I

1.Classification *INPE-COM.3/RPI*
C.D.U.: 631.551.5

2.Period

4.Distribution Criterion

internal [X]

external [ ]

3.Key Words (selected by the author)
SOYBEAN YIELD
MULTIPLE REGRESSION
WEATHER AND TECHNOLOGY EFFECTS

5.Report Nº
INPE-1647-RPI/014

6.Date
December, 1979

7.Revised by
Antonio Tebaldi Tardin

8. Title and Sub-title

REGRESSION ANALYSIS OF DIRA-RIBEIRÃO PRETO SOYBEAN YIELD (1956-1978) IN RELATION TO WEATHER AND TECHNOLOGY TREND

9.Authorized by
Nelson de Jesus Parada
Director

10.Sector *DSR/GAF* Code *30.312*

11.Nº of Copies *02*

12.Authorship *Sherry Chou Chen*

14.Nº of Pages *16*

15.Price

13.Signature of the responsible

16.Summary/Notes

A yield regression model was developed for soybeans (<u>Glycine max</u> (L.) Merr.) of DIRA-Ribeirão Preto, using total evaporation in December, relative humidity of January and a surrogate technology trend as indepedent variables. Eighty percent yield variation of soybeans from 1956 to 1978 was explained by the regression equation, which provides yield information in the beginning of February - two months after planting. The forecasted yield for crop year 1978-1979 is 1576.22 kg/ha if the past weather patterns and technology improvements continue. The results show that the standard monthly meteorological elements-relative humidity and total evaporation are better yield determinants for soybeans and corn than total precipitation and mean temperature which were used in the Thompson's models.

17.Remarks

The methodology and the yield estimate of crop year 1979-1980 will be presented to the Second "Symposium of Brazilian Agrimeteorology" (Pelotas - R.S - January, 1981)

INDEX

LIST OF TABLES .... *iv*

1. INTRODUCTION 1

2. MATERIAL AND METHODS 1

3. RESULTS AND DISCUSSION 2

REFERENCES 12

LIST OF TABLES

Table 1 - Estimated Soybean Production, Area and Yield of DIRA-RP by IEA (Instituto de Economia Agrícola)............ 5

Table 2 - Correlation Coefficients of Soybean Yield with Weather or Technology Variables............................ 6

Table 3 - Variables for Stepwise Multiple Regression Anal ysis 7

Table 4 - Summary Table of Stepwise Multiple Regression Analysis... 8

Table 5 - Analysis of Variance of Regression Model Y = a+b (VAR 32) + c (VAR 43) + d (VAR 10) 9

Table 6 - Soybean Yield Estimations by IEA (Instituto de Economia Agrícola) and Regression Model in DIRA-Ribeirão Preto (1956-1978)............................ 10

1. INTRODUCTION

The most important information to agribusiness strategy makers is timely and accurate statistics about crop production. Early prospective crop production may be determined considerably before harvest if the yield prospects are known. The major cause for season-to-season variation in crop yield is the fluctuation of weather and climate. Thus, substantial research has been carried out to predict yield based on past weather patterns (5,6,7). A primary consideration, in the development of an agrometeorological yield model, is that crop yield should be related to standard meteorological variables, which are highly correlated and readily available from a long recorded period. This prerequisite is necessary in order to provide the greatest potential for operational use of the weather-yield relationship. The regression methodology to predict crop yield, based on historic weather and yield data, has been widely used by NASA's LACIE (Large Area Crop Inventory Experiment) group. The predicted yield was, then, combined with the crop acreage estimated by LANDSAT satellite to provide crop production information (4).

The objective of this study was to explain soybean (Glycine max (L.) Merr.) yield variation from 1956 to 1978 of the agricultural district of Ribeirão Preto (DIRA-RP), using meteorological and trend factors. DIRA-RP was chosen as the study area not only due to its homogeneous climate, topography, soil type and farming pratices, but also because it was the test area of a crop inventory study using LANDSAT data processed by Instituto de Pesquisas Espaciais - INPE (1).

2. MATERIAL AND METHODS

Soybean is generally planted in the period from November to December and harvest in April and May in DIRA-Ribeirão Preto. Historic monthly meteorological data were provided by the Meteorological Service of the Agricultural Ministry. Yield data for soybeans since 1956 (Table I) were derived from the final estimates of crop production and area of the Instituto de Economia Agrícola (IEA).

