damage ($)
Tornado[a]	113	75 million
Lightning	150[b]	100 million[c]
Hail	--	284 million[d]
Hurricane[a]	75	500 million

Source: R. Davies-Jones and E. Kessler (1974)

Notes:

[a]Based on data from the Environmental Data Service, ESSA; applicable to period 1953-1972.

[b]Estimate based on data from National Center for Health Statistics (NCHS) applicable to period 1959-1965. Zegel (1967). Mogil (1977) estimates figures nearly double that reported by NCHS.

[c]Includes property damage by lightning caused building fires, $30,600,000 in 1967, according to Accident Facts, National Safety Council, Chicago, Ill., 1968, 96 pp. Other property loss includes forest fires, aircraft damage, disruption of electromagnetic transmissions and casualties to livestock. (ESSA 1969)

[d]Estimate for period 1958-1967 by Stanley Changnon, Illinois State Water Survey, Urbana, Ill. About 10% of Illinois losses represent property; the remainder is crop damage. (Changnon, Huff and Semonin, 1968)

fire stations, emergency operations centers, schools, office buildings, and residences make up this latter category.

Actuaries must make accurate estimations of damage potentials for insurance purposes.

Planners and designers need to make intelligent choices among such alternate courses of action as:

1. Ignoring the threat, if it is below some threshhold value.

2. Insuring against the threat.

3. Transfering the operation to an area of sufficiently low risk.

4. Construction of shelter areas (e.g., storm cellars) for occupants to enter upon receipt of tornado warnings.

5. Construction of a facility sufficiently strong as to survive contact with the strongest tornadic forces. This would include underground construction.

Costs and benefits for each such alternate course of action may be estimated, provided that an accurate estimate of the tornado risk level at each point in question is available. Once cost and benefit information is available, rational and objective decisions may be made.

Risk level for tornadoes is a function of frequency of occurrence per unit area and average area affected.

Review of Literature

Historically, there have been numerous attempts to quantify frequency of occurrence and/or area affected by tornadoes, and thus provide the basis for more accurate estimation of tornado risk levels. Court (1970) has summarized the more significant studies and/or representations of tornado incidence made between 1857 and 1970.

In general, the first studies were generalizations concerning entire regions. As more and better data became available, state by state comparison became the standard means of representing and comparing tornado incidence. Smaller areas, such as 5, 2, and 1 degree latitude/longitude quadrangles, and counties have all been used in an attempt to improve resolution and more accurately depict variations in tornado risk levels.

In his Climatology of the United States, Lorin Blodget (1857, p 403) wrote:

> The frequency and distribution of these tornadoes is a subject of practical interest, as in the case of the hurricanes. They occur over every part of the United States where the rain fall is abundant, and at the seasons of its greatest abundance. There are none on the great plains so far as known, at a distance from the Mississippi sufficient to reach the dry regions; they are most numerous in the Mississippi valley, and from this eastward they are quite equally distributed from Canada to Georgia. In the old forests, particularly of New York and Pennsylvania, the tracks of those which prostrated the older growth a century since may still be traced by the belt

> of trees of uniform size and peculiar aspect which grew up subsequently. The earth hillocks are thickly crowded over the soil of these belts, on all tracks where the soil readily turns up with the roots of trees. From the clue to frequency which such tracks give, these storms must be placed at very remote intervals for any one locality--the permanence of such a forest trace could be relied upon for at least five hundred years, and they now exist in only a few conspicuous lines, averaging fifty miles distance perhaps, and lying in threads of thirty to two hundred rods in width, and ten to fifty miles in length. The tracks are so narrow that great frequency would be required to mark the entire surface, yet it must be concluded that they are not more frequent than the hurricanes, while the space they cover is the smallest thread in comparison to those gigantic displays of atmospheric disturbance.

From Blodget's information, we may compute the area destroyed per tornado track (Pennsylvania/New York State) as being between 0.94 square mile (30 rods wide and 10 miles long) and 31.25 square mile (200 rods wide and 50 miles long). If these tracks occurred once per 50-mile square block, and represent one hundred years of history, the area destroyed per 10,000 square miles per year lies between 0.04 square miles and 1.25 square miles. These figures are surprisingly close to modern estimates. Howe (1974) lists average area destroyed per tornado as 0.37 square miles in New York State and 0.27 square miles in Pennsylvania. Kessler and Lee (1978) list tornadoes per 10,000 square miles per year as 0.70 for New York State and 1.22 for

Pennsylvania. From these figures we can estimate area destroyed per 10,000 square miles per year as 0.26 square mile in New York State and 0.32 square mile in Pennsylvania.

Only one other use of this tree windfall technique is on record. Burley and Waite (1965) cite a report by Increase A. Lapham (1872) in which he found, "About one chance in probability in 10,000 that any particular farm of 160 acres (in any one year) will be visited" by a tornado. This is 4 tornadoes per 10,000 square miles per year, which may be compared with a recent figure of 1.92 (Kessler and Lee, 1978).

The first comprehensive tornado tabulation was the "Report on the Character of Six Hundred Tornadoes" by John Park Finley (1884). It presented a detailed table of the time, place, and characteristics of 600 tornadoes and a map summarizing the total number of tornado reports, by states, and of the years for which information was available in each.

Cleveland Abbe (1888) recomputed Finley's figures to take into account state area and number of years for which tornadoes were reported. Finley continued to tabulate and publish tornado incidence/risk data. Court (1970) lists 7 additional publications: (1887, 1888a, 1888b, 1889, 1890, 1925, 1932).

Preston C. Day (1930) included a map by Keith B. Allen with data from 1481 tornadoes for the years 1916-1928, plotted in equal area squares of 10,000 square miles. Each track was counted in each square over which it passes, and the totals divided by 13 to give true annual tornado incidence per unit area. This was proclaimed by Court (1970) as the only such representation ever published.

Brown and Roberts (1935) reported a statistical study of 52 years of U.S. Weather Bureau data which included 3911 tornadoes. They stated:

> In addition, new studies for the entire time have been made of the hourly occurrence, the frequency, and the damage possibilities of different areas, and all the tornadic tracks of the half century have been plotted on one map. These latter studies are already serving as a basis in part for revision of tornado insurance rates.... One attempt to get a truer appreciation of the tornado-intensity of any given area was to take the total area of each state, in square miles, and divide by the number of tornadoes occurring within its boundaries.... In addition, the calculation of the "net" areas was made by eliminating areas not subject to tornadoes.

They eliminated 50 per cent of Michigan, 33 per cent of Nebraska, 50 per cent of Texas, 10 per cent of Tennessee, and 60 per cent of Minnesota. In a further report, Brown and Roberts (1937) submitted a county distribution map, showing "The results of the study upon smallest available unit". They were struck by the variability from county to county and further noted:

> It is obvious that the more densely populated the region is, the more complete and accurate will be the returns of tornado information, and that probably there is a more or less general relationship between the density of population and the reporting of tornadoes.

Beginning in 1951, a number of tornado data tabulations have been made using latitude/longitude quadrangles. (See Court (1970) pp 20-25). Summaries by coordinate quadrangles offer greater detail than summaries by states. However, area normalization should be made, since quadrangle area decreases as one proceeds poleward. The first such tornado incidence map by Fawbush, et al., (1951) does adjust for area and its figures are expressed as "total number of tornadoes per 50-mile square reported in the period 1920-1949. Unfortunately most other quadrangle summaries have not been area normalized.

Thom (1963) said that the probability of a tornado striking a point was the ratio of the mean area covered by tornadoes per year to the area over which the tornadoes might occur. He stated:

> If we take the mean path area of a tornado to be $\bar{z}$ in square miles and the mean number of tornadoes per year to be $\bar{t}$, then the average area covered by tornadoes per year will be $\bar{t}\ \bar{z}$...then the mean probability of a tornado striking a point in any year in a 1 degree square with $\bar{z}$, $\bar{t}$, and area A is
>
> $$P=\bar{z}\ \bar{t}\ /A$$

> If it is assumed further that $\bar{z}$ is invariant, we may substitute the value for $\bar{z}$ previously obtained and [the previous] equation becomes
> $$P = 2.8209\bar{t}/A$$

Skaggs (1970) applied Thom's procedure to reports for each of the states having the most tornadoes, 1950-1964. For Iowa, the "expected value of area affected by a tornado" was 2.82 square miles, as Thom had found. However, for Kansas, he found 3.12 square miles, for Minnesota 5.38 square miles and in the remaining 9 states, areas ranging from 2.66 in Michigan to 0.54 in New York and Pennsylvania.

Howe (1974) found the log-normal distributions (used by Thom to calculate his invariant $\bar{z}$ = 2.8209) did not apply in eight states which "comprised more than half the mid-continent area and about 40% of the reported tornadoes," and that damage area per tornado varied widely.

Reed (1971) presented data on distributions of deaths and losses in the U.S. He assumed that property values exposed to tornado damage risk were roughly proportional to annual bank deposits.

A detailed review of tornado risk analysis methodologies by Markee, et al. (1974), Dames & Moore (1975), Wen and Chu (1973), Garson, et al. (1975), McDonald, et al. (1975), and Abbey and Fujita (1975) were presented by Abbey (1976). Table 2 contains the summary of these efforts.

Table 2

Statistical Representation of Tornado Data in

Tornado Risk Models

Model	Tornado Data Representation
Markee, et al. (1974)	Mean damage area, Lognormal intensity distribution.
Dames & Moore (1975)	Intensity distribution represented by Gaussian, Gamma or Extreme Value type II. Occurrence modeled by Poisson or Weibull process.
Wen and Chu (1973)	Bivariate Log-normal distribution for area and intensity.
Garson, et al. (1975)	Multivariate log-normal joint probability, density function for intensity, path length, and path width.
McDonald, et al. (1975) (earlier Model)	Empirical area-intensity relationship. Empirical occurrence-intensity relationship.
Abbey and Fujita (1975)	Emprical area-intensity relationship.

Source: McDonald and Abbey (1977)

Sims and Baumann (1972) stressed indigenous psychology as a principle cause of high tornado death rates in the South. This view was contested by Davies-Jones, et al. (1973).

Kessler and Lee (1976) normalized coterminous state statistics of tornado frequency, deaths, and damage in terms of state areas, population, and property valuation to give insight into joint variability of tornado frequency and intensity and to improve estimates of the tornado hazard distribution. Kessler and Lee (1978) compared and related damage and deaths from tornadoes. In contrast with previous studies, little support was found for the concept of tornado deaths in southern states exceeding expectations based on tornado frequency and severity there. They found that tornadoes tended to be more intense east of the region in which they are most frequent.

Sadowski (1965) used 2 degree quadrangle tornado statistics to examine potential casualties from tornadoes.

In addition to these studies relating to tornado variability/risk level on a large scale, several investigators have examined the variability within a smaller area, usually a single state.

Snider (1977) examined Michigan tornado statistics. He concluded:

> 1. Present observing methods do not detect all tornadoes. The true number of tornadoes is several times greater that in any set of statistics so far obtained. Statistics from urban areas are more reliable than those from rural areas.
>
> 2. Tornadoes are more likely to touchdown (i) on smooth surfaces rather than on rough surfaces, (ii) in small towns rather than large cities, and (iii) over small bodies of water rather than adjacent land areas.

Asp (1956) studied Arkansas tornado statistics. He found more tornadoes reported in the more populated areas and the least in the rural, mountainous or forested sections. He found a positive correlation ($r = 0.45$) between the number of tornadoes and the population by counties. He stated:

> While it appears some areas in the state may be more subject to tornadoes than others, it is believed that the number of tornadoes of record so far is insufficient for an adequate sample on which to base definite statements.

Vaiksnoras (1972) tabulated tornado occurrences in Tennessee (1916-1970). He found a population bias and stated:

> The areal distribution of tornadoes common to Tennessee indicates that these storms are quite rare over terrain elevations above 2,000 feet and those few reported at higher elevations were mostly in valleys on the east side of mountain ranges. If one divides the state along the 86th Meridian, the western half has reported three times as many tornadoes as observed in the eastern half of the state.

Dailey (1970) discovered apparent differences in tornado distributions across Pennsylvania, and reasoned that:

> ...High tornado concentration areas are found in certain locales for the same reasons the urban centers developed there, simply as a result of the meteorological, geographic and orographic features of the area. The nearness to a water supply, a relatively level or rolling topography, and a mid-latitude climate are necessary ingredients for a large urban center--the same as for high tornado frequency.

He made no attempt to validate this interesting line of reasoning.

Wilson and Changnon (1971) summarized data on Illinois tornadoes in an update of an earlier study by Changnon and Stout (1957). Concerning spacial distribution they said:

> Accurate analysis of the spacial distribution of tornadoes within a given area such as a state is difficult and somewhat biased. The primary method of tornado observation, the personal observation, is dependent upon people being in the area of a tornado when it occurs, or being able to determine by storm damage patterns the existence of a tornado. The nonhomogeneous distribution of population in Illinois compounds the problem, as the chances of a tornado being reported, especially if it causes little damage, are somewhat proportional to the population density, particularly in rural areas...
> ...reveals a high degree of variability among counties in the number of tornadoes reported. Some of this variation is due to county size differences and some is due to differences in population among counties, with counties having

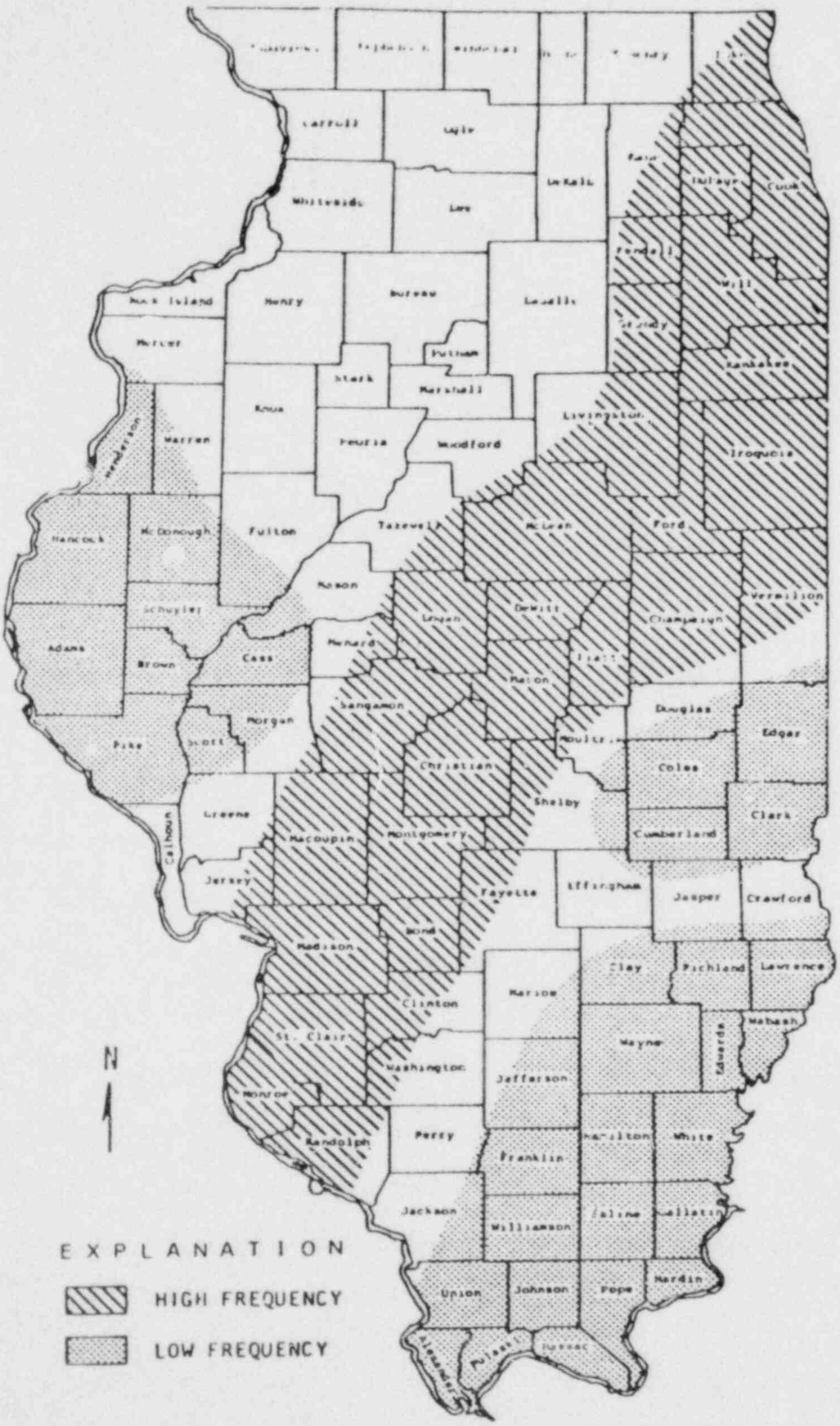
> higher population density usually exhibiting larger numbers of tornadoes. Chicago and St. Louis are evident as high occurrence areas...
> ...The most favored region for tornadoes extends from the southwestern to northeastern portions of the state. [Figure 1]... Areas with low tornado incidence are found in western Illinois and in the eastern part of the state, and in the southeast corner. Most of northern and northwestern Illinois is characterized by moderate frequencies.

Statistical studies of tornadoes in Iowa include the work of Spohn and Waite (1962) and Eshelman and Stanford (1977). Spohn and Waite (1962) noted:

> A certain amount of bias is evident in these data, as in those for any area, because of the uneven distribution of human population and related interest, and indirectly, perhaps, the distance from the collection center.

Eshelman and Stanford (1977) made an exhaustive survey of severe storm events in Iowa during 1974 and found much greater storm damage than revealed in the news media or in the Department of Commerce National Oceanic and Atmospheric Administration (NOAA) publication Storm Data. The latter reported 27 tornadoes in Iowa in 1974. By contrast, the survey by Eshelman and Stanford indicated 81 separate tornado touchdowns, a three to one ratio. These results are in complete agreement with the conclusions of Snider (1977) already cited. McDonald, et al. (1977) however, state; "Public awareness and population spread

Figure 1. Areas of Relative Tornado Frequency in Illinois



assure that virtually all significant tornadoes are now recorded". The validity of this statement obviously depends on one's definition of "significant".

