

General Disclaimer

One or more of the Following Statements may affect this Document

- This document has been reproduced from the best copy furnished by the organizational source. It is being released in the interest of making available as much information as possible.
- This document may contain data, which exceeds the sheet parameters. It was furnished in this condition by the organizational source and is the best copy available.
- This document may contain tone-on-tone or color graphs, charts and/or pictures, which have been reproduced in black and white.
- This document is paginated as submitted by the original source.
- Portions of this document are not fully legible due to the historical nature of some of the material. However, it is the best reproduction available from the original submission.

"Made available under NASA sponsorship
in the interest of early and wide dis-
semination of Earth Resources Survey
Program information and without liability
for any use made thereof."

LACIE-00505

JSC-11705

7.9-10129
CR-158138

LARGE AREA CROP INVENTORY EXPERIMENT (LACIE)



NASA NOAA USDA

A WHEAT YIELD MODEL FOR PUNJAB, INDIA

(E79-10129) LARGE AREA CROP INVENTORY
EXPERIMENT (LACIE). A WHEAT YIELD MODEL FOR
PUNJAB, INDIA (National Oceanic and
Atmospheric Administration) 34 p HC A03/HP
A01

N79-19426

Unclas
00129

CSCL 02C 63/43

NASA

National Aeronautics and
Space Administration

Lyndon B. Johnson Space Center
Houston, Texas 77058

JANUARY 1978

CENTER FOR CLIMATIC AND
ENVIRONMENTAL ASSESSMENT

Technical Note 78-2

A Wheat Yield Model for Punjab, India

by

Michael H. Procter

and

D. E. Umberger

Authorized by: *Jerry D. Hill*
Jerry D. Hill
Assistant Deputy Project Manager

Center for Climatic and
Environmental Assessment
Federal Building, Room 116
Columbia, Missouri 65201

January, 1978

A WHEAT YIELD REGRESSION MODEL

FOR PUNJAB, INDIA

by

Michael H. Procter and D. E. Umberger^{1/}

INTRODUCTION

This document discusses the results of efforts to use time series regression analysis on available historical weather and yield data for the purpose of estimating a model to forecast wheat yields in the State of Punjab, India.

PUNJAB BACKGROUND

Punjab is located in the northwestern corner of India, bounded by Pakistan on the west, the States of Rajasthan and Haryana on the south, Himachal Pradesh on the east, and Jammu and Kashmir on the north (Fig. 1). Punjab is one of India's smaller states by several measures--representing 1.6 percent of the geographical area, 2.6 percent of the cropped area, and 2.5 percent of the population; nevertheless Punjab produces about one-fourth of India's wheat (12, p. 21).

^{1/} Agricultural Economists, USDA LACIE Project Office, Columbia, Missouri.

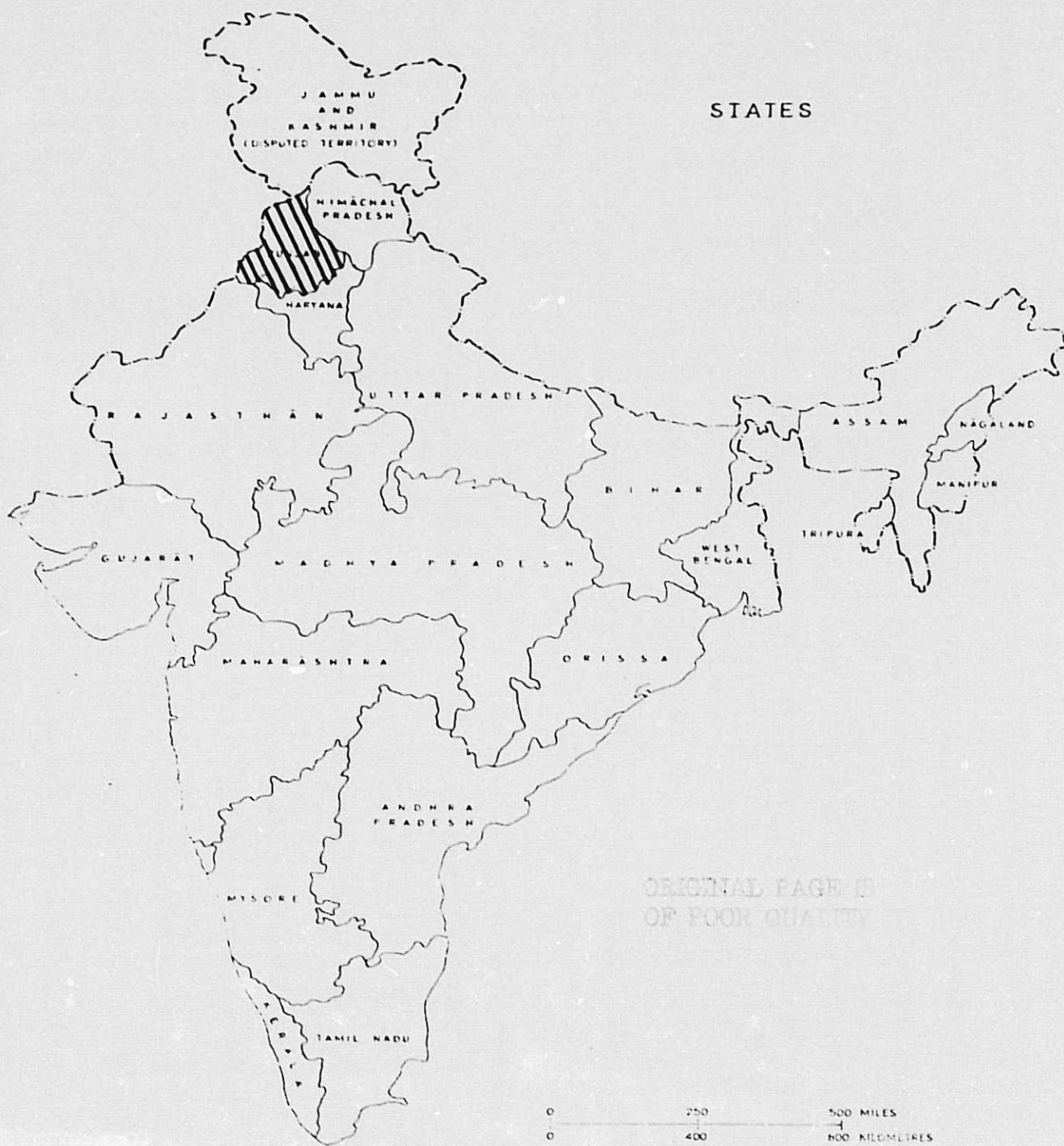


Figure 1. Map of India Showing Location of Modeled Area.