For model construction, correlation analysis were first carried out between historic yield and monthly meteorological data such as total evaporation, relative humidity, total precipitation and mean temperature. Yield was also correlated with solar radiation, which was estimated using the equation derived by Cervellini et al. (2). Besides using the original data of above mentioned meteorological factors, the transformed variables using the absolute value of the difference between each factor and its long-term normal (average from 1956 to 1978) were included for correlation analysis as well. All contributions to soybean yield by non-weather factors, such as better management, fertilization, disease resistant cultivars, weed and pest control, mechanization, were designated by a surrogate variable-technology trend. A series of successive numbers starting, from 1, was coded to each year for analysis (1 for 1956, 2 for 1957 ...., 23 for 1978). The variables, which correlated significantly to yield were then used as potential components to explain yield variability. Stepwise multiple regression of the SPSS (Statistical Package for Social Sciences) program, which selects the variable according to its contribution to yield variation, was used. This program adds one variable at a time to the regression and its importance to yield fluctuation is calculated.

3. RESULTS AND DISCUSSION

The correlation coefficients of soybean yield and various weather factors or technology trend from 1956 to 1978 are presented in Table II. Among the sixty-six variables analyzed, fourteen were significantly correlated. Linear technology trend has the highest correlation ($r=0.7320$) and total evaporation, which correlated significantly to yield in five of the six study months, was the most important monthly meteorological indicator of soybean yield. The significant correlations of the transformed variables of total evaporation in October, mean temperature in November and March, solar radiation and relative humidity in January, are worth noting. The original monthly meteorological factors and their transformations, which were used as indepedent variables for multiple regression analysis, are listed in Table III. The summary table of regression analysis (Table IV) shows that 53.59% yield variation during the 23-year study period was explained by technology trend (VAR 32). The linear yield increment is 24.34 kg/ha/year. The addition of VAR 43, i.e., absolute value of the difference between relative humidity of January and its long term average- 77.30, to the equation gave another 15.14% increment in explanation of yield variability. Inclusion of the third variable (VARIO)-total evaporation in December, also contributed an 11.53% improvement to the R^2 value of the equation. Tests of significance of multiple regression and partial regression coefficients of VAR 32, VAR 43 and VAR 10 are presented in Table V. The equation Y=1556.14+14.77(VAR 32)-46.46(VAR 43) - 2.22 (VAR 10) has R^2 value and standard error of estimation of 0.8026 and 107.79 kg/ha respectively.It is clear from the magnitudes of F statistics that all the coefficcients in the equation are statistically significant. The other eleven weather variables were not included in the equation owing to their small F values. The differences between the yield estimates by IEA and the regression equation range from 0.82 to 182.04 kg/ha (Table VI). Even though almost a 20% of yield variation was left unexplained., the selected equation approximates satisfactorily the yield fluctuation of soybeans from 1956 to 1978 (Fig.1). The simplicity of the regression model, which uses meteorological data from two successive months (total evaporation of December and relative humidity of January) as inputs, provides yield information in the beginning of February;two months after planting. The importance of total evaporation and relative humidity to explain the yield variation of corn have also been shown in a previous work by the author and Fonseca (3). These results suggest that total evaporation and relative humidity are better yield determinants for corn and soybeans than total precipitation and mean temperature, which were used in Thompson's models (6,7).

The expected soybean yield for crop year 1978-1979 is 1576.22 kg/ha if the past weather conditions and technology trend continue. However, in applying the regression model for yield forecasting beyond 1979, all the available historic data should be used in calculating the regression coefficients for technology trend, total evaporation and relative humidity. The addition of a greater number of yield observations over time may also lead to the inclusion of new terms in the equation, which would increase the percentage of explained soybean yield variation, and provide a more accurate yield forecasts.

TABLE 1

ESTIMATED SOYBEAN PRODUCTION, HARVEST AREA AND YIELD OF DIRA-RP BY IEA (INSTITUTO DE ECONOMIA AGRÍCOLA).