Carter (1970), in Georgia Tornadoes lists examples of spacial variability of tornado incidence statistics in Georgia, then states:

> These figures, although interesting, do not give an accurate picture of tornado distribution over Georgia. The number of reported tornadoes is affected by the size of the county and, more important, by the population density. There is little doubt that many tornadoes occur in the more remote and sparsely populated areas but these are never reported.

Agee (1970) studied statistics on Indiana tornadoes. He attempted to remove the problem of the population bias in tornado reporting by the use of a tornado index. He noted the effect of terrain on the areal distribution of tornadoes and suggested further study of these effects. Idso (1975) examined the effects of topography on vortex wind phenomena. Fujita (1973) studied the effects of surface friction on tornadoes, and found that the surface friction tends to diminish the intensity of the laboratory tornado. He postulated that the addition of heat over a large city further reduces the fury of a tornado, often down to zero, thus dissipating the entire funnel.

According to Darkow (1976)

> Within the larger scale (national-regional) variations in reported density, there continues to appear considerable subregional variability. This apparent subregional variability carries down to scales on the order of 50 km or less (county scale). It is not unusual to observe reported tornado density differences of a factor of 2 to 5 and occasionally as high as 10 between adjacent counties in areas of fairly uniform population density and physiographic characteristics...
> It is well known that the spacial variability in the number of reported tornadoes has been fraught with a marked but uncertain impact due to variations in population density, reporting procedures, reporting personnel, distance from NWS stations, etc.

Purpose of Thesis

The purpose of this thesis is to attempt to answer the questions:

1. Is the apparent subregional variability in tornado strike probabilities real?

2. If differences exist, are they significant?

3. If differences exist, what are the causes?

CHAPTER 2

DATA AND METHODS

Ideally it would be desirable to work with a data set listing actual tornado occurrences over an extended period of time. Such actual occurrences may be divided into two classes, detected and undetected. Intuitively, the percentage of undetected tornadoes should rise in sparsely populated regions.

The detected tornado class may be further subdivided into reported and unreported subclasses. Again, we might expect the percentage of detected, but unreported tornadoes to be higher in areas more distant from reporting stations, or where there is an absence of damage to persons or property.

A list of detected-reported occurrences may, unfortunately, include duplicate reports or reports of wind damage erroneously classified as tornadic damage.

The data set used in this study consisted of tornado reports, by county, for Missouri for the period 1916-1975. This data set has received a more thorough screening than any comparable data set. This screening is the result of the combined efforts, extending over many years, of the Missouri State Climatologist's office; the Atmospheric

Science Department of the University of Missouri-Columbia; NSSFC-NWS, Kansas City; the State Forecast office of NWS, St. Louis; and others. It is believed to be relatively free of duplicate reports and non-tornadic wind damage. Prior to 1954, the number of tornadoes reported per year is less than after that time. This is believed due largely to the use of press clippings and increased public awareness since the mid 1950's.

Figure 2 is a graphic representation of the number of tornadoes reported (1916-1975) in each county. Since Missouri counties vary in size from 267 square miles in Worth county to 1183 square miles in Texas county, it is necessary to convert the data to an equal area basis for comparison purposes. This was termed SDEN for Storm DENsity and is the number of tornadoes per 1000 square miles, calculated by:

SDEN = reported tornadoes x 1000 / County area

SDEN varied from 1.31 tornadoes per 1000 square miles in Wayne county to 64.13 tornadoes per 1000 square miles in St. Louis county. Figure 3 shows the value of SDEN in each county.

As a first approximation, assume that SDEN in each county differs by a random amount from the state average value of 14.35 tornadoes per 1000 square miles. This value

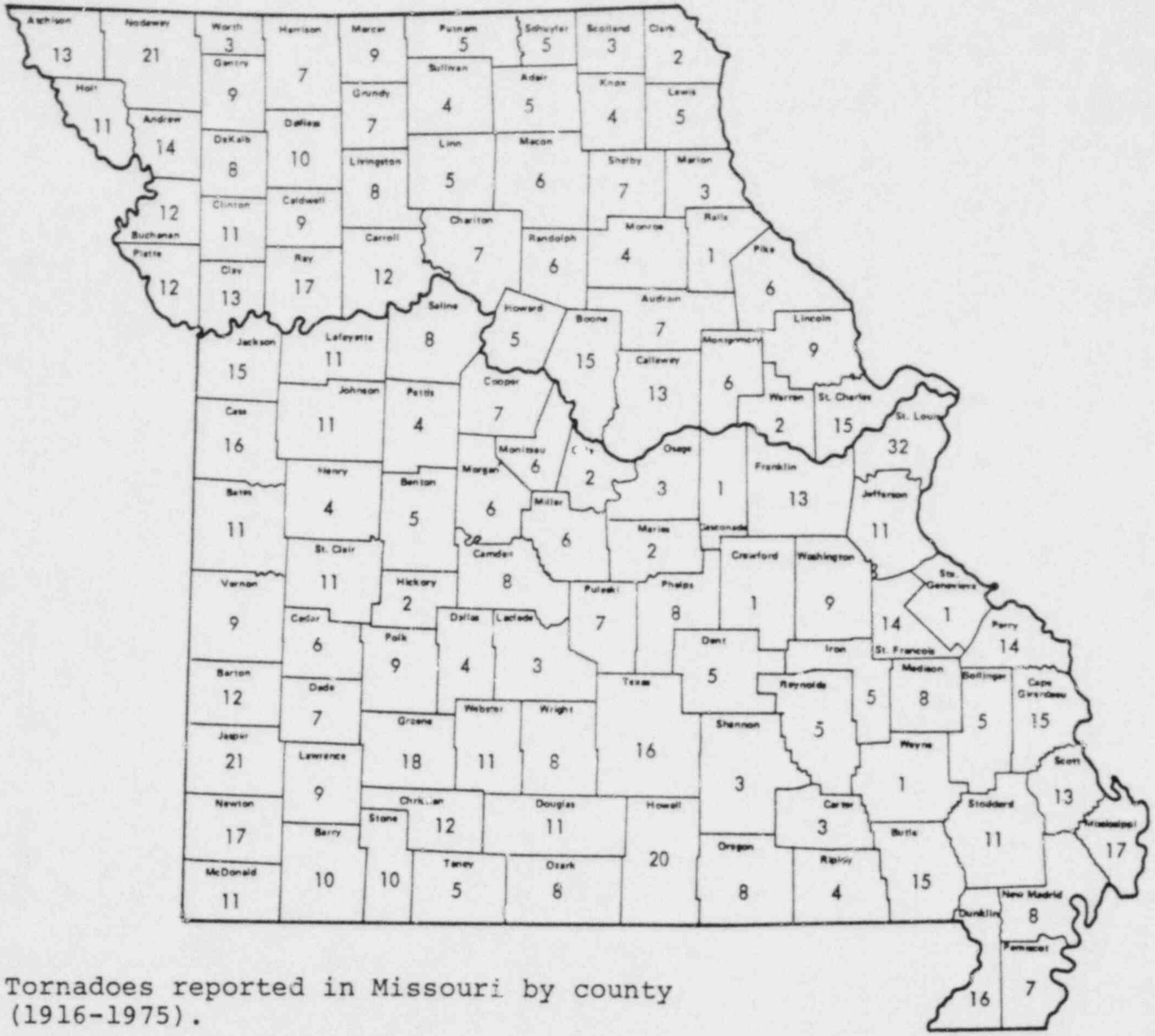


Figure 2. Tornadoes reported in Missouri by county (1916-1975).

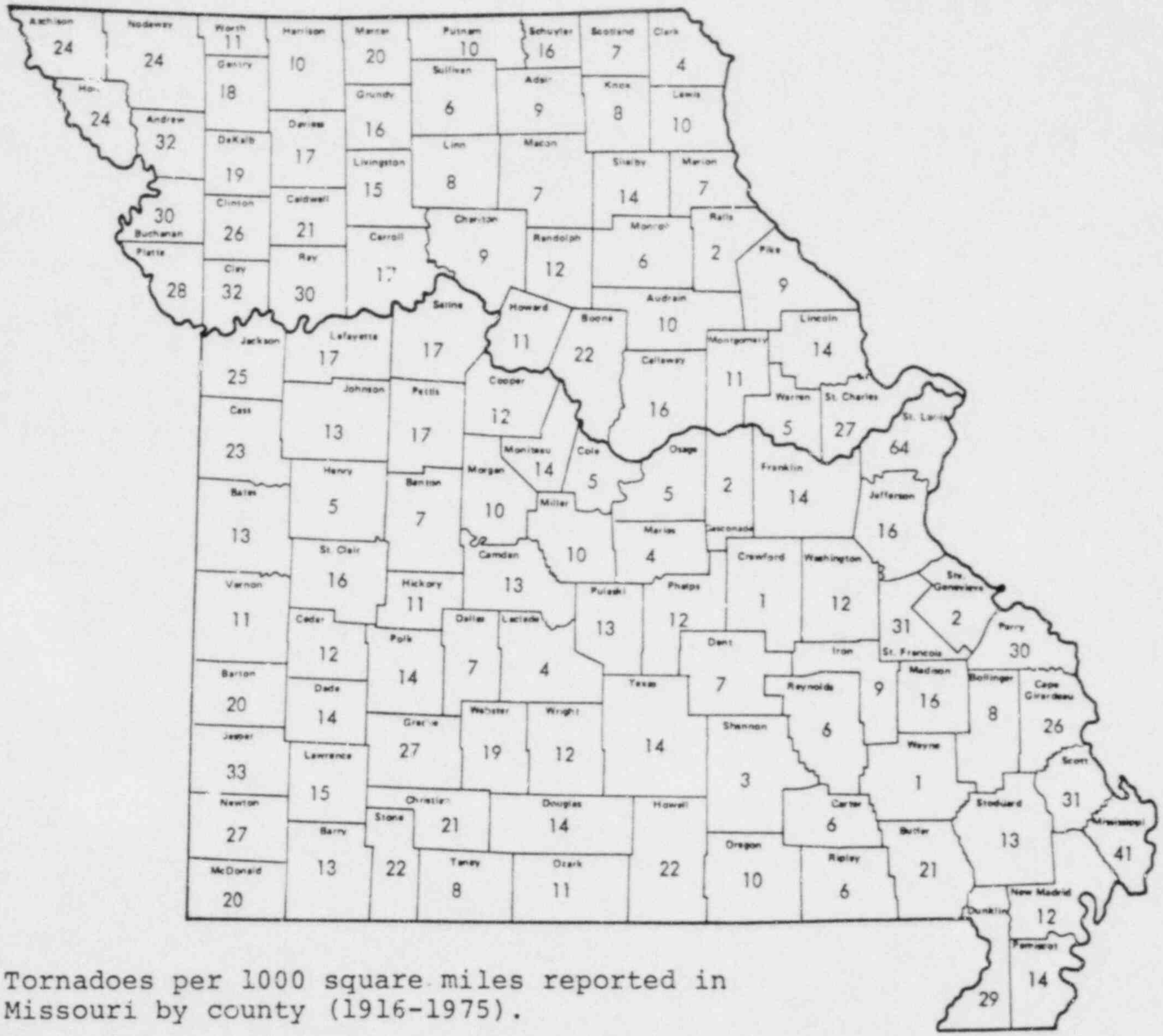


Figure 3. Tornadoes per 1000 square miles reported in Missouri by county (1916-1975).

was calculated by taking the total number of tornadoes in the data sample, multiplying by 1000 and dividing by the state area.

This corresponds to the model

$$SDEN = \overline{SDEN} + e \qquad \text{(model 1)}$$

Figure 4 shows the distribution of the departure terms in model 1 by county. Those less than -6.0 are shaded and those greater than 6.0 are diagonally lined. If the departures were purely random and the sample was sufficiently large, their distribution would be expected to be approximately Gaussian or normal, since the central limit theorm indicates the asymptotic distribution of this sum would be normal. The hypothesis Ho: distribution is Gaussian vs Ha: distribution is non-Gaussian can be tested by means of the chi-square goodness of fit test.

Divide the 114 observations into 6 classes with expected value 19, count the number in each class, and compute the statistic $(o_i - e_i)^2/e_i$ for each class. The test statistic is distributed, approximately, as a chi-square, where the $\Sigma(o_i - e_i)^2/e_i$ sum is over the classes. If the random variables are approximately normally distributed, this statistic will be approximately the chi-square distribution with (k-3) degrees of freedom, where k is the number of classes used in computing the chi-square

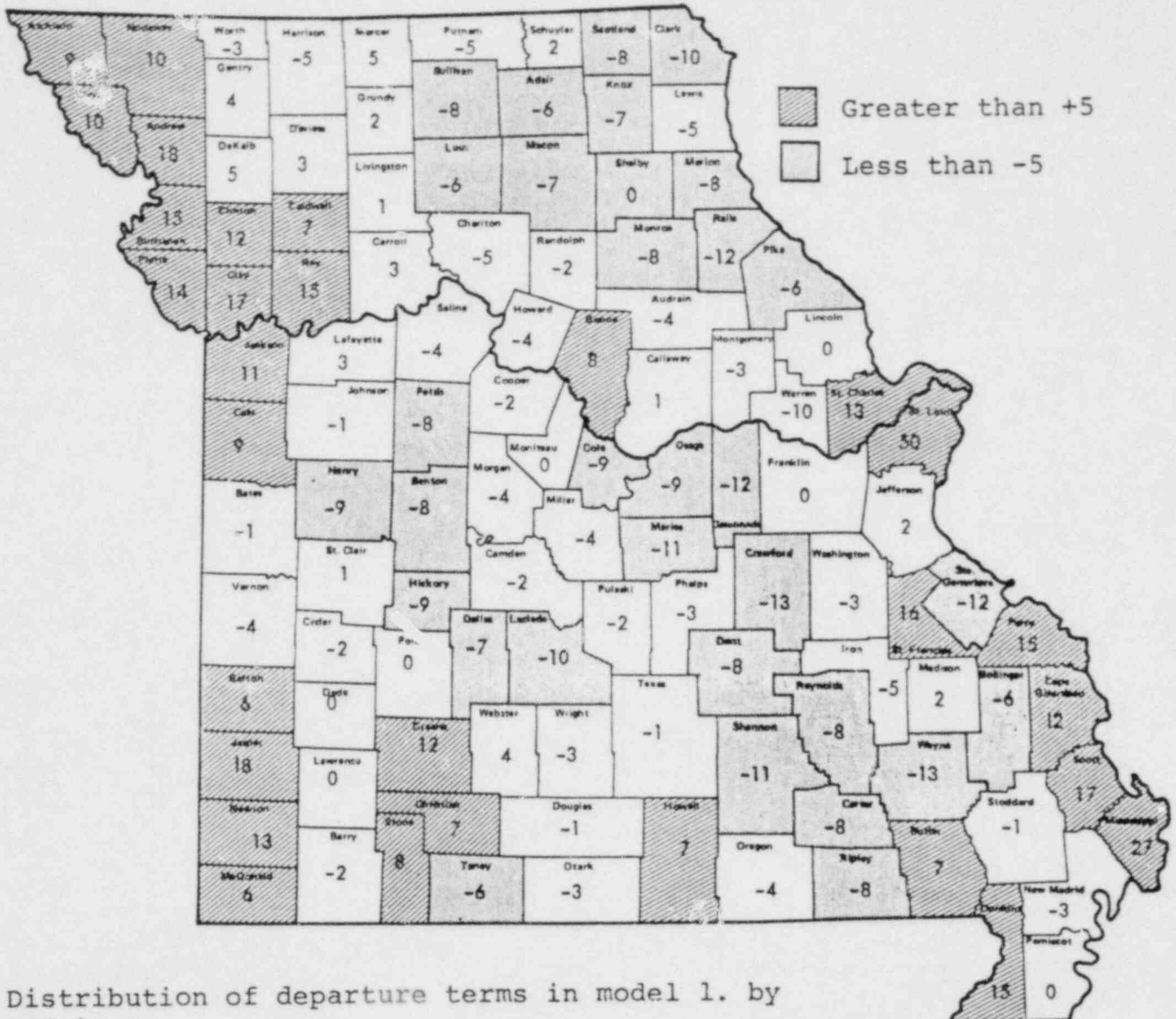


Figure 4. Distribution of departure terms in model 1. by county.

statistic. If the data come from any other distribution, the observed and expected values will tend to have poor agreement and the computed chi-square statistic becomes large. Large values of the chi-square statistic indicate rejection of the hypothesis of normality. Table 3 shows the results of this test.

Table 3

Chi-Square Goodness of Fit Statistics

for Model 1 Departure Terms

Class Limits	Observed	Expected	$(o_i - e_i)^2/e_i$
<5.065	12	19	2.5789
5.056 to 10.213	29	19	5.2632
10.213 to 14.355	27	19	3.3684
14.355 to 18.497	13	19	1.8947
18.497 to 23.655	12	19	2.5789
>23.655	21	19	.2105
	114	114	15.8947

The critical value for the chi-squared distribution with 3 degrees of freedom and a significance level of .005 is 12.941. Since our observed value is larger, we may reject Ho at the .005 level and conclude that the distribution of the departure terms is not Gaussian. The rejection of the null hypothesis indicates that the sample is not large enough for the distribution of SDEN to be indistinguishable from a normal distribution.

There are several statistical tools available that may help us gain insight into possible sources of the spacial variability of the storm density. One of these tools is that of regression analysis. If we wish to test the proposition that some independent variable X has a linear relationship with the dependent variable Y (in this case, the "Y" of interest is denoted as SDEN), we may fit the regression equation $Y = B_0 + B_1(X) + e$ and examine its ability to reduce the variations in SDEN through the computation of the coefficient of determination R^2.

$R^2 = 1\text{-}SSE/SSTO$ where

$SSE = (y_i - \hat{y})^2$, $SSTO = (y_i - \bar{y})^2$, y_i is the value of the dependent variable y on the ith observation, $\hat{y}$ is the value of the fitted value of y at the level X for the ith observation, and $\bar{y}$ is the mean of the y_is in the sample.

