Spatial Analysis of Agricultural *Cucurbita* Sp. Varieties in the Eastern Broadleaf Province

Kathleen M. Baker

Follow this and additional works at: https://scholarworks.wmich.edu/masters_theses

Part of the Geography Commons

**Recommended Citation**


https://scholarworks.wmich.edu/masters_theses/4789
SPATIAL ANALYSIS OF AGRICULTURAL CUCURBITA SP. VARIETIES IN THE EASTERN BROADLEAF PROVINCE

by

Kathleen M. Baker

A Thesis
Submitted to the Faculty of The Graduate College in partial fulfillment of the requirements for the Degree of Master of Arts Department of Geography

Western Michigan University Kalamazoo, Michigan August 1998
ACKNOWLEDGEMENTS

Thank you, first and foremost, to my friends and family who have added the word *Cucurbitaceae* to their vocabularies for my sake. My thesis advisor, Dr. Rolland Fraser, and committee members, Dr. Ilya Zaslavsky and Dr. Oscar Horst, have been marvelous, what can I say? Even when inedible cucurbits made you laugh, you tempered my crazy ideas with good sense. To the grad students, faculty and staff at Western, especially those of you who offered suggestions although pumpkins were far from your number one priority – you’ve been great, guys. May lightning never strike you. If any of you have had nightmares involving cucurbits, I take full responsibility and apologize deeply.

And, of course, I gratefully acknowledge my family: Gregory, who impressed the guys at Farm Bureau; Laura, who twirled the “dishrag” when the stress got high; Grandma, who faithfully watched my gourds; and my parents who taught me the importance of vegetables, God and simplicity. Thanks.

Kathleen M. Baker
The purpose of this research was to explain and quantify twentieth century agricultural distributions of the genus *Cucurbita* in the United States to determine the effect of temporal changes in economic utilization and a variety of historical, environmental and social circumstances on those distributions. Geographic information system (GIS) methods were combined with statistical software techniques to analyze the spatial and temporal trends of pumpkin and squash growth in one region (the Eastern Broadleaf Province, as defined by Bailey) of the United States from 1982-1992.

Counties with fewer frost free days, later spring frosts, more accumulated growing degree units throughout the growing season and less estimated change in population during the ten year period were more likely to report zero pumpkin and squash acreage. The increase in ornamental production within the Eastern Broadleaf Province was influenced by a combination of local factors, as opposed to broad climatological or social characteristics.
# TABLE OF CONTENTS

**ACKNOWLEDGEMENTS** .................................................................................................................................................................................. ii

**LIST OF TABLES** .................................................................................................................................................................................. v

**LIST OF FIGURES** .................................................................................................................................................................................. vi

**CHAPTER**

I. INTRODUCTION.......................................................................................................................................................................................................................................................... 1

Agricultural Trends .......................................................................................................................................................................................................................................................... 2

Research Objectives ................................................................................................................................................................................................................................. 4

II. NOMENCLATURE AND HISTORY OF THE GENUS CUCURBITA .................................................................................................................................................................................................................................................. 9

Categorization of Economic Utilization .......................................................................................................................................................................................................................................................... 18

* Cucurbita* as an Agricultural Genus .......................................................................................................................................................................................................................................................... 22

Summary and Conclusions ......................................................................................................................................................................................................................................................................................... 25

III. THE GENUS CUCURBITA AS AN UNITED STATES AGRICULTURAL CROP IN THE TWENTIETH CENTURY .......................................................................................................................................................................................................................................................................................... 26

Introduction .......................................................................................................................................................................................................................................................................................................... 26

Data ........................................................................................................................................................................................................................................................................................................ 28

Methods .................................................................................................................................................................................................................................................................................. 34

Results .............................................................................................................................................................................................................................................................................. 38

Discussion ..................................................................................................................................................................................................................................................................... 54
Table of Contents---continued

CHAPTER

IV. SPATIAL PATTERNS OF THE GENUS *CUCURBITA* IN THE EASTERN BROADLEAF PROVINCE, 1982-1992 .............................................. 57

    Introduction ............................................................................... 57
    Data ............................................................................................ 59
    Methods ..................................................................................... 64
    Results ........................................................................................ 78
    Discussion .................................................................................. 88

V. ANALYSIS OF ENVIRONMENTAL AND ECONOMIC FACTORS ON THE DISTRIBUTION PATTERNS OF THE GENUS *CUCURBITA* ......................................................... 92

    Introduction ............................................................................... 92
    Data ............................................................................................ 95
    Methods ..................................................................................... 98
    Results ........................................................................................ 105
    Discussion .................................................................................. 122

VI. SUMMARY AND CONCLUSIONS ..................................................... 126

BIBLIOGRAPHY ................................................................................. 132
LIST OF TABLES

1. Taxonomy and Economic Uses of the Genus *Cucurbita* .......................... 11

2. National and Province Farm Production Rate of Change, 1982-1992 ................................................................. 78

3. Spatial Autocorrelation Estimation Using Moran's Contiguity Coefficient, \( I \) ................................................................. 86

4. Correlation of Frost Free Days and Agricultural Values ...................... 107

5. Correlation of Date of Last Frost and Agricultural Values .................. 110

6. Correlation of Ten Month Growing Degree Units and Agricultural Values ................................................................. 114

7. Correlation of Population Estimates and Agricultural Values ............ 117

8. Spatial Regression Estimates ......................................................... 121
### LIST OF FIGURES

1. Hierarchy of Federal Data Sources ................................................................. 6
2. Squash, Pumpkins and Gourds.................................................................. 14
3. Presence of Pumpkin and Squash in Agricultural Data Sources,
   1920-1994.................................................................................................. 30
4. Squash as a Vegetable Crop in Historical Agricultural Census
   Data (U.S. Census Bureau 1909-1992)................................................. 39
5. Pumpkin as a Vegetable Crop in Historical Agricultural Census
   Data (U.S. Census Bureau 1909-1992)............................................... 40
7. Frozen Pumpkin and Squash, U.S. Packed and Processed
   (USDA 1925-1995).................................................................................. 43
8. Canned Pumpkin and Squash, U.S. Packed and Processed
   (USDA 1925-1995)................................................................................. 44
9. Moving Averages of *Cucurbita* Processing Data.................................. 45
10. Domestic Squash Shipments by Month (USDA 1925-1995).................. 47
11. Lag Correlation of Monthly Shipment Data........................................... 48
12. Squash Shipments and Arrivals (USDA 1925-1995).............................. 49
13. Pumpkin Arrivals (USDA 1925-1995)..................................................... 50
14. *Cucurbita* Acreage per Square Mile in the Contiguous 48
    States, 1992 (U.S. Census Bureau 1995)........................................... 51
List of Figures—Continued

15. Pumpkin Acreage per Square Mile in the Contiguous 48 States, 1992 (U.S. Census Bureau 1995) .......................... 52


17. National Hierarchy of Ecological Units, Eastern Broadleaf Province (U.S. Forest Service 1994) .......................... 60

18. Sections of the Eastern Broadleaf Province (U.S. Forest Service 1994) ...................................................... 62

19. Integration of Agricultural Census Records and Spatial Boundaries ............................................................... 65

20. Overlay of Province, Section, and County Boundaries ..................................................................................... 67

21. The Presence of Zeros in the Census of Agriculture (U.S. Census Bureau 1995) .................................................. 71

22. Neighbor Weighting Options for Estimating Spatial Autocorrelation ........................................................... 73

23. Rates of Change in Farms and Acreage Within the Eastern Broadleaf Province ............................................... 80

24. County Pumpkin Acreage Densities (U.S. Census Bureau 1982-1992) ............................................................... 81

25. County Squash Acreage Densities (U.S. Census Bureau 1982-1992) ................................................................. 82

26. Distribution of County Change in Acreage ........................................................................................................ 84

27. Extraction of Census Population Estimates and Integration With Spatial Boundaries ......................................... 100

28. Extraction and Conversion of Climate Station Data for Integration With Polygon Boundaries .......................... 101
List of Figures---Continued

29. Integration of Climate Data With Polygon Boundaries ......................... 102

30. Minimum Frost Free Days at a 10% Probability Threshold
(USCDC 1995) .............................................................................................. 109

31. Maximum Last Frost Date With a 90% Probability Threshold
(USCDC 1995) .............................................................................................. 112

32. Mean Growing Degree Units Calculated From a Base Temperature
of 55 F and Accumulated From January 1 to October 31
(USCDC 1995) .............................................................................................. 116

33. Rate of County Population Change, 1982-1992
(U.S. Census Bureau 1995) .......................................................................... 118

34. Scatterplots of Pumpkin Acreage and Population Changes ................. 119

35. Scatterplot of Log Transformed Variables ........................................... 120
CHAPTER I

INTRODUCTION

Throughout the United States, displays of pumpkins, squash, and gourds herald the coming of autumn. The plants belong to seven economically valuable species and species groups of the genus *Cucurbita*, and their brightly colored fruits have had cultural and economic importance in the America’s since before the beginnings of agriculture.

In popular perception, members of the gourd family (*Cucurbitaceae*) have long been associated with fertility (Ralf 1980). The frequency of their occurrence in prominent literary works and paintings illustrates their cultural significance around the world. Indigenous peoples of the tropics and subtropics relied upon the versatility of native species as ceremonial, medicinal and nutritious fruits. *Cucurbita* is the only genus of the gourd family (*Cucurbitaceae*) that is native to the American continents (Whitaker 1981) and its importance is evident from its early domestication and diversity of uses.

Historical utilization of *Cucurbita* ranges from processing the fruit as a pie filling, to extracting a vertebrate poison from its seeds, to drying the rinds
for ceremonial containers. Edible varieties include the summer squash, winter squash, marrows and pie pumpkins. Carving pumpkins and small, decorative gourds have popular appeal as ornaments in conjunction with Halloween, Thanksgiving and other fall celebrations.

Agricultural Trends

The use of the genus agriculturally, especially when combined with other principal crop plants in intercropping patterns (simultaneously cultivating more than one crop in a single field), has continued in Honduras and other Central American countries from pre-Columbian times (Powers and others 1994). In North America, by contrast, especially following the arrival of European colonists, the cucurbits as a family were associated with gardening, fruit growing and viticulture. They were considered to be horticultural plants, as opposed to agricultural species. It has been only since the beginning of the twentieth century that pumpkins, squash, cucumbers and melons have been considered agricultural crops in the United States. Because the United States Department of Agriculture (USDA) categorizes cucurbit fruits primarily as "Vegetables and Melons" in its statistical reporting, quantitative documentation of inedible cucurbit products is almost nonexistent (USDA 1920-1995). The strict focus on edible varieties has precluded research analyzing spatial and temporal patterns of Cucurbita as a
genus. While popular journals catalogue recipes, carving patterns and interesting facts, the more scholarly publications tend to focus on the archeology of domestication and the geography of wild species or early domesticates (Cutler and Whitaker 1961; Whitaker 1981; Nee 1990; Harlan 1992; Sauer 1993).

The 1980s and 1990s were a period of major transition for farming throughout the United States. Although trends in the structure and geography of agriculture have been researched with regard to field crops and livestock (National Farm Institute 1970; Hallam 1993; Hart 1998), the effects of these changes on vegetable and ornamental species are less documented. Recently, pumpkins and squash have shown an increase in economic importance as crop plants in the agricultural record. During the period 1982 to 1992, the number of farms in the United States decreased 14.09%, while farm acreage decreased 4.18%. During this same time period, squash acreage increased 35.87% and pumpkin acreage increased 143.45% (U.S. Census Bureau 1995). Simultaneously, imports of *Cucurbita* varieties also increased (CRS 1988).

Because of the growing economic importance of these vegetables, progenitor species as well as those with potential for new cultivars are being located and their ranges studied (GRIN 1998). Distribution characteristics of both edible and inedible *Cucurbita* varieties are vital to the implementation of
programs that seek to preserve gene sources of valuable varieties in their wild state. Gene curators, in the form of research cooperatives, such as the Cucurbit Genetics Cooperative (CGC), have been established to interface with the USDA through Crop Advisory Committees (CAC) (CGC 1998). Ongoing research surveys the location of agricultural crop plants as a way to determine the location of plants that retain their wild genes without cross pollination to cultivated varieties (CGC 1998). The gourd family is especially valuable not only as a source of human nutrition, but also for the variety of useful products that can be created from its fruits. Many species grow in arid regions and are a source of moisture for livestock and wildlife (Bates and others 1990).

Research Objectives

The purpose of this study is to explain and quantify agricultural distributions to determine the effect of temporal changes in economic utilization and a variety of historical, environmental and social circumstances on the spatial pattern of the genus *Cucurbita* in the United States. The four objectives are to: (1) investigate the spatial and temporal trends in the utilization of the genus *Cucurbita* as an agricultural crop in the United States in the twentieth century, (2) analyze the agricultural distribution of pumpkins and squash within one region of the United States during the period from
1982 to 1992 with respect to national trends in production and use, (3) quantify physical and economic factors affecting the spatial pattern of agriculturally grown pumpkins and squash within that region and (4) develop a methodological model of agricultural distribution analysis coupling geographic information systems (GIS) and spatial statistics.

This analysis of the genus *Cucurbita* integrates data from a variety of sources, which vary in their spatial resolution and content. Data are predominantly available as part of United States government censuses or annual records. Agricultural statistical data, environmental data, population data, and portions of the spatial data boundary files are produced and updated within federal agencies (Figure 1). National, state, and county boundary files were derived from private sources.

This thesis consists of six chapters which outline the history of the *Cucurbita* as an agricultural genus, the analysis of United States utilization patterns during the twentieth century and the quantification of county level distributions in one region of the country. Chapter II presents a brief overview of the genus *Cucurbita* and its position in the family *Cucurbitaceae* taxonomically (botanically and in the vernacular) and agriculturally. Actual and potential economic utilization of various valuable species and subspecies are also discussed with particular emphasis on the linkage between use patterns and common name. The history of economic uses is accompanied by
Figure 1. Hierarchy of Federal Data Sources.
a review of historical spatial trends as determined by archeologists and
ethno-botanists. This chapter elaborates on the historical characteristics and
economic importance of the genus *Cucurbita* that justify this research and
examines the goals of this study with respect to previous research endeavors.

Chapter III describes and analyzes the national statistics that quantify
*Cucurbita* production in the United States during the twentieth century.
Schulman, Zimmer, and Danaher (1994) recommend that geographical
analyses of agricultural data link various spatial resolutions. Following
chapters will analyze farm numbers and acreage at a smaller scale, but it is
first essential to understand the national trends in comparison to overall
farming characteristics before assessing more local patterns.

Agricultural distributions of the genus *Cucurbita* are analyzed at the
county level in Chapter IV within one ecologically defined region of the
United States. This portion of the study focuses on the time period from 1982
to 1992, a time of national agricultural change (Hallam 1993), with the
purpose of viewing the effects of this change on *Cucurbita* agricultural
varieties. Pumpkins, squash and gourds have received much attention in the
twentieth century through the popular press as plants with diverse edible
and inedible uses. However, the spatial and regional differences in this
pattern of popularity or the agricultural distribution of the species have not
been documented.
In an attempt to quantify the significance of various factors in relation to *Cucurbita* distributions, spatial statistics procedures are combined with geographic information system techniques in Chapter V. Both environmental and population variables are considered in the analysis of vegetable distributions in the United States, about which there is little existing literature. Conclusions of the study, a brief summary of the thesis and possible directions for future research are included in Chapter VI.
CHAPTER II

NOMENCLATURE AND HISTORY OF THE GENUS CUCURBITA

The following chapter surveys the historical geography and utilization of the genus Cucurbita with respect to both common and taxonomic names. The family Cucurbitaceae, frequently referred to as the gourd family or the squash family, is comprised of such diverse genera as Citrullus (watermelon), Cucumis (cucumber, muskmelon), Cucurbita (squash, pumpkin), Lagenaria (bottle gourd), Luffa (sponge gourd) and Momordica (balsam apple). With the exception of Cucurbita, all genus are native to the Old World. The genus Cucurbita is native to the America’s and no evidence of species elsewhere in the world predates the arrival of Europeans in 1492 (Harlan 1992).