Since 1947, Punjab's territorial boundaries have changed considerably. In 1947, more than half of the original state was allocated to the newly formed country of Pakistan. In 1956 the territory of Pepsu was merged with Punjab. Then, in 1966, Punjab was reorganized once again, this time on a linguistic basis. The State of Haryana and the union territory of Chandigarh were carved out of Punjab, and certain hill areas were transferred to Himachal Pradesh (12). These reorganizations have made analyses requiring comparisons over a period of time difficult since consistent statistical data in a long time series are often unavailable for the current State of Punjab.

Geology and Physiography

More than 90 percent of Punjab's area is a flat plain lying between 180 and 320 meters above sea level, formed by the deposition of alluvium, giving it deep and relatively fertile soils (12, pp. 22-29). Several rivers which originate in the Siwalik Hills and the Himalayas run through this region. The Siwalik Hills along the eastern border comprise about 10 percent of the state's area. They are composed of conglomerate, clay and silt soils, having the character of fluvial deposits. Water erosion has greatly modified the land surface in this area.

Punjab is partially drained by three rivers--Ravi, Beas, and Sutlej--whose flows originate in the high snow-clad mountains and are perennial in character. The lack of a deep, well-defined drainage channel in the plains region makes these rivers subject to frequent and extensive flooding. There are also scores of seasonal streams originating in the Siwalik Hills. Most of these streams are dry except for a

few months of the rainy season.

The Climate

Punjab is a sub-humid to semi-arid region. The average annual rainfall in Punjab varies from about 25 centimeters on the western side to over 100 centimeters near the Siwalik Hills (12). The highly seasonal occurrence of these rains is extremely important to crop production. The monsoon winds normally bring 70-80 percent of the annual rainfall during July, August, and September (Table 1). Occasionally, the monsoon carries into October. Most of the remaining rainfall comes during the winter months. The rainfall is highly variable over time, with the variability increasing in the winter months, when wheat is grown. Though the mean daily temperature almost never drops below 0°C, frost is common during winter nights and freeze sometimes damages the wheat crop. The highest mean monthly temperature occurs in June prior to the monsoon season (Table 2).

Cropping Practices

Intensive use of land is emphasized in Punjab. Over three-fourths of the area is under cultivation. This is largely possible because the Punjab plain is free from physical handicaps and irrigation is used to make up for the deficiency in rainfall. More than one-fourth of the net sown area produces more than one crop per year. Only in the districts of Ferozepur and Amritsar, in the relatively dry eastern region, is fallowing an important practice (12, pp. 27).

Cropping patterns vary within the state depending on local soil conditions and availability of moisture. About three-fifths of all crops

Table 1. Monthly Precipitation Data for Punjab, 1952-1976.

Year	MONTH											
	JAN	FEB	MARCH	APRIL	MAY	JUNE	JULY	AUG	SEPT	OCT	NOV	DEC
	-----millimeters-----											
1952	32	34	54	2	14	74	154	327	9	0	0	4
1953	77	3	0	10	2	70	357	244	92	2	2	13
1954	64	106	12	0	4	40	270	81	307	19	0	0
1955	41	3	31	27	23	35	189	264	184	407	0	8
1956	32	11	77	11	1	87	256	292	23	133	2	6
1957	147	10	44	15	12	31	203	277	108	24	15	44
1958	12	5	14	9	4	30	252	204	365	42	0	58
1959	49	31	14	4	27	36	260	331	222	18	33	0
1960	23	0	31	5	2	67	279	264	6	0	0	25
1961	72	52	7	21	5	58	273	313	144	24	8	13
1962	45	29	56	3	3	41	198	189	441	1	24	12
1963	0	12	50	5	25	32	141	259	65	1	12	28
1964	26	5	7	13	30	28	364	278	185	0	0	20
1965	22	36	25	46	50	4	250	76	9	10	8	1
1966	0	45	23	7	34	110	190	215	162	18	0	5
1967	0	21	93	2	4	11	148	318	802	13	11	67
1968	101	153	42	23	30	42	284	203	0	23	0	5
1969	7	26	29	15	24	12	201	99	162	0	12	0
1970	39	.17	9	1	14	105	78	263	129	5	0	0
1971	5	42	3	6	40	69	169	259	53	1	8	0
1972	15	13	12	6	2	51	190	147	37	3	9	14
1973	21	39	8	4	56	114	187	240	54	27	0	38
1974	2	7	2	0	10	65	85	136	28	0	0	10
1975	11	32	19	2	5	62	242	239	120	0	3	0
1976	30	25	2	13	30	45	162	365	165	15	0	0
Mean, 1952-76	35	30	27	10	18	53	215	235	155	31	6	15

Table 2. Monthly Mean Temperature Data for Punjab, 1952-1976.