YEAR	PRODUCTION (METRIC TONS)	HARVEST AREA (HA)	YIELD (KG/HA)
1956	147.00	140.36	1,047.30
1957	196.80	157.30	1,251.11
1958	175.20	142.78	1,227.06
1959	52.20	48.40	1,078.51
1960	1,398.96	1,205.16	1,160.80
1961	2,172.00	1,984.40	1,094.53
1962	2,646.00	2,129.60	1,242.48
1963	1,740.00	1,294.70	1,343.94
1964	2,431.14	2,359.50	1,030.36
1965	7,113.00	4,961.00	1,433.78
1966	13,374.00	8,857.20	1,509.95
1967	22,080.00	15,475.90	1,426.73
1968	33,060.00	25,482.60	1,297.35
1969	54,600.00	42,471.00	1,285.58
1970	83,700.00	56,918.40	1,470.52
1971	74,400.00	71,632.00	1,038.64
1972	175,200.00	100,000.00	1,752.00
1973	240,000.00	162,000.00	1,481.48
1974	309,840.00	211,000.00	1,468.43
1975	390,000.00	245,900.00	1,586.01
1976	336,000.00	184,000.00	1,826.08
1977	303,000.00	198,000.00	1,530.30
1978	396,300.00	252,000.00	1,572.61

TABLE 2

CORRELATION COEFFICIENTS OF SOYBEAN YIELD WITH WEATHER OR TECHNOLOGY VARIABLES

(DATA BASE = 1956 - 1978)

VARIABLE	SYMBOL	MONTH						
		O	N	D	J	F	M	O~M
Solar radiation (Cal/cm^2/day)	SR	0.1761	-0.1021	-0.3832	-0.2709	-0.0831	-0.1047	-0.1366
Total evaporation (mm)	E	-0.5783**	-0.4429*	-0.5907**	-0.5128*	-0.5183*	-0.2268	-0.6641**
Relative humidity (%)	RH	0.2753	0.3397	0.4009	0.2112	0.1725	0.1150	0.4994*
Total precipitation (mm)	P	-0.0316	0.2768	0.2833	0.0503	0.1737	-0.0081	-0.3617
Mean temperature (°C)	T	-0.3142	-0.2056	-0.4166*	-0.2151	-0.1551	0.1497	-0.3430
ABS$^+$ (SR - M^{++})		-0.4064	-0.1336	0.0535	-0.5601**	-0.0293	-0.2083	
ABS (E - M)		-0.4216*	-0.0874	-0.1936	-0.3270	-0.0123	0.1704	
ABS (RH - M)		-0.2486	0.0095	-0.1498	-0.5699**	0.0328	-0.0633	
ABS (P - M)		-0.2290	0.0723	-0.3882	-0.3010	0.0789	-0.0816	
ABS (T - M)		-0.1180	-0.4611*	-0.2940	-0.0882	0.3081	0.4140*	
Linear technology trend	TT	0.7320**						

ABS$^+$ = absolute value, M^{++} = the 23-year average of corresponding weather element

* significant at the 5% level, ** significant at the 1% level

TABLE 3

VARIABLES FOR STEPWISE MULTIPLE REGRESSION ANALYSIS

```
MULTIPLE REGRESSION RUN (STEPWISE)
VAR LABELS      VAR01, PRODUTIVIDADE HISTORICA/ VAR02, RAD SOLAR EM OUT/
                VAR03 , RAD SOLAR EM NOV/ VAR04, RAD SOLAR EM DEZ/ VAR05, RAD
                SOLAR EM JAN/ VAR06, RAD SOLAR EM FEV/ VAR07, RAD SOLAR EM MARC/
                VAR08, EVAP EM OUT/ VAR09, EVAP EM NOV/ VAR10, EVAP EM DEZ/
                VAR11, EVAP EM JAN/ VAR12, EVAP EM FEV/ VAR13, EVAP EM MARC/
                VAR14, UMID EM OUT/ VAR15, UMID EM NOV/ VAR16, UMID EM DEZ/
                VAR17, UMID EM JAN/ VAR18, UMID EM FEV/ VAR19, UMID EM MARC/
                VAR20,PO/VAR21,PN/VAR22,PD/VAR23,PJ/VAR24,PF/VAR25,PM/VAR26,TO/
                VAR27,TN/VAR28,TD/VAR29,TJ/VAR30,TF/VAR31,TM/VAR32,TT/VAR33,PSP/
DOCUMENT        DADOS OBTIDOS APARTIR DE 1956 ATE 1978 PARA O SOJA
COMPUTE         VAR34=VAR08+VAR09+VAR10+VAR11+VAR12+VAR13
COMPUTE         VAR35=VAR14+VAR15+VAR16+VAR17+VAR18+VAR19
COMPUTE         VAR36=VAR05-529.65
COMPUTE         VAR37=ABS(VAR36)
COMPUTE         VAR38=VAR07-505.91
COMPUTE         VAR39=ABS(VAR38)
COMPUTE         VAR40=VAR08-159.07
COMPUTE         VAR41=ABS(VAR40)
COMPUTE         VAR42=VAR17-77.30
COMPUTE         VAR43=ABS(VAR42)
COMPUTE         VAR44=VAR27-23.01
COMPUTE         VAR45=ABS(VAR44)
COMPUTE         VAR46=VAR31-23.10
COMPUTE         VAR47=ABS(VAR46)
```