The most recent classification of Cucurbita by Michael Nee (1990) divides the genus into thirteen species or species groups. Prior to this categorization, 20 to 26 individual species are typically cited (Cutler and Whitaker 1961; Bates and others 1990). Nee’s revised taxonomy, however, is widely accepted due to its completeness and the integration of biological evidence with historical phytogeography and the archeological domestication records. This taxonomy also takes precedence over other classifications in
recent United States government information resources (GRIN 1998) and so will be cited in this research. Updates regarding *Cucurbita* ssp. genes are available from the Cucurbit Genetics Cooperative (CGC) (Robinson and Hutton 1996).

Of the thirteen *Cucurbita* species groups, this paper focuses on the seven species and subspecies considered economically important by the United States Department of Agriculture (USDA), which include: (1) *C. argyrosperma* ssp. *argyrosperma*, (2) *C. ficifolia*, (3) *C. foetidissima*, (4) *C. maxima* ssp. *andreana*, (5) *C. maxima* ssp. *maxima*, (6) *C. moschata* and (7) *C. pepo* (GRIN 1998). Table 1 summarizes the taxonomy of economically viable plants, as catalogued by the Germplasm Research Information Network (GRIN). GRIN contains approximately 3000 vascular plants of world economic importance and is maintained by the USDA Agricultural Research Service (ARS) (USDA 1998).

While the botanical division of the genus involves between 13 and 26 species, the majority of *Cucurbita* varieties are commonly known as either pumpkins, squash, or gourds in common usage. Unfortunately, this terminology is virtually useless in identifying the botanical taxa of the plants (Cutler and Whitaker 1961). Common names not only fail to correspond with botanical nomenclature, but often refer to assorted varieties and cultivars within different species (Nee 1990).
Table 1

Taxonomy and Economic Uses of the Genus *Cucurbita*

<table>
<thead>
<tr>
<th><em>Cucurbita</em> ssp. Varieties</th>
<th>Selected Common Names</th>
<th>Economic Uses</th>
</tr>
</thead>
<tbody>
<tr>
<td><em>Cucurbita argyrosperma</em></td>
<td>Ayote, green-stripe cushaw, Japanese pie pumpkin, silver-seed gourd, white cushaw</td>
<td>Human Food</td>
</tr>
<tr>
<td>ssp. <em>argyrosperma</em></td>
<td></td>
<td></td>
</tr>
<tr>
<td>var. <em>argyrosperma</em></td>
<td></td>
<td></td>
</tr>
<tr>
<td>var. <em>callicarpa</em></td>
<td></td>
<td></td>
</tr>
<tr>
<td>var. <em>palmeri</em></td>
<td></td>
<td></td>
</tr>
<tr>
<td>var. <em>stenosperma</em></td>
<td></td>
<td></td>
</tr>
<tr>
<td><em>Cucurbita ficifolia</em></td>
<td>black-seeded squash, chilacayote, fig-leaf gourd, Malabar gourd</td>
<td>Human Food</td>
</tr>
<tr>
<td><em>Cucurbita foetidissima</em></td>
<td>buffalo gourd, calabazilla, calabacilla amarga, chilicote, Missouri gourd</td>
<td>Human Food</td>
</tr>
<tr>
<td><em>Cucurbita maxima</em></td>
<td></td>
<td></td>
</tr>
<tr>
<td>ssp. <em>andreana</em></td>
<td></td>
<td>Gene Source</td>
</tr>
</tbody>
</table>
Table 1 - Continued

<table>
<thead>
<tr>
<th>Cucurbita ssp. Varieties</th>
<th>Selected Common Names</th>
<th>Economic Uses</th>
</tr>
</thead>
<tbody>
<tr>
<td><em>Cucurbita maxima</em></td>
<td>banana squash, buttercup squash, giant pumpkin, Hubbard squash, turban squash, winter squash</td>
<td>Human Food</td>
</tr>
<tr>
<td>ssp. <em>maxima</em></td>
<td></td>
<td></td>
</tr>
<tr>
<td><em>Cucurbita moschata</em></td>
<td>ayote, butternut squash, cheese pumpkin, citrouille, golden cushaw, pumpkin, squash, winter crookneck squash</td>
<td>Human Food</td>
</tr>
<tr>
<td><em>Cucurbita pepo</em></td>
<td>acorn squash, cocozelle, fordhook squash, marrow, ornamental gourd, pattypan squash, pumpkin, scallop squash, spaghetti squash, straightneck squash, summer crookneck squash, table queen squash, zucchini</td>
<td>Human Food</td>
</tr>
<tr>
<td>var. <em>fraterna</em></td>
<td></td>
<td>Animal Food**</td>
</tr>
<tr>
<td>var. <em>ozarkana</em></td>
<td></td>
<td>Environmental Use</td>
</tr>
<tr>
<td>var. <em>ovifera</em></td>
<td></td>
<td>Gene Source</td>
</tr>
<tr>
<td>var. <em>pepo</em></td>
<td></td>
<td>Vertebrate Poison</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Medicines</td>
</tr>
</tbody>
</table>

* Summary of information obtained from the USDA Germplasm Resources Information Network, created and maintained by the Agricultural Research Service

**Economic Uses not listed in GRIN
Diversity of fruit morphology and pigmentation is characteristic of the *Cucurbita* cultivars and similar appearing fruits are often produced from diverse species. Differentiation between pumpkin, squash and gourd tends to be associated with the physical appearance of the fruit or the uses for which the fruit is suited (Sauer 1993). Although taxonomy, common name and use are intertwined to a considerable degree, common terminology is defined and discussed with respect to fruit characteristics and taxonomy before the economic uses are described in detail. Fruits from some of the more common *Cucurbita* varieties are shown in Figure 2.

In general, all *Cucurbita* fruits prepared as a vegetable are considered squash. Squash include both the winter squash, whose fruit is eaten when completely ripe, and the summer squash, whose fruit is eaten when immature. The flesh is milder in flavor, but more diverse in color and texture, than that of the pumpkins (Tapley 1936; Cutler and Whitaker 1961). Squash rinds range in color from yellows and greens to oranges, and can be striped or mottled. There is also a marked diversity in the shape of fruit from the various cultivars (Sauer 1993). Among the most popular squash varieties are *C. maxima* ssp. *maxima* (Hubbard, buttercup), *C. moschata* (butternut), *C. pepo* var. *ovivera* (acorn, pattypan) and *C. pepo* var. *pepo* (zucchini, spaghetti squash) (Bailey 1937; Gurney 1986; Stokes 1998; W. Atlee Burpee and Co. 1998). In some instances, certain varieties of *C. argyrosperma* ssp. *rgyrosperma*,
Figure 2. Squash, Pumpkins and Gourds.
C. ficifolia and C. foetidissima are locally considered squash, but they are less economically important as a source of human food (GRIN 1998).

The term “pumpkin” generally refers to any of the orange-colored, round fruits of a Cucurbita variety. Although pumpkins are also used as a food source for both humans and livestock, their flesh is coarser in texture, and in flavor, than those fruits classified as squash. Unlike the squash, they are eaten only when completely ripe, and rarely used as a vegetable (Cutler and Whitaker 1961).

Pie pumpkins are edible members of the genus Cucurbita and originally included only those edible pumpkin varieties that were baked and sweetened for use in bakery products such as pies, breads and cookies. Contemporary commercial processors, however, use both pumpkin and winter squash from the species C. moschata and C. pepo in processed pie fillings, as well as the cushaw varieties of C. argyrosperma var. argyrosperma. The flesh of popular pie pumpkin cultivars, such as ‘Buckskin’, ‘Sugar Pie’, and ‘Green Striped Cushaw’ are touted as sweeter and less stringy than that of other pumpkins, and the fruits are traditionally among the smaller pumpkin varieties with dark orange flesh (Gurney 1996; Stokes 1998).

In addition to their edible uses, pumpkins are harvested and used as ornamental fruits for the holidays of Halloween and Thanksgiving as well as other autumn celebrations. Well known pumpkin cultivars for carving and
display, such as ‘Kentucky Field’ and ‘Howden’, generally belong to the taxa *Cucurbita pepo* var. *pepo* (Tapley 1937). Occasionally some varieties of *Cucurbita moschata* are also known as pumpkins (GRIN 1998).

The larger carving pumpkin cultivars, such as ‘Connecticut Field’ belong to the taxa *C. maxima* ssp. *maxima*. Giant pumpkins, ‘Prizewinner’ and ‘Dill’s Atlantic Giant’, grown for the tremendous size of their fruits, are the largest of the cucurbits and also belong to this taxa (Nee 1990). Since most economically important members of *C. m. var. maxima* are considered squash, however, there is ongoing debate concerning whether giant pumpkins are, in fact, pumpkins. For general purposes, if the giant fruit is deep yellow or orange in color, it is considered a pumpkin. If it remains a dark green or light greenish yellow when ripe, it is considered a squash (Royte 1997).

Similar to the squash, at least in variety of shape and coloration, are the gourds. In general, gourds have thicker rinds than either pumpkins or squash and can be dried for practical, ceremonial, and ornamental uses (Conniff 1987). When referring to the genus *Cucurbita*, the term “ornamental gourd” is applied to the small inedible, hard rind fruits that are grown from the *pepo* and *ovifera* varieties of *C. pepo* (Bailey 1937). Cultivars of *C. argyrosperma* ssp. *argyrosperma*, *C. ficifolia*, and *C. foetidissima*, such as the silver-seed gourd, the Malabar gourd and the buffalo gourd, respectively, have both edible and inedible uses (GRIN 1998).
Although the terms pumpkin and squash are at least specific to the *Cucurbita* genus, the word “gourd” more frequently leads to confusion since the family *Cucurbitaceae* is commonly referred to as the gourd family. The fruits of at least six cucurbit species, including *Lagenaria siceraria*, the large African bottle gourd, and *Luffa cylindrica*, the sponge gourd, are known as gourds in addition to those within the genus *Cucurbita* (Bailey 1937; Tapley 1937). The calabash, the fruit of *Crescentia cujete*, is also inappropriately referred to as a gourd. The tree bears no relation to the family *Cucurbitaceae*, and instead is a relative of the catalpa (Tapley 1937; Cutler and Whitaker 1961).

There have long been attempts to standardize common names with taxonomic names. In the United States, this would involve calling *C. maxima*, “squash”, and *C. pepo* and *C. moschata*, “pumpkin”, regardless of fruiting characteristics. However, standardization attempts have failed so far due to the centuries of history associated with round, orange pumpkins, the historical “pompions”. Also hindering standardization are regional and international discrepancies in nomenclature that persist even within attempts at uniformity. In Australia, for example, the recommendation was made to limit “pumpkin” to *C. maxima*, and “squash” to *C. pepo*, which would create a terminology opposite that of the United States (Tapley 1937).
Consumer demand as well as common name is based on the physical characteristics of *Cucurbita* fruit. Ornamental gourds are used for purely decorative purposed because of the variation in their color, shape and texture. Equally diverse are winter and summer squash, which supplement autumn displays but are primarily considered vegetable crops (Stokes 1998). Although typically orange and round, even pumpkin purchases are dictated to some extent by holiday fads. Trends in carving style and media coverage influence the shape and color of pumpkins bought for carving. Pumpkins as ornamentals are often chosen for their subtle variation in color and shape that lead to aesthetically pleasing holiday displays. Pie pumpkins are chosen in a similar manner and, like squash, are used as ornamentals prior to cooking.

Categorization of Economic Utilization

Because of the confusion surrounding *Cucurbita* common names, various attempts have been made to correlate botanically correct names with economic uses. Utilization tends to relate to the nutritional, chemical, or botanical characteristics of the seeds, rind, vegetation and fruit of the plant and often are identified only by common name in historical literature. In the following discussion, the GRIN database was the dominant reference for taxonomic correctness as it integrates much of the research in this area (Bates
and others 1990; Nee 1990). Where GRIN did not catalogue a use pattern documented elsewhere, other sources are referenced.

As a human food source, the fruit of *Cucurbita* varieties classified as both pumpkins and squash are edible and have been used as both fruits and vegetables throughout the Americas. Although squash are considered vegetables in statistical documents, they are the fruit of the *Cucurbita* plant. As the Supreme Court ruled in 1893, “in the common language of the people, tomatoes and all other garden crops that are served with dinner, and not dessert, are vegetables” (Missler 1997). The baked fruits of both pumpkin and squash plants are a popular ingredient in bakery products, soups, and preserves.

The seeds and seed oil of pumpkins and squash are also edible and nutritious. In fact, it was probably the nutritional value of the seeds that led to early domestication prior to the development of fleshier cultivars (Harlan 1992). Because of their high protein content and the physical, sensory and mineral properties of these proteins, pumpkin seeds have been studied as an additive in Bologna sausages. Similar to soy products, pumpkin seed proteins can act as binders and extenders in meat systems to enhance functional properties and reduce cost, as well as recover nutrients from the waste of the fruit and vegetable canning industry (Mansour and others 1996). The large yellow flowers of the *Cucurbita* are also edible, and considered a
delicacy in Mexico. Some cultivars are grown simply for their quality of flowers (Harlan 1992).

*C. ficifolia* and *C. foetidissima* are used as forage and fodder crops for livestock, in addition to their capacity as a human food source. Both of these species tend to grow wild in the drier regions of the United States. Other varieties of *Cucurbita* designated as pumpkins or squash are also used as fodder, but the practice is much less common than it was earlier in the twentieth century. Livestock are generally fed the waste or extra fruit from processing and ornamental production. Pumpkin varieties, rather than squash, are more likely to be used as fodder, but this may be the result of changes in terminology. In California, for example, during the early part of the century, "all squash used as stock feed shall be reported as pumpkins" (Burbank 1921, p 101).

Cucurbitacin, found in the vines, leaves, roots or seeds of most cucurbits, is a bitter, poisonous substance (Bates and others 1990). Although many varieties of the family *Cucurbitaceae* are edible, some varieties of *C. pepo* were used as vertebrate poisons by Native Americans. The strong chemical properties that led to its use as a native poison also contributed to its popularity as a folk medicine among native peoples and early settlers. More recent research has shown that pumpkin seeds may, in fact, have medicinal properties that can be used to combat prostatitis (Home remedies 1995).
In keeping with the properties of cucurbitacins, pumpkins (botanical taxa unavailable) can be used as a mild insecticide. Acetone extracts from pumpkin seeds will kill mosquito larvae, and freshly cut pumpkin leaves, if rubbed on cattle or horse, deter flies (USDA 1952). *C. foetidissima* (buffalo gourd), conversely, has potential as an insect attractant. Its processing byproducts can be used to bait the *Diabrotica* beetle. The buffalo gourd is also being studied as a potential fuel source. A fuel alcohol project at the New Mexico Energy R&D Institute (New Mexico State University, Las Cruces) has utilized the roots for this purpose (Gathman and Bemis 1990).