Year	MONTH											
	JAN	FEB	MARCH	APRIL	MAY	JUNE	JULY	AUG	SEPT	OCT	NOV	DEC
	-----Degree Centigrade-----											
1952	15.1	18.3	21.7	29.8	33.1	32.9	31.1	29.4	30.3	25.7	19.5	15.0
1953	12.4	17.0	23.4	27.7	32.5	33.8	30.5	29.5	28.5	25.2	18.8	16.5
1954	12.0	16.1	20.1	27.1	32.5	33.9	30.9	31.1	28.7	22.8	18.4	13.8
1955	12.3	15.6	22.3	24.4	28.4	33.6	31.0	28.8	27.8	23.2	18.4	14.0
1956	12.7	15.8	20.2	26.9	33.5	32.5	29.2	28.9	29.7	23.7	18.0	14.6
1957	12.4	13.8	19.6	24.6	29.0	32.3	31.5	29.2	27.6	24.6	18.9	14.3
1958	14.3	15.6	21.5	28.9	30.9	33.8	30.7	29.8	27.8	24.7	19.0	14.4
1959	12.4	13.9	20.8	26.6	30.4	33.8	30.5	29.5	28.7	25.9	18.7	14.1
1960	12.1	17.7	19.4	25.3	31.8	34.1	30.8	29.9	29.5	24.9	18.0	13.8
1961	13.3	13.3	20.7	25.6	30.7	33.6	30.6	29.6	29.3	24.6	17.1	11.8
1962	11.4	15.8	19.6	27.1	30.8	33.0	31.4	29.9	26.8	23.8	17.9	13.2
1963	11.7	16.6	19.7	25.6	29.3	33.6	31.3	29.0	28.1	25.7	19.5	13.8
1964	10.6	14.4	21.4	26.8	29.0	33.2	29.0	29.7	28.0	24.3	17.5	13.4
1965	13.9	14.8	19.6	23.4	28.8	33.8	30.1	29.3	28.9	25.6	20.1	13.8
1966	13.7	17.4	20.5	26.0	30.8	32.4	30.7	30.4	27.9	24.5	18.1	13.0
1967	11.8	17.0	19.4	25.2	29.8	34.3	30.8	29.4	28.4	24.6	18.5	13.9
1968	11.6	13.6	19.9	26.2	29.2	33.9	30.6	29.6	30.6	24.0	18.8	13.9
1969	15.6	15.4	22.1	25.1	31.6	33.9	31.1	30.5	29.0	25.8	19.8	13.9
1970	13.2	15.1	19.6	27.3	31.8	31.7	32.1	29.6	29.4	25.8	17.8	13.7
1971	11.8	15.9	21.8	29.2	30.5	33.0	30.6	29.4	28.6	25.2	19.3	14.6
1972	13.2	14.4	21.0	25.4	31.8	33.8	31.6	30.0	28.2	24.2	19.4	14.5
1973	12.8	16.8	20.0	28.4	32.3	34.0	31.6	30.0	30.0	25.0	18.6	11.8
1974	12.0	13.8	21.8	28.4	31.6	32.5	31.2	30.8	30.0	24.4	18.5	12.8
1975	12.0	14.0	19.8	25.5	32.4	32.2	28.8	30.0	28.2	25.4	18.1	14.1
1976	13.7	15.4	18.6	24.0	31.5	33.0	30.7	28.5	28.4	25.3	18.3	14.2
Mean 1952-76	12.8	15.5	20.6	26.4	31.0	33.3	30.7	29.7	28.7	24.8	18.6	13.9

are grown during the winter season. Cereals account for three-fifths of total crop acreage. Wheat is the dominant crop, accounting for about two-fifths of the annually cropped area, but maize, rice and bajra are important. Wheat is grown in the winter season, being planted in October and November and harvested in late April and May.

Punjab plays a critical role in India's wheat economy. Normally, enough wheat is produced in Punjab to satisfy the demands of the local population and provide a surplus for the rest of India. For example, in 1970-71, Punjab contributed two-thirds of the national wheat surplus (12). In 1977 Punjab produced 23 percent of India's wheat production of 26 million metric tons (calculated from table 3). For the period 1967-77 Punjab produced an average of 22 percent of India's wheat production. This production required an average 13 percent of India's harvested acreage, giving Punjab the highest average yield of any wheat growing state in India.

WHEAT YIELD TRENDS

In the mid 1970's Punjab wheat yields are more than twice those of the 1950's (Table 3). The most rapid rate of advance in yields occurred from 1966 to 1972. Since 1972, the annual rate of yield increase appears to be much slower, but the 1977 yield estimate is still a record.

Year to year variations in yields have been irregular but increases have outweighed declines by almost 2 to 1, the average increase being 1.5 quintals/hectare while the average decrease is only .84 quintals/hectare. Alternatively expressed, increases have been as

Table 3. Area, Yield, and Production of Wheat in India and Punjab, 1952/53-1976/77.^{1/}

Year	India			Punjab		
	Area	Yield	Production	Area	Yield	Production
	(1,000 Ha)	(T/Ha)	(1,000 MT)	(1,000 Ha)	(T/Ha)	(1,000 MT)
1952/53	9,828	.763	7,502	1,120	1.098	1,230
1953/54	10,681	.750	8,017	1,162	1.093	1,270
1954/55	11,136	.803	8,913	1,236	1.086	1,342
1955/56	12,297	.708	8,707	1,321	.880	1,162
1956/57	13,590	.699	9,463	1,344	1.024	1,376
1957/58	11,858	.666	7,893	1,329	.998	1,327
1958/59	12,603	.786	9,934	1,479	1.075	1,590
1959/60	13,170	.780	10,252	1,432	1.046	1,498
1960/61	12,960	.847	10,993	1,438	1.243	1,788
1961/62	13,521	.794	10,752	1,465	1.236	1,811
1962/63	13,590	.793	10,776	1,559	1.211	1,888
1963/64	13,498	.730	9,853	1,543	1.254	1,934
1964/65	13,422	.913	12,257	1,625	1.495	2,428
1965/66	12,656	.824	10,424	1,527	1.253	1,913
1966/67	12,838	.887	11,393	1,606	1.520	2,441
1967/68	14,998	1.103	16,540	1,804	1.858	3,352
1968/69	15,958	1.169	18,651	2,086	2.167	4,520
1969/70	16,626	1.209	20,098	2,162	2.220	4,800
1970/71	18,240	1.307	23,833	2,299	2.380	5,145
1971/72	19,139	1.380	26,410	2,335	2.405	5,618
1972/73	19,881	1.271	24,735	2,401	2.233	5,368
1973/74	18,583	1.172	21,778	2,338	2.216	5,181
1974/75	18,010	1.338	24,104	2,213	2.395	5,300
1975/76	20,112	1.409	28,336	2,402	2.375	5,705
1976/77 ^{2/}	19,800	1.313	26,000	2,470	2.401	5,930

^{1/} For 1952/53-1965/66 the source of the India data was USDA, ERS Wheat Situation, various issues and the Punjab data was summarized from (5). For 1966/67-1976/77 data for India and Punjab were taken from a memo dated June 21, 1977 from John Parker of USDA/ERS. His source was IN 7037 dated May 13, 1977 from the Agricultural Attache in New Delhi.

^{2/} USDA, ERS estimate

much as 25 percent while decreases have been no greater than 18 percent. In 16 of the 24 years changes in wheat yields from the previous year were less than ten percent. Five of the eight largest deviations from the previous year occurred at a time when technology factors were important--1965 to 1969.

TECHNOLOGY AND YIELDS

The factors responsible for the rapid increase in yields are largely independent of weather. In fact, Punjab yields reached their highest levels during a period (1969-75) when average annual precipitation was only 42 to 77 percent of the 1953-76 average. Three factors-- an increasing share of irrigated acreage, the adaptation of high yielding wheat varieties, and increased usage of commercial fertilizers-- are primarily responsible for the rapid rise in yields since 1965-66.