R^2 may be interpreted as the proportionate reduction of total variation associated with the use of the independent variable X. Thus, the larger is R^2, the more is the total variation of Y reduced by introducing the independent variable X. The lower limit of R^2 is 0, when there is no linear relation between X and Y in the sample data and the independent variable X is of no help in reducing the variation in the observations Y. The upper limit for R^2 is 1, which occurs when all observations fall on the fitted regression line. Here the independent variable X accounts for all variation in the observations Y.

This technique assumes that a linear relationship exists. If in fact a real, but nonlinear relationship exists, the apparent percentage of variability explained (R^2) will be less than actual. In this case an appropriate transformation may linearize the regression function. Examples of such transformations often found useful include logarithmic and reciprocal transformations. (See Natrella, 1963).

There are various nonparametric methods utilizing rank correlation that do not depend on a linear relationship.

According to Conover (1971),

> By tradition, a measure of correlation between X and Y should satisfy the following requirements, in order to be acceptable:
> 1. The measure of correlation should assume only values between -1 and +1.
> 2. If the larger values of X tend to be paired with the larger values of Y, and hence the smaller values of X and Y tend to be paired together, then the measure of correlation should be positive, and close to +1.0 if the tendency is strong. Then we would speak of a positive correlation between X and Y.

3. If the larger values of X tend to be paired with the smaller values of Y and vice versa, then the measure of correlation would be negative and close to -1.0 if the tendency is strong. Then we would say that X and Y are negatively correlated.
4. If the values of X appear to be randomly paired with the values of Y, the measure of correlation should be fairly close to zero. This should be the case where X and Y are independent and possibly some cases where Y and Y are not independent. We then say that X and Y are uncorrelated, or have no correlation or have correlation zero.

Numerous measures of correlation have been devised which satisfy these requirements for acceptability. A survey article by Kruskal (1958) discusses many of these. The three measures cited in this and following chapters are: Pearson's product moment correlation coefficient, Spearman's Rho, and Kendall's Tau. The following discussion relating to these three is paraphrased from Conover (1971).

The most commonly used measure of correlation is Pearson's product moment correlation coefficient, denoted by r and defined by:
$r = \Sigma(x_i-\bar{x})(y_i-\bar{y})/(\Sigma(x_i-\bar{x})^2 \ \Sigma(y_i-\bar{y})^2)$ where $\bar{x}$ and $\bar{y}$ are the sample means. r however, is a random variable, and as such has a distribution function. Unfortunately, the distribution function of r depends on the bivariate distribution of (X,Y). Therefore, r has no value as a test statistic in nonparametric tests unless the distribution of (X,Y) is known.

The measure of correlation as given by Spearman 1904) is designated by Rho, and defined (if no ties) by $Rho = (R(x_i)-(n+1)/2)(R(y_i)-(n+1)/2)/n(n^2-1)/12$ where $R(x_i)$ is the rank of x_i with $r(x_i)=1$ for the smallest x_i observed and $R(y_i)$ is the rank of y_i, with $R(y_i)=1$ for the smallest y_i observed.

The Spearman rank correlation is sometimes used as a test for independence between two random variables, although it is insensitive to some forms of dependence.

The last measure of correlation considered is that of Kendall's Tau. Like Spearman's Rho, it is based on the rank order of the observations and the distribution of the measure does not depend on the distribution of X and Y if they are independent and continuous. The distribution approaches the normal distribution quite rapidly, so that the normal approximation is better for Kendall's Tau than it is for Spearman's Rho, when the null hypothesis of independence between X and Y is true.

Kendall's Tau is defined by the equation $Tau = N_c - N_d/n(n-1)2$ where N_c denotes the number of concordant pairs and N_d denotes the number of discordant pairs. Two observations are called concordant if both members of one observation are larger than their respective members of the other observation. A pair of observations is called discordant if the two numbers in one observation differ in opposite directions (one negative and one positive) from the respective members of the other observation. Pairs with ties between respective members are neither concordant nor discordant.

If all pairs are concordant, Kendall's Tau equals 1.0. If all pairs are discordant, the value is -1.0. Kendall's Tau may also be used as a test statistic to test the null hypothesis of independence between X and Y.

All three correlation coefficients were calculated for tornado density with each of the independent variables considered. If one or more of these coefficients indicated a statistically significant relationship, it was included in a simple linear regression model. Sometimes the correlation coefficients indicated significant correlation, but the r^2 value computed for the linear regression model indicated less statistical association.

In this event, non-linear least squares procedures may be employed; or data transformations on the independent variable may linearize the relationship. In this study, data transformations were employed as the method of choice.

Chapters three and four describe the results of this process for each of several classes of independent variables. Chapter five examines the effect of combining factors into multiple regression models.

CHAPTER 3

DEMOGRAPHIC FACTORS

Independent variables which measure population related or caused statistics are grouped under the heading "demographic factors". They include population density, wealth or value at risk, distance to reporting station distances and education level.

Population Factors

As noted in the literature review, many investigators have commented on an apparent relationship between population density and reported tornado density. Such a relationship is reasonable, since a tornado occurring in a sparsely inhabited region would have a low probability of being observed or detected. With population increases, the probability that a tornado event would be detected should rise until such time as no significant gaps in observer coverage exist. Intuitively then one might expect that even in an area of uniform tornado occurrence density one might observe:

1. Low population densities to be associated with low reported tornado densities.

2. High population densities to be associated with high reported tornado densities.

3. A non-linear relationship between population density and reported tornado density, given human propensity to unevenly populate an area. By the time all areas have a population density sufficiently high to detect essentially all tornadoes, some subareas have achieved large "surpluses" of population density that contribute nothing to improved detection.

From U.S. Census data (Goldfield, 1967) taken 1960 to 1964, several population density measures were computed for each county.

These included:

1. Total population density (DENT)

   DENT = POPT/AREAT

2. Effective population density (EDEN)

   EDEN = (AREAU x FU) + (AREAR x FR)/AREAT

3. Rural population density (DENR)

   DENR = POPR/AREAR

4. Home density (HDEN)

   HDEN = Homes/AREAT

5. Family density (FDEN)

   FDEN = Families/AREAT

where:

AREAU = urban area

AREAR = rural area

AREAT = total are

DENU = urban population density, defined as POPU/AREAU

FU = either POPU or M, whichever is smaller

FR = either POPR or M, whichever is smaller

POPU = urban population

POPR = rural population

POPT = total population

M = maximum effective value, set at values from 50 to 200.

Correlation coefficients and associated significance levels are in Table 4. A very high significance level was obtained for all of the measures tested. Rural population density and effective population density had the highest correlation coefficients of any of the measures. Of these two measures, rural population density was selected since it was easier to compute and had very nearly identical correlation coefficients with effective population density. Pearson's, Spearman's and Kendall's correlation coefficients for EDEN6 (M=100) and DENR were .972, .997, and .964, respectively.

Linear regression analysis techniques were applied to the model

$$\text{SDEN} = B_0 + B_1(\text{DENR}) + e \qquad \text{(model 2)}$$

Table 4

Correlation Coefficients for SDEN with Various

Population Density Measures

Population Density Measure	Pearson's r	Pearson's Sig. Lvl	Spearman's RHO	Spearman's Sig. Lvl	Kendall's TAU	Kendall's Sig. Lvl
DENT	.543	.0001	.519	.0001	.368	.0001
DENR	.619	.0001	.575	.0001	.405	.0001
EDEN1 (M=50)	.597	.0001	.576	.0001	.405	.0001
EDEN2 (M=60)	.579	.0001	.574	.0001	.403	.0001
EDEN3 (M=70)	.576	.0001	.575	.0001	.404	.0001
EDEN4 (M=80)	.573	.0001	.576	.0001	.405	.0001
EDEN5 (M=90)	.582	.0001	.576	.0001	.406	.0001
EDEN6 (M=100)	.596	.0001	.576	.0001	.407	.0001
EDEN7 (M=125)	.621	.0001	.573	.0001	.404	.0001
EDEN8 (M=150)	.633	.0001	.574	.0001	.405	.0001
EDEN9 (M=200)	.638	.0001	.573	.0001	.403	.0001
FDEN	.534	.0001	.537	.0001	.379	.0001
HDEN	.517	.0001	.534	.0001	.377	.0001

which resulted in an R^2 value of .383. Several data transformations of the independent variable were tried in an attempt to increase the R^2. The transformed independent variables were all of a form such that population density difference had a diminishing impact on reported tornado density as population densities increased. Results are shown in Table 5.

Table 5

R^2 Values Corresponding to Various Transformed Population Variables

Transformed Independent Variable	R^2
$(DENR)^{.2}$	.407
$(DENR)^{.3}$	.408
$(DENR)^{.4}$	.407
(DENR)$^{.}$[illegible]	.405
Log(DENR)	.403
DENR x exp(.001 x DENR)	.380
DENR x exp(.002 x DENR)	.377

The transformed variable with the highest R^2 value was used in model 3.

$$SDEN = B_0 + B_1(DENR)^{.3} + e \qquad \text{(model 3)}$$

Residuals from such a model represent the difference between the actual tornado density and that predicted by the model. Residuals from model 3 are shown in Figure 5. The model underestimates tornado densities along the western edge of the state, especially in the northwest, and along the Mississippi from St. Louis County to Mississippi County. The model overestimates tornado density over much of the remainder of the state, especially in the south-central and northeast portions.

Significant improvements in the model require incorporation of additional factors.

Distance to Reporting Station

Court (1970), Dames & Moore (1975), and others have noted that tornado reporting is dependent on the proximity of National Weather Service (NWS) observational facilities. This is plausible, since it is less convenient and more expensive to report or investigate tornado occurrences at greater distances.

Figure 6 shows the Missouri counties assigned to each of five NWS offices. The average distance from the assigned NWS station to each county was computed and designated ASD.

Figure 7 shows the areas closest to each of the NWS offices without regard to county boundaries. The average

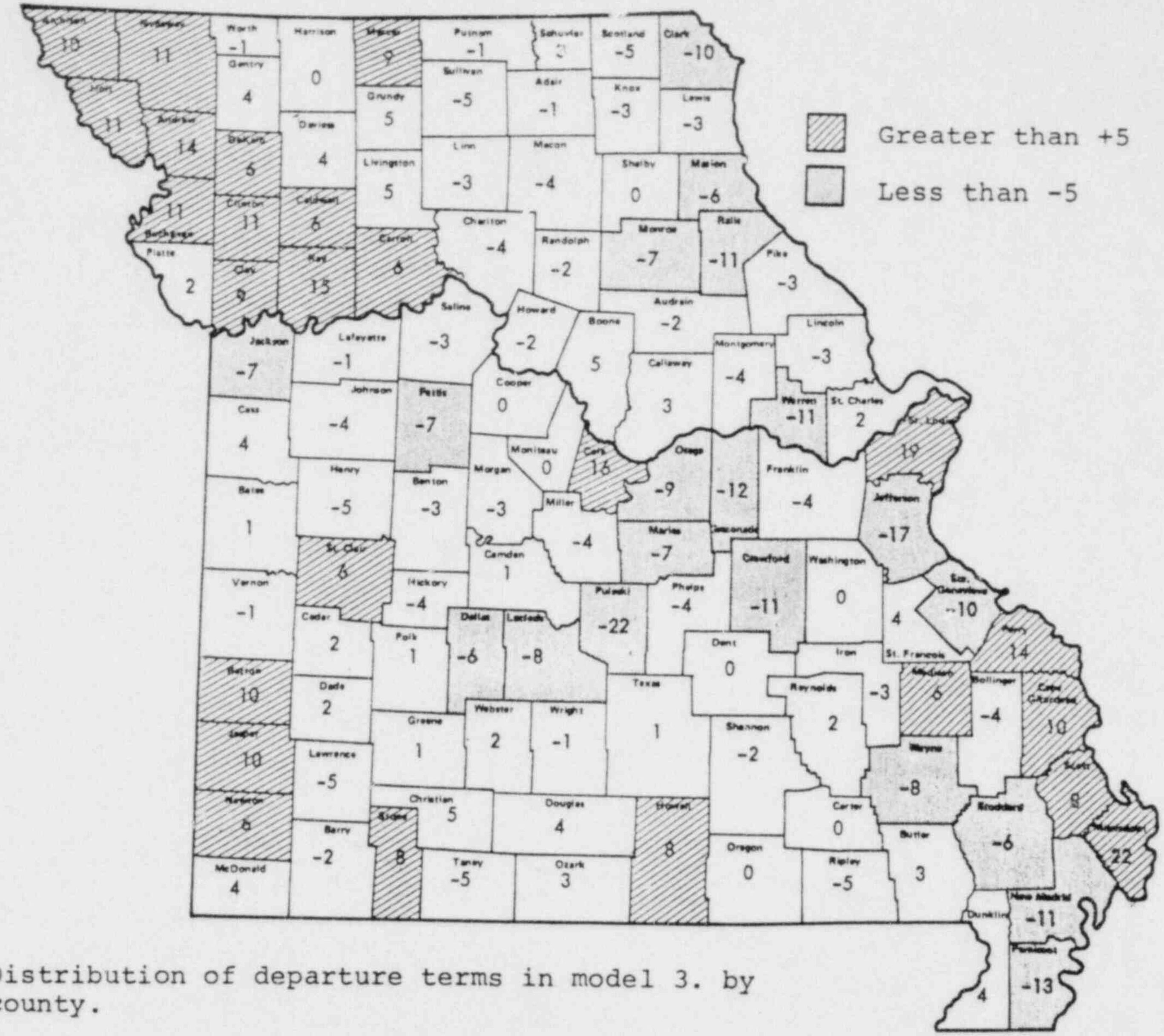


Figure 5. Distribution of departure terms in model 3. by county.

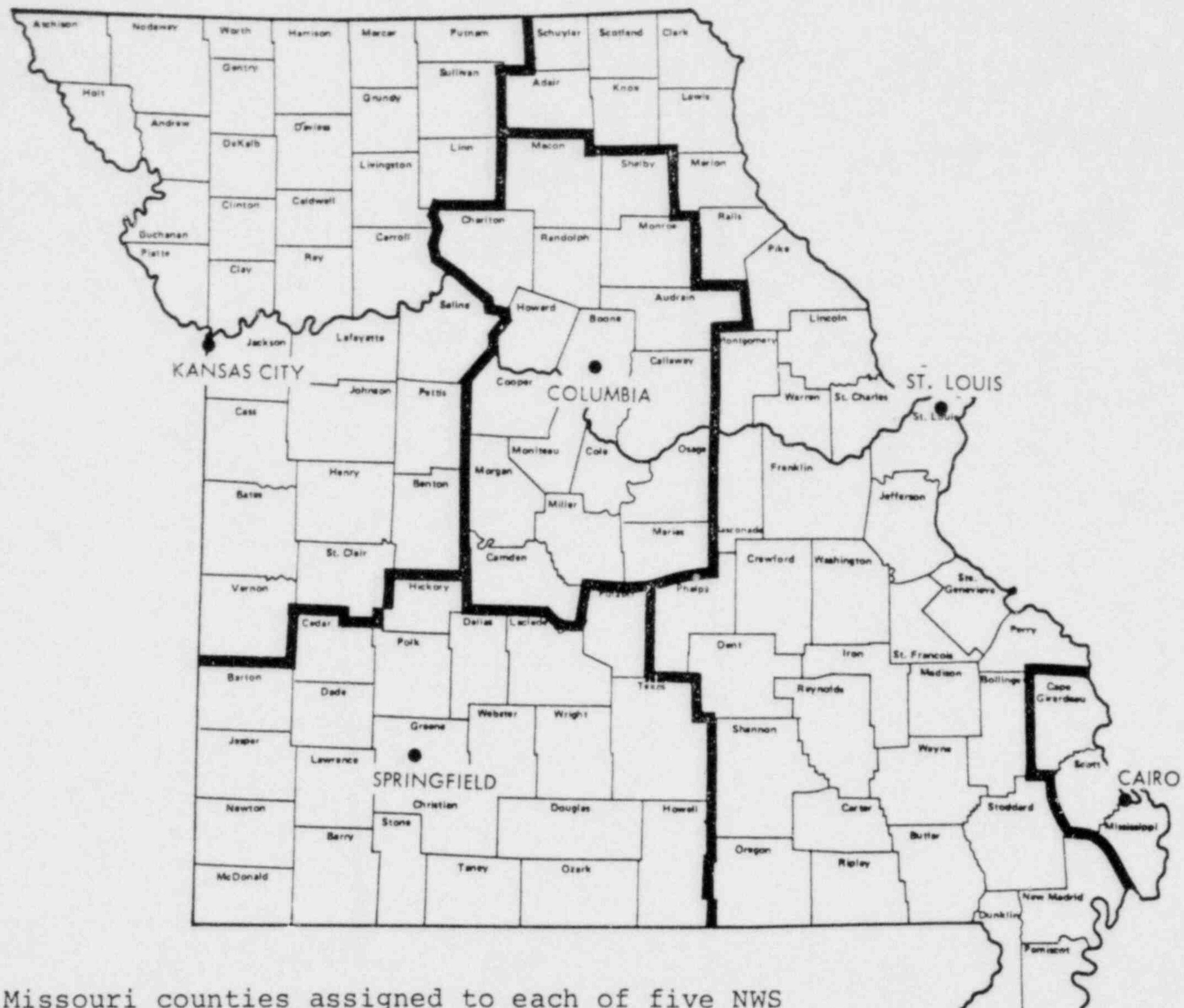


Figure 6. Missouri counties assigned to each of five NWS offices.