A study by maverick scientist, Alan Hirsh, links *Cucurbita’s* chemical properties to the ancient association of the family *Cucurbitaceae* with fertility. American males rated the combined scent of pumpkin pie and lavender the most sexually stimulating aroma (Vitez 1997).

Economically valuable biological properties involve the physical appearance of a cultivar’s fruits, the hardiness of the vine and the importance of the plant’s genes to the development of crop species in the United States. Environmental utilization of *Cucurbita* includes the practice of using *C. ficifolia* as a graft stock for less hardy varieties of greenhouse cucurbits, such as cucumbers. Ornamental uses, such as seasonal sales of squash, pumpkins and gourds in the United States, most prominently of the variety *C. pepo* var. *pepo* for carving and decoration are also considered “environmental”
according to the GRIN. First hand accounts from nineteenth century New England state that farmers used the vines of pumpkins (probably *C. moschata* or *C. pepo*) as a material to bank their houses for increased protection from cold winter temperatures (Sloane 1962).

Also considered economically valuable are potential gene sources and progenitor plants. Those species containing important gene sources are being located and preserved. While the majority of *Cucurbita* varieties are known only in cultivation, those with important gene sources still grow wild (Bates and others 1990). Unfortunately, the progenitors of some economically important cultivars have still not been identified (Nee 1990).

*Cucurbita* as an Agricultural Genus

The family *Cucurbitaceae* played an integral role in the developing agriculture of the tropics and subtropics around the world. It is comprised of more than eight hundred species, some of which are utilized by indigenous cultures on every continent. On the American continent only *Capsicum* (peppers) and *Solanum* (potatoes) can compare with the five domestications from the genus *Cucurbita*. Cutler and Whitaker (1961), among others, rate the cucurbits as more important than any other cultivated plant group in the Americas, with the exception of *Zea* (maize).
The cultural significance of the genus *Cucurbita* in America dates back to its early domestication and diversity of uses. Native people grew pumpkins, squash and gourds "in every part of the country where they are now grown" (Hedrick 1950, p 14). As part of the "bean, corn and squash complex", they supplied the nutritional base for most of the pre-Columbian continent (Harlan 1992).

The genus includes both annual and perennial plants. The more mesophytic species are annuals or short term perennials, lacking the storage roots characteristic of the arid zone perennials (Whitaker and Bemis 1975). Early cultures, seeking a food source as well as a plant with versatility in use, cultivated those with sweet, fleshy fruits, nutritious seeds and hard durable rinds. In all, five species of *Cucurbita* were domesticated from among the mesophytic annual members of the genus (Harlan 1992). Archeological records support the fact that *Cucurbita* was an early domesticate as seeds or rind have been found with beans below the earliest evidence of corn on at least two sites (Cutler and Whitaker 1961) making the genus possibly the earliest domesticated plant on the American continents.

Geographically, the wild cucurbits and progenitors of contemporary cultivars cluster in Mexico, northern Central America and southwestern North America. The earliest cultivated species was probably *C. ficifolia* (chilicayote), and evidence from coastal Peru has been dated at 3000 B.C.
Earlier seed fragments of *C. pepo* from Ocampo Caves, Tamaulipas (Nee 1990) and Guila Naquitz Cave, Oaxaca (Harlan, 1992) were dated at 7000 BC and 8000 BC, respectively. The seed remnants in inhabited caves suggest that prior to the advent of agriculture, the plants were culturally important to hunter-gatherer societies (Whitaker 1981).

Three progenitors have been identified with relative certainty, but at least two others remain illusive. For the most part, wild *Cucurbita* and early domesticates cluster in the southwestern United States, Mexico and Central America and northern South America. The exception to this pattern is *C. andreana*, the possible progenitor of *C. maxima*, which grows in northern Argentina and Bolivia and was possibly transported from an earlier Mexican center by humans (Nee 1990).

Although *Cucurbita* is the only native genus, *Lagenaria sciceraria*, the bottle gourd, was also present at the Tamaulipas site (Cutler and Whitaker 1961). This cucurbit, recognized more for the durability and utility of its hard rind than for its potential as a food source, likely drifted on ocean currents from the Old World during pre-agricultural times and was quickly utilized by coastal populations (Harlan 1992). With the advent of agriculture, species from both the species *Cucurbita* and *Lagenaria* spread rapidly throughout the American tropics and subtropics (Sauer 1969).
From the beginnings of agriculture, *Cucurbita* domesticates spread along trade routes throughout the North America with maize and beans species. With the advent of Europeans in 1492, Native Americans readily accepted the seeds of cucurbit domesticates from Europe, including those of the cucumber, the watermelon, and the muskmelon (Hill 1952). Their familiarity with the utility of cucurbits led to the swift spread of these newly acquired species. Early American settlers, likewise, adopted the American cucurbits and planted them as an integral part of their kitchen gardens (Hedrick 1950).

The lack of correlation between common names and taxonomic terminology has long plagued the genus *Cucurbita*. The variety of economic uses and the extensive connection of these with common name led to early attempts at standardization and eventual cataloguing efforts that matched nomenclature with functionality. While use patterns have been listed with corresponding names, there has been little effort to quantify them within the United States. Neither has there been an effort to link changing utilization with changes in *Cucurbita* distributions, although much has been written concerning the archeological evidence for cucurbit domestication and spread. A transition in use or consumer preference could alter spatial patterns of demand and the environmental influences on *Cucurbita* distributions.
CHAPTER III

THE GENUS CUCURBITA AS AN UNITED STATES AGRICULTURAL CROP IN THE TWENTIETH CENTURY

Introduction

This chapter links the common names used in statistical data with temporal changes in the economic utilization of the genus. In the United States, farm size has been increasing throughout the century (National Farm Institute 1970). Mechanization has allowed farmers to cultivate more land in the same time frame. Increasingly, especially in the 1980s, small family farms have been unable to compete in the presence of ever growing corporate farming ventures (Hallam 1993). In the 1980s and 1990s, although farm size has been increasing, total farm acreage has decreased (U.S. Census Bureau 1995). Most notably, farmland has been lost to urban sprawl during this period. Increase in taxes near residential areas, accompanied by rising land values, has forced public and private organizations to think of preserving farmland (Klinge and Bartholic 1998). Farmers wishing to survive in this more competitive environment have shifted to a greater diversity of crops as
well as seasonal and specialty crops that yield more dollars per acre (USDA 1992).

The most recent U.S. Census of Horticultural Specialties was conducted in 1988, and released in 1991 (U.S. Census Bureau 1991). This census showed that greenhouse products, including vegetable crops, produced on farms with over $2000 in horticultural sales had not increased since the 1978 census. Although significant increases did take place in some southern states, such as Mississippi and Florida, production declined in some of the most predominant horticultural states, such as California. This downsizing was specifically attributed to some operations switching to other specialty crops such as ornamentals (Hickman 1992).

Although the genus *Cucurbita* is not specifically considered a Horticultural Specialty crop, unless the plants are started and sold from greenhouses, the census provides coverage of cucumbers, tomatoes, peppers and other vegetables that are indicative of the overall vegetable production of the country. A congressional hearing on the changing structure of the vegetable industry also took place in 1988 (U.S. House Committee on Agriculture 1988). The hearing stressed the significance of decreasing United States vegetable production and cited increasing fruit and vegetable imports as the primary reason for decline. Increasing imports have been attributed to five factors: (1) the rise of the dollars exchange rate against the currencies of
agricultural exporters, (2) subsidies and promotional programs created by foreign governments, (3) shortages in domestic production due to bad weather and crop disease, (4) trend toward globalization of both business and agriculture and (5) changes in demographics and lifestyle in the U.S. that have created a greater demand for fruit and vegetable products (GAO 1988).

Because of the concern for the vegetable industry in the United States and the economic value of *Cucurbita* as both a vegetable and as an ornamental, this genus lends itself to analysis of use changes that may have occurred in the United States during the 1980s and 1990s. The purpose of this chapter is to analyze, through the use of national statistics, the economic use patterns of pumpkins and squash that reflect the agricultural trends of this century. This analysis will provide the historical data necessary for the smaller scale research in the following chapters with emphasis on the time period of 1982 to 1992, which covers three agricultural censuses during a time of transition in the vegetable industry. It focuses on the consumer demand of *Cucurbita* products as recorded by common name in national statistics.

**Data**

Data were extracted solely from the United States government censuses and annually reported statistics. These statistics are federally
maintained, and have been collected consistently over most of the twentieth century. Although they are derived in different departments (Figure 1), many of the same definitions and biases affect the data sources, so they can be easily compared with one another. At the national, state and county levels, federally maintained data continue to be the leading source of information for agricultural research, which allows for easier comparison between studies.

As the most prominent source of data concerning agricultural production in the United States and the changing structure of agriculture (Schulman and others 1994), the census of agriculture was used extensively in this research. During the twentieth century, there will be fifteen agricultural censuses. At this time, data from the 1997 census is not yet available, so analysis of census data will take into account only the first fourteen. Census years include: 1909, 1919, 1929, 1939, 1949, 1954, 1959, 1964, 1969, 1974, 1978, 1982, 1987, and 1992. Prior to 1949, the census took place only once per decade, but since then censuses have been conducted approximately every five years. A comparison of census years with the availability of other data sources is shown in Figure 3.

Although the definition of a farm has changed throughout the history of agriculture, in the United States a farm is "any establishment from which $1000 or more of agricultural products were sold or would normally be sold during the year" (NASS 1998). The USDA standardized this definition in
Figure 3. Presence of Pumpkin and Squash in Agricultural Data Sources, 1920-1994.
1975. National totals for acreage and number of farms reporting pumpkins or squash as a vegetable crop are available for each of the fourteen census years. In the years 1939, 1949 and 1954, pumpkins are also reported as a fodder crop with both farm and acreage data.

The annual *Agricultural Statistics* published by the USDA National Agricultural Statistics Service (NASS) were also used to trace temporal changes in various vegetable statistics at the national level from the 1920’s when vegetables first appeared in the data, to 1994. Various governmental restructuring efforts have resulted in an assortment of names for the board and service agencies responsible for the production of annual reports. Table numbers, and occasionally historical data values, are therefore inconsistent between volumes. In the case of data disagreement, the values from later publications were assumed to be correct and earlier data, considered preliminary, was ignored for statistical purposes.

Statistics were derived from the following seven tables: (1) Vegetables, canned: United States pack; (2) Vegetables, canned: Civilian per capita consumption; (3) Vegetables, frozen: Commercial pack; (4) Vegetables, frozen: Civilian per capita consumption; (5) Vegetables and melons: Arrivals at four markets, by commodities; (6) Vegetables and melons, fresh: Total reported domestic rail, truck, and air shipments; and (7) Vegetables, commercial: Acreage, production, and value of principal crops. The seventh table, unlike
the former six, reports only total vegetable values and was used exclusively for comparison with annual squash and pumpkin values. Other annual data that portrayed overall vegetable production in the United States was not used in this research due to the numerous discrepancies between publications.

Quantities of commercially packed product are reported in cans or by weight, and consumption data is calculated from these numbers. The total apparent civilian consumption of pumpkin and squash is divided by the adjusted United States population. Estimates are adjusted for the underenumeration of children below the age of five and decreased for the members of the armed forces assumed to be consuming exclusively military supplies. In the 1960s, can size changed from the #2 can (weighing approximately 5 lbs.) to the #303 (2 lbs.) can (Veggies Unite! 1998). However, data relating to previous years did not change with the can size to correspond with the new weight equivalencies. This discrepancy may result from governmental restructuring that occurred in the early 1960s, when the USDA’s Statistical Reporting Service was created (NASS 1998).

The transport of vegetable products by railroad, boat, air and truck are recorded by commodity (as a vegetable and melon). Eight markets report annual vegetable arrivals based on shipping quantity. They include (1) Baltimore, (2) Boston, (3) New York, (4) Philadelphia, (5) Atlanta, (6) Chicago, (7) Los Angeles and (8) San Francisco. Monthly shipment totals of fresh
vegetables and melons are simply catalogued by commodity for domestic railroad, air and truck shipments. Squash and pumpkins are almost uniformly registered as arriving by motor truck, with the few exceptions arriving by rail. Interestingly, arrivals were consistently much higher than shipments but this may be due to the fact that arrivals are not specified as domestic. If this is the case, increasing imports could skew the arrival data. Both shipments and arrivals are recorded by the thousand hundredweight (.cwt).

As discussed in Chapter II, there is no nationally agreed upon definition for the terms “pumpkin” and “squash”. Consequently, no definition of these terms accompanies the census of agriculture, but both pumpkin and squash are reported as distinct vegetable crops. The annual statistics have different categories for the genus *Cucurbita* in different data tables. For both the shipments and arrivals data, pumpkins and squash are reported separately. The table of commercially canned vegetables contains only the entry, “Pumpkins and Squash,” while the frozen vegetable table reports both “Pumpkin and cooked squash” and “Squash, summer”. As the canned vegetable data does not report miscellaneous vegetables, it is unclear whether the value of canned product reflects summer squash, winter squash and pumpkins, or just the latter two. For purposes of this research, it is assumed that the processed, canned category of “Pumpkins and Squash”
includes all pumpkins and winter squashes frozen as a vegetable or as a pie pumpkin. This assumption excludes summer squash. Because there is a separate category for summer squash in the frozen packed data, it is assumed that quantities were insufficient for a separate category in the canned data and that values were simply not recorded.

Methods

The investigation of federal statistics includes both time series analysis and exploratory data analysis techniques. Temporal trends in processing and consumption data are compared with production data to identify possible changes in utilization, especially with respect to the possible growth of the ornamental industry in recent decades. Spatial patterns of cucurbit acreage in the 1990s are surveyed as a foundation for choosing a regional study area in the succeeding chapter.

The presence or absence of Cucurbita data in agricultural sources as compared to other vegetable crops is shown in time line format in Figure 3. The availability of data sources is compared, as is the duration of pumpkin and squash as a vegetable in national statistics. Because the data was generally stated in yearly totals, analysis of most records required the creation of standard bar charts and line graphs. Values from each of the seven tables were compiled into a spreadsheet software environment. Data
analysis techniques are based on years that contained data for each pair of variables analyzed. Where possible, the attempt was made to use units that are comparable between tables.

Although frozen packed pumpkin and squash are reported by weight in pounds, the canned product is reported by number of cases, equivalent to 24 cans each. To facilitate easier comparison, both U.S. processed and packed vegetable categories are presented in this research by weight, as opposed to volume. Processed weights for years in which can size is undetermined are plotted along two lines based upon both the reported and the adjusted can size.

