

**A DATABASE PROVIDING COTTON VARIETY
ACREAGE DATA AT THE COUNTY LEVEL
FOR 1950-95
A. F. Robinson
USDA, ARS
College Station, TX**

Abstract

Data obtained from the Agricultural Marketing Service (AMS, USDA), the National Agricultural Statistics Service (NASS, USDA), and the U. S. Census Bureau (U. S. Dept. Commerce), were compiled into a database that provides acres planted to each of 60 major cotton varieties during the interval 1950-95. Acres for each variety are given each year at the state level for all years from 1950 to 1995, and each year at the county level from 1979 to 1995; county level data prior to 1979 are given only at 5-year intervals from 1952 to 1972, due to unavailability of percentage estimates for specific varieties in intervening years.

Introduction

Cotton variety acreage data are of interest to persons involved in diverse aspects of cotton production, research, and development. Disconnected information regarding cotton variety acreage is available from several sources. A single source providing the comprehensive information needed to answer many kinds of questions related to the geographic distribution of varieties planted through time is not available.

Last year, the ARS cotton nematology program at College Station, TX, initiated a study to find out whether gradual but important changes in resistance to the major nematode pests of cotton has occurred during the last 40 or 50 years, and considered that this question could be answered by directly comparing reproduction by root-knot and reniform nematodes on a suitable number of currently and historically grown varieties in the same series of experiments. Ideally, the varieties tested would account for most of the land in cotton production in each of the last 4 or 5 decades. In order to identify as objectively as possible the varieties that we should test, I undertook the task of compiling data on acreage planted to specific Upland cotton varieties between 1950 and 1995.

Data Sources

Since 1950, the Agricultural Marketing Service (AMS) annually has published a booklet entitled "Cotton Varieties Planted", which provides estimates of percentages of cotton acreage planted to specific varieties at the county, state, and national levels. Data on more than 400 varieties have been

given since 1950 (Table 1). Prior to 1979, county level data were provided only at 5-year intervals, starting in 1952. The AMS publication provides no acreage data per se, and the percentage data provided cannot be used by itself to tally acres for specific varieties, or to map the geographic distribution of varieties through time. The National Agricultural Statistics Service (NASS, USDA), however, has collected information every year on the total acreage of Upland and of Pima cotton at the county level for many years, and this can be used in conjunction with the AMS data. County level data on cotton acreage is also independently collected by the U. S. Census Bureau.

Data Acquisition

The AMS generously provided a complete set of "Cotton Varieties Planted" for 1950-95. The national office of NASS recently put downloadable files with all county level data on Upland cotton for 1972-95 on the Internet and the Washington office assisted me in downloading. Data for years prior to 1972 was not available from Washington. I therefore contacted the NASS offices in 16 cotton-producing states, and they kindly provided hard copies of all available data between 1950 and 1971. For some years in some states NASS data did not exist and it was necessary to obtain missing estimates, either directly from the U.S. Census Bureau in Washington, D. C., or by interpolation (Fig. 1).

Description of the Database

At this point, the database consists of two total-acreage files and two corresponding percentage-acreage files for each of 60 varieties. One file of each pair gives county level data every 5 years and the other gives state level data every year. Multiplying corresponding elements in the total-acreage and variety-percentage files, generates variety-acreage data. I have used a simple, custom-written GWBASIC program to do this for the graphs and maps illustrated. Data in all county level files are stored in FIPSCODE format so that they can be mapped. The maps illustrated were generated from variety-acreage data files by Mapviewer, an inexpensive mapping package (Golden Software, Golden, CO). Similar maps and graphs of total acres planted through time for 30 of the 60 varieties were made. These provide a quick view of the history as well as geographic distribution of each variety. Year-acre totals also were tallied for all varieties to obtain a ranking and relative impact value for each.

Limitations of the Database

The original AMS and NASS cotton data was provided by cotton gins and classing offices. The U.S. Census Bureau data is derived from questionnaires. During the 1950's and 1960's, "Cotton Varieties Planted" is in some cases unclear. In early years, state and national estimates were provided only for the leading 10 or so varieties; several variety

names in the early 1950's are ambiguous. In some county level data sets within state districts or classing offices, a significant percentage of all cotton acreage is given as "other", or "unreported". During some years in some states, NASS data was nonexistent and acreage values had to be interpolated. Finally, there are undoubtedly uncorrected errors in the keypunching done at College Station. Overall, however, it is my opinion that the database as a whole is reasonably accurate, and that its merits and potential applications far outweigh its inherent limitations, if used judiciously.