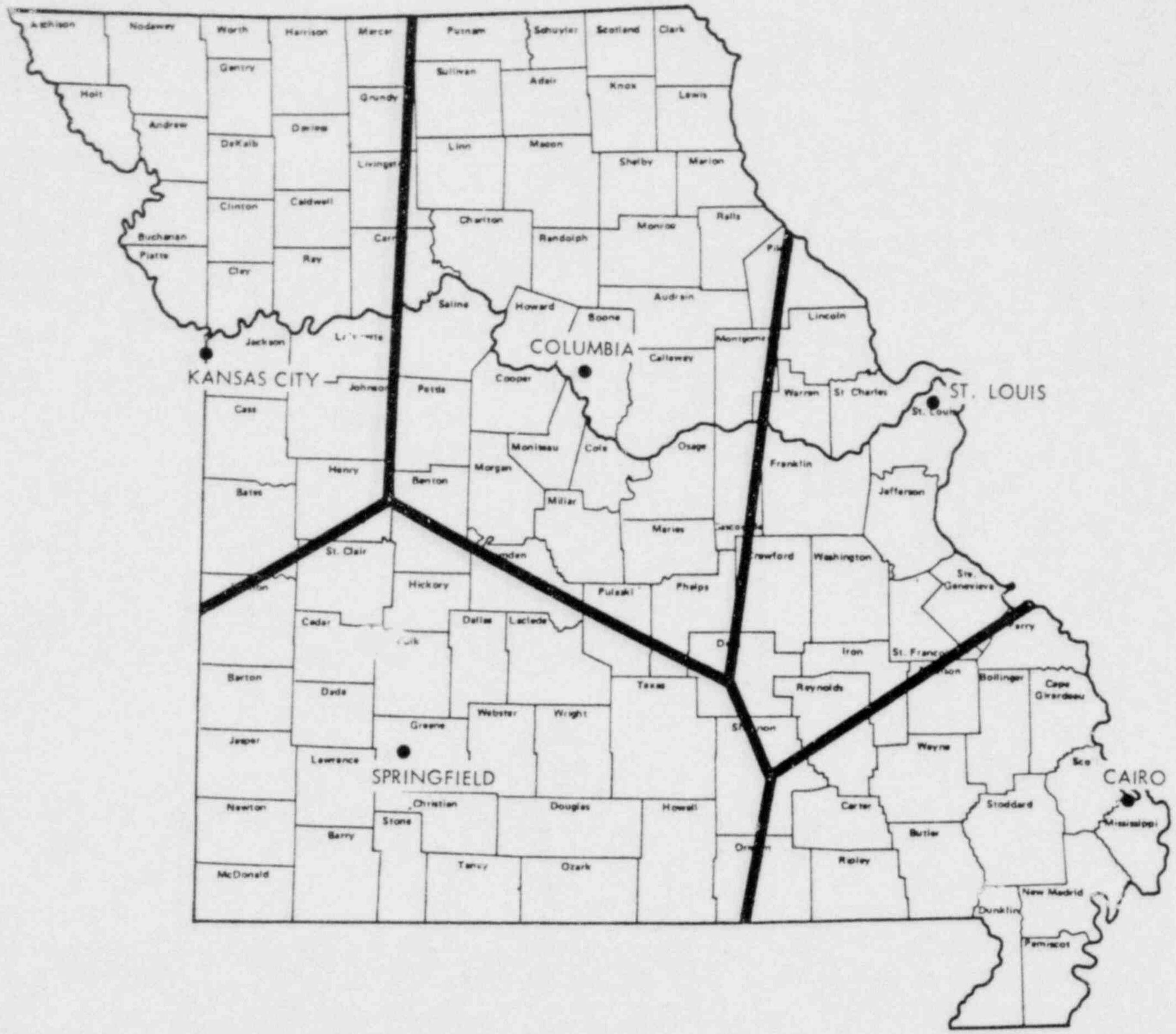


Figure 7. Missouri counties closest to NWS offices

distance from each county to its closest reporting station was computed and designated CSD. In 84 cases, ASD was equal to CSD, leaving 30 cases in which ASD exceeded CSD.

The average distance from a point (NWS station) to an area may be approximated in several ways. In first approximation, we may take the distance from the geographical center of the area to the point. In practice, this gave lower correlations than expected, and led to the search for a better method. For some purposes, the average of the highest and lowest value leads to a satisfactory approximation of the average. This idea was adapted by computing the average of the minimum distance from the NWS station to the county (taken as zero if the point was within the county), and the maximum distance from the station to any part of the county. This required 288 measurements, 2 per county for CSD, and 2 for each of the 30 values of ASD not equal to CSD. A third method, of course, would involve the division of each county into several smaller areas of equal size, determination of the distance from the geographical center of each to the given point, and division by the number of these smaller areas. If this practice were continued, and the number of divisions became large, we would expect the resulting "average distance" to approach a limiting value. In practice, the number of divisions need not be excessively large to approximate the true average distance. However, as few as 10 divisions per county would require 1440 measurements. In addition, the irregular shape and placement of Missouri counties was a significant disadvantage of this method.

A hypothetical county with sides of 25 miles was created on paper. Its area of 625 square miles was chosen to approximate the average Missouri county area (607 square miles). A hypothetical NWS station was systematically moved through and around the county. The "average distance" from the NWS station to the county was approximated by the following three methods:

1. Summation of distances from centers of each square mile in the county to the station, divided by the number of square miles (625). This was designated DAV, and taken as essentially equal to the true distance, and used as a standard of comparison with the results of methods 2 and 3 which follow.

2. Distance from geographical center of county to station, designated DCM.

3. Average of the minimum distance from the station to the county and the maximum distance from the station to the county. This was designated AVD.

The maximum difference between DAV and DCM (method 2) was 9.6 miles, with station located at the county center.

The magnitude of these errors confirmed the need for a more accurate method.

The maximum difference between DAV and AVD (method 3) was 1.4 miles, at the corners of the county. Approximately 4 percent of the area closer than 26 miles had a difference exceeding 1.0 miles.

Method 3 was used in the computation of CSD and ASD, because of its relative simplicity and its accuracy. Three weighted averages of CSD and ASD were also computed:

$$AVD1 = \frac{(3 \times CSD) + ASD}{4}$$

$$AVD2 = \frac{CSD + ASD}{2}$$

$$AVD3 = \frac{CSD + (3 \times ASD)}{4}$$

Correlation coefficients and associated significance levels were computed for each of these five measures with SDEN. They are shown in Table 6.

All measures tested had highly significant correlation coefficients with storm density. When these measures, together with their reciprocals, squares, and logarithms were entered as independent variables, along with SDEN as a dependent variable in a stepwise regression procedure, the

Table 6

Correlation Coefficients for SDEN with Various Distance to Reporting Station Measures

Distance Measure	Pearson's r	Pearson's Sig. Lvl	Spearman's RHO	Spearman's Sig. Lvl	Kendall's TAU	Kendall's Sig. Lvl
CSD	-.389	.0001	-.357	.0001	-.252	.0001
AVD1	-.386	.0001	-.363	.0001	-.254	.0001
AVD2	-.371	.0001	-.363	.0001	-.254	.0001
AVD3	-.351	.0001	-.366	.0001	-.255	.0001
ASD	-.331	.0001	-.364	.0001	-.251	.0001

reciprocal of AVD1 was selected as the best single variable with an R^2 value of .256.

Figure 8 shows residuals of this model

$$SDEN = B_0 + B_1(1/AVD1) + e \qquad \text{(model 4)}$$

The residuals represent the difference between tornado densities predicted by the model and reported tornado densities.

Value at Risk

In addition to the factors already considered, the property value in a county may be related to tornado reporting. It would appear that a tornado causing damage to highly valued property would be much more likely to be

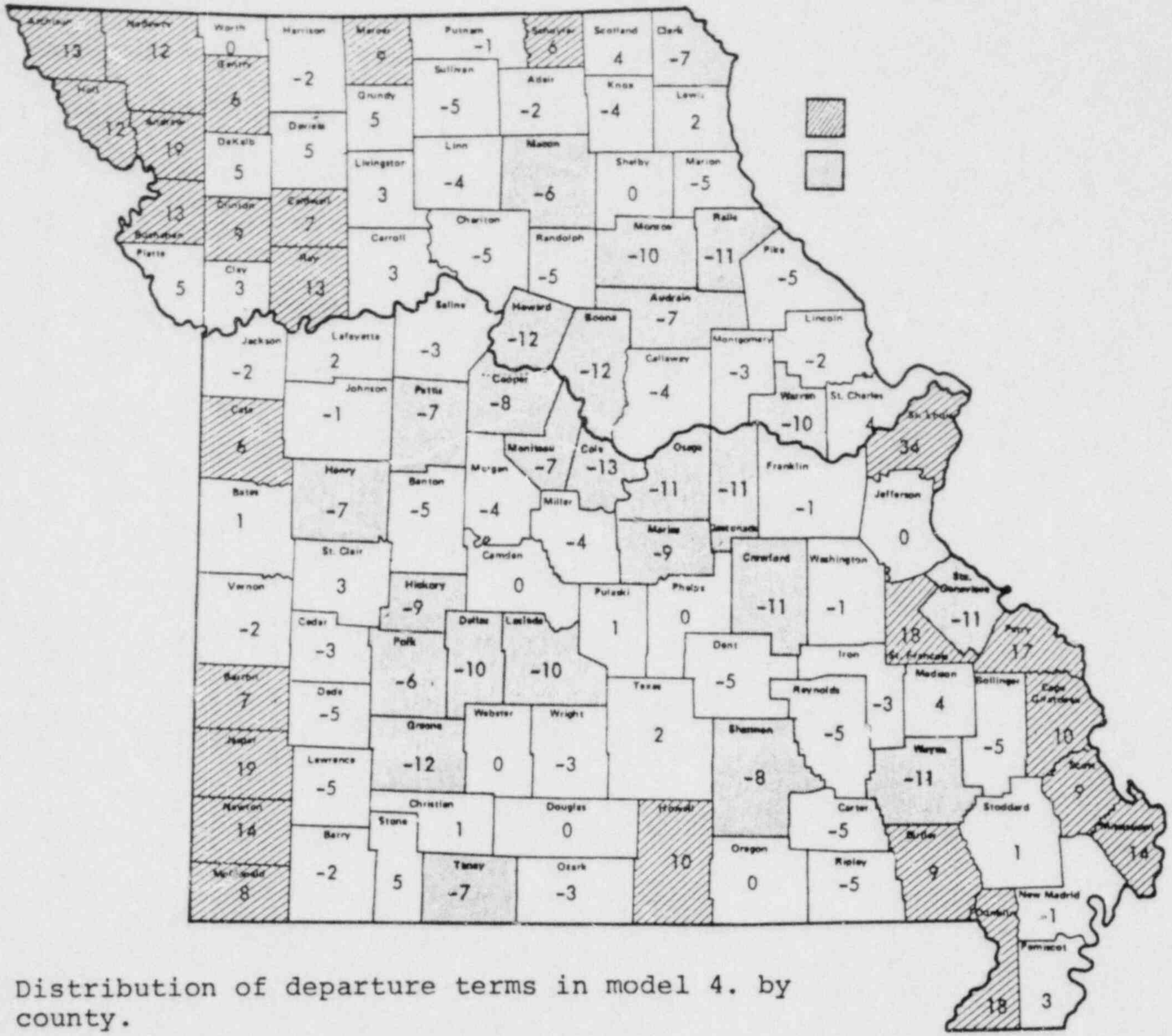


Figure 8. Distribution of departure terms in model 4. by county.

reported than a similar tornado causing only negligible damage to low valued property.

Reed (1971) studied destruction probabilities. He found that direct determination of both value of property in a county and tornado caused property damage was difficult to determine. He assumed that tornado damage reporting errors could be averaged out, and that annual bank deposits were at least roughly proportional to property values.

Examination of the 1960 U.S. Census data yielded several figures that could be used as an index to the property value at risk in a county. These included:

1. Aggregate income, representing "The amount received by all income recipients 14 years old and over from all income sources" (symbolized INCOME).

   A. Area normalized aggregate income (INKDEN)

      INKDEN = INCOME/AREAT

   B. Population normalized aggregate income, income per capita (INKCAP)

      INKCAP = INCOME/POPT

   C. Family income (INKFAM)

      INKFAM = INCOME/FMLY

2. Home index (symbolized HNDX). This index reflects the extent to which occupied housing units have certain specified types of home equipment. It was derived by adding

together the percents of occupied units with clothes washing machine, home food freezer, air conditioning, television set, telephone, and automobile. The highest possible index is 600.

3. Bank deposits, including time, demand, interbank, and governmental deposits (symbolized B).

4. Bank time deposits, consisting of deposits subject to withdrawal only after 30 days notice (symbolized BTI).

5. Bank demand deposits, consisting of those which can be withdrawn without notice (symbolized BD).

6. The sum of time and demand deposits (symbolized BTO).

7. Area normalized bank deposits (bank deposits per square mile):

A. BDEN = B/AREAT

B. BTIDEN - BTI/AREAT

C. BDDEN = BD/AREAT

D. BTODEN = BTO/AREAT

8. Bank deposits per capita:

A. BCAP = B/AREAT

B. BTICAP = BTI/POPT

C. BDCAP = BD/POPT

D. BTOCAP = BTO/POPT

The correlation coefficients of these value at risk indices with SDEN are shown in Table 7.

Table 7

Correlation Coefficients for SDEN with Various Value at Risk Measures

Value at Risk Measure	Pearson's		Spearman's		Kendall's	
	r	Sig. Lvl	RHO	Sig. Lvl	TAU	Sig. Lvl
INCOME	.505	.0001	.419	.0001	.291	.0001
INKDEN	.525	.0001	.490	.0001	.343	.0001
INKCAP	.465	.0001	.266	.0042	.185	.0036
INKFAM	.077	.4172	.256	.0061	.171	.0071
HNDX	.292	.0016	.244	.0089	.170	.0077
B	.309	.0008	.429	.0001	.298	.0001
BTI	.408	.0001	.417	.0001	.294	.0001
BD	.298	.0013	.418	.0001	.28[illegible]	.0001
BTO	.345	.0002	.493	.0001	.296	.0001
BDEN	.342	.0002	.498	.0001	.352	.0001
BTIDEN	.443	.0001	.456	.0001	.321	.0001
BDDEN	.331	.0003	.489	.0001	.342	.0001
BTODEN	.380	.0001	.493	.0001	.350	.0001
BCAP	.102	.2823	.146	.1222	.093	.1411
BTICAP	.124	.1884	.149	.1136	.092	.1451
BDCAP	.044	.6427	.102	.2792	.068	.2823
BTOCAP	.098	.2997	.128	.1757	.081	.2003

It is obvious that none of the value at risk measures involving bank deposits per capita have sufficient correlation with SDEN to justify inclusion in a linear regression model. The remaining measures were introduced into separate linear models as independent variables, with SDEN as the dependent variable. R^2 values are shown in Table 8.

Table 8

R^2 Values Corresponding to Value at Risk Variables

Independent Variable	R^2
INCOME	.255
INKFAM	.006
INKCAP	.216
INKDEN	.276
BTIDEN	.196
BTI	.166
BTODEN	.144
BDDEN	.110
B	.096
BD	.089
HNDX	.085

These relatively low R^2 values suggested that the linear relationship was not significant. The possibility of a non-linear relationship led to several data transformations in an attempt to linearize the data. The highest R^2 value, .381 was obtained from the model

$$SDEN = B_0 + B_1(INKDEN)^{.3} + e \qquad \text{(model 5)}$$

Residuals from this model are shown in Figure 9. Residuals greater than or equal to +6 are found in the west especially the northwest, and east of a north-south line through St. Charles County. Residuals less than or equal to -6 are found predominately in central and northeastern counties.

Education

An additional factor which may well have a bearing on the likelihood that a sighted tornado will be reported, is the educational level of the observer. The 1960 U.S. census report included two items which can be used as a measure of educational level:

1. Median school years completed, defined as "The value dividing the county population (25 years of age and over) into two equal groups, one half having completed more, and the other half less schooling than the median," (Goldfield, 1967, pp xxi). This was symbolized as SYR.

2. Percentile of persons 25 years of age and over who had at least completed high school. This was symbolized as PCHS.

Table 9 shows correlation coefficients and significance levels of these variables with SDEN.

It may be inferred that there is indeed a statistical relationship between educational level and reported

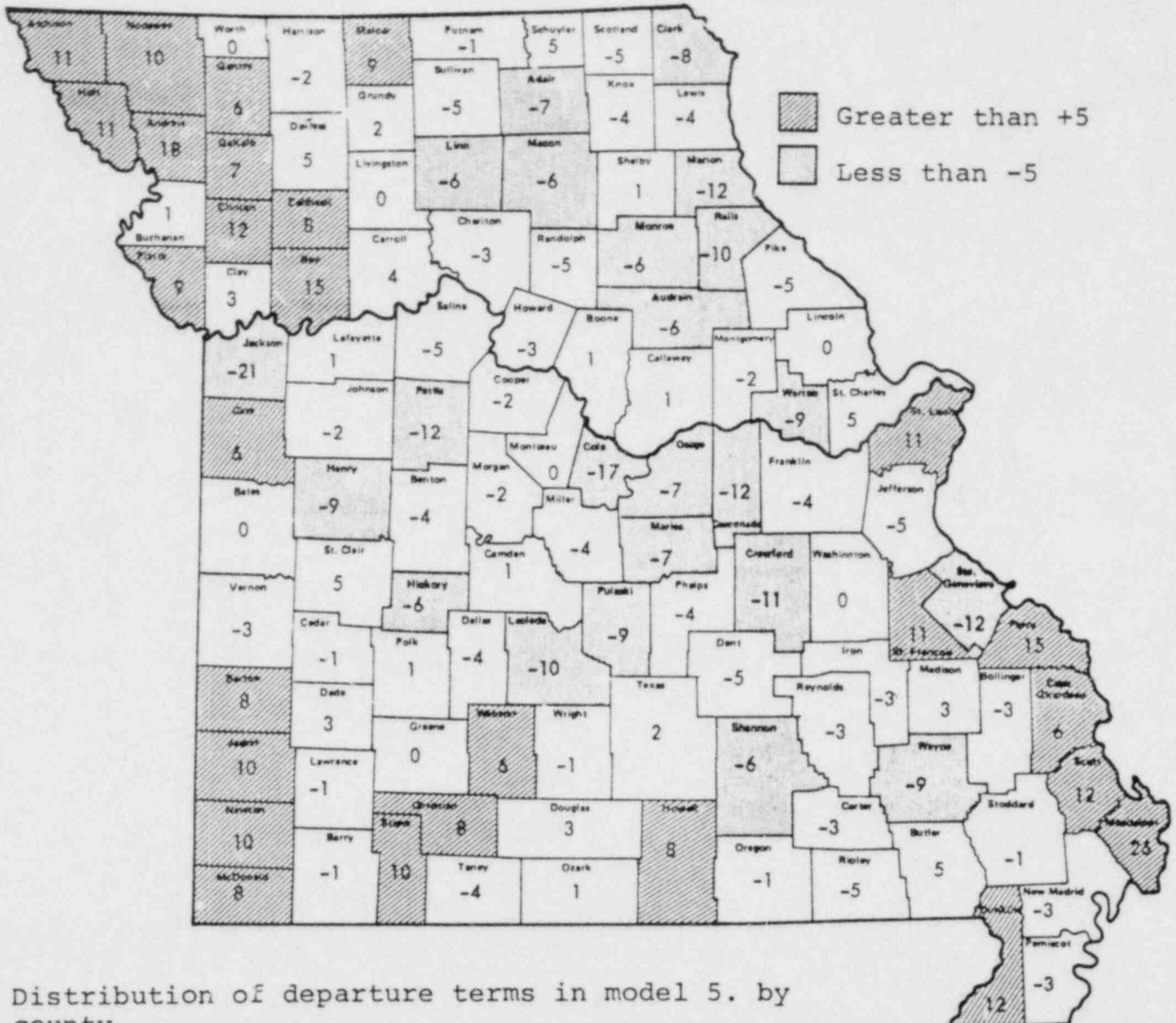


Figure 9. Distribution of departure terms in model 5. by county.