The time period covered by the data set makes it susceptible not only to recording errors, but also to extremes in climate that may have affected agricultural productivity for only one year. Moving averages are computed for processing data in order to create generalized temporal trend lines and to smooth the extreme values in the data. The moving average converts the initial time series \( x_t \) to another time series \( y_t \) using a smoothing filter which included the surrounding ten values of each point, so that:

\[
y_t = \frac{1}{2q+1} \sum_{r=-q}^{q} x_{t+r}
\]  

(1)

where \( q \) was five (Chatfield 1984).
SPSS 7.5® spatial statistics software was used to generate correlation matrices to determine the strength of the relationship between values involving processed weight, acreage, and total United States vegetable production. Pearson's correlation coefficient $r$ was calculated for each pair $x, y$ using the formula:

$$r = \frac{n \sum_{i=1}^{n} Y_i X_i - (\sum_{i=1}^{n} Y_i)(\sum_{i=1}^{n} X_i)}{\sqrt{n \sum_{i=1}^{n} Y_i^2 - (\sum_{i=1}^{n} Y_i)^2} \sqrt{n \sum_{i=1}^{n} X_i^2 - (\sum_{i=1}^{n} X_i)^2}}.$$ \hspace{1cm} (2)

Because $r$ accurately represents only linear relationships, scatterplots were created in SPSS® and SPLUS® to display the linear tendency of the data.

Time series lag correlations were also computed to determine the oscillation of monthly shipping records and other annual data. Lag correlation assumes that in a continuous data sequence the correlation decreases with temporal distance so that the distribution of $X(t_1)$ and $X(t_2)$ depends only on $(t_2 - t_1)$, which is the lag. This assumption would result in an output graph with a smooth decrease in correlation through time. Any oscillations in the output, therefore, result from data periodicity.

$$\mu(t) = E(X_t)$$ \hspace{1cm} (3)
is the mean function, and the autocovariance function, \( \gamma(t_1, t_2) \), can be written as \( \gamma(\tau) \), so that the autocovariance coefficient at lag \( \tau \), is:

\[
\gamma(\tau) = E[(X(t) - \mu)(X(t + \tau) - \mu)].
\]  

(4)

The autocorrelation between \( X(t) \) and \( X(t + \tau) \) can be measured with the autocorrelation function:

\[
\rho(\tau) = \frac{\gamma(\tau)}{\gamma(0)},
\]  

(5)

which standardizes the autocovariance function (Chatfield 1984).

The uneven spacing of agricultural census years makes series analysis difficult. Instead, census data were used to construct bar graphs showing temporal progression in production. State totals for census categories were also used to spatially represent general trends in \textit{Cucurbita} agriculture in the United States. Pumpkin and squash acreages from the 1992 census were mapped, as were total cucurbit acres, by linking spatial boundary files (United States 1997) with census data using Federal Information Processing Standard (FIPS) codes.
Results

Changing economic impacts of the genus *Cucurbita* on the vegetable agriculture in the United States are evident from both historical agricultural census data and statistical data from the USDA. The census data show that pumpkins have lagged behind squash in terms of both number of farms and acreage for most of the century. It was not until the end of the 1980s that pumpkins began to approach quantities similar to those of squash in both of these categories. Squash farms (Figure 4) increased most dramatically in the 1940s and 1950s, but declined beginning with the 1964 census. The 1969 and 1974 census required drastically different criteria for vegetable farms, which resulted in the tremendous decline in squash farms for those years. For the remaining decade of the study period, however, the number of squash farms remained relatively constant. Squash acreage and the number of squash farms increased during the first half of the twentieth century. Since then farm numbers have declined, but acreage has been on the rise since the late 1970s.

Increasing at a much lower rate than squash, pumpkin farms (Figure 5) reached a high point in the 1992 census with 9,530 farms, surpassing the number of squash farms, 9,008, for the first time in an agricultural census. Pumpkin acreage also increased steadily throughout the century, although at
Figure 4. Squash as a Vegetable Crop in Historical Agricultural Census Data (U.S. Census Bureau 1909-1992).
Figure 5. Pumpkin as a Vegetable Crop in Historical Agricultural Census Data (U.S. Census Bureau 1909-1992).
a slower rate than that of squash. The census data from 1987 and 1992 show substantial increase over previous years.

In addition to being reported as a vegetable, pumpkins were also reported as a fodder crop in some censuses of agriculture. During the 1950s, (Figure 6) pumpkins as a vegetable crop were increasing steadily, while fodder pumpkins made their last appearance in the census of agriculture in 1954. In that year, only 133 farms were using pumpkins for feed, while 4,882 farms were growing pumpkins as a vegetable. The decrease in the number of farms growing fodder pumpkins quite possibly suggests that they were increasingly being grown primarily as a vegetable crop with leftovers being fed to livestock. In the earlier censuses of agriculture, both pumpkins and squash are reported in very small quantities (although they were probably...
grown throughout the country). This lack of representation in the data is largely due to the fact that most pumpkins and squash were raised as home garden plants (Hedrick 1950), so they were used fresh or stored for home use. Because of the definitions of farms and crops, in the earlier census any crop that was raised in quantities of less than one acre was excluded, even if it was sold for a profit. Vegetables grown for home use were also not reported in the census. In the 1909 census, of those reporting vegetable sales, most were under $500, although most farms of the period had extensive vegetable gardens (USDA 1936). Therefore, low numbers at the beginning of the century are likely to be misleading and incomparable to later census years unless changing status of vegetables in the economic structure of the agricultural community are taken into account.

In the late 1920s and early 1930s, the USDA began reporting quantities of processed, cooked pumpkin and squash that was being frozen and canned on an annual basis. Figures 7 and 8 show temporal changes in the quantities of U.S. packed Cucurbita vegetables. The effect of variations in the data due to the growing conditions of specific years as well as discrepancies in reporting has been reduced with a moving average (Figure 9). Both the bar charts and the moving average lines show an increase in the amount of frozen pumpkin and squash throughout the extent of the data. Canned weights were much more consistent and were no longer reported following 1982.
Figure 7. Frozen Pumpkin and Squash, U.S. Packed and Processed (USDA 1925-1995).
Figure 8. Canned Pumpkin and Squash, U.S. Packed and Processed (USDA 1925-1995).
Figure 9. Moving Averages of *Cucurbita* Processing Data.
Data concerning the civilian per capita consumption of both frozen and canned pumpkin and squash corresponded fairly well with the processed weights for the years in which data are available. Frozen consumption began to increase in the 1960s, as frozen weight increased, while canned consumption remained fairly consistent.

While records of production and processing extend to the early part of the 1900s, transport of the genus *Cucurbita* was not recorded until the latter portion of the century. Shipments of vegetable products have been reported since 1978, but pumpkins are only listed in the data from 1978 until 1980, probably because of their low numbers and occurrence in only one or two months per year. It is unclear in the data definition whether pumpkins shipped after 1982 are recorded as squash or are simply not recorded at all. Squash shipment data for the period 1978 to 1993 (Figure 10) show a distinctive pattern with two peak shipping periods. One peak in early spring, around April, and another in the month of November, probably corresponding to Thanksgiving holiday sales. These squash peaks relate to the harvest dates of growing locations that annually produce both one crop, harvested in the fall, and two crops, generally of summer squash. Pumpkin shipments by comparison, although available for only three years, have only one much smaller peak in October (100-200 cwt.), relatively few shipments in September (1-3 cwt.), and no shipments recorded for the rest of the year.
Figure 10. Domestic Squash Shipments by Month (USDA 1925-1995).
Pumpkins are generally grown in the more northern states, where a shorter growing season allows only one harvest.

A time series lag correlation was run on monthly squash shipments for the entire fourteen year period, with 168 data points, in SPSS®. The output graph (Figure 11) clearly shows the biannual oscillations, as well as the expected decline in correlation with temporal distance. The lag correlation shows no change from when pumpkins were shipped as pumpkins, nor does it seem to be following any pattern other than the yearly oscillation.
In 1982, pumpkins and squash both became vegetables for which transport data was reported at eight specific cities throughout the U.S. These records contain total truck, air, train, and boat shipments to representative port cities. Both squash arrivals and squash shipments (Figure 12) show a slight decline in shipping at the end of the 1980s. Although beyond the scope of this chart, both increased again in 1994 and 1995. Pumpkins (Figure 13) in both data sets exhibit a more steady increase over the time period, but at a smaller magnitude, as can be seen by the scale of the charts.

Although the national statistics can be displayed to visualize the temporal trends of the agricultural use of the genus *Cucurbita*, there is no spatial component to this data. Some annual tables record regions of
production and processing, but data values are inconsistent and the spatial delineation of the regions changes from publication to publication making the information uninterpretable. Because the census data can be tied to regional boundaries at various scales, maps created in ArcView® for state level data reveal general spatial patterns for the genus Cucurbita (Figure 14). States with higher Cucurbita acreage cluster along the West Coast as well as in the Midwest, Northeast, and Southeast. When the genus is mapped after being divided into the categories of pumpkin and squash (Figures 15 and 16) the results yield similar patterns. Pumpkin acreage, however, is densest in the more northern states and lower in the southeast, while squash densities are highest around the coasts and in Michigan.
Figure 14. *Cucurbita* Acreage per Square Mile in the Contiguous 48 States, 1992 (U.S. Census Bureau 1995).
Figure 15. Pumpkin Acreage per Square Mile in the Contiguous 48 States, 1992 (U.S. Census Bureau 1995).
Figure 16. Squash Acreage per Square Mile in the Contiguous 48 States, 1992 (U.S. Census Bureau 1995).
Discussion

Although yearly variations are large, pumpkin and squash processing, as well as farms and acreage continued to increase at a fairly smooth rate until the 1970s. The only other apparent trend during this period is a definite increase in pumpkin and squash production between 1939 and 1954 (Figure 6). Although civilian consumption remained relatively steady for these years, packing increased dramatically during WWII indicating the increase in military consumption of canned and frozen squash and pumpkin (USDA 1950-51). A movement toward increased efficiency in the use of vegetable matter advocated the economic utilization of oil pressed from seeds removed at packing facilities as a food source and as a factory lubricant. In fact, American processors were berated for failing to use the 500 tons of oil that could be recovered annually from the waste of the pumpkin and squash processing industry at a time when European countries were actively involved in recovering vegetable byproducts (USDA 1950-51).

Following the slight increase to the 1970s, processing and civilian per capita consumption began to level off, as did the increase in squash farms and acreage. Pumpkin farms and acreage, however, continued to increase. The period 1982 to 1992 revealed an especially dramatic increase of 143% in the pumpkin acreage of the United States.
At the start this period of increase, the presence of pumpkin and squash in the agricultural record began to change (Figure 3). Annual data for consumption and canned *Cucurbita* products were no longer recorded, while shipping and arrival at principal ports for both pumpkin and squash were added to the annual statistics. It is interesting to note some dissimilarities in the data availability for pumpkin and squash as opposed to other vegetable crops such as tomatoes, cucumbers, and broccoli. In regards to arrivals, pumpkin and squash arrivals have only been reported since 1982, although arrivals for other vegetables have been recorded since the 1930s (USDA 1920-1995).

Both the data recording classifications and the agreement between production and processing data undergo a change in the 1980s and 1990s. It is probable that the trend of vegetable growers switching to more lucrative ornamental crops (Hickman 1992) is responsible for the vast increase in pumpkin farms and acreage. The low number of pumpkin arrivals, even while production is comparable to that of squash, makes it possible that arrivals are following the trend of shipping records (USDA 1920-1995). Only those pumpkins transported in late September and October, primarily for ornamental use, would then be recorded. Squash shipments and arrivals, therefore, would include summer and winter squashes as well as pumpkins grown for edible purposes. This would also make sense with regard to the
appearance of *Cucurbita* products, which is often the criterion for giving them the name "pumpkin" or "squash" (Sauer 1993). Many pie pumpkin cultivars, such as 'Buckskin' and 'Green Striped Cushaw', grown strictly for edible purposes are not round or orange, and might easily be considered squash (Stokes 1998).

Because of the differences in the data, I would hypothesize that edible pumpkin products and those fruits being marketed as "pumpkin" for human consumption, regardless of common name or taxonomic variety, are being reported as squash for shipping purposes. Conversely, pumpkins in the shipping records indicate varieties grown for primarily ornamental purposes for Halloween and associated fall decorating. This would also correspond with national trends relating to the stability or decrease of vegetable production, and simultaneous increase in ornamental production (Hickman 1992). A test for such hypotheses could examine consumer demand for use-specific *Cucurbita* varieties by linking seed sale records with spatial boundary files.
CHAPTER IV

SPATIAL PATTERNS OF THE GENUS CUCURBITA IN THE EASTERN BROADLEAF PROVINCE, 1982-1992

Introduction

In the 1980s American farms underwent a crisis when many farmers, especially those being pressured by urbanization, began looking at seasonal and specialty crops with which to supplement their regular income (USDA 1992). From 1982 to 1992, as shown by national data in Chapter III, there was a tremendous increase in pumpkin acreage in the United States which may have been the result of increasing ornamental sales. In fact, although there was a 4% decrease in overall farm acreage in the U.S during this period, vegetable acreage increased just over 13%. Squash acreage increased almost 36%, and pumpkin acreage, in only 10 years, increased 143% (U.S. Census Bureau 1995).

Finch and Baker (1917) document the spatial patterns of total vegetable acreage in the US with a dot map derived from the 1909 agricultural census, showing a concentration of acreage around urban areas. They also map specific vegetable crops, including muskmelons and cantaloupes, and
watermelons, among the cucurbits. Pumpkins and squash, however, are not mapped at this time. Neither are these vegetables recorded in other agricultural mapping efforts throughout the century.

After preliminary analysis in Chapter III of the distribution of _Cucurbita_ acreage from the 1992 U.S. Agricultural Census (Figures 14, 15, and 16) areas of increased production were evident along the West coast and the New England coast as well as in portions of the Southwest, Southeast, and Midwest. Popular publications often make statements regarding the distribution of economically important _Cucurbita_ varieties (Royte 1997). However, there have been no spatial analyses of pumpkins and squash at a level that would support any observations of their current pattern or change over the past decade.

This chapter examines the distribution of agricultural _Cucurbita_ species at the county level within one ecological region of the United States. The ecoregion is defined by Bailey (1992) and based upon the homogeneity of the physical environmental. Descriptions of spatial patterns in this region will be explained with respect to the national trends outlined in the previous chapter. The subsequent chapter will explore the connection of these patterns to environmental and economic factors.
Data

The National Hierarchy of Ecological Units, first created by Robert Bailey for the USDA Forest Service (1975), divides the country into regions at various levels of resolution based on homogeneity of several physical factors including climate, physiography, water, soils, hydrology, and potential natural communities (Bailey 1995). Bailey's ecoregion map was chosen instead of similar schemes, such as Omernick's ecoregion, Köppen climate classification modified by Thornthwaite, or U.S. physiographic provinces for two reasons. The ecoregion boundaries combine essential growth factors and are compatible with general patterns of cucurbit agriculture in 1992 and the factors used in this study. Since the boundaries were developed by the Forest Service under the auspices of the USDA they have been widely used in research and accepted as a base unit by other federal agencies (EPA, etc.).

In the hierarchy, domains are the largest unit and are classified by sweeping climatic characteristics. Divisions delineate, within the larger domain, vegetational affinities as classified by water deficit estimates and winter temperature patterns. Divisions are further subdivided into provinces (boundaries shown in Figure 17) which are the smallest unit at the ecoregion scale. These vegetational regions are characterized by homogenous continental weather patterns and soil type. At subregion resolution, sections
Figure 17. National Hierarchy of Ecological Units, Eastern Broadleaf Province (U.S. Forest Service 1994).
are characterized by underlying geology and a similarity in geomorphic processes.

