E83-10138

CR-169775

AgRISTARS

EW-J2-04358
JSC-18572

A Joint Program for Agriculture and Resources Inventory Surveys Through Aerospace Remote Sensing

Early Warning and Crop Condition Assessment

September 1982

IMPUTING HISTORICAL STATISTICS, SOILS INFORMATION, AND OTHER LANDSAT-USE DATA TO CROP AREA

PREPRINT FOR THE ANNUAL MEETING OF THE AMERICAN STATISTICAL ASSOCIATION
- SURVEY RESEARCH METHODS SECTION -
CINCINNATI, OHIO, AUGUST 16-19, 1982

Lyndon B. Johnson Space Center
Houston, Texas 77058

IMPUTING HISTORICAL STATISTICS, SOILS INFORMATION, AND OTHER LAND-USE DATA TO CROP AREA

Charles R. Perry, Jr., Ronald W. Willis, and Lyle Lautenschlager

U.S. Department of Agriculture

ABSTRACT

Today, in foreign crop condition monitoring, use is routinely made of satellite acquired imagery. To facilitate interpretation of this imagery, it is advantageous to have estimates of the crop types and their extent for small area units, i.e., grid cells on a map represent, at 60° latitude, an area nominally 25 by 25 nautical miles in size. The feasibility of imputing historical crop statistics, soils information, and other ancillary data to crop area for a province in Argentina is studied.

1. INTRODUCTION

The Foreign Agricultural Service (FAS) of the U. S. Department of Agriculture (USDA) estimates crop acreages and production for foreign countries. Input into these estimates is provided by the Foreign Crop Condition Assessment Division (FCCAD) of FAS, currently located in Houston near the Johnson Space Center. The FCCAD is responsible for assessing the change in area planted to various crops in important foreign agriculture areas. The primary data source used by FCCAD analysts to assess the areal change is Landsat Multispectral Scanner (MSS) data (Wenderoth and Yost, 1972). Four of these satellites have been put into orbit, the latest being in July of this year. The design and orbital pattern of Landsats 1, 2, and 3 enabled each to collect data over the same spot every 18 days. At times, two Landsats were operating concurrently, but staggered, so that data could be collected every 9 days (providing cloud cover did not interfere). Landsat-4 is designed to collect data every 16 days over the same spot. Each Landsat MSS image covers an area of 100 by 100 nautical miles.

The FCCAD analysts monitor the majority of the crop region within a country, and a great deal of time is required to analyze each Landsat MSS image. Thus, only certain high density or highly variable areas are closely monitored to determine if a shift in cropping patterns has occurred. A grid mesh system, which is rectangular to a polar stereographic projection of the Northern Hemisphere [Earth Satellite Corporation (1976)], was selected as the basis for choosing the highly variable areas. The spacing between successive grid points at 60° latitude is nominally 25 nautical miles. The grid cells were further divided into four quadrants for sampling purposes.

A data set containing soils, agricultural density, and meteorological data for each quadrant is part of the FCCAD geo-gridded data base. However, historical agricultural area statistics are available only at the state or country level for most countries. Therefore, to estimate the crop mix in a particular grid cell quadrant, the state or country level historical statistics must be apportioned to the grid cell quadrant.

The purpose of this study was to determine if data elements stored in the geo-gridded data base could be used to apportion historical agricultural area statistics to the 12.5- by 12.5-nautical-mile grid cell quadrant.

The scope of this study was limited to the use of historical agricultural area statistics for one year for four crops (corn, wheat, soybeans, grain sorghum) from the Province of Cordoba, Argentina; all grid cell quadrants which lie within the boundaries of that province; and the associated data available from the geo-gridded data base. Although there are 444 grid cell quadrants in the province, only 434 were used, as 10 quadrants were cloud covered on the imagery used to determine the agricultural density.

2. DATA SET

HISTORICAL AGRICULTURAL AREA STATISTICS

Historical agricultural area statistics at the province (state) and partido (county) level were available from the Province of Cordoba, Argentina, for four crops for one year. The crops were corn, wheat, soybeans and grain sorghum with the crop year being 1978/79 for corn, soybeans, and grain sorghum and 1977/78 for wheat. The partido-level area statistics for each crop in each partido were initially apportioned equally to each quadrant in the partido. This procedure provided the initial figure for each crop in each quadrant for use in the correlation analysis.