Table 9

Correlation Coefficients for SDEN with Education Level Indicators

Variable	Pearson's r	Pearson's Sig. Lvl	Spearman's RHO	Spearman's Sig. Lvl	Kendall's TAU	Kendall's Sig. Lvl
SYR	.408	.0001	.331	.0003	.237	.0003
PCHS	.363	.0001	.342	.0002	.236	.0002

tornado density. When linear regression models were tested, however, the R^2 values were relatively low. A value of .166 was obtained for SDEN with SYR, and .134 for SDEN with PCHS. Since the R^2 value is a measure of linear correlation, several transformations were performed in an attempt to linearize the relationship and increase the R^2. The highest R^2 value (.175) was obtained with the model

$$SDEN = B_0 + B_1(SYR)^2 + e \qquad \text{(model 6)}$$

The residuals of this model, when plotted against SDEN, show a tendency for negative residuals for small values of SDEN, and positive residuals for large values of SDEN. This indicates that additional factors must be used to adequately explain the variations in tornado density. Distribution of residuals in model 6 is shown in Figure 10.

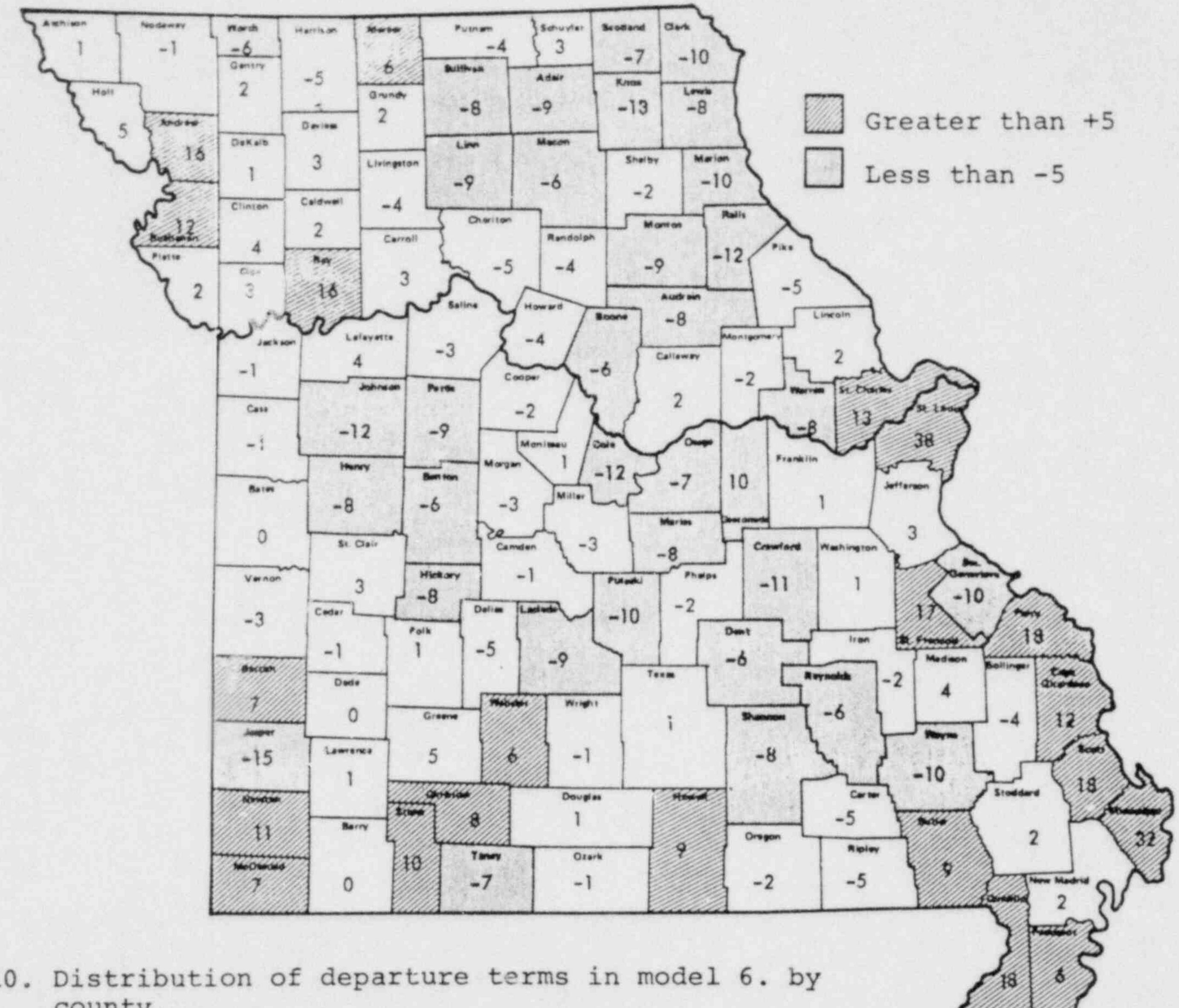


Figure 10. Distribution of departure terms in model 6. by county.

Area

Dames & Moore (1975) included county area as an independent variable with other independent variables in a stepwise regression procedure, with tornado frequency as the dependent variable. They state:

> Furthermore, the correlation between tornado incidence and the function of area was found to be much weaker than the other two independent variables (as would be expected, since tornado incidence per unit area should be independent of area.

For completeness sake, county area (AREAT), rural area (AREAR), ratio of urban area to total county area (RAREA), and ratio of farm land to total area (FRPC) were correlated with SDEN. Results are shown in Table 10.

Table 10

Correlation Coefficients for SDEN with Various Area Related Variables

Independent Variable	Pearson's r	Pearson's Sig. Lvl	Spearman's RHO	Spearman's Sig. Lvl	Kendall's TAU	Kendall's Sig. Lvl
AREAT	-.238	.0106	-.225	.0162	-.154	.0155
AREAR	-.295	.0014	-.256	.0059	-.179	.0048
RAREA	.538	.0001	.340	.0002	.248	.0002
FRPC	.064	.5011	.142	.1320	.100	.1159

The ratio of farm land to total area (FRPC) is obviously not significantly correlated with SDEN. The correlation of RAREA with SDEN may be population related since counties with high values of RAREA would clearly be those with large urban areas and hence, relatively high population densities. Historically, counties were formed with roughly equal populations, so low population densities would be expected with large area counties.

Since low population densities are associated statistically with low reported tornado densities, the small negative correlations of area with tornado density are reasonable.

Linear regression analysis was performed with each of the area related independent variables and SDEN.

Table 11

Linear Regression R^2 Values for SDEN with Area Related Variables

Independent Variables	R^2 Values
FRPC	.004
AREAT	.056
AREAR	.087
RAREA	.289

These values are compatible with values in Table 10

CHAPTER 4

GEOGRAPHIC EFFECTS

Location

When tornado incidence per unit area is plotted on a map of the United States, regional variations become apparent with a relative maximum area located to the southwest of Missouri, near central Oklahoma. It would be desirable to separate this large scale effect from the subregional variability that is the subject of this study.

The latitude (LAT) and longitude (LONG) of the geographical center of each county were correlated with SDEN. Results are shown in Table 12.

Table 12

Correlation of SDEN with LAT and LONG

Location Variable	Pearson's r	Pearson's Sig. Lvl	Spearman's RHO	Spearman's Sig. Lvl	Kendall's TAU	Kendall's Sig. Lvl
LAT	-.020	.8340	-.023	.8047	-.015	.8122
LONG	.121	.1981	.286	.0021	.221	.0005

The lack of significant correlation of latitude with tornado density indicates the absence of any large scale north-south bias in the data sample. Longitude, however, does appear to have significant correlation with tornado density. This correlation does not appear to be linear, since testing the regression model

$$SDEN = B_0 + B_1(LONG) + e$$

resulted in a R^2 value of .015.

When a simple curvilinear model

$$SDEN = B_0 + B_1(LONG) + B_2(LONG)^2 + e \quad \text{(model 7)}$$

was evaluated, a multiple R^2 value of .325 was obtained. Figure 11 shows the residuals of this model plotted by county. A plot of longitude and SDEN is examined (Figure 12), the lowest values of SDEN appear to be between 91 and 92 degrees west longitude. This area includes Clark, Ralls, Montgomery, and Gasconade counties. All of which have both low tornado incidence per unit area and low population densities. To the west of this area, we find a steady increase in SDEN values. This is believed to be due, in part, to the large scale general increase in tornado incidence centered in Oklahoma and Kansas. Increases to the east are reflections of high values found in nearly all counties along the Mississippi River from St. Louis county southwards.

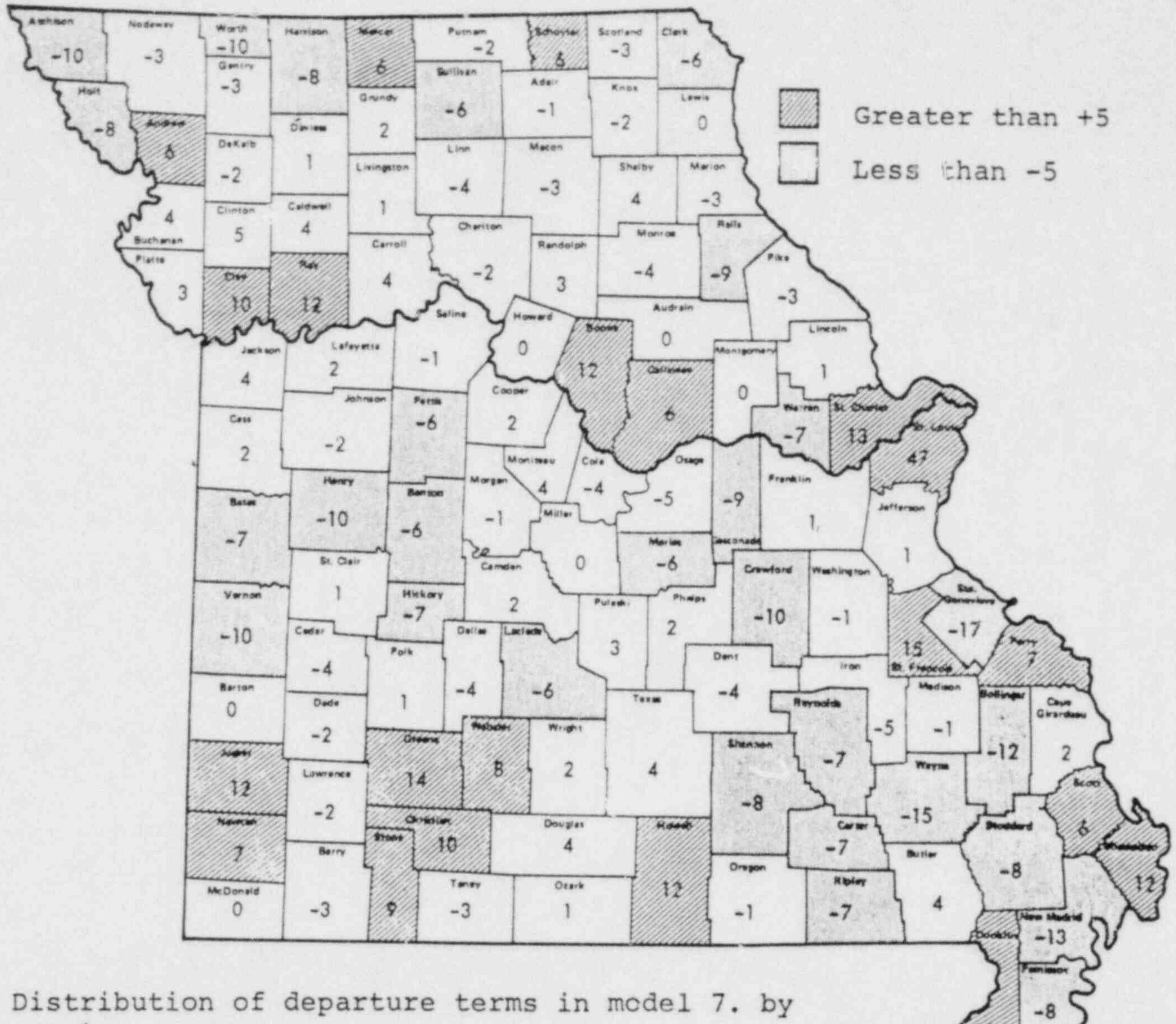


Figure 11. Distribution of departure terms in model 7. by county.

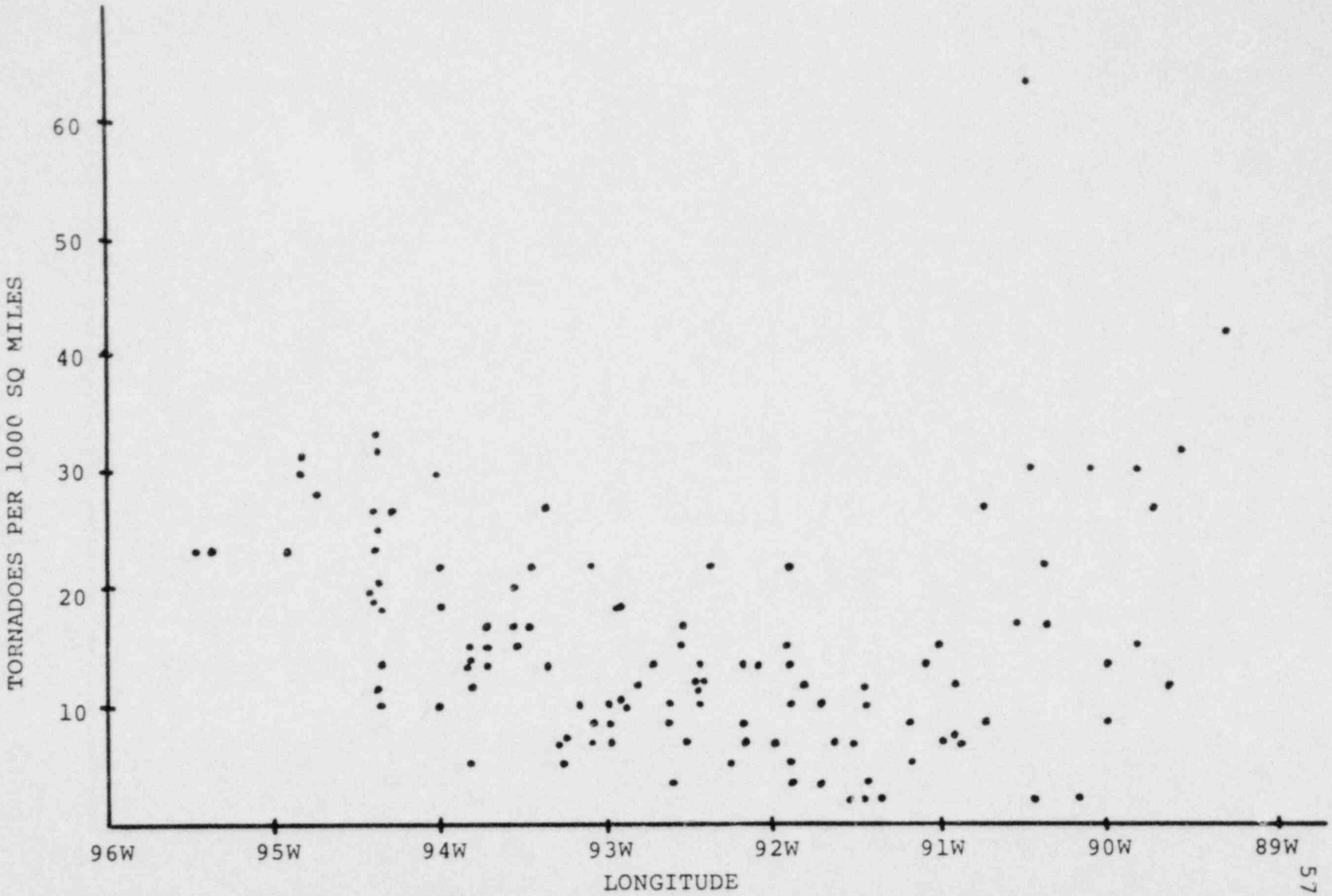


Figure 12. Plot of SDEN vs LONG

Terrain

Four terrain related variables were considered to be of possible significance. These included elevation, roughness, east-west slope, and north-south slope. Computation of each of these variables required the input of elevations for a large number of points. As the number of data points increases, accuracy improves, as does the labor of data extraction. A grid size of 5 minutes of latitude and 5 minutes longitude was selected. This required a total of 2788 data points. Elevations were interpolated from the seventeen U.S. Geological Survey maps (scale 1:250,000) covering the various sections of Missouri. These data points were then used to compute:

1. The average elevation of each county (symbolized AVE).

2. A roughness parameter (taken as the standard deviation of the elevations of all data points in the county) (symbolized RUF).

3. The north-south slope, taken as the weighted average of the slopes of all columns of gridpoints within a county (symbolized NS).

4. The east-west slope, taken as the weighted average of the east-west slopes of all rows of gridpoints within a county (symbolized EW).

Correlation coefficients of these variables with SDEN are shown in Table 13.