TABLE 4

SUMMARY TABLE OF STEPWISE MULTIPLE REGRESSION ANALYSIS

SUMMARY TABLE

VARIABLE		MULTIPLE R	R SQUARE	RSQ CHANGE	SIMPLE R	B	BETA
VAR32	TT	0.73207	0.53593	0.53593	0.73207	17.68756	0.53195
VAR43		0.82906	0.68735	0.15142	-0.56993	-8.89090	-0.08458
VAR10	EVAP FM DEZ	0.89591	0.80266	0.11531	-0.59079	-0.09640	-0.01654
VAR09	EVAP FM NOV	0.91040	0.82883	0.02617	-0.44297	3.02548	0.53427
VAR12	EVAP FM FEV	0.92549	0.85653	0.02770	-0.51835	-2.18887	-0.21883
VAR28	TD	0.93400	0.87235	0.01582	-0.41663	-29.16938	-0.10563
VAR35		0.94009	0.88377	0.01142	0.49949	5.54180	0.37532
VAR11	EVAP FM JAN	0.95188	0.90607	0.02230	-0.51283	8.22048	0.86053
VAR41		0.96323	0.92781	0.02174	-0.42168	-3.26062	-0.34175
VAR45		0.96543	0.93206	0.00425	-0.46112	-58.89232	-0.15574
VAR47		0.96839	0.93779	0.00573	0.41408	96.65939	0.20418
VAR37		0.97262	0.94598	0.00819	-0.56018	-2.60220	-0.38908
VAR34		0.97586	0.95230	0.00632	-0.66412	-0.74598	-0.48114
(CONSTANT)						-830.75087	

TABLE 5

ANALYSIS OF VARIANCE OF REGRESSION MODEL Y = a+b (VAR 32) + c (VAR 43) + d (VAR 10)

ANALYSIS OF VARIANCE	DF	SUM OF SQUARES	MEAN SQUARE	F
REGRESSION	3.	898053.49671	299351.16557	25.75992
RESIDUAL	19.	220795.39018	11620.81001	

MULTIPLE R	0.89591
R SQUARE	0.80266
STANDARD ERROR	107.79986

---------------- VARIABLES IN THE EQUATION ----------------

VARIABLE	B	BETA	STD ERROR B	F
VAR32	14.77513	0.44436	3.94902	13.999
VAR43	-46.46224	-0.44202	11.17883	17.275
VAR10	-2.22458	-0.38158	0.66765	11.102
(CONSTANT)	1554.14736			

------------- VARIABLES NOT IN THE EQUATION -------------

VARIABLE	BETA IN	PARTIAL	TOLERANCE	F
VAR08	-0.04745	-0.06885	0.41555	0.086
VAR09	0.21196	0.36416	0.58249	2.752
VAR11	0.01563	0.02791	0.62873	0.014
VAR12	-0.16847	-0.32711	0.74400	2.157
VAR28	0.13169	0.16664	0.31598	0.514
VAR34	0.03766	0.04653	0.30115	0.039
VAR35	0.07491	0.12678	0.56525	0.294
VAR37	-0.04929	-0.06765	0.37176	0.083
VAR41	-0.09710	-0.17935	0.67326	0.598
VAR45	-0.03126	-0.05496	0.60997	0.055
VAR47	0.05817	0.11593	0.78385	0.245