Improved Varieties

Punjab farmers rapidly adopted the high yielding dwarf varieties developed from strains that were received in 1963 from Dr. E. N. Borlaug in Mexico (12, pp. 68). From essentially no commercial plantings in 1965, Punjab farmers increased the share of high yielding dwarf wheat to total wheat acreage harvested to almost 90 percent in 1976 (Table 4).

Exploiting the genetic yield potential of the dwarf varieties requires sound agronomic practices. Without sufficient water, a high level of soil fertility, and pest and disease controls, the dwarf wheats yield little more than traditional varieties. Because these traditional wheat varieties are preferred by Indian consumers, the

Table 4 -Percentage of total wheat area planted to high yielding varieties, Punjab and All India, 1967-1977.

Year ending March 31 ^{1/}	Percentage of total wheat area planted to high yielding varieties	
	Punjab	: All India
1967	3.7	4.2
1968	35.4	19.6
1969	48.5	30.0
1970	69.5	30.1
1971	69.1	35.9
1972	72.6	41.1
1973	78.5	51.2
1974	84.3	58.7
1975	88.4	62.3
1976	85.8	67.9
1977	86.0	75.7

Source: John Parker, USDA, ERS Highlight, "India's 1977 Wheat Harvest Estimated at 26 Million Tons", (Washington, D.C.:June 21, 1977), p. 3.

^{1/} Minimal in 1966 and zero prior to 1966.

dwarf wheats are price discounted in the local markets. Thus, before farmers adopted the new dwarf varieties, potential yield increases had to be large enough to offset the price differential and the costs of commercial inputs.

Irrigation

Punjab farmers were already practicing a fair amount of irrigation in 1965 (Table 5). Thus, one of the necessary conditions for realizing the high yielding capability of the dwarf wheats was already available. Moreover, the availability of dwarf wheats enhanced the potential benefits from irrigation and Punjab farmers increased the share of wheat acreage harvested from irrigated land from 50 percent in 1965 to over 80 percent in 1969--the latest year for which data is available (Table 5).

Wells and tube-wells supply water for 55 percent of the irrigated acreage with canals providing the remainder (12, p. 105). Although the number of wells in Punjab exceeds 200,000, private tube-wells are becoming increasingly important--exceeding 232,000 in 1972 (12, pp. 108-9). Most of western Punjab is blessed with a plentiful supply of underground water. Most tube-wells are relatively small and shallow, providing underground water for drinking as well as for irrigation. Most are farmer owned, making them a more reliable source of supply for the individual than the government owned canals. Because seven to eight irrigations applied at particular stages of plant growth, e.g., tillering, grain formation, and grain filling are required for maximum yields from the Mexican varieties, supply reliability is important. Since tube-wells are close

Table 5-Percentage of total wheat area under irrigation, Punjab, India, 1953-1969 1/

Year ending March 31	:	Percentage of total wheat area irrigated
1953	:	58.80
1954	:	63.65
1955	:	61.42
1956	:	57.22
1957	:	56.11
1958	:	56.31
1959	:	54.32
1960	:	54.69
1961	:	57.65
1962	:	59.08
1963	:	57.98
1964	:	62.02
1965	:	62.27
1966	:	72.31
1967	:	75.99
1968	:	75.01
1969	:	83.09

1/ Aggregated from district data reported in Bulletin on Wheat Statistics in India (Districtwise), India Directorate of Economics and Statistics, Ministry of Agriculture, (New Delhi:1972), pp. 50-65.

to the fields being irrigated, percolation losses and the need for an extensive water distribution system are reduced relative to canals.

Increases in the number of tube-wells appears to be limited more by a lack of power for pumping than by an inadequate supply of water (12, pp. 131-139). In 1969 there were over 53,800 pending applications for new electrical hook-ups--the major mechanical power source for pumping. The per capita electric consumption for irrigation increased from 0.03 KWH in 1951 to 10.86 KWH in 1968 (12, p. 131). "The phenomenal increase in electrical consumption in the agricultural sector is attributed to the fact that the need to extend irrigational facilities became very acute with the introduction of the high-yielding dwarf wheat varieties (12, p. 133)."

Fertilizer

Given adequate soil moisture the dwarf wheats are highly responsive to nitrogenous fertilizers. Alternatively, a high level of soil fertility is required to achieve the potential yields of the dwarf wheats. Farmyard and green manures generally cannot supply the nitrogen and phosphorus needs for obtaining high yields. Commercial fertilizers are used to meet the bulk of nutrient needs. Although statistics on fertilizer use are not available by crop in Punjab, total consumption of fertilizers including manures is estimated to have increased from 5,000 tons in 1950-51 to 856,000 in 1969-70 (12, pp. 124-5). Of the 1969-70 total, applied nitrogen (N) accounted for 144,000 tons, phosphorus (P_2O_5) 28.5 tons, and potassium (K_2O) 9,000 tons. By 1973-74 consumption of N was 242,000 tons, P_2O_5 was 73,000 tons and K_2O was 23,000 tons (6, pp. I-72).

Although rising rapidly, commercial fertilizer use remained less than one half of estimated needs. Thus, substantial wheat yield increases remain possible in Punjab.

The increased wheat yields were achieved through a combination of improved practices. High yielding dwarf wheat, fertilizer, and irrigation have played complementary roles in increasing Punjab wheat yields. Other factors have also contributed. Increased efforts have also been devoted to evolving varieties resistant to pests and diseases. Use of chemicals as a means of pest control has also increased in Punjab.

PREVIOUS STUDIES

Two previous studies seeking to explain yield/weather relationships at the state level for Punjab were reviewed. Dayal (4), working with yield and precipitation data from 1948/1949 to 1960/61 period, tested a model of the form:

$$\hat{Y} = a + b_1X_1 + b_2X_2 + b_3X_3 + b_4X_4 + rT$$

where Y = yield per acre,

X_1, \dots, X_4 = Average rainfall in the months of

December, January, February and March,

respectively and

T = time variable

Rainfall in January and February was found to have a positive effect on yield, while December and March precipitation entered the estimated equation with negative coefficients. Though estimates were generally

consistent with expectations, only the trend coefficient (a proxy for technological change) was significantly different from zero. A semi-log transformation improved variable significance somewhat, but overall explanatory power was still considered relatively low ($R^2 = .75$).