Table 13

Correlation of SDEN with Terrain Variables

Terrain Variables	Pearson's r	Pearson's Sig. Lvl	Spearman's RHO	Spearman's Sig. Lvl	Kendall's TAU	Kendall's Sig. Lvl
AVE	-.067	.4784	.111	.2395	.087	.1620
RUF	-.204	.0291	-.172	.0666	-.111	.0808
NS	.052	.5829	.105	.2640	.072	.2589
EW	.290	.0017	.271	.0035	.187	.0032

When each of the independent variables was used in a simple linear regression analysis with the dependent variable SDEN, the following results were obtained:

Table 14

Linear Regression R^2 Values for SDEN With Terrain Variables

Independent Variable	R^2
NS	.003
AVE	.004
RUF	.042
EW	.084

Figure 13 shows the residuals of the regression with the variable EW.

$$SDEN = B_0 + B_1(EW) + e \qquad \text{(model 8)}$$

The explained variance may be low for two reasons:

1. They are based on elevation values taken at points 5 nautical miles apart in the north-south direction and approximately 4 nautical miles part in the east-west direction.

2. They consider only the terrain within the county and neglect the terrain over which the tornado-bearing storms travel enroute to the county.

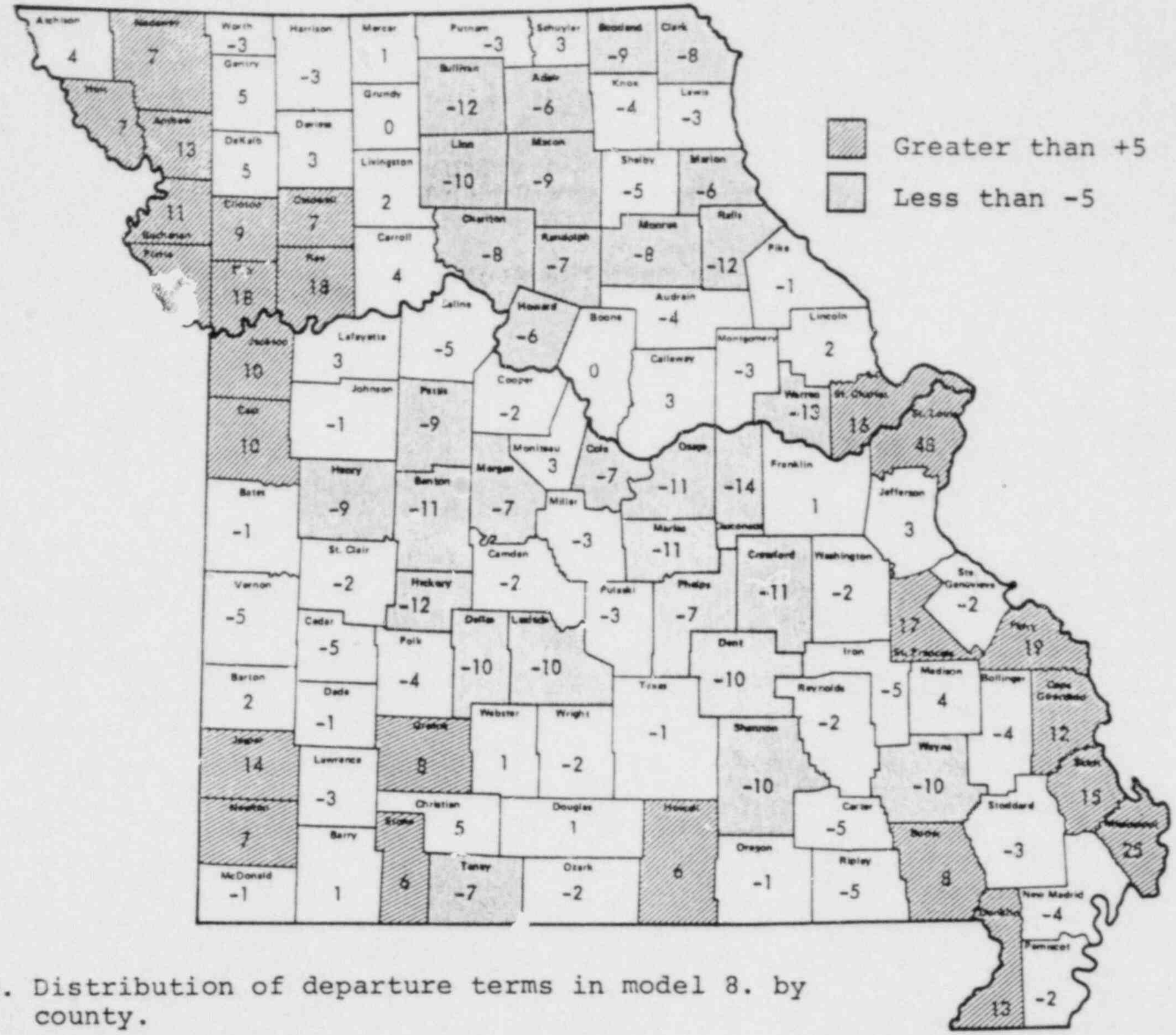


Figure 13. Distribution of departure terms in model 8. by county.

CHAPTER 5

COMBINED EFFECTS

Multiple Regression Models

Chapters 3 and 4 examined regression models involving one independent variable and tested the power of each model to predict the value of the response variable storm density, SDEN. These models could explain, at best, less than 41 per cent of the observed variability in the dependent variable.

A multiple regression model combining several independent variables would be expected to predict the response variable better than any of the independent variables alone. The coefficient of multiple determination denoted by R^2, is defined as follows:

$$R^2 = 1 - SSE/SSTO$$

where $SSE = \Sigma(Y_i - \hat{Y}_i)^2$, and $SSTO = \Sigma(Y_i - \bar{Y})^2$.

Neter and Wasserman (1974) state that:

> 2. A large R^2 does not necessarily imply that the fitted model is a useful one. For instance, observations may have been taken at only a few levels on the independent variables. Despite a high R^2 in this case, the fitted model may not be useful because most predictions would require extrapolations outside the region of observations....
> 3. Adding more independent variables to the model can increase R^2 and never reduce it, because SSE can never become larger with more

> independent variables and SSTO is always the same for a given set of responses. Since R^2 often can be made large by including a large number of independent variables, it is sometimes suggested that a modified measure be used which recognizes the number of independent variables in the model. This adjusted coefficient of multiple determination, denoted R_a^2 is defined:
>
> $$R_a^2 = 1-((n-1)/(n-p))\ SSE/SSTO$$
>
> This adjusted coefficient of multiple determination may actually become smaller when another independent variable is introduced into the model, because the decrease in SSE may be more than offset by the loss of a degree of freedom in the denominator n-p.

When the one selected independent variable from each of the various types of variables considered in the previous chapters were considered as separate variables in the multiple regression model

$$\begin{aligned} SDEN = {} & B_0 + B_1(DENR)^{.3} + B_2(RAVD1) + B_3(SYR)^2 + \\ & B_4(AREAR) + B_5(LONG) + B_6(LONG)^2 + B_7(EW) \\ & + e \qquad \text{(model 9)} \end{aligned}$$

an R^2 value of .649 and an R_a^2 value of .622 were obtained. Figure 14 depicts the residuals of this model. The residuals show considerably less organization than previous plots. Most of the negative values are scattered through the central and southeastern counties, while most of the positive values are scattered in the northern and south-central counties.

Although this model represented a significant improvement over any of the previous single regression models, the

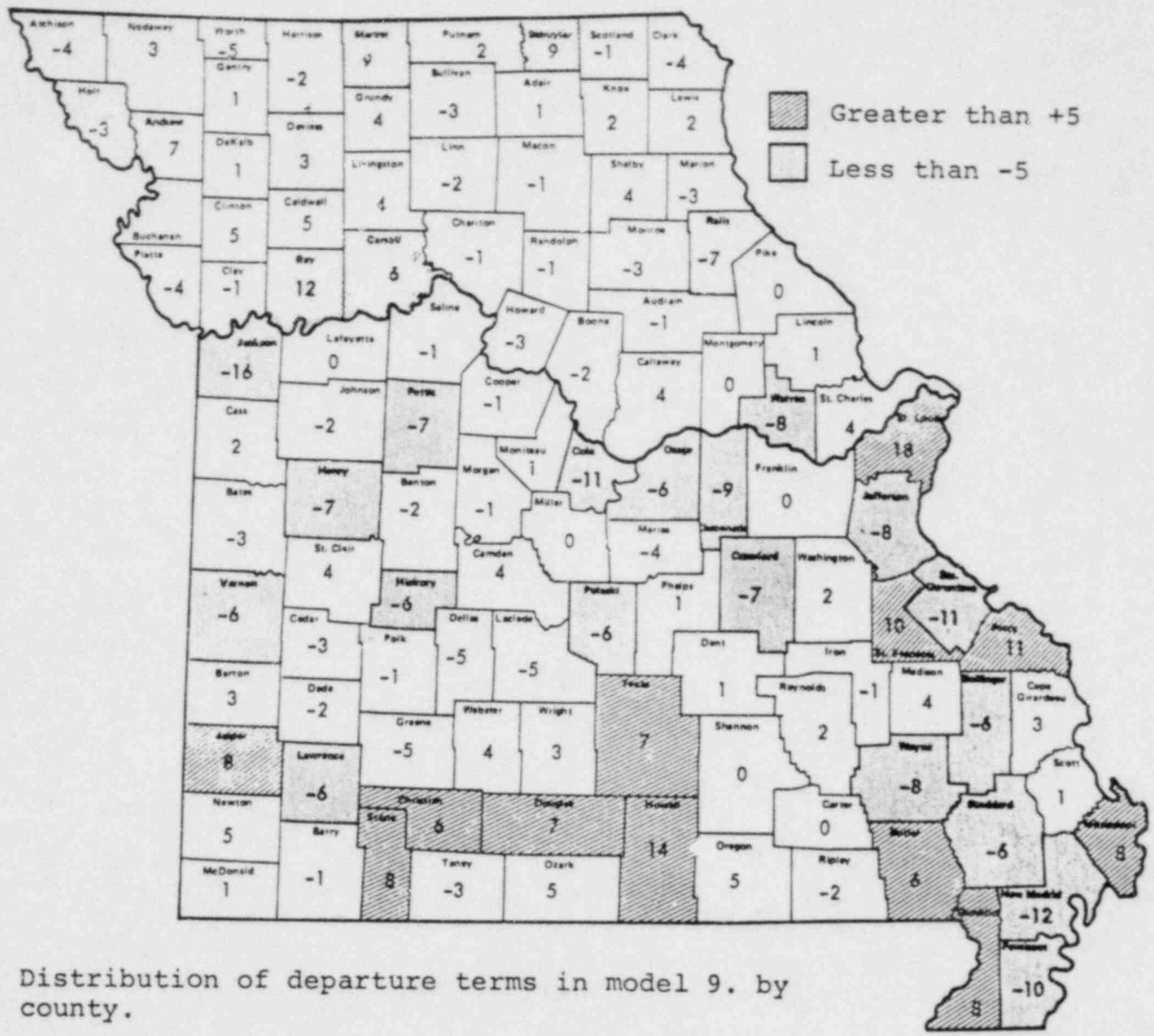


Figure 14. Distribution of departure terms in model 9. by county.

possibility of further improvement was investigated through the use of stepwise regression techniques. When 43 independent variables, consisting of all variables found significant in Chapters 3 and 4 were used in a multiple regression model, R^2 and R_a^2 values of .818 and .710 were obtained. When forward stepwise regression techniques were used with the same 43 independent variables, the models with highest R^2 values for equal numbers of independent variables were studied. Table 15 gives the R^2 and R_a^2 values for models with 1 to 15 independent variables.

Initially, as more independent variables are added to the model, the R^2 and R_a^2 values increase rapidly. After the eighth variable, little improvement is noted with the addition of further terms, and R_a^2 shows actual decreases after the 14th term. Models 10, 11, 12, and 13 are as follows:

$$\text{SDEN} = B_0 + B_1(\text{DENR}) + B_2(\text{RASD}) + B_3(\text{LONG}) + B_5(\text{LONG})^2 + e \qquad \text{(model 10)}$$

$$\text{SDEN} = B_0 + B_1(\text{HDEN}) + B_2(\text{DENR})^{.1} + B_3(\text{RCSD}) + B_4(\text{RAVD1}) + B_5(\text{BDDEN}) + B_6(\text{LONG}) + B_7(\text{LONG})^2 + B_8(\text{NS}) + e \qquad \text{(model 11)}$$

$$\text{SDEN} = B_0 + B_1(\text{DENR}) + B_2(\text{HDEN}) + B_3(\text{DENR})^{.8} + B_4(\text{RCSD}) + B_5(\text{RAVD1}) + B_6(\text{BDDEN}) + B_7(\text{LONG}) + B_8(\text{LONG})^2 + B_9(\text{NS}) + e \qquad \text{(model 12)}$$

Table 15

R^2 and R_a^2 Values for Multiple Regression Models

R^2	R^2	R_a^2	R_a^2	Number of Variables
.4074	---	.4074	---	1
.4585	.0511	.4536	.0462	2
.4980	.0395	.4889	.0353	3
.6608	.1628	.6515	.1626	4
.6745	.0137	.6626	.0111	5
.6891	.0146	.6747	.0121	6
.6959	.0068	.6789	.0042	7
.7503	.0544	.7338	.0549	8
.7678	.0175	.7501	.0163	9
.7754	.0076	.7560	.0059	10
.7816	.0062	.7604	.0044	11
.7849	.0033	.7617	.0013	12
.7877	.0028	.7625	.0008	13
.7903	.0026	.7630	.0005	14
.7922	.0019	.7628	-.0002	15

$$SDEN = B_0 + B_1(DENR) + B_2(HDEN) + B_3(DENR)^{.9} + B_4(RCSD) + B_5(RAVD1) + B_6(SQHXD) + B_7(PCHS) + B_8(LONG) + B_9(LONG)^2 + B_{10}(NS) + e \quad \text{(model 13)}$$

Residuals from models 10, 11, 12, and 13 are depicted on Figures 15, 16, 17, and 18, respectively. Table 16 lists values of coefficients, rounded to 5 significant figures, associated with models 10, 11, 12, and 13. The coefficients associated with population density in models 10 and 11 are positive, thus as rural population density increases, storm density should increase. This is in complete agreement with results expected from Chapter 3. In models 12 and 13, the contribution to SDEN by rural population density is positive.

The contribution of HDEN is positive in all models, as expected. The coefficients associated with the value at risk variables BDDEN and SQHDX are small and negative. These results may be due to the correlation between DENR and the value at risk variables. The Kendall's tau correlation coefficient associated with DENR and BDDEN is .4653, with a significance level of .0001, and the Kendall's tau correlation coefficient associated with DENR and SQHDX is .2001, with a significance level of .0017.

The coefficient associated with PCHS is positive, as expected from Chapter 3. The coefficients associated with

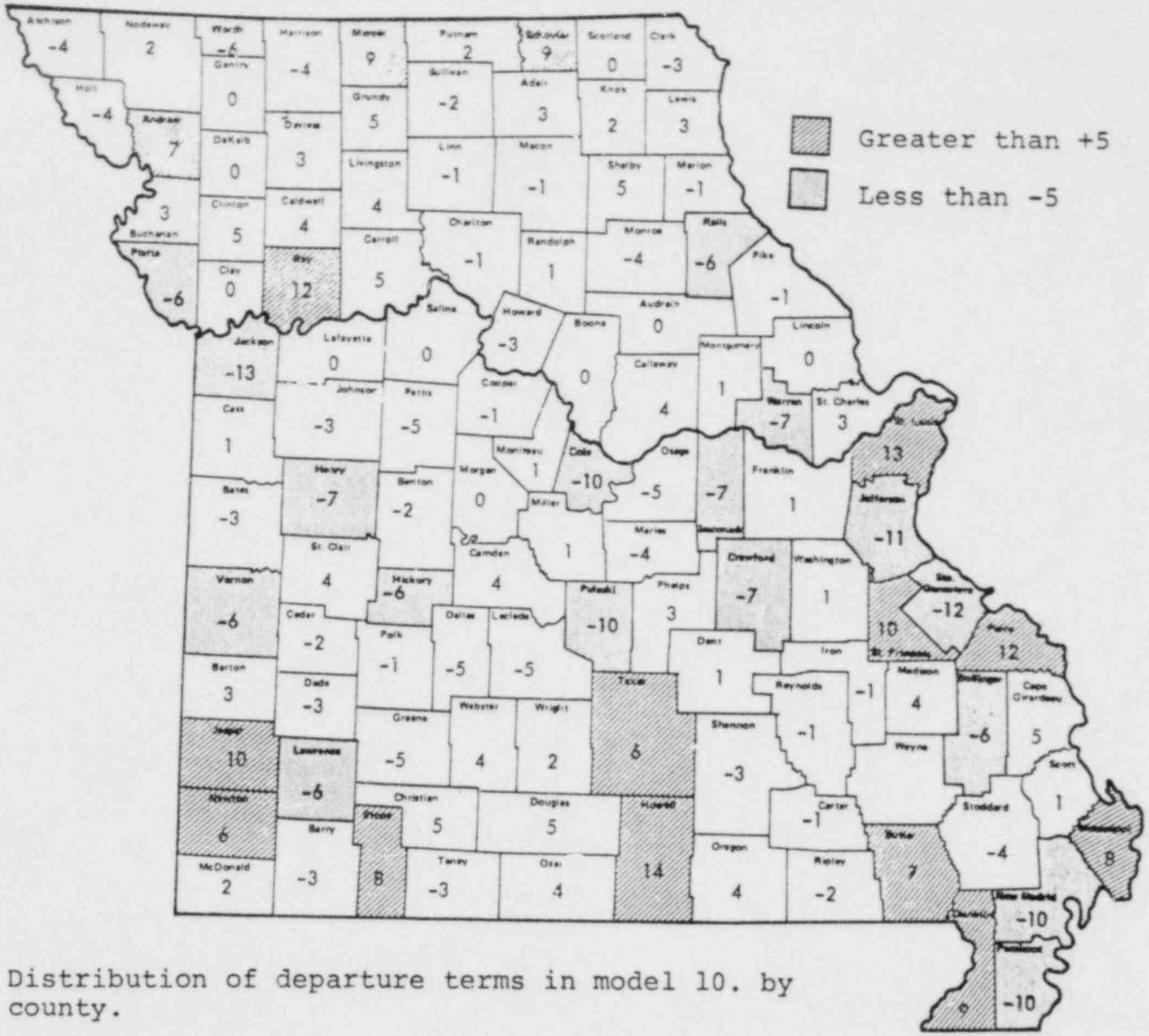


Figure 15. Distribution of departure terms in model 10. by county.