The Eastern Broadleaf Province is a part of the Humid Temperate Domain and the Hot Continental Division in the Hierarchy of Ecological Units. The thirteen sections of the Eastern Broadleaf province are shown in Figure 18 and include: (1) the Central Till Plains, Beech-Maple Section; (2) the Central Till Plains, Oak-Hickory Section; (3) the Erie and Ontario Lake Plain Section; (4) the Interior Low Plateau, Bluegrass Section; (5) the Interior Low Plateau, Highland Rim Section; (6) the Interior Low Plateau, Shawnee Hills Section; (7) the Lake Agassiz, Aspen Parklands Section; (8) the Minnesota and Northeast Iowa Morainal, Oak Savannah Section; (9) the North Central U.S. Driftless and Escarpment Section; (10) the Ozark Highlands Section; (11) the South Central Great Lakes Section; (12) the Southwestern Great Lakes Morainal Section; and (13) the Upper Gulf Coastal Plain Section.

This research examines spatial patterns at the county level within one ecoregion province. The province level of the hierarchy represents interaction between various components of the physical environment and delineates a contiguous area, which minimizes boundary errors. A personal interest in the Midwest as well as statements made by growers of giant pumpkins about the possible optimization of potential growth along the 43rd parallel of latitude contributed to the choice of the Eastern Broadleaf Province as the
Figure 18. Sections of the Eastern Broadleaf Province (U.S. Forest Service 1994).
region of study. Because wild species extend only to the fringes of this
ecoregion (Nee 1990), interaction between population-oriented, economic
aspects of the spatial distribution with those factors relating to the physical
geography of the area should be more clearly shown than in other provinces.
The relationship of county *Cucurbita* production with these factors will be
more closely examined in the following chapter.

The ecoregion boundary coverage, which was produced and
distributed by the USDA Forest Service, was downloaded as an ARC/INFO®
interchange file (1994). The county boundary file was selected from an
These boundaries were linked with ten variables from the 1992 *Agricultural
Census* CDROM (U.S. Census Bureau 1995) at the national, state and county
level for the three most recent agricultural census years, including 1982, 1987
and 1992. Census of agriculture records, with table number, utilized in this
study are as follows: (1) Farms (number), 010001; (2) Land in farms (acres),
010002; (3) Land used for vegetables (farms), 290001; (4) Land used for
vegetables (acres), 290002; (5) Vegetables harvested (farms), 290005; (6)
Vegetables harvested (acres), 290006; (7) Pumpkins, Harvested (farms),
290157; (8) Pumpkins, Harvested (acres), 290158; (9) Squash, Harvested
(farms), 290173; and (10) Squash, Harvested (acres), 290174.
Methods

Following the selection of data sources, the spatial boundaries and database files were integrated for analysis. A flow chart of this procedure is shown in Figure 19. In order to determine the statistical reporting units enclosed by the Eastern Broadleaf Province, an overlay of province and county boundaries was performed in PC ARC/INFO®. The coverage of Ecological Units was re-projected from the Lambert Conformal Conic projection to the Geographic Reference System (decimal degrees as units), which was consistent with that of the county boundaries. Using ArcView®, the two layers were overlaid and the county coverage was spatially queried to determine the counties entirely or partially contained within the Eastern Broadleaf Province. The province consists of portions of 649 counties, representing 16 states: (1) Alabama, (2) Arkansas, (3) Illinois, (4) Indiana, (5) Iowa, (6) Kansas, (7) Kentucky, (8) Michigan, (9) Minnesota, (10) Missouri, (11) New York, (12) Ohio, (13) Oklahoma, (14) Pennsylvania, (15) Tennessee, and (16) Wisconsin. Agricultural census data files for each of these states were imported to a Microsoft Access® database and queried for the ten specified records before the resulting tables were appended to one another.

The table of counties was imported to Access® from ArcView®. Using Federal Information Processing Standard (FIPS) codes, the database was
Figure 19. Integration of Agricultural Census Records and Spatial Boundaries.
queried for all records with data pertaining to the province's 649 counties. This resulting table was queried for each of the ten census record numbers, so that each could be connected with polygon boundaries and mapped independently of other variables. Each of the final ten tables contained values for the entire province for the years 1982, 1987, and 1992. Agricultural data were then joined to a spatial boundary file at the county level by FIPS codes, using ArcView's® open database connectivity capabilities.

The boundary file of United States counties was intersected with province and section boundaries in ARC/INFO® and the derived polygons not contained within the province were deleted in ArcView®. With the overlay of section boundaries, the number of polygons within the province increased to 825, from the 649 original counties (Figure 20). Although total vegetable production, among other variables, is available at the zip code scale, county level statistics are the smallest spatial unit at which the production of pumpkin and squash are recorded. Agricultural data were joined to county fragments with the assumption that farms and acreage were uniformly distributed within the county (Chrisman 1997) as there were no data available for patterns at a smaller scale. The percentage of county area in each county fragment was calculated and this same percentage of the Cucurbita variable value was assigned to the polygon. One peculiarity in the
Figure 20. Overlay of Province, Section, and County Boundaries.
data related to St. Louis City. The city has its own FIPS code and boundaries within the county data set, but agricultural census data is not recorded for the city of St. Louis. Therefore, all agricultural data values for St. Louis City (FIPS 29510) are encoded as zero.

In order to understand the distributions of agriculturally defined Cucurbita at the county level within the Eastern Broadleaf Province the statistical data were analyzed with respect to its spatial association. Maps were created for each variable queried from the Agricultural Census CDROM and mathematical operations to determine derive new variables from these files were performed in Microsoft Access® and ArcView®. The spatial autocorrelation in the data was computed using SPLUS® spatial statistics software.

National, province, and sectional means were compared for each agricultural record. Computations were accomplished with Access®, ArcView®, and Excel®. National totals were available from the census data. Province and section totals and means, however, were computed by summarizing county polygon values within each spatial area. Rates of change at these larger scales were computed for each census of agriculture category, $x$, where:
While Excel was used to graph statistical results, choropleth maps were created in ArcView® showing the density of pumpkin, squash, and *Cucurbita* production per county in 1992. In order to demonstrate areas of greatest change which might correspond with an increase in ornamental sales, rate of change was computed from 1982 to 1992 for these categories for each of these categories at the county level, \( x \), using absolute change per county area:

\[
\frac{x_{1992} - x_{1982}}{Area}
\]  

(7)

As increasing farm size has been one of the more significant agricultural trends on the national level, farm size was also computed at the county level, as was change in farm size with the formula:

\[
\Delta Farm Size = \frac{A_{1992} - A_{1982}}{F_{1992} - F_{1982}} 
\]

(8)

where \( A \) is acres and \( F \) is farms and \( Area \) is the county area in square miles.
The high number of zeroes in the data prevented percentage rate of change calculations at the county level. The distribution of zeroes in the data is shown in Figure 21 and shows both a lack of production and lack of reporting. Unlike other variables in the agricultural census, zero values did not result from the suppression of information for privacy of individuals.

Because the number of zeroes impacted the outcome of analyses, preliminary measures involved rescaling techniques to alleviate the problem. In addition to being treated as an actual zero value, zeroes were given a value of 999,999 to determine if an abnormally high value would affect outcomes differently. The value of 999,999 was chosen, simply because it would be easily distinguishable after further analysis. Since the majority of zeroes indicated that no data were available, instead of an actual zero value, numeric values could be manipulated at this point, as long as they remained uniquely identifiable. Another rescaling technique, recommended by Cressie (1993) for county level lattice data structures:

\[
\frac{1000 \times (x + 1)}{Area} \quad (9)
\]

simultaneously creates densities per county area while rescaling the county agricultural census values, \( x \).
Figure 21. The Presence of Zeroes in the Census of Agriculture (U.S. Census Bureau 1982-1992).
In addition to the problem of zero values, a standard difficulty in the creation and interpretation of geographic models is the non-random spatial clustering of regression residuals. To determine the extent that this may be a problem in a data set, estimates of spatial autocorrelation are computed to quantify the clustering of data prior to regression modeling. In this particular study, the complex shape of the ecoregion was also recognized as a potential source of bias during analysis. Both the irregular boundaries and the interference of Great Lake shorelines impacted the number of polygon neighbors generated in ARC/INFO®. The relative geometric simplicity of county boundaries integrated with the more fluid outline of the ecoregion resulted in small partial-county polygons whose centroids were placed in the center of the polygon, as opposed to the center of the county.

In order to estimate the spatial relationship between the counties, spatial weights, assigned by two different methods, were compared using the procedures summarized in Figure 22. In boundary weighting, each neighbor of a polygon is weighted by the length of its shared boundaries. Polygon neighbors were calculated in ARC/INFO®, but weights were assigned using the SPLUS® spatial module. Since the S+GISLINK software for transfer of data files between ARC/INFO® and the SPLUS® spatial module was discontinued, the polygon neighbor list (.pnl) and polygon attribute table
Figure 22. Neighbor Weighting Options for Estimating Spatial Autocorrelation.
(.pat) from ARC/INFO® were imported to SPLUS® as ASCII files. At this point, the removal of extraneous ARC/INFO® polygons such as polygons with negative area (including the universal polygon) and those with duplicate identifiers was essential. With the pair of edited tables, SPLUS® creates a spatial neighbor object in which the length of shared boundaries from the .pnl is used to assign neighbor weights for the analysis of data from the .pat file.

The second weighting scheme was based on a distance-decay correlation function applied to the distance between county centroids (Cressie 1993; Mathsoft 1996). The \( x, y \) coordinates of county centroids, located at the geometric center of each county, were added to the polygon attribute table in UNIX ARC/INFO®. Polygons were considered neighbors if their centroids, recorded as latitude and longitude coordinates, were within one degree (approximate distance within the ecoregion) of each other. Ideally, the radius defining neighboring centroids would be larger, but the one degree limit was based on the computational capacity of the computers available for this research. After deleting self-neighbors, the procedure resulted in 12,624 polygon neighbors. The function for estimating correlation between neighbor pairs with irregular lattice data structure, such as county boundary files, recommended by Cressie (1993) is as follows:
where $\rho$ is the constant to be estimated, $d_{ij}$ is the distance between the county centroids, $n_i$ is the area of the county $i$, $C(k) =$ \[ \max \{ d_{ij}^{-k} : j \in N_i; i = 1, \ldots, n \} \] (a scaling factor), and $N_i$ is the set of neighbors of region $i$. The neighbor weights for each pair are estimated as $c_{ij}\rho$. Because Cressie (1993) recommends that the best model results are obtained when $k = 1$, neighbor weights for the spatial neighbor object in S-PLUS are estimated by:

\[
c_{ij} = \begin{cases} \rho \left( d_{ij}^{-k} / C(k) \right) \left( n_j / n_i \right)^{1/2} & j \in N_i \\ 0 & \text{otherwise.} \end{cases}
\]

 Polygon attribute tables were imported, and spatial neighbor tables were created three separate times in the S-PLUS spatial module. The first spatial neighbor object used boundary weighting, while zeroes retained their zero value. The second also incorporated zero values, but used the distance decay function as a weighting scheme. Distance weighting was also used in the importation of a .pat file, in which all zeroes had been converted to the value of 999,999.
Polygon attribute tables were imported, and spatial neighbor tables were created three separate times in the S-PLUS spatial module. The first spatial neighbor object used boundary weighting, while zeroes retained their zero value. The second also incorporated zero values, but used the distance decay function as a weighting scheme. Distance weighting was also used in the importation of a .pat file, in which all zeroes had been converted to the value of 999,999.

Once the neighbor objects were created in S-PLUS, Moran's contiguity ratio could be computed to analyze spatial autocorrelation. Both Moran's contiguity ratio ($I$) and Geary's $c$ tests for spatial autocorrelation are appropriate for data on the ordinal and interval scale. Although Moran's $I$ shows sensitivity to outliers in the estimation of the mean, the $c$ test squares deviations from the mean and may be more easily skewed (Haining 1990). Since the county agricultural data contained numerous outliers, the $I$ test was chosen as more resistant. In the interpretation or Moran's ratio, $I$ values farther from zero indicate greater spatial dependence in the data. The formula for $I$ is:

$$I = \frac{\sum_{i=1}^{n} \delta_{ij} \hat{z}_i \hat{z}_j}{\sum_{i=1}^{n} \sum_{j=1}^{n} \delta_{ij} \hat{z}_i^2}$$

(12)

where $n$ is the number of observations. $i$ and $j$ are two counties, so that:
\[ \delta_{ij} = \begin{cases} 1 & \text{if } i \text{ and } j \text{ are contiguous (} i \neq j) \\ 0 & \text{otherwise} \end{cases} \] (13)

The replacement of \( \delta_{ij} \) with \( w_{ij} \), such that:

\[ w_{ij} \geq 0 \text{ if } i \text{ and } j \text{ are contiguous (} i \neq j) \]
\[ = 0 \text{ otherwise} \] (14)

permits differential weighting between contiguous neighbors as specified in the .pnl file. The mean of \( n \) observations \( \{y_i\} \) is \( \bar{y} \), so that \( z_i = y_i - \bar{y} \). \( A \) is the total number of joins in the county system and is equivalent to \( \frac{1}{2} \Sigma_{(2)} \delta_{ij} \)

where:

\[ \Sigma_{(2)} = \Sigma_i \Sigma_{j(i \neq j)}. \] (15)

When the cross-product \( z_i z_j \) is computed, clustered spatial patterns will tend to yield positive values, while alternating patterns tend to be negative.

The spatial autocorrelation procedure was not only necessary to estimate spatial dependency, but also was used to evaluate the impact of zero values on that estimation. As previously mentioned, three different spatial neighbor objects incorporated changes in weighting schemes and zero values. In addition, Cressie’s (1993) rescaling technique was also used in the estimation of Moran’s \( I \). The rescaling created a county density ratio for the
different agricultural production categories, but did not require the creation of a separate spatial neighbor object.

Results

In the United States, the period from 1982 to 1992 was a time of change for agriculture. Rates of change for total farms, vegetable farms and \textit{Cucurbita} farms at both the national and province level are available in Table 2. The

Table 2

National and Province Farm Production Rate of Change, 1982-1992

<table>
<thead>
<tr>
<th></th>
<th>United States</th>
<th>Eastern Broadleaf Province</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Number of Farms</td>
<td>Total Acreage</td>
</tr>
<tr>
<td>Total Change</td>
<td>-0.1409</td>
<td>-0.0418</td>
</tr>
<tr>
<td>Land in Vegetables</td>
<td>-0.1033</td>
<td>0.1305</td>
</tr>
<tr>
<td>Vegetables, Harvested</td>
<td>-0.1033</td>
<td>0.1356</td>
</tr>
<tr>
<td>Pumpkins, Harvested</td>
<td>0.8359</td>
<td>1.4345</td>
</tr>
<tr>
<td>Squash, Harvested</td>
<td>0.1050</td>
<td>0.3587</td>
</tr>
</tbody>
</table>
total number of farms in the United States decreased 10% while acreage decreased 4%. Consequently farm size increased, and megafarms started appearing, changing the structure and focus of American farming practices (Hallam 1990). During the same period in the Eastern Broadleaf Province, both acreage and farms decreased more than the national average (Figure 23), with the exception of only one and two sections respectively.

In keeping with this decreasing trend, farms growing and harvesting vegetables decreased in both the nation (10%) and the province (5%), whereas acreage in the province increased at a higher rate (15%) than the nation as whole (13%). At the national level, the number of farms harvesting pumpkins and the acreage of harvested pumpkins and squash all increased at a rate higher than for vegetables. Farms and acreage of both these vegetable products increased even more dramatically in the Eastern Broadleaf Province. Pumpkin acreage in the Eastern Broadleaf Province increased over 250%. Pumpkin farms, squash farms, and squash acreage also increased more dramatically at the province scale between 1982 and 1992. Increasing mean farm size in the United States, has been mirrored in the increase of pumpkin farm size both on a national scale and within the thirteen Eastern Broadleaf Province sections.