Arguing that monthly data might be too highly aggregated for effective analysis, the author attempted to estimate yield from 13 years of fortnightly precipitation data. While the reestimated equation still had relatively low explanatory power ($R^2 = .74$), Dayal observed that his results showing positive effects of rainfall on yield from the second fortnight in December through the second fortnight in February and negative effects in the first fortnight in December and in March "by and large confirm the views of agronomists" (4, p. 53).

Das (3), working with a considerably more extensive dataset, reported improved results in his study. Using yield, precipitation, maximum temperature, minimum temperature, and humidity data from 1918 through 1965 (with 1966 through 1970 data reserved for model validation), Das employed linear correlation methods to determine, from all possible seven to ninety day periods in the growing season, those periods in which the weather variables had the most significant impact on yield. Results for the final equation selected are shown in Table 6. All coefficients are significantly different from zero at the 5 percent level. Occasions of daily temperature less than 42°F appears to be the most important weather variable. Results of extra-sample prediction tests (1966-1970) are shown in Table 7. They appear to be remarkably good considering the rather dramatic technological changes during this period.

Table 6. Das Yield Regression Equations

Regression Equations ^{1/}	R ₂	CV
<p>1. Using 1918-1965 data</p> $\hat{Y} = 1083.338 + 4.330X_2^* - 4.591X_3$ <p style="text-align: center;">(2.36) (2.23)</p> $- 10.872X_4 + 3.969X_5$ <p style="text-align: center;">(3.76) (1.99)</p> $+ 6.557X_6 - 10.124X_7 + 11.025X_8$ <p style="text-align: center;">(6.41) (5.01) (10.45)</p>	86.2	15%
<p>2. Using 1949-1965 data</p> $\hat{Y} = 1164.272 + 4.114X_2 - 9.041X_3$ <p style="text-align: center;">(1.64) (1.67)</p> $+ 9.138X_4 - 11.814X_5 + 7.053X_6$ <p style="text-align: center;">(1.73) (2.62) (5.17)</p> $- 7.163X_7 + 9.808X_8$ <p style="text-align: center;">(1.97) (6.83)</p>	87.9	11%

Y = Yield (pounds/acre)

X₂ = Rainfall (in.), Sept. 13 - Oct. 16

X₃ = Mean maximum temperature (°F) Dec. 22 - Jan. 20

X₄ = Mean minimum temperature (°F) Dec. 2 - Dec. 29

X₅ = Mean minimum temperature (°F) Jan. 24 - Feb. 3

X₆ = Mean relative humidity (0830 hrs) Feb. 8 - Feb. 24

X₇ = Occasions of temperature less than 42°F

X₈ = Technological trend

^{1/} t-values are in parenthesis

Table 7. Results of Extra-sample Prediction Tests

Year	Reported Yield	Estimated Yield Equation 1	% Difference	Estimated Yield Equation 2	% Difference
1966	1102	1116	1	1171	6
1967	1375	1418	3	1376	0
1968	1655	1751	6	1781	8
1969	1931	1871	3	1729	11
1970	1978	2082	5	2156	9

Anticipating criticism that the yield data obtained before 1949 might not be comparable to more recent yield data due to a change in estimating procedures, Das reestimated the equation using post-1949 data only. Regression results and extra-sample predictions are shown in Table 6 and 7. Three variables, mean minimum and mean maximum temperatures, are no longer significant at the five percent level, and, perhaps more importantly, two coefficients--those for the mean minimum temperature during December 2 to December 29 and during January 24 to February 3 exhibit a change in sign. Moreover, the estimated errors associated with extra sample predictions from this model are considerably larger than those from the model employing the complete time series.

While these results can be attributed in part to a loss of information due to the smaller sample size, the overall effect is to cast some doubt on the validity of the model, since only three coefficients (technological trend, mean relative humidity during February 8-22 and temperatures below 42⁰F) remain significant and consistent in sign.

DEVELOPING A REGRESSION MODEL

Because of the limited success of the Das and the Dayal regression models and the need to update their technology terms, an attempt was made to redevelop a predictive Punjab wheat yield model. The linear hypothesis

was:

$$Y = a + \sum_{j=1}^m b_j T_{i,j} + \sum_{k=1}^n c_k W_{i,k} + e_i$$

where

i = year;

j = technology variable, $j = 1, \dots, m$;

k = weather variable, $k = 1, \dots, n$;

Y_i = yield for the i th year;

$T_{i,j}$ = technology variables for i th year, which were limited by lack of data to $T_{i,1}$, the trend for the i th year (1953 = 1, 1954 = 2, ... 1972 = 20, 1973 = 21, 1974 = 21, 1975 = 21, 1976 = 22), and $T_{i,2}$, the proportion of wheat acreage harvested in high yielding dwarf varieties;

$W_{i,k}$ = k th weather variable for the i th year; and

e_i = error term.

DATA CONSTRAINTS

Yield estimates in quintals per hectare (Table 6), were available from USDA for 1952/53 through 1975/76. Data for years after 1966/67 were available from John Parker, ERS, USDA Washington, D.C. and were revised from earlier estimates.

The Technology Variables

The availability of or lack of data on technology factors limited the technology variables to a measure of the proportion of the harvested wheat acreage planted to high yielding dwarf varieties, $T_{i,2}$. The effect of other technology factors was estimated by using time as a surrogate variable. Several alternative formulations were tested but the one selected was $T_{i,1}$. Except for the period 1973-75, this time variable assumes a linear increase in technology. The period 1973-75 was subject to shortages of and a rapid rise in the prices of fuel and fertilizer adversely affecting farmers' usage of fertilizer and irrigation.

The Weather Variables

Aggregated estimates of monthly mean temperature (in tenths of °C) Table 2 and monthly precipitation (mm) Table 1 were available for the present State of Punjab from 1967 onward from India sources. Prior to that time Punjab included the present State of Haryana, and no aggregated meteorological data were separately available for the area comprising the present State of Punjab. Consequently, the mean of the monthly minimum and maximum data for seven individual stations reporting in the Monthly Weather Review of India (7) was constructed by personnel of the Center for Climatic and Environmental Assessment, Columbia, Missouri to estimate monthly mean temperature and monthly precipitation for crop years 1953 through 1967. While coverage was generally good, there were instances when only a few of the stations reported all the information required. The methods and stations used by India personnel to construct the more recent (post 1967) data are unknown but they could differ causing an unknown but probably small inconsistency in the data series.