AGRICULTURAL DENSITY

Using historical Landsat data, Dr. Fred Westin, South Dakota State University, determined the quadrant-level agricultural density and divided it into five categories based on the percentage of cultivated crop land. The categories and percent ranges were: 1 = 80 to 100 percent; 2 = 60 to 80 percent; 3 = 40 to 60 percent; 4 = 5 to 40 percent; and 5 = less than 5 percent. Neither the categories nor the percentage ranges provided a usable value with which to work. Consequently, the midpoint of each range was selected to replace the category. The categories were replaced as follows: 1 with 0.90; 2 with 0.70; 3 with 0.50; 4 with 0.225; and 5 with 0.025. Also stored in the data base with the categories were the reasons for the different rankings for each quadrant; e.g., "hvg" indicates hilly with vegetation. Approximately 60 different reasons were available. The reasons were used only as an aide throughout the analysis.

SOILS DATA

The soil type chosen to represent the soils in the quadrant was the predominant soil within the quadrant. Associated with the soil types were five categories: texture, depth, drainage, salinity and slope. Each soil type received a rating from one to four in each of the five categories with one being the best and four the worst. The values of each category were summed for each soil type and then ranked, with the lowest total receiving a ranking of one. Dummy values were also added for the five categories. The category or categories with the highest value were deemed to be the limiting factor(s) for that soil type. For example, a soil type might have these ratings: texture 1, depth 2, drainage 3, salinity 1, and slope 3. The limiting categories would be drainage and slope. Therefore, the dummy values would be: texture 0, depth 0, drainage 1, salinity 0, and slope 1.

AGRO-PHYSICAL UNITS

Agro-physical units (APU's) for Argentina were delineated by H. Edward Bulloch, of the FCCAD. An APU is a relatively homogeneous area similar in climate, soils, and crops. Each quadrant lies within only one APU depending on which APU was predominant in the quadrant. All APU's were grouped in similar climatic regimes and then numbered within the regime for analysis purposes.

3. ANALYSIS

To evaluate the apportioning procedure, accuracy was selected as the primary criterion for determining its general performance. Secondary criteria considered were efficiency and repeatability.

The accuracy criterion was selected as the primary criterion because if the apportioning procedure did not accurately determine the area to be apportioned to each grid cell quadrant, then the efficiency and repeatability criteria were moot questions to be considered. The apportioning procedure should accurately apportion the crop area to each grid cell quadrant, since the analyst will use this as a basis for evaluating subsequent crop mixes in that grid cell quadrant.

GRAPHS

Scatter plots were created to help determine visually if any relationship existed between the crops and any of the other variables described in Part 2. No clear relationship was indicated in any of the graphs. Therefore, no variables were deleted from consideration for use in the apportioning procedure.

APPORTIONMENT AND CORRELATION

Partido-level Statistics Equally Apportioned

The initial apportioning was to allocate equally the partido-level statistics of each of the four crops to each grid cell quadrant. This was accomplished by dividing the partido-level area statistics by the number of grid cell quadrents in the partido. A Spearman correlation was computed between the apportioned amount and the different soil variables. No variable showed a high correlat.on for any of the four crops.

Partido-level Statistics Apportioned Using Ag Density

The next apportioning procedure utilized the quadrant level ag density to apportion the partido-level statistics. A Spearman correlation was also computed in this apportionment (Willis, 1982). Ag density had the highest correlation with all four of the crops. This was expected because the ag density was used in the apportioning procedure. No other variables showed a high correlation for any of the crops.

Province-level Statistics Apportioned Using Ag Density

The next apportioning procedure analyzed was to use province-level statistics and apportion them to the quadrant using ag density. A Spearman correlation was computed between the apportioned amounts and the soil variables. No variable other than ag density indicated a very high correlation with all of the crops - an expected result since ag density was used in the apportionment.

Province-level Statistics Using Equal Apportionment

The last apportionment procedure analyzed was to apportion equally each crop to the partido based on partido size. A Spearman correlation was computed. None of the soil variables showed a very high correlation with any of the the four crops.