Figures 24 and 25 show the county acreage densities for each of the three censuses of agriculture records of pumpkins and squash. In 1982, both
Figure 23. Rates of Change in Farms and Acreage Within the Eastern Broadleaf Province.
Figure 24. County Pumpkin Acreage Densities (U.S. Census Bureau 1982-1992).
Figure 25. County Squash Densities (U.S. Census Bureau 1982-1992).
vegetables were more prevalent in counties that lie in the north central portion of the province, around the shores of the Great Lakes. Through 1987 and 1992, the distributions broaden to include an increasing number of counties in the southern sections, with increases also to the west and east. Increases in pumpkin farms occur also predominantly along the lake plains, but there are also clusters of increase apparent in southern Indiana and the Mississippi River valley.

The high number of zero values in the data required the use of absolute measures of increase per area, as opposed to rates. Figure 21 shows spatial clustering of zeroes in the data. In the lightest counties, zero pumpkin acres were reported for each of the three agricultural census years. The darkest colors indicate that pumpkin acreage was reported for during the entire decade covered by the three years of census data. These areas tend to correspond with areas of the highest pumpkin acreage, while areas that reported pumpkins in only one or two of the census years are evident at the edges of the distributions.

Counties with the most increase in acreage per area occur around the lake plains and into Minnesota, where values were initially high (Figure 26). Following the national trend of increasing farm size, all section means showed an increase in the size of pumpkin farms over the 10 year period. A distinct band from the northwest to southeast of sections where farm size
Figure 26. Distribution of County Change in Acreage (U.S. Census Bureau 1995).
growth was less is clearly visible, especially when sections are divided at the sectional mean increase of 136%.

Moran's I contiguity coefficient estimates, calculated for the purpose of quantifying the spatial clustering, are summarized in Table 3. The initial spatial neighbor object was created with the boundary weighting scheme and total county acreage values for pumpkins and squash (Cucurbita) harvested. Spatial autocorrelation was computed for each of the agricultural census years, 1982, 1987 and 1992, with zero values retained. Although all three I's were very low, 1992 Cucurbita acreage showed the highest value. For this reason, 1992 acreage for pumpkins, squash and Cucurbita (pumpkins and squash) was the category used for subsequent contiguity calculations.

Weighting polygon neighbors by the distance decay function consistently resulted in higher spatial autocorrelation values than weighting by shared boundary length. This result was expected because of the higher number of neighbors per polygon considered by this method. Due to the fact that 999,999 is a more significant outlier in the data set than the value zero, spatial autocorrelation estimates involving these artificially inflated zero values were also inflated. Regardless of zero value, count data (acres) showed more spatial autocorrelation than that of densities (acres per county area). When zeroes were rescaled to densities, using Cressie's equation, spatial autocorrelation was estimated with almost the exact figures of density
Table 3
Spatial Autocorrelation Estimation Using Moran’s Contiguity Coefficient, $I$

<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Total Value</td>
<td>Boundary Weights</td>
<td>Zeroes as 0</td>
<td>N/A</td>
<td>N/A</td>
<td>0.1142</td>
<td>0.0892</td>
<td>0.0776</td>
</tr>
<tr>
<td>Total Value</td>
<td>Distance Decay</td>
<td>Zeroes as 0</td>
<td>0.1206</td>
<td>0.0925</td>
<td>0.1368</td>
<td>N/A</td>
<td>N/A</td>
</tr>
<tr>
<td>Density Per Area</td>
<td>Distance Decay</td>
<td>Zeroes as 0</td>
<td>0.1414</td>
<td>0.0813</td>
<td>0.1361</td>
<td>N/A</td>
<td>N/A</td>
</tr>
<tr>
<td>Total Value</td>
<td>Distance Decay</td>
<td>Zeroes as 999,999</td>
<td>0.2214</td>
<td>0.2042</td>
<td>0.2810</td>
<td>N/A</td>
<td>N/A</td>
</tr>
<tr>
<td>Density Per Area</td>
<td>Distance Decay</td>
<td>Zeroes as 999,999</td>
<td>0.1533</td>
<td>0.1754</td>
<td>0.1707</td>
<td>N/A</td>
<td>N/A</td>
</tr>
</tbody>
</table>
Table 3—Continued

<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>$1000^* (x + 1)$</td>
<td>Distance Decay</td>
<td>Zeroes Rescaled</td>
<td>0.1411</td>
<td>0.0794</td>
<td>0.1353</td>
<td>N/A</td>
<td>N/A</td>
</tr>
</tbody>
</table>
per area with zeroes unchanged. A decrease of less than five thousandths was evident in the $I$ value when calculations included this scheme.

Discussion

As part of the spatial analysis of county values within the Eastern Broadleaf Province, this research quantifies the impact of spatial proximity on *Cucurbita* production. The creation of a spatial neighbor object and the estimation of Moran's contiguity ratio require the analysis of weighting schemes and neighborhood relationships, which reveal basic information about the characteristics of the lattice within which the data is represented. The difference in the number of polygon neighbors associated with the data under various neighborhood schemes and the optimization of centroid distance for decay modeling become important steps toward goals expressed in subsequent chapters, as well as in the computation of spatial autocorrelation estimates. Neighbor identification and weighting procedures are initial requirements of spatial regression modeling, and are integral in the larger goal of establishing interaction between spatial statistics software and GIS.

The proximity of all the calculated $I$ values to zero indicates that there is very little spatial autocorrelation in this data set. Spatial proximity, therefore, plays little role in understanding the pattern of spatial distribution
of *Cucurbita* by county. The highest contiguity estimation was 0.2810, from the 1992 total *Cucurbita* (pumpkins and squash) acreage per county when zeroes were reassigned a value of 999,999. Since this estimate was inflated by artificially high zero values, it should be regarded with caution. The predominance of positive values is indicative of a tendency toward a clustering as opposed to an alternating pattern, but the quantification of this trend is not highly significant. County acreage, therefore, is more likely to be determined by a local combination of physical factors and consumer characteristics rather than general, broad-scale patterns.

Similar to national trends, an increase in pumpkin acreage took place during the ten-year period in all thirteen sections of the province, but increases were at a greater rate than at the national level. Pumpkins tend to be grown farther north than squash in the United States as a whole, as well as within the ecoregion. The highest mean increases occurred in the northeastern sections. While estimated Moran's *I* values remain low at a smaller scale, continued increases throughout the province reiterate the importance of local characteristics in *Cucurbita* distributions. The pattern could be based on small-scale environmental factors or, more likely, the economic impacts of population centers in this portion of the country. Regions of high *Cucurbita* production occur near the high population areas of Minneapolis-St.Paul, Chicago and Detroit where urban pressure would be
greater. These factors will be analyzed in relation to *Cucurbita* distributions in the next chapter.

Maps of *Cucurbita* census data and derived variables tend to show corresponding areas for both production and changes in production. These results are highly dependent on the fact that counties in which there have been *Cucurbita* increases are those counties in which values have been consistently recorded throughout the time period. However, since the vast majority of provinces within the county do not report values for pumpkins and squash, the increase in production as compared to the national increase, is even more remarkable. Whether this recording tendency demonstrates actual areas of concentrated production is unclear. Cultural factors, such as the awareness of the crops as viable economic plants, may impact the frequency of their presence in agricultural documents. In either case, areas along the lake plains show patterns of *Cucurbita* production that are far from the norm in this ecological province.

*Cucurbita* farms and acreage tend to be denser around the margins of the Great Lakes, as well as in southeastern Minnesota. Although choropleth maps of the region show these areas of high production, they are not evident from the spatial autocorrelation estimation. The shape of the lake shore boundaries and the lack of neighbor polygons for those counties lying along the lake plains creates a pattern that is not easily documented by spatial
association formulas. Increasing the polygon neighbor list by lengthening the centroid distance limit could possibly result in higher correlations, but software limitations restricted additional testing. Further research could establish the Great Lakes as polygon connectors, as opposed to barriers, for the purpose of determining neighbor polygons. Counties along coastal regions could then be assigned topological neighbors which bordered the same lake margins.

The unique shape of the ecoregion also influenced assignment of neighbor polygons to each county. The inclusion of values from counties outside province boundaries as polygon neighbors, so that counties closer to ecoregion boundaries would give a more accurate representation of their neighborhood production, could increase the accuracy of the techniques in this study. Alternative techniques might involve the creation of a province cartogram based on county attributes (such as travel time between centroids, etc). Analysis procedures could then be completed after the generation of a polygon neighbor list computed from the centroids of adjusted county boundaries.
CHAPTER V

ANALYSIS OF ENVIRONMENTAL AND ECONOMIC FACTORS ON DISTRIBUTION PATTERNS OF THE GENUS CUCURBITA

Introduction

Spatial statistics and GIS techniques were integrated in this chapter for the purpose of quantifying the influence of environmental and economic factors on the distribution of Cucurbita crops. The Eastern Broadleaf Province, as defined by the USDA Forest Service (1994), delimits the study region because of its environmental homogeneity, departure from national agricultural means, and tremendous increase in Cucurbita acreage and farm numbers from the period 1982 to 1992.

Because county values are arranged by their state affiliation in the U.S. Census of Agriculture, research often focuses on states or portions of states with specific characteristics. Leaders in horticultural specialty crops are generally designated by state and include California, Ohio, Pennsylvania and New York. North Carolina, Texas, Florida and Indiana are also major horticultural producers (U.S. Census Bureau 1988; Hickman 1992). The states that lead the nation in horticultural specialty products are also those with
large *Cucurbita* production values (U.S. Census Bureau 1995). Attempts at regional analyses, and even regional data values, have resulted in countless delineations of the United States, each containing a different combination of states (USDA 1920-95). This approach fails to give appropriate attention to environmental and economic factors that cross state boundaries, and, perhaps more importantly, fails to consider state area or population in calculations of economic importance. This investigation seeks to provide a deeper level of analysis than is possible from state defined regions, which seem to predominate the field.

Gains and increases from the 1988 Census of Horticultural Specialty crops were attributed, rightly or wrongly, to state education efforts and state cooperative extension staff (Hickman 1992). During the congressional hearing on the structure of the fruit and vegetable industry both the witnesses and submitted materials came predominantly from the state of California. Although there was some auxiliary representation by Florida, Michigan and Texas (U.S. House Committee on Agriculture 1988), testimonies repeatedly elaborated on state associated conditions rather than regional characteristics. This analysis seeks to eliminate from its methods the state-based analysis that seems to predominate the field.

Economic and environmental factors, independent of political boundaries, were chosen as variables with the attempt to arrive at possible
relationships to explain the increase in the production of pumpkin and squash. In addition, this portion of the research strives to establish a methodology suitable for incorporating diverse national data sources with regions defined by attribute as opposed to state affiliation. Among the Cucurbita production values, pumpkin acreage in 1992 changed most significantly during the decade and is highlighted during this chapter.

Various regional climate characteristics have been considered important to Cucurbita distributions. Growers of giant pumpkins, the orange "volkswagen size" fruits of Cucurbita maxima, have proposed that prime pumpkin growing occurs along the 43rd parallel of latitude (Royte 1997). While growers can modify other growing requirements, such as water, nutrient availability and temperature, sunlight hours can not be as easily manipulated. Devoted growers, such as Tony Ciliberto, even admit to removing stands of trees that decreased total daily sunlight by 30 minutes (Vitez 1997). Supposedly, the 43rd parallel of latitude balances the appropriate temperature range and sunlight requirements for pumpkin growth. Contrary to this belief, however, seed catalogues (Gurney 1996; Stokes 1998), seed packets, nurseries, and gardening books, cite the USDA hardiness zones as the key to planting locations. These zones correspond to contours of minimum winter temperatures. Both of these approaches are far from ideal, because of their narrow focus and failure to incorporate other
factors impacting growth. Regional temperature characteristics are important factors affecting plant growth and fruit production (Klages 1942; Rosenberg and others 1993; Geiger and others 1995). However, latitude and minimum winter temperature do not adequately represent thermal characteristics that may influence the agricultural patterns of the genus *Cucurbita*.

Data

This analysis explores the relationship of both temperature related environmental factors, such as frost characteristics and growing degree units, and county population characteristics derived from population estimates with *Cucurbita* values for the period 1982 to 1992. As in earlier portions of this study, data were extracted from government sources containing statistics for the entire United States. Data pertaining to the 649 county polygons were acquired from these source tables and connected to spatial point and polygon themes using ArcView’s® open database connectivity in conjunction with Microsoft Access® databases.

Environmental data were taken from “Climatic Summaries for Selected Sites, 1951-1980” (Climatography of the U.S. No. 20), a file within the *U.S. Divisional and Station Climatic Data and Normals* CDROM (USCDC 1995). This file contains data from 1,879 stations established at locations with
population greater than 5,000 people. Consequently, fewer data points are located in areas of low population density, but the population structure of the study region insures that it is sufficiently covered with sites. Data concerning frost and growing degree units were extracted for each station from the ASCII files using programs written in Fortran.

The tables extracted for frost data included the predicted dates of first and last frost at different levels of probability, as well as the probable duration of days when the temperature exceeded various frost related values. The estimation of this freeze data, computed by the USCDC (1995) was based on the work of Thom and Shaw (1958), Thom (1959), and Vestal (1971). For any preselected temperature, \( A \) is the event that the temperature will be reached during the cold season and \( B \) is the event that the temperature will be reached a certain date after the beginning of the cold season. \( B \) is predetermined at the probability levels of 0.1, 0.2, 0.3, 0.4, 0.5, 0.6, 0.7, 0.8, and 0.9 and \( P(A) \) is estimated by dividing the number of years that the temperature was reached by the total number of years in the climatic sample. Probabilities were calculated for the temperatures of 36, 32, 28, 24, 20, and 16 degrees Fahrenheit.

To calculate the date prior to which a preselected temperature has a certain probability of being reached, Vestal (1971) suggests the formula:
where \( x \) is the number of days after January 1 by which a temperature will be reached with a specific probability \( B \), \( t \) is the number of \( N(0,1) \) standard deviations associated with \( P(B/A) \), \( s \) is the standard deviation of the day numbers of event \( B \), and \( \bar{x} \) is the mean day number of event \( B \). This approach better estimates the frost characteristics of a region that does the minimum winter temperature used in the USDA Hardiness Zone delineation.

Growing degree units were also available in the *Climatic Data and Normals* CD and were computed at base temperatures that correspond to common phenological cycles in the United States including (in degrees Fahrenheit): 40, 45, 50, 55, 57, 60, 65, 70. Two bases were truncated, with adjustments of daily maximum and minimum temperatures to 48/86 and 50/86 if temperatures were lower than the low threshold, or higher than 86. Growing degree units were calculated from daily station extremes for each base temperature with the following equation:

\[
\frac{GDU}{N} = \frac{\sum GDU_i}{N} = \frac{\sum [(TMAX_i + TMIN_i)/2 - BASE]}{N} \tag{17}
\]

where \( N \) is the number of years in the record (in this case, \( N = 30 \)), \( GDU_i \) is the growing degree units from the year \( i \), \( TMAX \), and \( TMIN \), are the
maximum and minimum temperature for the date from the year $i$, and $BASE$ is the base temperature from which the growing degree days were calculated.