The basic weather data, consisting of monthly temperature and monthly precipitation were used to derive monthly weather variables consisting of an aridity index, a monthly precipitation departure from normal, and a monthly temperature departure from normal. The normal is the average value for the period 1952-1953 to 1975-76. The aridity index, also expressed as departure from normal, is defined as monthly precipitation minus potential evapotranspiration (P.E.T.) Thornthwaite's procedure (9) was used for estimating P.E.T. Because the index was not included in the final model only the long-term average values are plotted in Figure 2. The need for irrigation is

millimeters

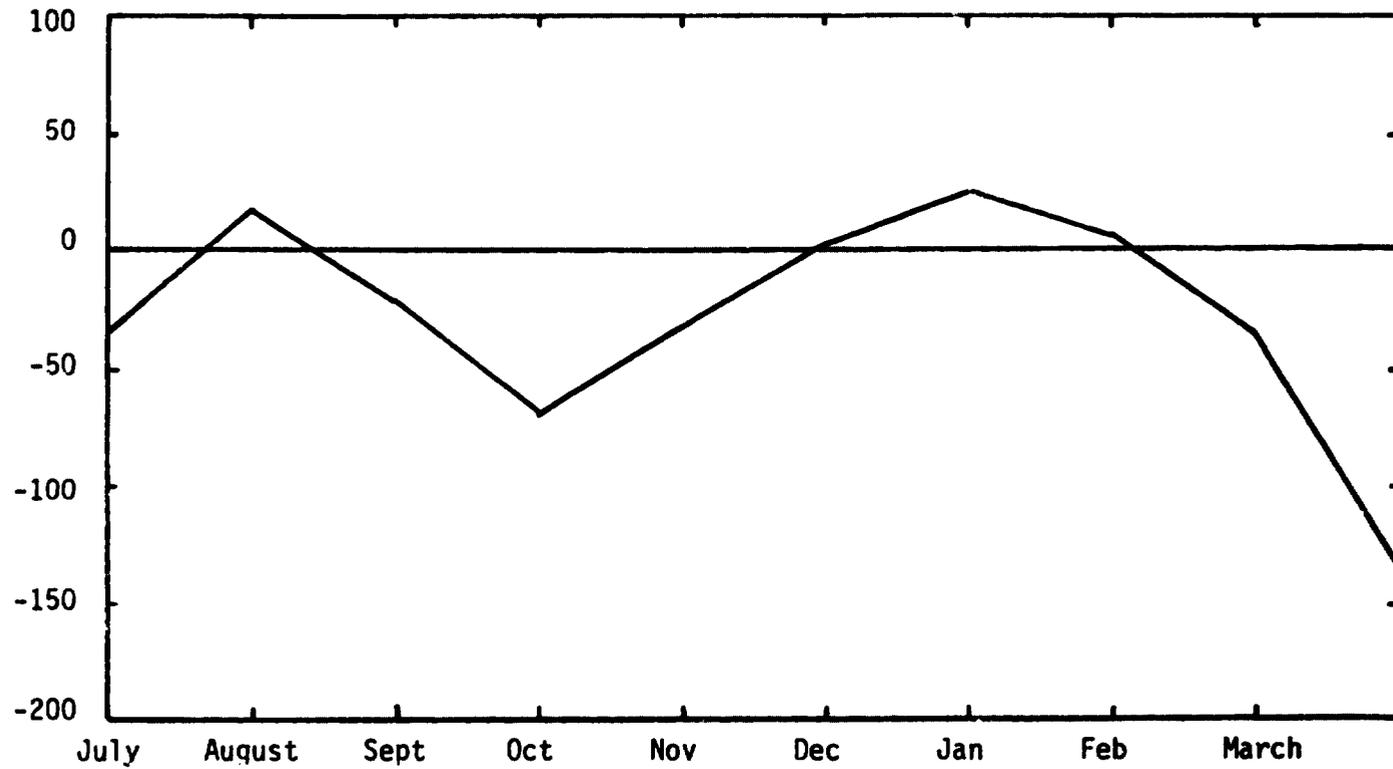


Figure 2. Monthly Average Precipitation minus Potential Evapotranspiration, 1952/53-1975/76.

shown by observing that during most of the year P. E. T. exceeds monthly average precipitation.

RESULTS OF THE ANALYSIS

Selection of Variables

The basic guidelines used in selecting the final model were:

1. The signs of the coefficient are agronomically feasible and significantly different from zero at the .15 level.
2. The final "full season" model explains as much of the yield variability as possible.
3. The variables in each truncation are a time related subset of the variables in the final "full season " model.

The selection of variables was an iterative procedure using the stepwise and maximum R^2 statistical procedures to assist the researchers' judgement concerning the proper variable combinations. Limited agronomic information on the mathematical nature of the causal relationship of weather factors to wheat yields limited the complexity of the mathematical model forms tested. For most weather variables quadratic and linear terms were analyzed; but if the estimated coefficients were not significantly different from zero at the .15 level of confidence, the signs were not in the expected direction, or the predicted yields were extremely sensitive to values of the variable near the end points of the data set, they were omitted from the final model forms.

The selection of variables began with the yield trend. The values of the variables used to develop the models are given in Table 8. Observation of the yield data and a literature review (1,2,10,11,12) indicated the importance of technology factors. In fact, two factors, high yielding varieties and the time surrogate, explained over 97% ($R^2 > .97$) of the variation in the yield terms and reduced the standard error of the fitted equation to less than 1.0 quintals per hectare (Table 9). Hence, little variation remained to be explained by weather variables.

The selection of weather variables began with analyzing pre-season variables such as July through October or November precipitation. Other monthly variables were added if they improved the fit of the equation and met the other statistical and agronomic tests.

An analysis of the aridity index showed that P.E.T. was normally greater than precipitation for every month of the year. Based on this analysis and a suggestion by Mather (9, p. 108) that Thornthwaite's expression for P.E.T. was of questionable accuracy in climates where humidity changes markedly from month to month, the aridity index was dropped after preliminary analysis indicated poor model results.

Monthly precipitation variables also gave poor results. Estimated precipitation coefficients either had wrong signs, had unrealistic values, or were statistically insignificant. For some months a few extreme observations, e.g. abnormally high precipitation values for October in 1955 and 1956, were probably responsible for the wrong signs. Another problem was that precipitation for the 1968-1976 period was consistently below the mean precipitation for the period 1953-1976. The below "normal" consistency in the precipitation data for 1968-1976 introduced a high

Table 8. Values of Punjab Model Variables.