Partido Totals for Two Province-Level Procedures

Tables 1-4 list the actual partido statistics and the values for each partido for each of the province-level apportionment procedures for each crop. The partido total for each procedure was determined by summing the values apportioned to each quadrant in the partido. The totals at the bottom of each column indicate that no difference existed between the two procedures. The major problem lies at the partido level. Partidos with no area for a crop were apportioned some area of that crop, and partidos with a large actual area for a crop were invariably underapportioned. The soybean crop is a prime example (Table 2).

Approximately one-half of the partidos had soybean area; yet the apportioning procedure using province-level statistics allocated some soybean area to every partido. Similarly, partidos with large actual areas were allocated much less than the actual area.

Table 5 shows the Spearman correlation coefficients computed at the partido level between the actual statistics and the two estimates using province-level data. The coefficients ranged from .134 to .915. The correlation coefficients appeared quite high for sorghum, but when comparing the estimates there were some large differences.

ORIGINAL PAGE IS
OF POOR QUALITY

4. CONCLUSIONS AND RECOMMENDATIONS

CONCLUSIONS

Apportioning of province-level statistics was very poor. Although the province total was apportioned to the quadrants, the individual partido totals varied greatly from the actual partido areas. This procedure would lead to inaccurate analyst conclusions concerning cropping practices. An analyst could conclude that a major shift from one crop to another has occurred within a partido while in actuality no such shift occurred. Therefore, because very few countries publish partido-level statistics and the apportioning procedure employing province-level statistics was very poor, some other method or procedure must be devised to apportion province, or some higher hierarchial level, statistics to the quadrant.

RECOMMENDATIONS

Because of the low correlation between the soils information and the crops, the first recommendation is that more detailed information about the soils be obtained. These variables could then be studied to determine their value in an apportioning procedure.

Quadrant level ag density is also required in any procedure, and the ranges used in the categories in this study may be too broad, especially in the fourth category (5 to 40 percent). The second recommendation is that categories should be broken down into five-percent ranges. Although the small ranges are highly desirable, they may be unfeasible due to limitations of present technology.

5. REFERENCES

Earth Satellite Corporation, 1976, "Earthsat Spring Wheat Yield System Test 1975," NASA/JSC NAS9-14655.

Wenderoth, Sondra and Edward Yost, 1972, "Multispectral Photography for Earth Resources," NASA/JSC NAS9-11188.

Willis, Ron, 1982, "Apportioning Historical Agricultural Statistics Using Geo-gridded Data Base Elements," FCCAD TM 15.

---

*Charles R. Perry, Jr. and Lyle Lautenschlager are with the Statistical Reporting Service (SRS) assigned to the Early Warning/Crop Condition Assessment (EW/CCA) project within the Agriculture Resources Inventory Surveys Through Aerospace Remote Sensing (AgRISTARS) program. Ron Willis is with the Foreign Agriculture Service (FAS) assigned to the Foreign Crop Condition Assessment Division (FCCAD). All are located at 1050 Bay Area Blvd., Houston, Texas 77058.

TABLE 1. Actual partido statistics for sorghum vs the two province-level apportionment procedures based on ag density and partido size.

	ACTUAL	AG DENSITY	SIZE
Calamuchita	3,600	3,001	18,680
Capital	1,000	2,348	1,698
Colon	5,700	12,003	10,189
Cruz Del Eje	1,000	14,938	30,567
General Roca	68,000	67,451	61,134
General San Martin	21,800	30,529	22,076
Ishilin	1,500	5,153	20,378
Juarez Celman	47,000	41,749	30,567
Marcos Juarez	32,000	64,711	47,548
Minas	0	5,153	15,283
Pocho	1,700	5,349	15,283
Pres. Roque Saenz Pena	60,000	49,642	39,058
Punilla	1,000	11,546	11,887
Rio Cuarto	100,000	91,522	73,021
Rio Primero	23,000	32,682	32,265
Rio Seco	1,000	29,290	49,247
Rio Segundo	57,000	33,465	25,472
San Alberto	1,000	11,938	13,585
San Javier	0	5,349	6,793
San Justo	172,000	103,067	76,417
Santa Maria	8,000	18,787	13,585
Sobremonte	300	391	10,189
Tercero Arriba	39,000	21,723	16,982
Totoral	4,100	9,067	11,887
Tulumba	3,300	10,763	37,359
Union	84,000	55,383	45,850
TOTAL	737,000	737,000	737,000

TABLE 2. Actual Partido statistics for soybeans vs the two province-level apportionment procedures based on ag density and partido size.