TABLE 6

SOYBEAN YIELD ESTIMATIONS BY IEA (INSTITUTO DE ECONOMIA AGRÍCOLA) AND REGRESSION MODEL* IN DIRA-RIBEIRÃO PRETO (1956~1978)

SEQNUM	ESTIMATED YIELD (KG/HA) I E A	REGRESSION MODEL	RESIDUAL
1	1047.310	956.1542	91.15579
2	1251.110	1165.684	85.42622
3	1227.060	1218.874	8.185957
4	1078.510	1073.199	5.310949
5	1160.810	1263.431	-102.6206
6	1094.540	1150.800	-56.26024
7	1242.490	1405.380	-162.8900
8	1343.940	1371.156	-27.21649
9	1030.360	999.4521	30.90793
10	1434.150	1382.942	51.20841
11	1509.960	1440.620	69.34036
12	1426.730	1283.720	143.0098
13	1297.360	1466.731	-169.3715
14	1285.580	1284.757	0.8228012
15	1470.530	1471.659	-1.129005
16	1038.640	1210.451	-171.8111
17	1752.000	1594.985	157.0150
18	1481.480	1545.438	-63.95828
19	1468.440	1458.110	10.32979
20	1586.010	1519.691	66.31907
21	1826.090	1644.043	182.0466
22	1530.300	1608.485	-78.18511
23	1572.620	1640.256	-67.63647

*Estimated Yield (kg/ha) = 1556.14 + 14.77 (VAR 32) - 46.46 (VAR 43) - 2.22 (VAR 10)