Year	Yield ^{1/}	Constant	Trend Term	Log (HYV) ^{2/}	November Temperature ^{3/}	December Temperature ^{3/}	March Temperature ^{3/}
1953	10.98	1	1	0.0000	1.6	1.7	2.8
1954	10.93	1	2	0.0000	0.9	3.2	-0.5
1955	10.86	1	3	0.0000	0.5	0.5	1.7
1956	8.80	1	4	0.0000	0.5	0.7	-0.4
1957	10.24	1	5	0.0000	0.1	1.3	-1.0
1958	9.98	1	6	0.0000	1.0	1.0	0.9
1959	10.75	1	7	0.0000	1.1	1.1	0.2
1960	10.46	1	8	0.0000	0.8	0.8	-1.2
1961	12.43	1	9	0.0000	0.1	0.5	0.1
1962	12.36	1	10	0.0000	-0.8	-1.5	-1.0
1963	12.11	1	11	0.0000	0.0	-0.1	-0.9
1964	12.53	1	12	0.0000	1.6	0.5	0.8
1965	14.94	1	13	0.0000	-0.4	0.1	-1.0
1966	12.53	1	14	0.0000	2.2	0.5	-0.1
1967	15.44	1	15	1.5476	0.2	-0.3	-1.2
1968	18.58	1	16	3.5946	0.6	0.6	-0.7
1969	21.67	1	17	3.9020	0.9	0.6	1.5
1970	22.20	1	18	4.2556	1.9	0.6	-1.0
1971	23.80	1	19	4.2499	-0.1	0.4	1.2
1972	24.05	1	20	4.2986	1.4	1.3	0.4
1973	22.33	1	21	4.3758	1.5	1.2	-0.6
1974	22.16	1	21	4.4462	0.7	-1.5	1.2
1975	23.95	1	21	4.4931	0.6	-0.5	-0.8
1976	23.75	1	22	4.4636	0.2	0.8	-2.0
1977	24.01	1	23	4.4636	0.4	0.9	1.4

^{1/} Yield data are in quintals per hectare

^{2/} Percentage of total wheat area under high yielding varieties
 $\text{Log (HYV)} = \log \text{transformation [Natural log (\% HYV + 1.0)]}$

^{3/} Temperatures are expressed as departures from normal in degrees celsius

Table 9. Wheat Model for Punjab, India

Variable	Normal ^{2/}	Truncation Coefficients ^{1/}			
		Trend	November	December	March
Constant	1.00	9.35414 (17.59)	9.48247 (15.87)	8.97609 (12.91)	8.69926 (12.34)
Trend 1953-1977	12.23	0.26838 (4.39)	0.26391 (4.21)	0.31360 (4.38)	0.35772 (4.68)
Log (% HYV) ^{3/}	1.71	1.82742 (9.13)	1.84557 (8.92)	1.73554 (7.94)	1.61645 (7.06)
Nov Temperature (°C) DFN ^{4/}	17.89		-0.14495 (-0.51)	-0.32927 (-1.06)	-0.47921 (-1.49)
Dec Temperature (°C) DFN	13.34			0.37853 (1.35)	0.46548 (1.67)
Mar Temperature (°C) DFN	20.64				0.27304 (1.43)
R Squared =		0.97209	0.97243	0.97474	0.97718
F Value =		254.2	184.6	154.4	136.2
Degrees of Freedom =		22	21	20	19
Standard Error (Q/Ha) =		0.98945	1.00659	0.98722	0.96259

Standard Deviation of Yields = 5.67 Q/Ha

^{1/} t-values are in parentheses

^{2/} Normals and Estimates Based on Data from 1953-1977

^{3/} Percentage of high yielding varieties under cultivation. $\text{Log (\% HYV)} = \text{Natural log (\% HYV + 1.0)}$

^{4/} DFN = Deviation from Normal

degree of multicollinearity between precipitation and the trend factors.^{1/}

With high multicollinearity, the estimated parameters are likely to have an unsatisfactory, low degree of precision. A potential solution to the multicollinearity problem is the acquisition of more data, such as recent district (cross-section) data. However, new data is not immediately available. Also, because approximately 80 percent of Punjab wheat acreage is irrigated, precipitation may be a relatively unimportant contributor to wheat yield levels. Consequently, the best short-term remedy to the problems of poor data and multicollinearity appears to be to leave precipitation variables out of the model. In fact, in testing precipitation variables, none were found both significantly different from zero and theoretically acceptable. For example, preseason precipitation, while significant, consistently entered the equation with a negative sign. Similarly, while a few quadratic precipitation terms (squared departure from normal) for some months were significantly

^{1/} Johnston (8, pp. 161) states the main consequences of multicollinearity as:

1. The precision of estimation falls making it difficult or impossible to separate the relative influences of the various independent variables. "The loss of precision has three aspects: specific estimates may have very large errors; these errors may be highly correlated, one with another; and the sampling variances of the coefficients will be very large.

2. Investigators may incorrectly "drop variables from an analysis because their coefficients are not significantly different from zero, but the true situation may be not that a variable has no effect but simply that the set of sample data does not enable..." picking it up.

3. "Estimates of coefficients become very sensitive to particular sets of sample data, and the addition of a few more observations can sometimes produce dramatic shifts in some of the coefficients.

different from zero (.20 level), equations containing them performed poorly in extra-sample prediction tests (1975-77).

The Model

Separate predictive models or truncations were developed for four periods of the crop year. Truncations were: (a) preseason, consisting of parameters for the time trend and high yielding varieties; (b) the addition of November and December temperatures, and (c) the addition of March temperature for the "full-season" model. The coefficients and statistics of the final "full-season" model and selected truncations are shown in Table 9. Except for the November truncation the coefficient of determination was increased and the standard error was decreased as the season progressed.

Although various time surrogate variables, including piecewise, linear, and quadratic terms, were tested, the technology terms chosen for the final model were a log transformation of percentage of wheat area harvested in high-yielding varieties, and a trend term which is allowed to "level off" in crop years 1973/74 and 1974/75.