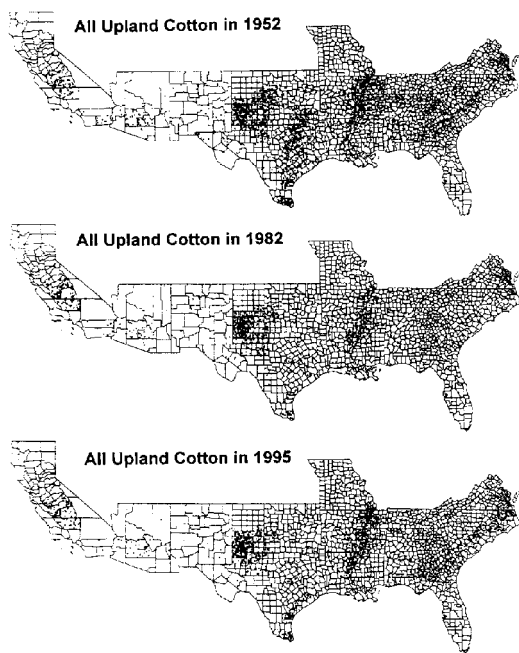


Figure 1. Examples of plots of total Upland cotton in the United States. One dot equals 10,000 acres. Within each county, dots are randomly distributed and may not accurately reflect the distribution inside that county.

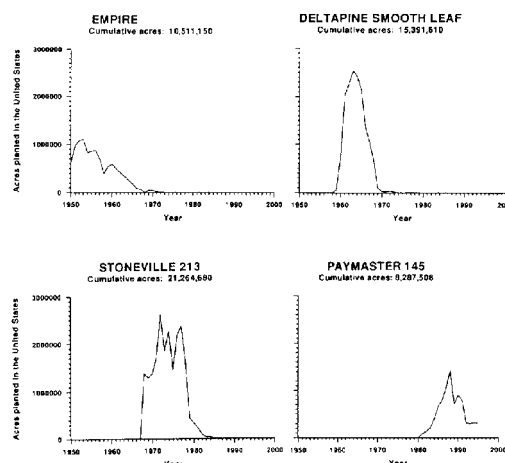


Figure 2. Examples of graphs showing acreage of Upland cotton varieties in the United States through time..

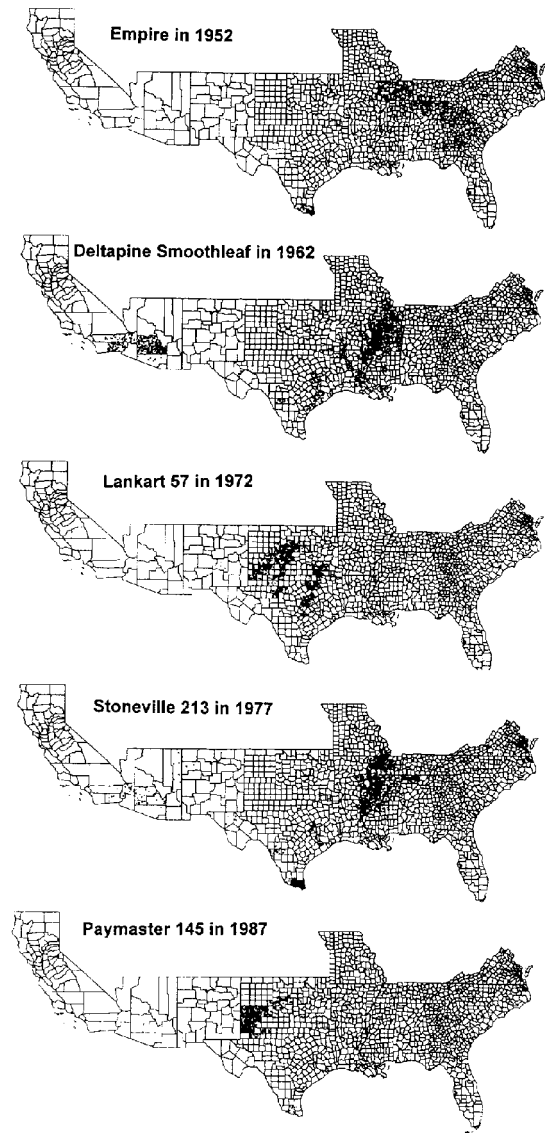


Figure 3. Examples of plots of specific Upland cotton varieties. One dot equal 1,000 acres. Dots are randomly distributed within each county and may not accurately represents the distribution inside the country.