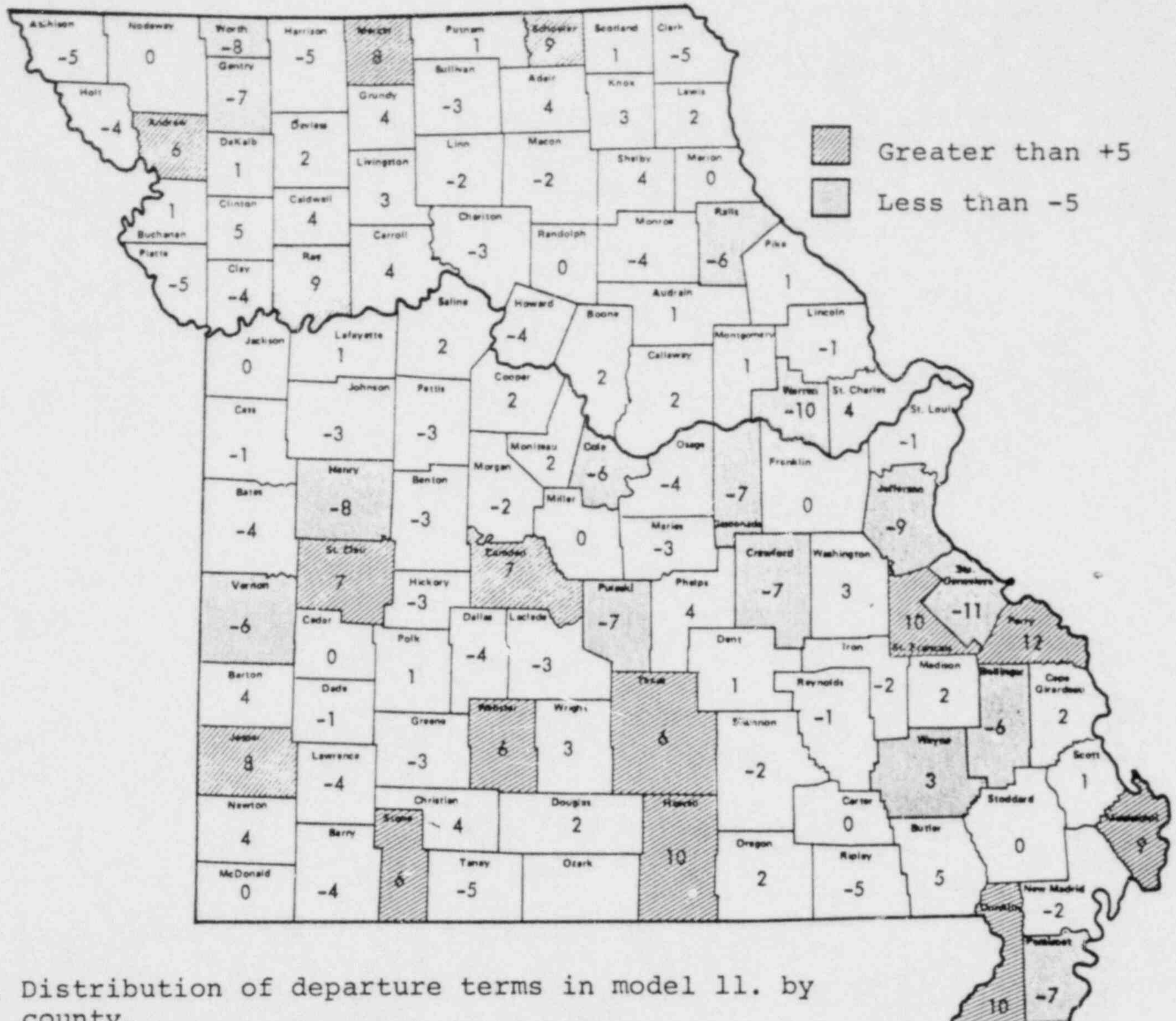


Figure 16. Distribution of departure terms in model 11. by county.

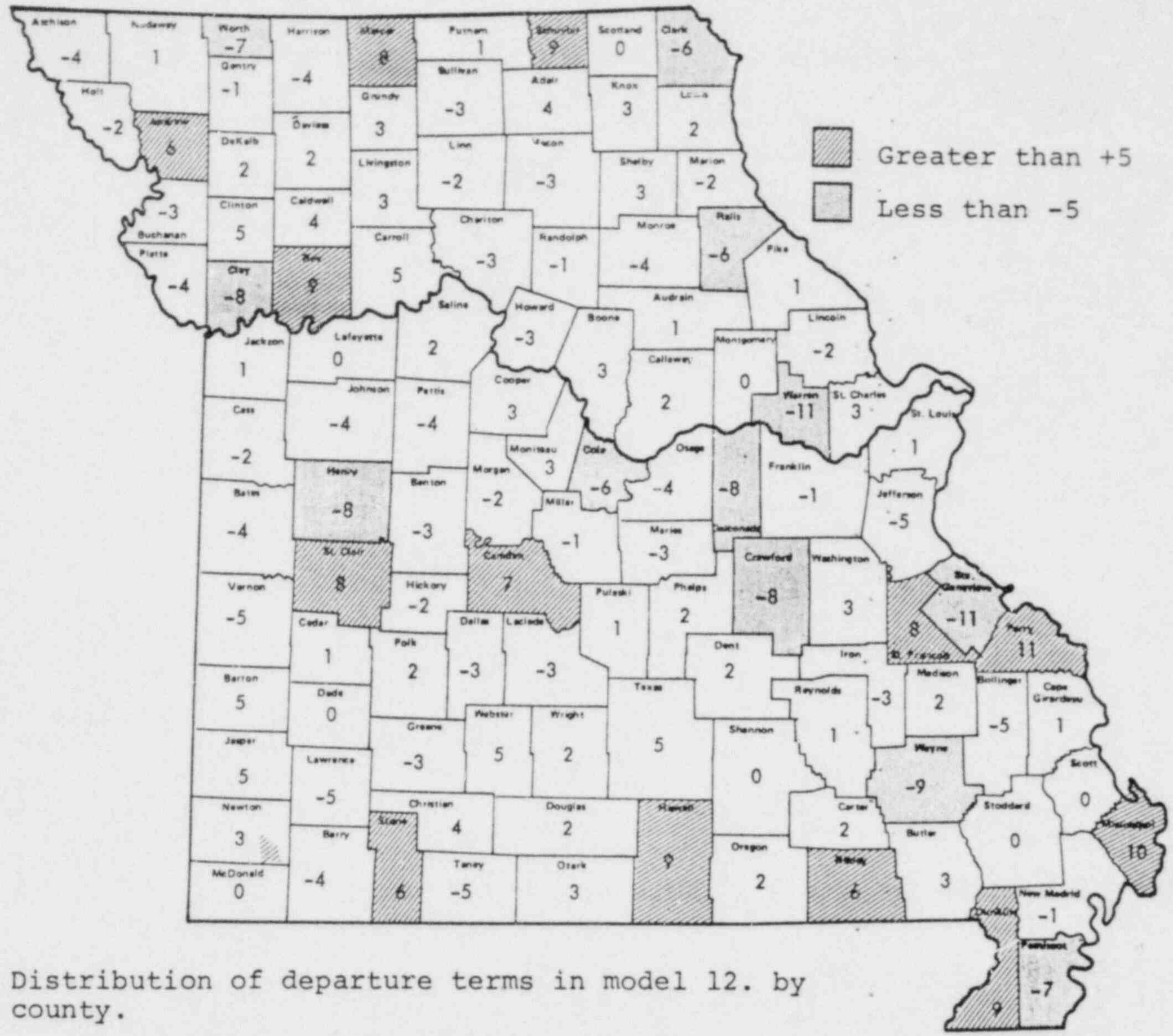


Figure 17. Distribution of departure terms in model 12. by county.

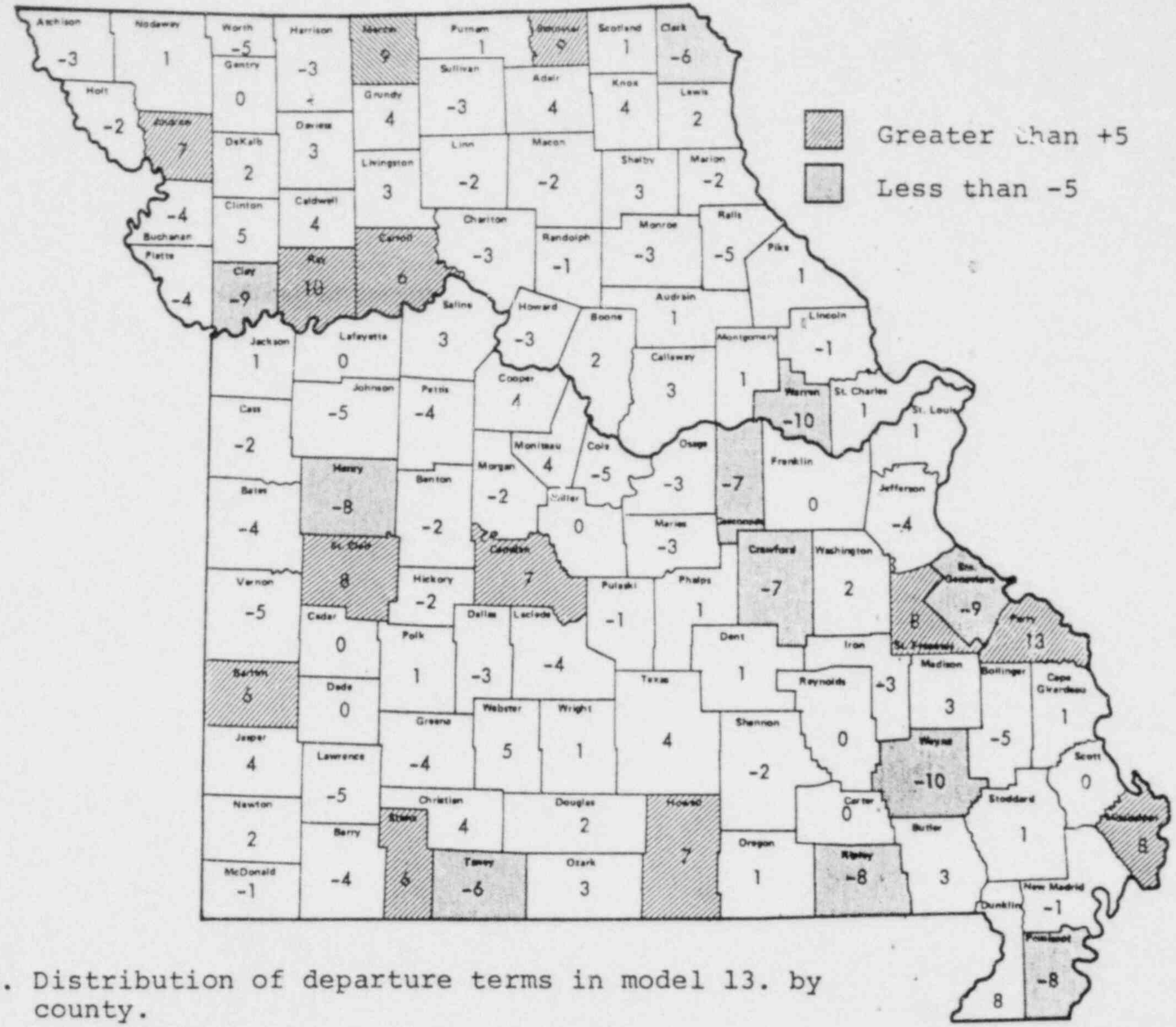


Figure 18. Distribution of departure terms in model 13. by county.

Table 16

Multiple Regression Model Coefficients

Coefficient	Model 10	Model 11	Model 12	Model 13
B_0	15515.	14867.	13385.	13606.
DENR	.2216	--	-2.2166	-5.1679
$DENR^{.1}$	--	.43493	--	--
$DENR^{.8}$	--	--	5.9946	--
$DENR^{.9}$	--	--	--	8.4225
HDEN	--	.14263	.27030	.30100
BDDEN	--	-.04074	-.06772	-.07569
SQHDX	--	--	--	-1.2509
PCHS	--	--	--	7.9015
RCSD	--	-1161.6	-1405.6	-1564.7
RAVD1	--	1269.9	1462.6	1610.4
RASD	171.45	--	--	--
NS	--	11.218	12.736	11.904
LONG	-337.06	-323.96	-290.82	-295.13
$LONG^2$	1.8308	1.7588	1.5783	1.6016

distance to reporting station are such that short distances to reporting station yield larger contributions to SDEN in the model than greater distances. The coefficient associated with NS is positive in all cases, indicating that slope increasing toward the north is associated with increased SDEN.

The coefficients associated with the quadratic term involving longitude indicate, in each case, increased values of SDEN along the eastern and western boundaries of Missouri, and minimum contributions to SDEN in the vicinity of 92.1 degrees west longitude.

As more terms are added to the multiple regression models, the pattern of residuals loses more and more of its apparent organized nature and becomes more random in appearance. The initial variability can be partially explained by variables considered here, and the rest as a random component. With an R^2 of .75 or greater (models 11, 12, 13), most of the non-random component has hopefully been removed. An unknown, but large portion of that remaining is believed to be of a random nature. Other variables, as yet unmeasured or at least untried, might reduce the variation more.

The error term in regression analysis is assumed to be normally distributed for test statistics to have the correct distribution. This is often justified in real world situations since the error terms frequently represent the effects of many factors not included in the model. Unless the departures from normality of the error terms are serious, the actual confidence coefficients and risks of errors will be close to the levels for exact normality. (Neter and Wasserman, 1974). Figure 19 is a histogram of the residuals from model 13. Values for a normal distribution with a mean of 0.0 and a standard deviation of 4.6328, the same as that found in the observed residuals, is superimposed on the figure for comparison purposes. The symmetry of the histogram and its close resemblance to the normal distribution, particularly in the tails is evident.

We may utilize a chi-square goodness of fit test, with Ho: distribution of residuals is normal vs Ha: distribution of residuals is not normal. Table 17 shows the results of this test. The histogram and chi-square values are indicative of at least approximate normality of the departure terms, as assumed by the multiple regression model, model 13.

Estimated Tornado Occurrences

Regression models can be used to estimate the tornado densities that would have been reported if uniform demographic conditions had existed. This, of course, assumes that the multiple regression model is correct within the range of values of the predictor variables. If large and uniform population densities existed throughout the state,

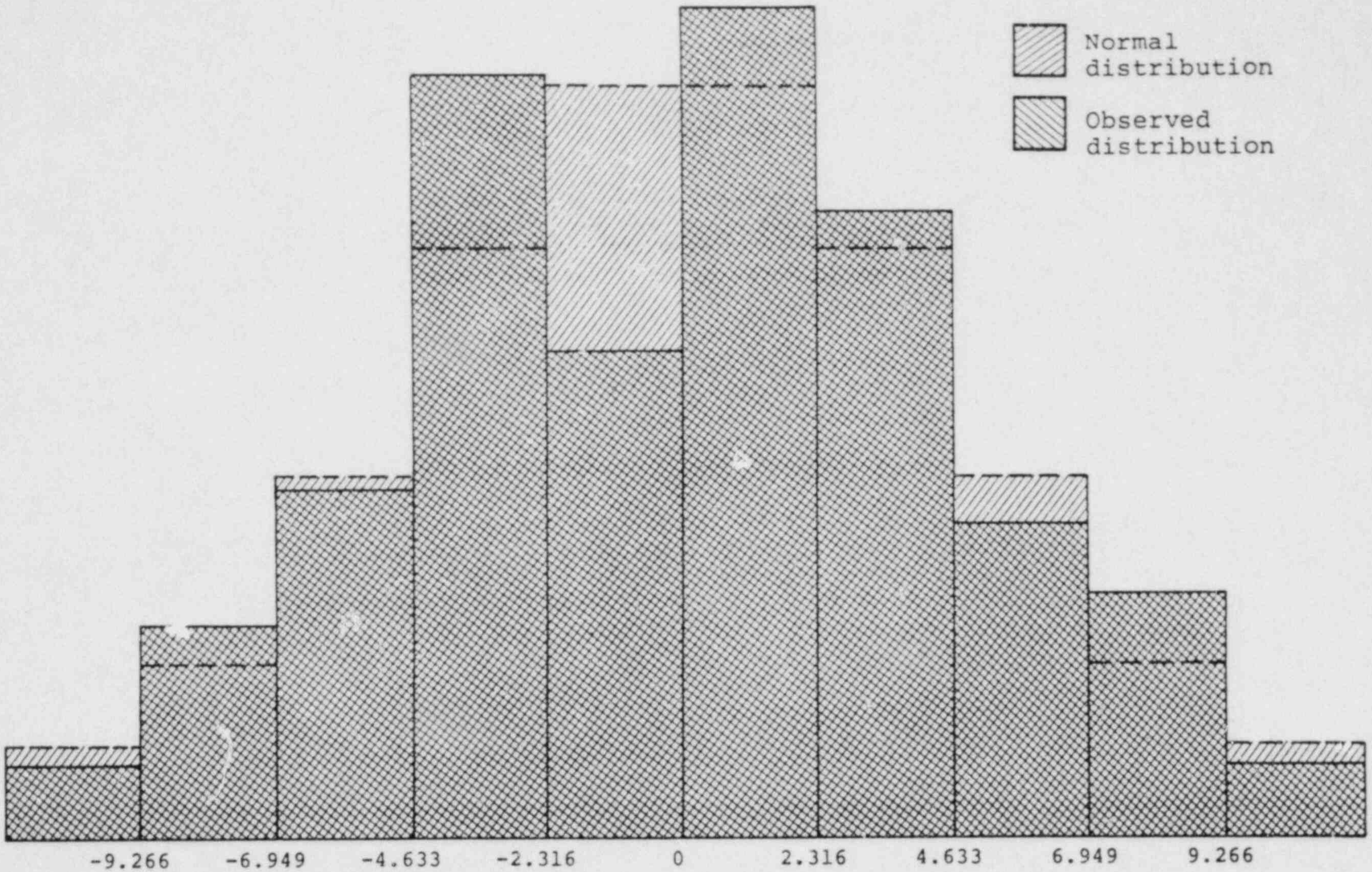


Figure 19. Histogram of residuals from model 13.