Because the U.S. Census of Agriculture and Census of Population take place during different years, the population estimates for 1982, 1987, and 1992 at the county level were used to correspond with Agricultural Census years. These populations are estimated by the Census Bureau and are based on revised populations for previous years, as well as from the results of special censuses and test censuses (U.S. Census Bureau 1998).

Methods

Neighbor weighting schemes established in the previous chapter were integrated with factors pertaining to economic and environmental structure of the province. Economic factors included population density and population density rate of change. Environmental variables included growing degree units, the frost free period, and probability of first frost as calculated in the climatic normals from the Climatic Data Center (USCDC 1995). After Pearson’s correlation coefficient was computed following the procedures established for variable pairs in Chapter III, scatter plots were created to determine the linearity of the relationship. Surveys of initial factor maps in relation to *Cucurbita* distributions and spatial regression procedures were used to quantify the relationships between variables and distribution.
In a manner similar to that used for the agricultural data (Figure 27), U.S. Census Bureau population estimates were connected with county boundaries. Because files for each annual period were organized individually, the necessary years were initially assembled into a spreadsheet containing all counties within the 16 states. This table was then queried for the appropriate 649 counties after being exported to the Access® database described in Chapter IV. The county data was then linked to spatial boundary files from the database to ArcView®.

Since climate normals were recorded at fixed points, their integration with polygon boundaries and associated attributes was more complex (Figure 28 and 29). The points were initially interpolated into grid coverages, which were summarized within county polygon zones. Station identification points were plotted for the entire country with extracted latitude/longitude coordinates (USCDC 1995). Those stations within the northeastern portion of the United States were selected for use in this study. Since the area is considerably larger than the Eastern Broadleaf Province boundaries, interpolation errors resulting around the boundaries of the point layer do not interfere with contour positions within the ecoregion. Both the growing degree unit and the freeze data were queried for these stations from within
Figure 27. Extraction of Census Population Estimates and Integration With Spatial Boundaries.
Figure 28. Extraction and Conversion of Climate Station Data for Integration With Polygon Boundaries.
Figure 29. Integration of Climate Data With Polygon Boundaries.
the Access® database and surfaces of 0.075 degree grid cells were
interpolated. The surface interpolation used inverse distance weighting,

\[ V_j = \frac{\sum_{i}^{n} \frac{V_i}{r_{ij}^p}}{\sum_{i}^{n} \frac{1}{r_{ij}^p}} \]  

(18)

an inverse power function, where \( n \) is the number of polygon neighbors of a
specific county and \( r \) is the distance between the centroids of polygon \( j \) and
its neighbor, \( i \). \( P \) is the power of the function, in this case \( P = 2 \), and \( V \) is the
attribute value assigned to polygon \( i \). The weighting of environmental data
was based on the values of a predetermined number of neighbors, since
determining neighbors radially would have limited the station data utilized
in less populated regions. The climatic surfaces were overlaid with the
polygon boundary file and grid cell values within each county were
summarized so that the resulting maximum, minimum, and mean values
could be linked to the polygon boundary file.

In surface interpolation from station location coordinates, neighbor
weighting was limited to the closest three station points. This limit resulted
in a surface quite similar to surfaces tested with greater neighbor definitions.
However, using three neighbors accentuated local patterns, and decreased
smoothing of contours. It also decreased the influence of lake boundaries on
lake shore counties, as grid cells were likely to find three neighbors without crossing a lake.

The date of last frost was extracted from the climatic normals CD by month and by day. These month and day columns were converted to day counts from January 1. Because late frosts often have an impact on the feasibility of growing specialty, fruit, and vegetable crops, the date of last frost in the spring, with a 90% probability was determined from the CD and correlated with county agricultural values. The number of frost free days at the 10% probability interval contained values for each of the selected 1,260 points. This representation gave a better impression of frost patterns throughout the region than did other probability values after preliminary analysis. As many of the higher probabilities lacked values for more southern station identifiers, surface estimation errors were greater, rendering the results uninterpretable.

Growing degree unit surfaces were calculated from cumulative growing degree units from January to October, using a base of 55 degrees. This time period spans the growing season for pumpkins, up to the period of increasing consumer demand in October when most shipping takes place.

Following their extractions from larger data files, the applicable factor values from frost, growing degree unit, and population tables were linked with agricultural data and county data in an Access® database. Pearson’s
correlation coefficient (see Chapter III) was calculated in SPSS® to determine the strength of the linear relationship between *Cucurbita* production data and other factors. Both S-PLUS® and SPSS® generated pairwise scatter plots of the more strongly correlated relationships to insure linearity. Tables containing the most strongly related factors were linked to county boundary files after being imported to ArcView®, where they were joined to county centroid values for use in neighbor determination. The table was then exported as a .dbf file and imported into the S-PLUS® spatial module.

Neighbor weights were computed based on the procedure followed in Chapter IV for the estimation of spatial autocorrelation. These weights were then used in spatial autoregression formulas to estimate the association between external factors and with *Cucurbita* distribution patterns.

**Results**

Climate related variables consistently showed a stronger relationship with absolute county values than with any change in agricultural values from 1982 to 1992. In contrast, population estimation correlation showed higher $r$ values when paired with changes instead of absolute production. In all cases, when zeroes in the county agricultural census data were given the artificial value of 999,999, correlations were also higher. This can be explained
by the fact that the 999,999 value represented a much stronger outlier in the distribution than zero, thus inflating correlations.

Table 4 reports data values from correlations between the number of frost free days, summarized by county polygon using a 10% probability estimation and agricultural pumpkin values. Although all values of Pearson's correlation coefficient (r) approached zero for variables relating to the change in agricultural production from 1982-1992, and therefore showed little relationship between the variables, some significance was shown by mean and minimum values. When correlated with the 1992 county totals of pumpkin and squash acreage, r values were higher. In both cases, whether zeroes were treated as 0 or 999,999, the highest correlation was shown between the minimum frost free days at 10% probability and 1992 county pumpkin acreage. Minimum values, as shown in Figure 30, increase southward, with few exceptions throughout the ecoregion.

The maximum date of last frost showed the strongest correlation of the 90% probable last frost dates (Table 5) with 1992 pumpkin acreage (r = 0.209, when zeroes retained their zero value). This variable is also indicative of later frosts that may shorten the growing season. Larger maximum values indicate counties where last frost is later than in other counties, especially when the last frost has a 90% chance of having already occurred. A map of the values (Figure 31) by county polygon appears to be the inverse of the previous
<table>
<thead>
<tr>
<th>Change in County Agricultural Values</th>
<th>Mean</th>
<th>Maximum</th>
<th>Minimum</th>
<th>Range</th>
<th>S. D.</th>
<th>Sum</th>
</tr>
</thead>
<tbody>
<tr>
<td>Total Farms</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Farms</td>
<td>0.106*</td>
<td>0.92*</td>
<td>0.114**</td>
<td>-0.076</td>
<td>-0.082*</td>
<td>-0.078</td>
</tr>
<tr>
<td>Acreage</td>
<td>0.069</td>
<td>0.064</td>
<td>0.078</td>
<td>-0.047</td>
<td>-0.040</td>
<td>-0.068</td>
</tr>
<tr>
<td>Vegetables, Harvested</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Farms</td>
<td>0.082*</td>
<td>0.070</td>
<td>0.092*</td>
<td>-0.074</td>
<td>-0.077</td>
<td>-0.065</td>
</tr>
<tr>
<td>Acreage</td>
<td>0.082*</td>
<td>0.070</td>
<td>0.092*</td>
<td>-0.074</td>
<td>-0.077</td>
<td>-0.065</td>
</tr>
<tr>
<td>Pumpkins, Harvested</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Farms</td>
<td>0.088*</td>
<td>-0.067</td>
<td>-0.099*</td>
<td>0.103**</td>
<td>0.115**</td>
<td>0.081*</td>
</tr>
<tr>
<td>Acreage</td>
<td>-0.032</td>
<td>-0.020</td>
<td>-0.041</td>
<td>0.066</td>
<td>0.074</td>
<td>0.061</td>
</tr>
<tr>
<td>Squash, Harvested</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Farms</td>
<td>-0.083*</td>
<td>-0.077</td>
<td>-0.085*</td>
<td>0.029</td>
<td>0.026</td>
<td>0.026</td>
</tr>
<tr>
<td>Acreage</td>
<td>-0.012</td>
<td>-0.010</td>
<td>-0.023</td>
<td>0.042</td>
<td>0.039</td>
<td>0.040</td>
</tr>
</tbody>
</table>
Table 4—Continued

Frost Free Days, 10% Probability Threshold

<table>
<thead>
<tr>
<th>1992 County Agricultural Values</th>
<th>Mean</th>
<th>Maximum</th>
<th>Minimum</th>
<th>Range</th>
<th>S. D.</th>
<th>Sum</th>
</tr>
</thead>
<tbody>
<tr>
<td>Zeroes as Squash</td>
<td>-0.185**</td>
<td>-0.165**</td>
<td>-0.201**</td>
<td>0.120**</td>
<td>0.109**</td>
<td>0.049</td>
</tr>
<tr>
<td>Zeroes as Pumpkin</td>
<td>-0.122**</td>
<td>-0.095*</td>
<td>-0.140**</td>
<td>0.147**</td>
<td>0.115**</td>
<td>0.077</td>
</tr>
<tr>
<td>Zeroes as Pumpkin</td>
<td>0.379**</td>
<td>0.361**</td>
<td>0.391**</td>
<td>-0.112**</td>
<td>-0.090*</td>
<td>-0.080*</td>
</tr>
<tr>
<td>Squash</td>
<td>0.259**</td>
<td>0.235**</td>
<td>0.275**</td>
<td>-0.135**</td>
<td>-0.109</td>
<td>-0.055</td>
</tr>
</tbody>
</table>

** Correlation is significant at the 0.01 level (two-tailed).
* Correlation is significant at the 0.05 level (two-tailed).
Figure 30. Minimum Frost Free Days at a 10% Probability Threshold (USCDC 1995).
Table 5

Correlation of Date of Last Frost and Agricultural Values

<table>
<thead>
<tr>
<th>Change in County Agricultural Values</th>
<th>Mean</th>
<th>Maximum</th>
<th>Minimum</th>
<th>Range</th>
<th>S. D.</th>
<th>Sum</th>
</tr>
</thead>
<tbody>
<tr>
<td>Total Farms Farms</td>
<td>-0.109**</td>
<td>-0.108**</td>
<td>-0.095*</td>
<td>-0.044</td>
<td>-0.046</td>
<td>-0.136**</td>
</tr>
<tr>
<td>Total Farms Acreage</td>
<td>-0.074</td>
<td>-0.074</td>
<td>-0.068</td>
<td>-0.022</td>
<td>-0.013</td>
<td>-0.102*</td>
</tr>
<tr>
<td>Vegetables, Harvested Farms</td>
<td>-0.084*</td>
<td>-0.086*</td>
<td>-0.077</td>
<td>-0.032</td>
<td>-0.029</td>
<td>-0.106**</td>
</tr>
<tr>
<td>Vegetables, Harvested Acreage</td>
<td>-0.084*</td>
<td>-0.086*</td>
<td>-0.077</td>
<td>-0.032</td>
<td>-0.029</td>
<td>-0.106**</td>
</tr>
<tr>
<td>Pumpkins, Harvested Farms</td>
<td>0.091*</td>
<td>0.089*</td>
<td>0.082*</td>
<td>0.025</td>
<td>0.029</td>
<td>0.124**</td>
</tr>
<tr>
<td>Pumpkins, Harvested Acreage</td>
<td>0.040</td>
<td>0.033</td>
<td>0.041</td>
<td>-0.020</td>
<td>-0.019</td>
<td>0.070</td>
</tr>
<tr>
<td>Squash, Harvested Farms</td>
<td>0.076</td>
<td>0.084*</td>
<td>0.064</td>
<td>0.065</td>
<td>0.062</td>
<td>0.077</td>
</tr>
<tr>
<td>Squash, Harvested Acreage</td>
<td>0.024</td>
<td>0.024</td>
<td>0.019</td>
<td>0.015</td>
<td>0.013</td>
<td>0.039</td>
</tr>
</tbody>
</table>
Table 5—Continued

<table>
<thead>
<tr>
<th>1992 County Agricultural Values</th>
<th>Mean</th>
<th>Maximum</th>
<th>Minimum</th>
<th>Range</th>
<th>S. D.</th>
<th>Sum</th>
</tr>
</thead>
<tbody>
<tr>
<td>Zeroes as Pumpkin</td>
<td>0.198**</td>
<td>0.209**</td>
<td>0.188**</td>
<td>0.071</td>
<td>0.061</td>
<td>0.112**</td>
</tr>
<tr>
<td>Zeroes as Squash</td>
<td>0.133**</td>
<td>0.144**</td>
<td>0.110**</td>
<td>0.105**</td>
<td>0.074</td>
<td>0.119**</td>
</tr>
<tr>
<td>Pumpkin</td>
<td>-0.376**</td>
<td>-0.381**</td>
<td>-0.366**</td>
<td>-0.062</td>
<td>-0.038</td>
<td>-0.211*</td>
</tr>
<tr>
<td>Squash</td>
<td>-0.258**</td>
<td>-0.264**</td>
<td>-0.246**</td>
<td>-0.064</td>
<td>-0.040</td>
<td>-0.157**</td>
</tr>
</tbody>
</table>

** Correlation is significant at the 0.01 level (two-tailed).
• Correlation is significant at the 0.05 level (two-tailed).
Figure 31. Maximum Last Frost Date With 90% Probability Threshold (USCDC 1995).
variable and shows their strong relationship with one another. In general, last frost date was more strongly related to pumpkin and squash acreage than frost free days.

Growing degree units (Table 6) accumulated to October demonstrated a stronger correlation with 1992 pumpkin and squash acreage than did the frost related variables. Figure 32 shows the decrease in growing degree units as counties increase in latitude. The minimum, maximum, and mean accumulated growing degree units, as measured from a base of 55, showed a negative correlation with pumpkin acreage of -0.23. The inverse relationship indicates that when fewer growing degree units are accumulated in a county, it is likely to have more pumpkin and squash acreage.