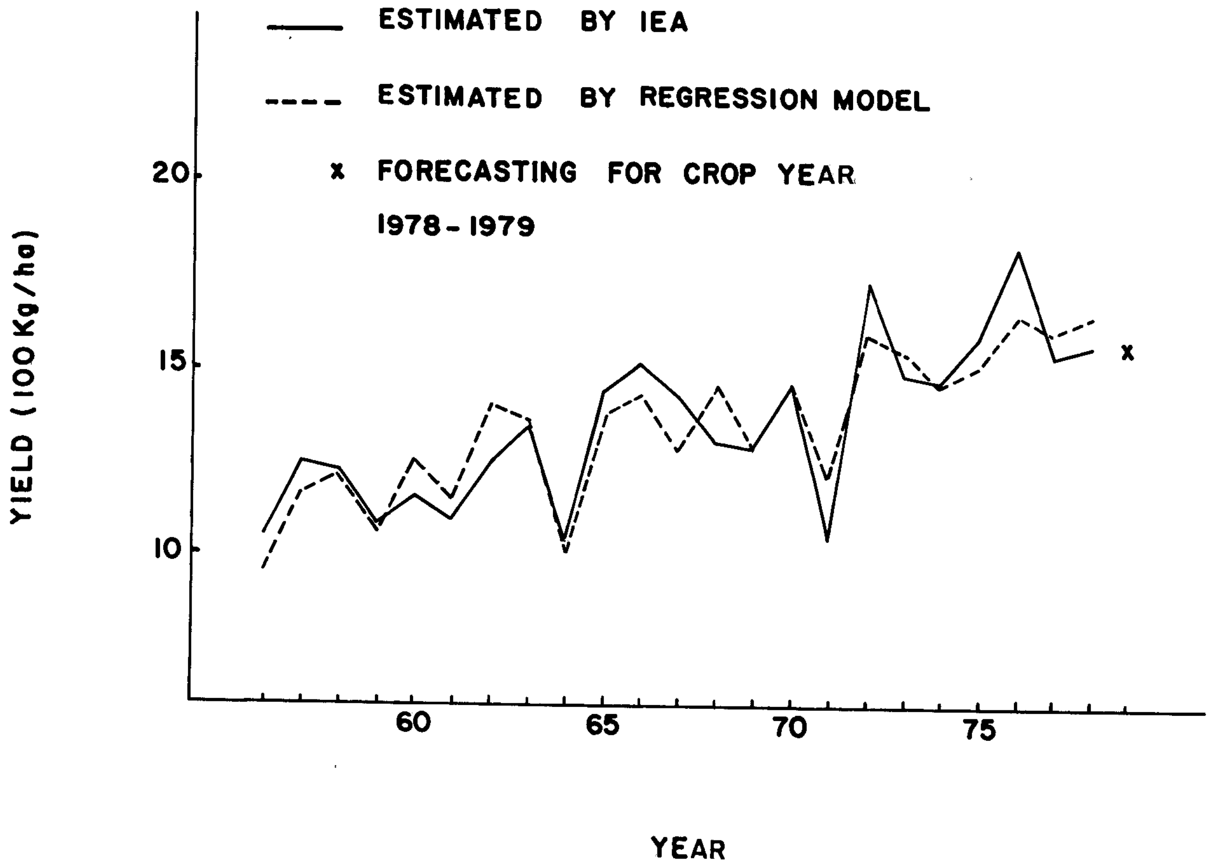

Fig.1 - Soybean yield estimates by IEA (Instituto de Economia Agrícola) and regression model in DIRA-Ribeirão Preto (1956-1978).

REFERENCES

BATISTA, G.T.; MENDONÇA, F.J.; LEE, D.C.L.; TARDIN, A.T.; CHEN, S.C.; NOVAES, R.A.; *Uso de sensores remotos a bordo de satélite e aeronave na identificação e avaliação de áreas de culturas para fins de previsão de safras*. São José dos Campos, INPE, 1978. (1229-NTI/103).

CERVELLINI, A.; SALATI, E.; GODOY, H. *Estimativa da distribuição da energia solar no Estado de São Paulo.* Bragantia 25(3): 31-40, 1966.

CHEN, S.C.; FONSECA da L.B.; *Corn yield model for Ribeirão Preto,São Paulo State,Brazil.* Agricultural Meteorology (in press).

MACDONALD, R.B.; MALL, F.G.; *LACIE - an experiment in global crop forecasting.* NASA report 1978 (JSC-13747).

RUNGE, E.C.A.; ODELL, R.F.; *The relationship between precipitation, temperature and yield of soybeans on the agronomy south farm, Urbana Illinois.* Agronomy Journal 52:245-247, 1960.

THOMPSOM, L.M.; *Weather and technology in the production of corn in the U.S. Cornbelt.* Agronomy Journal 61:453-456, 1969.

THOMPSOM, L.M.; *Weather and technology in the production of soybeans in the central United States.* Agronomy Journal 62:232-236, 1970.

TABLE 2

CORRELATION COEFFICIENTS OF SOYBEAN YIELD WITH WEATHER OR TECHNOLOGY VARIABLES
(DATA BASE = 1956 - 1978)

VARIABLE	SYMBOL	MONTH						
		O	N	D	J	F	M	O~M
Solar radiation (Cal/cm^2/day)	SR	0.1761	-0.1021	-0.3832	-0.2709	-0.0831	-0.1047	-0.1366
Total evaporation (mm)	E	-0.5783**	-0.4429*	-0.5907**	-0.5128*	-0.5183*	-0.2268	-0.6641**
Relative humidity (%)	RH	0.2753	0.3397	0.4009	0.2112	0.1725	0.1150	0.4994*
Total precipitation (mm)	P	-0.0316	0.2768	0.2833	0.0503	0.1787	-0.0081	-0.3617
Mean temperature (°C)	T	-0.3142	-0.2056	-0.4166*	-0.2151	-0.1551	0.1497	-0.3430
ABS$^+$ (SR - M^{++})		-0.4064	-0.1336	0.0535	-0.5601**	-0.0293	-0.2083	
ABS (E - M)		-0.4216*	-0.0874	-0.1936	-0.3270	-0.0123	0.1704	
ABS (RH - M)		-0.2486	0.0095	-0.1498	-0.5699**	0.0328	-0.0633	
ABS (P - M)		-0.2290	0.0723	-0.3882	-0.3010	0.0789	-0.0816	
ABS (T - M)		-0.1180	-0.4611*	-0.2940	-0.0882	0.3081	0.4140*	
Linear technology trend	TT	0.7320**						

ABS$^+$ = absolute value, M^{++} = the 23-year average of corresponding weather element

* significant at the 5% level, ** significant at the 1% level