Table 17

Chi-square Goodness of Fit Statistics for

Model 13 Departure Terms

Class Limits	Expected	Observed	$(o_i - e_i)^2/e_i$
<-4.480	19	18	.053
-4.480 to -1.997	19	24	1.310
-1.997 to 0.0	19	12	2.580
0.0 to 1.997	19	20	.053
1.997 to 4.480	19	22	.474
>4.480	19	18	.053
	114	114	4.526

Critical Value	Significance Level	Degrees of Freedom
4.11	.25	3
4.53	.22	3
6.25	.10	3

and sufficient reporting stations existed such that distance to reporting station was small in each county, most tornadoes occurring within the state should be detected and reported. This can be modeled by substituting appropriate values of the demographic predictor values into a multiple regression model such as 11 or 13. If values within the range of those reported are chosen, the potentially large errors associated with extrapolation may be avoided.

If the 95th percentile values for DENR (47.35 people per square mile), HDEN (29.74 homes per square mile), HNDX (393), BDDEN ($71,262.10 per square mile) and the 5th percentile value for CSD, ASD, and AVD1 (20.3 miles) are chosen, models 11 and 13 become:

$$\text{SDEN} = 14937.91 - 323.959\ \text{LONG} + 1.7588\ \text{LONG}^2 + 11.217\ \text{NS} + e \qquad \text{(model 11a)}$$

$$\text{SDEN} = 13616.75 - 295.131\ \text{LONG} + 1.6016\ \text{LONG}^2 + 11.9042\ \text{NS} + e \qquad \text{(model 13a)}$$

When these values are calculated, converted to estimated tornado strikes per county, and summed over the entire state, the total estimated tornado strikes become 1662 (model 11a) and 1658 (model 13a). With the demographic assumptions made in these models, estimated tornado strikes exceed reported tornado strikes by approximately 68 per cent.

More extreme values of demographic values may be chosen, however, the results are subject to increasing uncertainty at values outside the range of the observations used to estimate the regression equation parameters. Neter and Wasserman (1974) state:

> Another caution deals with inferences pertaining to levels of the independent variables which fall outside the range of observations. ...If the [independent variable] level falls far beyond the range of past data, extreme caution should be exercised since one cannot be sure that the regression function which fits past data is appropriate over the wider range of the independent variables.

With these cautions, it may still be of interest to estimate the tornado strikes that would be predicted by an appropriate multiple regression model with optimum demographic conditions. If the highest reported values of DENR, HDEN, and BDDEN, and lowest reported values of CSD, ASD, and AVD1 are substituted into model 11, model 11b is obtained:

$$\text{SDEN} = 14951.13 - 323.959\ \text{LONG} + 1.7588\ \text{LONG}^2 + 11.217\ \text{NS} + e \qquad \text{(model 11b)}$$

With this model and the same procedure as above, total estimated tornado strikes become 2574. This exceeds reported tornado strikes by 159 per cent. These calculations suggest that actual tornado strikes exceed reported tornado strikes by a large margin. This is in qualitative agreement with the results of Eshelman and Stanford (1977) who found a 3 to 1 ratio between actual tornadoes in Iowa in 1974 and those appearing in Storm Data. Snider (1977) also concluded that the true number of tornadoes was several times larger than the number reported.

Summary and Conclusions

Statistically significant correlations were found between several independent variables and the response variable, tornadoes per 1000 square miles (SDEN). These independent variables were demographic and geographic in nature.

Combined in appropriate multiple regression models, they accounted for over 75 per cent of the apparent subregional variability in Missouri tornado strike probability.

Conclusions that may be drawn from this study include:

1. Less than twenty-five per cent of the subregional variability is not explained by demographic and geographic variables. This unexplained portion may be due to random noise in the data sample, independent variables not yet studied, or a combination of these factors.

2. This study has demonstrated that a significant portion of the variability in reported tornado density is accountable in terms of population density and other demographic factors. In general, high population densities are associated with the higher densities of reported tornadoes. Historically, Missouri counties with low population densities have reported relatively lower numbers of tornadoes per unit area. It is reasonable to infer that such counties might well have reported larger numbers of tornadoes if optimal demographic conditions, including population density had existed. If the assumptions that led to multiple regression models such as 11 or 13 are correct, a significant number of tornado strikes have occurred that were either not detected or not reported.

Suggestions for Further Research

1. The unexplained variability may be due, in part, to random noise in the data sample. A longer period of record, combined with careful screening of the entire data collection may serve to reduce this source of variability.

2. The grid used in computing terrain variables suffered from 2 defects: wide spacing (5 nautical miles north-south and 4 nautical miles east-west) and the errors associated with manual data extraction. A solution to these problems would be the use of National Cartographic Information Center (NCIC) digital terrain data tapes. Elevation data points on these tapes are separated by approximately 200 feet and were extracted from the primary base maps by objective techniques, thus eliminating both wide spacing and manual data extraction. The NCIC digital terrain tapes could be used to compute the required terrain variables for each county, and for the area adjacent to the county. This adjacent area should be considered, since orographic effects may significantly alter a thunderstorm as it approaches a county, either suppressing or enhancing its tornado generating potential.

3. Meteorological variables have been considered only indirectly in this study. Their explicit inclusion may be of assistance in explaining a portion of the subregional variability.

LITERATURE CITED

Abbe, Cleveland, 1888: Recent Progress in Dynamic Meteorology. Record of Science for 1887 and 1888, Gen. App. to Annual Report of the Board of Regents of the Smithsonian Institute to the end of June 1888, 355-424.

Abbey, R. F., Jr. and T. T. Fujita, 1975: "Use of tornado path lengths and gradations of damage to assess tornado intensity probabilities." Preprints, Nineth Conference on Severe Local Storms, American Meteorological Society.

Agee, Ernest M., 1970: "The climatology of Indiana tornadoes." Proceedings of the Indiana Academy of Science for 1969. Volume 79, 1970, pp 299-308.

Asp, M. O., 1956: "Geographical distribution of tornadoes in Arkansas." Monthly Weather Review, 84, 143-145.

Blodget, Lorin, 1857: Climatology of the United States. Philadelphia: Lippincott, 536 pp.

Brown, Charles W., and W. O. J. Roberts, 1935: "The distribution and frequency of tornadoes in the United States from 1880 to 1931." Trans. Am. Geophys. Union 16, 151-158.

_______, 1937: "The areal frequency of tornadoes in the United States by counties, 1880-1931." Trans. Am. Geophys. Union 18, 144-146.

Burley, M. W., and P. J. Waite, 1965: "Wisconsin tornadoes." Trans. Proc. Wisconsin Acad. Sci., Arts and Letters 54 1-35.

Carter, Horace S., 1970: Georgia Tornadoes. ESSA Technical Memorandum EDSTM 16. Silver Spring, Md.

Changnon, S. A., F. A. Hoff, and R. G. Semonin, 1968: "METROMEX: An investimation of inadvertent weather modification." Bulletin of the American Meteorological Society, 52, 958-967.

Changnon, S. A., Jr. and G. E. Stout, 1957: Tornadoes in Illinois. Circular 63, State of Illinois Department of Registration and Education. Urbana State Water Survey Division.

Canover, W. J., 1971: Practical Nonparametric Statistics. New York: John Wiley & Sons, Inc.

Court, Arnold, 1970: Tornado Incidence Maps. ESSA Technical Memorandum ERLTM-NSSL 49, Norman, Oklahoma, National Severe Storms Laboratory.

Dailey, Paul W., Jr., 1970: Tornadoes in Pennsylvania. Information Report Number 63, Institute for Research on Land and Water Resources. The Pennsylvania State University, University Park, Pennsylvania.

Dames & Moore, 1975: A Meteorological and Engineering Approach to the Regionalization of Tornado Wind Criteria for Nuclear Power Plant Design. Prepared for United States Energy Research and Development Administration Division of Reactor Research and Development. AEC Contract No. AT(11-1)-2396.

Davies-Jones, Robert, Joseph Golden, Joseph Schaefer, Ronald H. Pine, H. E. Landsberg, and L. Pedersen, 1973: Correspondence, Science, 180, 544-548.

Davies-Jones, Robert and Edwin Kessler, 1974: "Tornadoes." Chapter 16 from Weather and Climate Modification. John Wiley & Sons.

Day, P. C., 1930: "Weather Bureau gets data on behavior and effects of tornadoes." Yearbook of Agriculture, 1930. Washington, D. C., U.S. Government Printing Office, 530-534.

Eshelman, S. F. and J. L. Stanford, 1977: "Tornadoes, funnel clouds and thunderstorm damage in Iowa during 1974." Iowa State Journal of Research, 51, 327-361.

ESSA, 1969: Lightning. PI 660024, Public Information Office, Environmental Science Services Administration, U.S. Dept. of Commerce.

Fawbush, Ernest J., Robert C. Miller, and Lloyd G. Starrett, 1951: "Methods of forecasting tornado development." Bulletin of the American Meteorological Society, 32, 1-9.

Finley, J. P., 1884: "Report on the character of 600 tornadoes. Papers of the Signal Service, No. 7, 29 pp. (First publication, 1882, withdrawn for correction.)

_______, 1887: "Tornadoes: What they are and how to observe them, with practical suggestions for the protection of life and property." (Insurance Monitor Press, New York, N.Y.), 196 pp.

_______, 1888a: "The relation of tornado regions to areas of low pressure." Am. Meteorol. J., 5, 83-87.

_______, 1888b: "Tornadoes: What they are and how to escape them." (J. H. Soule, Washington, D. C.) 90 pp. (Also published as Signal Corps Circ. No. 22).

_______, 1889: "State tornado charts." Am. Meteorol. J., 5, 446-476, 501-507, 545-551.

_______, 1890: "Tornadoes." (First Prize Essay), Am. Meteorol. J., 7, 165-179.

_______, 1925: "Weather, crop and social insurance in the United States through climatological investigations by the Finley system, 1880-1925." National Storm Insurance Bureau, New York, N.Y., iii + 39 pp.

_______, 1932: "Climatological business charts for the protection and benefit of commerce, aviation and agriculture and the promotion of insurance against weather hazards." National Storm and Aviation Insurance Bureau, New York, N.Y., (typescript), 6 pp.

Fujita, T. Theodore, 1973: "Tornadoes around the world." Weatherwise, 26:2 (April 1973) 56-62.

Garson, R. C., J. Morla-Catalan, and C. A. Cornell, 1975: "Tornado design winds based on risk." Journal of Structural Division, Proceedings of American Society of Civil Engineers, 101, 1883-1897.

Goldfield, Edwin D., ed., 1967: County and City Data Book, 1967. U.S. Bureau of the Census. Washington, D. C., U.S. Government Printing Office.

Howe, George M., 1974: "Tornado path sizes." Journal of Applied Meteorology, 13:3 (April 1974) 343-347.

Idso, Sherwood B., 1975: "Whirlwinds, density currents, and topographic disturbances: A meteorolgical melange of intriguing interactions." Weatherwise, 28:2 (April 1975) 61-65.

Kessler, Edwin and J. T. Lee, 1978: "Distribution of the tornado threat in the United States." Bulletin of the American Meteorological Society, 59, 61-62.

_______, 1976: Normalized Indices of Destruction and Death by Tornadoes. NOAA Tech. Memo. ERL-NSSL, 47 pp.

Kruskal, W. H., 1958: "Ordinal measures of association." Journal of the American Statistical Association, 53, 814-861.

McDonald, James R. and Robert F. Abbey, Jr., 1977: "Advancements in tornado risk analysis." Preprints, Tenth Conference on Severe Local Storms, American Meteorological Society.

McDonald, J. R., K. C. Mehta, and J. E. Minor, 1975: Development of Windspeed Risk Models for the Savannah River Plant Site. Report for E. I. DuPont de Nemours Company under purchase order Number AXC 672-W.

Markee, E. H., Jr., J. G. Beckerley and K. E. Sanders, 1974: "Technical basis for interim regional tornado criteria." U.S. Atomic Energy Commission, Office of Regulation, Washington, D. C.

Mogil, H. Michael, Marjorie Rush and Mary Kutka, 1977: "Lightning--an update." Preprints, Tenth Conference on Severe Local Storms, American Meteorological Society.

Moore, H. E., 1958: Tornadoes over Texas, A Study of Waco and San Angelo in Disaster. Austin, University of Texas press, 334 pp.

Natrella, Mary G., 1963: Experimental Statistics. National Bureau of Standards Handbook 91. Washington, D. C., U.S. Government Printing Office.

Neter, John and William Wasserman, 1974: Applied Linear Statistical Models. Homewood, Il., Richard D. Irwin, Inc.

Reed, Jack W., 1971: "Some averaged measures of tornado intensity based on fatality and damage reports." Preprints, Seventh Conference on Severe Local Storms, American Meteorological Society.

Sadowski, Alexander, 1965: "Potential casualties from tornadoes." Paper presented at the 244th National Meeting, American Meteorological Society, Reno, Nevada.

Sims, John H. and Duane D. Baumann, 1972: "The tornado threat: Coping styles of the North and South." Science, 176, 1386-1392.

Skaggs, Richard H., 1970: "Analysis and regionalization of the diurnal distribution of tornadoes in the United States." Monthly Weather Review, 97, 103-115.

Snider, C. R., 1977: "A look at Michigan Tornado statistics." Monthly Weather Review, 105, 1341-1342.

Spohn, H. R. and P. J. Waite, 1962: "Iowa tornadoes." Monthly Weather Review, 90, 398-406.

Thom, H. C. W., 1963: "Tornado probabilities." Monthly Weather Review, 91, 730-736.

Vaiksnoras, John V., 1972: "Tornadoes in Tennessee (1916-1970)." Nashville, Tennessee, the University of Tennessee Institute for Public Service, Civil Defense University Extension Program.

Wen, Y, K. and S. L. Chu., 1973: "Tornado risks and design wind speed." Journal of Structural Division, Proceedings of American Society of Civil Engineers, 99, 2409-2421.

Wilson, John W. and Stanley A. Changnon, Jr., 1971: Illinois Tornadoes. Circular 103. State of Illinois Department of Registration and Education. Urbana, Illinois State Water Survey.

Zegel, Ferdinand H., 1967: "Lightning deaths in the United States; a seven-year survey from 1959 to 1965." Weatherwise, 20, 169-173, 179.

NRC FORM 335 (7-77)

U.S. NUCLEAR REGULATORY COMMISSION

BIBLIOGRAPHIC DATA SHEET

1. REPORT NUMBER (Assigned by DDC)
NUREG/CR-2358

4. TITLE AND SUBTITLE (Add Volume No., if appropriate)
Subregional Variability in Missouri Tornado Statistics

2. (Leave blank)

3. RECIPIENT'S ACCESSION NO.

7. AUTHOR(S)
S.W. Jamison and G.L. Darkow

5. DATE REPORT COMPLETED
MONTH September YEAR 1981

9. PERFORMING ORGANIZATION NAME AND MAILING ADDRESS (Include Zip Code)
Department of Atmospheric Science
University of Missouri
Columbia, Missouri 65211

DATE REPORT ISSUED
MONTH November YEAR 1981

6. (Leave blank)

8. (Leave blank)

12. SPONSORING ORGANIZATION NAME AND MAILING ADDRESS (Include Zip Code)
Division of Health, Siting and Waste Management
Office of Nuclear Regulatory Research
U.S. Nuclear Regulatory Commission
Washington, D.C. 20555

10. PROJECT/TASK/WORK UNIT NO.

11. CONTRACT NO.
FIN B6196
NRC-04-77-055

13. TYPE OF REPORT
technical

PERIOD COVERED (Inclusive dates)
April 1977 - June 1979

15. SUPPLEMENTARY NOTES

14. (Leave blank)

16. ABSTRACT (200 words or less)

The reality of subregional variability in tornado occurrence density as evidenced in the county to county variability in Missouri is examined. Reported tornadoes for the period from 1916 through 1975 were used. Demographic and geographic factors known tp impact on tornado reporting efficiencies and accuracies are related to county tornado report densities by step-wise multiple linear regression techniques. The analysis suggests that over 75 percent of the county to county apparent variability in reported tornado densities in Missouri is explainable in terms of variability in population density, other related demograpbic variables and regional scale geographic factors. The remaining 25 percent of unexplained variance appears quite randomly distributed and may be due to random noise in the initial report data or to physical and geographic factors not included in the study. The analysis suggests that during the 1916-1975 period actual tornado occurrences in Missouri exceeded reported occurrences by at least 65 %. Most of these reporting deficiencies are manifest in the reports prior to 1954, the start of the press clipping service era. Although the population-induced biases in reported tornado densities during the post-1954 reporting period were not examined in this study, they are likely still present to a lesser extent and should be taken into account in interpreting subregional scale variability in reported tornado densities.

17. KEY WORDS AND DOCUMENT ANALYSIS

17a. DESCRIPTORS

17b. IDENTIFIERS/OPEN-ENDED TERMS

18. AVAILABILITY STATEMENT
Unlimited

19. SECURITY CLASS (This report)
Unclassified

21. NO. OF PAGES

20. SECURITY CLASS (This page)
Unclassified

22. PRICE
$

NRC FORM 335 (7-77)

UNITED STATES
NUCLEAR REGULATORY COMMISSION
WASHINGTON, D. C. 20555

OFFICIAL BUSINESS
PENALTY FOR PRIVATE USE, $300

POSTAGE AND FEES PAID
U.S. NUCLEAR REGULATORY
COMMISSION

U.S. MAIL

120555064215 2 ANRBR6
US NRC
ADM DOCUMENT CONTROL DESK
PDR
016
WASHINGTON DC 20555