Population parameters were also correlated with both change in farms and change in acreage for each of the categories extracted for the Census data files (Table 7). Because of the similarity in data values, harvested vegetables were used in correlation estimates, while the vegetables grown were not. Population \( r \) values were much higher than those of the environmental variables, but were more related to the change in distribution than to the actual county acreage values. Most highly correlated with estimated change in population from 1982 to 1992 (Figure 33) were harvested pumpkins. Both change in farms (-0.720) and change in acreage (-0.946) per area showed
Table 6
Correlation of Ten Month Growing Degree Units and Agricultural Values

<table>
<thead>
<tr>
<th>Change in County Agricultural Values</th>
<th>Mean</th>
<th>Maximum</th>
<th>Minimum</th>
<th>Range</th>
<th>S. D.</th>
<th>Sum</th>
</tr>
</thead>
<tbody>
<tr>
<td>Total Farms</td>
<td>0.099*</td>
<td>0.093*</td>
<td>-0.103**</td>
<td>-0.057</td>
<td>-0.051</td>
<td>-0.035</td>
</tr>
<tr>
<td>Farms</td>
<td>0.074</td>
<td>0.071</td>
<td>0.077</td>
<td>-0.037</td>
<td>-0.043</td>
<td>-0.033</td>
</tr>
<tr>
<td>Acreage</td>
<td>0.078</td>
<td>0.072</td>
<td>0.082*</td>
<td>-0.056</td>
<td>-0.068</td>
<td>-0.032</td>
</tr>
<tr>
<td>Vegetables, Harvested</td>
<td>-0.092*</td>
<td>-0.087*</td>
<td>-0.095*</td>
<td>0.043</td>
<td>0.045</td>
<td>0.037</td>
</tr>
<tr>
<td>Farms</td>
<td>-0.053</td>
<td>-0.052</td>
<td>-0.053</td>
<td>0.005</td>
<td>0.006</td>
<td>0.030</td>
</tr>
<tr>
<td>Acreage</td>
<td>-0.046</td>
<td>-0.039</td>
<td>-0.050</td>
<td>0.071</td>
<td>0.081*</td>
<td>0.015</td>
</tr>
<tr>
<td>Pumpkins, Harvested</td>
<td>-0.039</td>
<td>-0.039</td>
<td>-0.040</td>
<td>0.003</td>
<td>-0.003</td>
<td>0.016</td>
</tr>
</tbody>
</table>
Table 6—Continued

<table>
<thead>
<tr>
<th>1992 County Agricultural Values</th>
<th>Growing Degree Units Accumulated Until October</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Mean</td>
</tr>
<tr>
<td>Zeroes as Pumpkin</td>
<td>-0.232**</td>
</tr>
<tr>
<td>Squash</td>
<td>-0.171**</td>
</tr>
<tr>
<td>Zeroes as 999,999 Pumpkin</td>
<td>0.429**</td>
</tr>
<tr>
<td>Squash</td>
<td>0.310**</td>
</tr>
</tbody>
</table>

** Correlation is significant at the 0.01 level (two-tailed).
* Correlation is significant at the 0.05 level (two-tailed).
Figure 32. Mean Growing Degree Units Calculated From a Base Temperature of 55°F and Accumulated From January 1 to October 31 (USCDC 1995).
Table 7

Correlation of Population Estimates and Agricultural Values

<table>
<thead>
<tr>
<th>Change per County Area</th>
<th>Change in Farms</th>
<th>Change in Acreage</th>
<th>Change in Acreage (- Wayne Cnty)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Total Farms</td>
<td>0.079</td>
<td>0.138</td>
<td>-0.542</td>
</tr>
<tr>
<td>Vegetables, Harvested</td>
<td>0.444</td>
<td>0.444</td>
<td>-0.461</td>
</tr>
<tr>
<td>Pumpkins, Harvested</td>
<td>-0.720</td>
<td>-0.946</td>
<td>0.589</td>
</tr>
<tr>
<td>Squash, Harvested</td>
<td>0.431</td>
<td>-0.023</td>
<td>0.721</td>
</tr>
</tbody>
</table>

strong negative correlations with the population variables. A scatter plot of these variables (Figure 34), however, illustrates the influence of one outlier on the strength of this relationship. Removal of Wayne County, MI, the county that includes much of urban Detroit, creates a completely different scatter plot (also Figure 34) and correlation estimate (0.589). Outliers continue to influence these results. Because of the extreme range of values, represented by both the pumpkin densities, and the population density estimates, the log of these variables were correlated. In this case, Pearson’s correlation
Figure 33. Rate of County Population Change, 1982-1992 (U.S. Census Bureau 1998).
Counties in Eastern Broadleaf Province

Pumpkin Acreage Change per County Area

Counties in Eastern Broadleaf Province without Wayne County, MI

Pumpkin Acreage Change per County Area

Figure 34. Scatterplots of Pumpkin Acreage and Population Changes.
coefficient was 0.5, as well. The scatter plot of log-transformed data is available in Figure 35.

![Log of Pumpkin Acreage Change](image)

**Figure 35.** Scatterplot of Log Transformed Variables.

In summary, when zeroes were left unchanged, the following four pairs of variables showed the highest correlation values: (1) 1992 pumpkin acreage and minimum frost free days, -0.20; (2) 1992 pumpkin acreage and maximum last frost date, 0.21; (3) 1992 pumpkin acreage and accumulated growing degree units, -0.23; and (4) change in pumpkin acreage and change in population, -0.95 (0.59 without Wayne County). Thus, environmental
variables were more strongly linked with actual production values, while population showed a stronger relationship with change in acreage.

Since population estimates seemed strongly correlated with change in harvested pumpkin acres, these variables were spatially linked using the Access® database and ArcView®. The theme attribute table was then exported to SPLUS®, for spatial regression analysis. Spatial regression estimates the amount of the relationship that is determined by spatial proximity. Spatial regression values were very close to zero (Table 8), and

<table>
<thead>
<tr>
<th>Change in Pumpkin Acreage</th>
<th>Population Rate of Change</th>
</tr>
</thead>
<tbody>
<tr>
<td>0.0248</td>
<td>0.0235</td>
</tr>
<tr>
<td>N/A</td>
<td>0.0268</td>
</tr>
</tbody>
</table>

showed that relationship between variables cannot be attributed to the spatial pattern of their joint distribution. However, the relatively high values of non-
spatial correlation estimates were encouraging from the perspective of this study’s hypothesis.

Discussion

The practical implementation of GIS and statistical software in conjunction with agricultural census data was hampered more by data characteristics than by application difficulties. Although census data are widely used because of the size and uniformity of its coverage (Evans 1981), the disadvantage is the number of non-numeric answers that must be represented numerically during statistical procedures. The presence of zero values was especially noticeable in this data set, because vegetable and specialty crops are traditionally of lower economic importance than the more well known field crops. Zero values influenced the outcome of every step in the procedure, especially in correlation calculations and the determination of neighbor weighting that heavily impacted the spatial autocorrelation and regression estimates. As discussed in Chapter IV, rescaling techniques such as those advocated for lattice data by Cressie (1993) can be used to reassign zero values. In this analysis, it was not only the zero values, but also the frequency of the zero values that were a problem when performing traditional statistical methods and spatial statistical techniques. No matter what value was reassigned to the zero values, they were frequent enough to
skew the remaining counties. Removing all zeroes, as well as other obvious outliers such as Wayne County (Figure 34), would reduce the data set to such an extent that conclusions would be extremely biased.

Zeroes did not necessarily indicate a lack of pumpkin and squash production in the county. Some counties did have zero production but a majority of the counties reporting a zero value simply did not record a production figure. In the 1982 census, many counties were labeled as "data not available", but this categorization gradually was phased out of use and by the 1992 census was rarely found in the record. Approximately 130 counties reported *Cucurbita* production for all three censuses. Advanced techniques for assigning values to the categories recorded as zero during importation or the use of imputation procedures could possibly alleviate some of these problems.

Even with different treatment of zero values, however, the results of this analysis do indicate that pumpkins are most frequently grown where factors affecting temperature and length of growing season are in balance. In general, pumpkin and, to a lesser extent, squash acreage was weakly correlated with frost probabilities and growing degree units. The affect of the zeroes on the Pearson's $r$ values is clearly shown by changing the zero value and from pairwise scatterplots. The influence of zero values was strong enough to change the direction of the relationship in nearly every situation.
(Tables 4, 5 and 6). As this is the case, correlation estimates can be interpreted to mean that in counties with fewer frost free days, later last frost dates and more accumulated growing degree units, there is more likely to be a zero value recorded for pumpkin and squash acreage.

Because pumpkins and squash have not shown a tremendous increase in shipments or arrivals to correspond with the escalation of acreage and farm numbers, *Cucurbita* products are being used on a more local scale. This would indicate that economic factors, such as population, as opposed to purely environmental variables, are driving the distribution of *Cucurbita* production.

The change in acreage and farm numbers, in fact, shows a slightly stronger relationship with estimated population change than with climatological variables. As the most prominent outlier in the population data set, Wayne County's large population decline over the ten year period, coupled with an impressive increase in pumpkin acreage suggest that previous urban pressure may have been too great for farming in the earlier years of the study period (Figure 34). With the decline of population, pumpkins increased dramatically. Other relative outliers in the population data, shown following the removal of Wayne County and subsequent change in the direction of the relationship, also show that those counties with a
greater change in population have a greater increase in pumpkin acreage (Figures 34 and 35).

Population factors were also extensively influenced by the frequency of zero values. Those counties with the greatest changes in population also had significant changes in pumpkin acreage, and those with zero, or extremely low, acreage showed less population change per area. This could indicate that pumpkin farming and changes accompanying the increase in ornamental production may relate to local population characteristics within the broader limits of environmental suitability.
CHAPTER VI

SUMMARY AND CONCLUSIONS

As possibly the first domesticated plant in the America's, the genus *Cucurbita* has long been lauded for its economic potential. In the United States pumpkin and squash production have recently been increasing with some regions of the nation showing dramatic changes within the ten year period from 1982 to 1992. In the Eastern Broadleaf Province, pumpkin acreage increased nearly 253% during the decade and the number of pumpkin farms increased 85%. In response to the increasing production of the genus *Cucurbita* in the United States and the innumerable uses of the plant, this research examined the temporal changes in utilization patterns.

The following is a summary of the outcomes of this research both with respect to *Cucurbita* distributions and the methodology used. Also included are suggestions for further research.

Increases in the number of farms and harvested acres of pumpkin and squash correspond with a period in the data record in which the consumption of pumpkins, as well as the canned production of squash and pumpkin products, stopped being reported in the agricultural statistics. This change,
coupled with an increase in shipping data and a leveling off of frozen processing estimates seems to indicate that a greater percentage of pumpkins are being harvested for inedible purposes. A transition of this nature would correspond with other structural changes in the United States agricultural industry during the 1980s and 1990s marked by increases in urban pressure and the decline of the small farm (Hallam 1993; Hart 1998; Klinge and Bartholic 1998). An increase in ornamental use of pumpkins, squash, and gourds and the increase of the harvesting of these fruits would also correspond with the increase in specialty, seasonal, and ornamental crops grown to provide a greater income to smaller farmers. Census data also seems to indicate that most *Cucurbita* fruits produced for human food are being reported as squash in the statistical record. Pumpkin farms and acreage, therefore, would to some extent represent inedible utilization.

The spatial analysis in this study focused on one region of cucurbit production in the United States, specifically, the Eastern Broadleaf Province. A tendency for high production values around the Great Lakes is visible in this ecological region. Three areas, including the (1) Erie and Ontario Lake Plain Section, (2) South Central Great Lakes Section, and (3) Southwestern Great Lakes Morainal Section consistently show evidence of higher *Cucurbita* production than national and regional means. An increase in farm numbers and acreage is also concentrated in these regions.
Throughout the United States and within the province boundaries, pumpkins are cultivated farther north than squash. As shown in Chapter V, date of last frost, length of the frost free period, and cumulative growing degree units showed evidence of weak correlation with 1992 acreage values extracted from the United States agricultural census. Changes in distribution from 1982 to 1992 were more strongly correlated with county population changes. In fact, counties that recorded strong change in population often recorded a strong change in pumpkin acreage.

The integration of statistical and spatial data through database management and spatial statistics software yielded little information relating to the spatial relationship between variables. The spatial autocorrelation was negligible, but the irregular shape of the ecoregion province boundaries affected the calculation of Moran's contiguity coefficient, $I$.

While visualization is essential in the field of geography, quantification of spatial patterns is useful in the analysis of lattice data structures and for comparison with similar distributions (Haining 1990; Cressie 1993). Disadvantages to the methods applied in this research involve the affect of missing or zero values and the shape of ecoregion boundaries, which do not conveniently align with political units. While environmental and population values are delimited by curvilinear shapes, counties and other land survey units, used by federal agencies for data collection and reporting, are more
geometrically simple. A categorical analysis of this data, in conjunction with statistical procedures, could be used to more accurately take zero records into account. In addition, the inclusion of counties outside the province boundaries would potentially increase the feasibility of spatial autocorrelation and regression values. Further analysis should adjust this study's methodology to accommodate the zero characteristics and spatial characteristics of the region of investigation.

Vegetable and ornamental crops, which are undergoing a transition in economic importance due to changes in foreign relation policies and United States population structure, need to be studied further with respect to recent changes in agriculture in the United States (U.S. House Committee on Agriculture 1988). Because of the local importance of *Cucurbita* crops and limited data on the production of pumpkin and squash in the agricultural census, the spatial analysis of these crops could easily lead to integrated research across spatial scales, as advocated by (Schulman and others 1994). On a larger scale, there is potential for a global study of the geography of *Cucurbita* utilization patterns. Numerous analyses of particular species include global perspectives that have been neglected in this research, but these have not been comprehensively summarized with respect to geography (Whitaker 1981; Harlan 1992; Yang and Walters 1992; Sauer 1993). At a smaller scale, the comparison of *Cucurbita* as a horticultural plant and as an
agricultural genus may reveal both regional customs and cultural use patterns.

Continuing research of the spatial distribution of the genus *Cucurbita* as an agricultural species for both edible and ornamental purposes could also lead to a methodology for the delineation of wild species in relation to domesticated cultivars. The preservation of genes as potential for economic resources in the future increasingly requires research in this direction, especially with the recent concern for the effects of global climate change on agricultural distributions (CGC).

On a more sociological note it is possible that distributions of agricultural production are corresponding with locations where festivals or contest-winning pumpkins are being grown. John Fraser Hart, currently one of the most prominent agricultural geographers in the United States, has documented the phenomena of field crop and livestock production concentrating around locations of dominant producers (Hart 1998). The available evidence does not indicate whether this pattern is also affecting less economically valuable crops, such as vegetables, ornamentals, and specialty crops.

In conclusion, *Cucurbita* species are important indicators in the changing scale and structure of U.S. agriculture. As farmland increasingly diminishes and small farms in areas of high population continue to decrease,
the genus has significant research potential because of its utilization as both a vegetable and an ornamental crop plant. Within the Eastern Broadleaf Province, the increase in ornamental production during the period 1982-1992 seems most affected by a combination of local factors as opposed to broad climatological patterns, social characteristics, or spatial proximity and adjacency.


Langevin D. 1993. Ten steps to a giant pumpkin. World Class Pumpkins.
Available: http://www.athenet.net/~dang/tips/10steps/10steps.html [1997, April 10].


----. 1998. S+GISLINK / ArcInfo / ArcView. Data Analysis and Products Division (dougj@statsci.com). E-mail to Leonard Pierce (leonard.pierce@wmich.edu); February 19, 1998.


of Commerce, Bureau of Census, Data User Services Division.

[Online data file]. Available: 

---. 1998. Estimates of the population of counties: Annual time series, July 1,
1990 to July 1, 1997. [Online state data files]. Available: 

[USDA] United States Department of Agriculture. 1920 through 1995, annual

---. 1950-51. Crops in peace and war. 81st Congress. 2nd Session. House
Document No. 691. Yearbook of agriculture series. Washington, DC:

---. 1952. Insects. 82nd Congress. 2nd Session. House Document No. 413.

products from U.S. agriculture. Yearbook of agriculture series.
p.

[1998, March 18].

U.S. Forest Service. 1994. Ecoregions and subregions of the United States
[Online ARC/INFO coverage]. Available: 

fruit and vegetable industry: Hearing before the Subcommittee on
Domestic Marketing, Consumer Relations, and Nutrition, 100th Cong., 2nd