	ACTUAL	AG DENSITY	SIZE
Calamuchita	2,400	1,059	6,590
Capital	0	829	599
Colon	400	4,234	3,594
Cruz Del Eje	0	5,270	10,783
General Roca	0	23,795	21,567
General San Margin	4,200	10,770	7,788
Ischilin	0	1,818	7,189
Juarez Celman	16,000	14,728	10,783
Marcos Juarez	82,000	22,829	16,774
Minas	0	1,818	5,392
Pocho	0	1,887	5,392
Pres. Roque Saenz Pena	0	17,513	13,779
Punilla	0	4,073	4,194
Rio Cuarto	6,000	32,287	25,760
Rio Primero	3,000	11,530	11,382
Rio Seco	0	10,333	17,373
Rio Segundo	29,000	11,806	8,986
San Alberto	0	4,211	4,793
San Javier	0	1,887	2,396
San Justo	2,000	36,360	26,959
Santa Maria	22,000	6,628	4,793
Sobremonte	0	138	3,594
Tercero Arriba	61,000	7,663	5,991
Totoral	0	3,199	4,194
Tulumba	0	3,797	13,180
Union	32,000	19,538	16,175
TOTAL	260,000	260,000	260,000

TABLE 3. Actual partido statistics for corn vs. the two province-level apportionment procedures based on ag density and partido size.

	ACTUAL	AG DENSITY	SIZE
Calamuchita	8,500	2,687	16,728
Capital	2,000	2,103	1,521
Colon	6,500	10,749	9,124
Cruz Del Eje	1,700	13,378	27,373
General Roca	11,500	60,404	54,747
General San Martin	5,000	27,340	19,770
Ischilin	2,000	4,615	18,249
Juarez Celman	25,000	37,387	27,373
Marcos Juarez	174,000	57,951	42,581
Minas	2,000	4,615	13,687
Pocho	50,000	4,790	13,687
Pres. Roque Saenz Pena	7,000	44,455	34,977
Punilla	2,800	10,340	10,645
Rio Cuarto	226,000	81,960	65,392
Rio Primero	4,000	29,267	28,894
Rio Seco	4,000	26,229	44,101
Rio Segundo	11,000	29,968	22,811
San Alberto	14,000	10,690	12,166
San Javier	2,000	4,790	6,083
San Justo	4,000	92,299	68,433
Santa Maria	9,000	16,824	12,166
Sobremonte	1,000	351	9,124
Tercero Arriba	17,000	19,453	15,207
Totoral	7,000	8,120	10,645
Tulumba	3,000	9,639	33,456
Union	60,000	49,596	41,060
TOTAL	660,000	660,000	660,000

TABLE 4. Actual partido statistics for wheat vs. the two province-level apportionment procedures based on ag density and partido size.

	ACTUAL	AG DENSITY	SIZE
Calamuchita	3,600	2,036	12,673
Capital	0	1,593	1,152
Colon	500	8,143	6,912
Cruz Del Eje	0	10,135	20,737
General Roca	22,000	45,760	41,475
General San Martin	10,500	20,712	14,977
Ischilin	0	3,496	13,825
Juarez Celman	15,000	28,324	20,737
Marcos Juarez	180,000	43,902	32,258
Minas	0	3,496	10,369
Pocho	0	3,629	10,369
Pres. Roque Saenz Pena	27,000	33,679	26,498
Punilla	0	7,833	8,065
Rio Cuarto	10,000	62,091	49,539
Rio Primero	6,500	22,172	21,889
Rio Seo	0	19,871	33,410
Rio Segundo	16,000	22,703	17,281
San Alberto	0	8,099	9,217
San Javier	0	3,629	4,608
San Justo	22,000	69,924	51,843
Santa Maria	2,900	12,746	9,217
Sobremonte	0	266	6,912
Tercero Arriba	21,000	14,737	11,521
Totoral	0	6,152	8,064
Tulumba	0	7,302	25,346
Union	163,000	37,573	31,106
TOTAL	500,000	500,000	500,000

TABLE 5. Correlation coefficients computed at the partido level between actual data and two apportionment estimates.

	Sorghum	Soybeans	Corn	Wheat
Ag Density	.915	.282	.558	.457
Size	.800	.134	.498	.365