Monthly temperatures were the only weather variables included in the predictive models. In the "full-season" model, with the exception of the April variable which was significant at the .15 level, linear temperature departures from normal were different from zero at the .05 level of significance ($pr>|T|<.05$). Because the introduction of November temperature to the preseason truncation increased the standard error of fit, the November truncation is not recommended for prediction

purposes. The November temperature variable was retained because it was important to the fit of the data in truncations made later in the season.

Quadratic temperature terms were tested for possible inclusion in the final model. However, the stepwise and maximum R^2 selection procedures did not provide consistent results with the "best" model changing from one procedure to the other and selected variables changing as model size was increased. Also, coefficient signs were often questionable in the selected models. Also, extra sample prediction tests (1975-77) performed on the "best" models indicated errors were larger than for the linear model selected. Because of these problems, quadratic terms were excluded from the model.

Test Results

The results of a eleven year "jackknife" test of the models "predictive" capability for the trend truncation and the March truncation are shown in Table 10. The "jackknife" test was made by holding one year (the year of prediction) from the data set and a "predictive" model was developed from the remaining data set. A prediction was made for the year held out of the data set and the process repeated for the other ten years until predictions for eleven years were available.

"Predictive" errors for the March truncation range from an underestimate of 1.5 quintals per hectare in 1970 and 1975 to an overestimate of 1.6 quintals per hectare in 1968. Predictive errors for the trend truncation range from an overestimate of 1.8 quintals per

Table 10. "Jackknife" test results for Punjab, India.

Year : (Y) :	TRUNCATION ^{1/}								
	Trend		November		December		March		
: Yield:	\hat{Y}	S_p	\hat{Y}	S_p	\hat{Y}	S_p	\hat{Y}	S_p	
------(Quintals/Hectare)-----									
1966	12.5	13.3	1.002	13.2	1.026	13.2	1.009	13.3	0.984
1967	15.4	16.3	0.998	16.3	1.013	16.2	0.998	16.0	0.980
1968	18.6	20.4	0.941	20.4	0.954	20.4	0.931	20.2	0.923
1969	21.7	21.7	1.013	21.7	1.031	21.6	1.013	21.9	0.987
1970	22.2	21.9	1.011	21.8	1.027	21.5	1.000	20.7	0.931
1971	22.4	22.2	1.012	22.3	1.031	22.5	1.013	23.1	0.979
1972	24.1	22.4	0.954	22.3	0.961	22.5	0.956	22.6	0.936
1973	22.3	23.1	1.001	23.0	1.023	23.3	0.994	23.1	0.977
1974	22.2	23.2	0.989	23.2	1.006	22.6	1.009	23.0	0.978
1975	24.0	23.1	0.998	23.1	1.017	22.8	0.982	22.4	0.933
1976	23.8	23.4	1.010	23.5	1.030	23.9	1.013	23.6	0.986
1977	24.0	23.6	1.010	23.7	1.029	24.2	1.012	25.2	0.965

Bias ^{2/}	:	-0.117		-0.108		-0.125		-0.158	
RMSE ^{2/}	:	0.935		0.932		0.917		1.077	
$\overline{S_p}^2$:	0.995		1.012		0.994		0.963	

^{1/} \hat{Y} is the predicted yield for the year and S_p is the standard error of prediction.

^{2/} Bias = $\frac{\sum(Y-\hat{Y})}{N}$, RMSE = $\sqrt{\frac{\sum(Y-\hat{Y})^2}{N}}$ and $\overline{S_p} = \frac{\sum S_p}{N}$ where the summation runs from 1966 through 1977 and N = 12.

hectare in 1968 to an underestimate of 1.7 quintals per hectare in 1972. The test results for the 1966-77 test period indicate that the November truncation has the lowest mean bias and the December has the smallest RMSE. None of the weather models appear to significantly improve the trend results. The 1977 results must be considered preliminary as the USDA estimate is subject to revision when better data becomes available from India.

Although Das (3) indicated that low temperatures during parts of January and February were important, the coefficients of mean monthly temperatures for these months were insignificant; and consequently, they were left out of the model. Also, an April variable had an insignificant coefficient, and provided poorer predictions in most years than the March truncation in the twelve year "jackknife" test. It was omitted. Supporting the omission of the April variable is the fact that most of the grain is mature and harvest is normally well underway by the last week of April in Punjab.

Summary and Conclusions

Based on an analysis of available time series data, the major components of the year to year yield variability in the Punjab appear to be related to technology changes, not weather variability. Two technology factors, the share of total wheat acreage harvested in high yielding dwarf varieties and a time surrogate for increases in the proportion of acreage under irrigation, increased commercial fertilization, and other management improvements, explained 97 percent of the variability in the 1953-1976 yield series.

If major yield-weather relationships exist, more complex model formulations (than linear multivariate), and a more detailed set of data on other environmental factors probably will be needed to sort out the relationships. For example, no meaningful analysis of the potential impact of freezing temperatures can be made without daily minimum and maximum temperature data. No consistently useful and statistically significant precipitation/yield relationships were discovered using monthly precipitation data. The large proportion of wheat land irrigated (80%) in Punjab helps to explain this finding. Much of the needed soil moisture probably comes from irrigation facilities, reducing the need for rainfall which is normally very light during the growing season.

The strong technology trend in the data also caused difficulty in statistically selecting weather variables. With technology explaining 97 percent of the yield variability little remains to be explained by weather.

REFERENCES

1. Dabrymple, Dana G., Development and Spread of High-Yielding Varieties of Wheat and Rice in the Less Developed Nations, FAER 95, USDA-ERS. Aug. 1976.
2. Dagli, V., A Regional Profile of Indian Agriculture, Bombay, Vora and Co. Publishers Private Limited, 1975.
3. Das, J. C., Forecasting Wheat Yield In India with the Help of Weather Parameters, Part I-Punjab and Haryana, Sc. Rep. No. 40, Meteorological Office, Poona-5, Feb. 1971.
4. Dayal, R., "Impact of Rainfall on Crop Yield and Acreage." Indian J. of Agric. Econ. 20(3), 1965, pp. 48-54.
5. Directorate of Economics & Statistics, Ministry of Agriculture, Government of India, Bulletin on Wheat Statistics in India (Districtwise), New Delhi, India, Jan. 1972.
6. The Fertilizer Association of India, Fertilizer Statistics: 1973-74, 19th Ed. Near Jawaharla Nehru University, New Delhi, Dec. 1974.
7. India Meteorological Department, India Weather Review, Monthly Weather Report, Poona, India, various issues.
8. Johnston, J., Econometric Methods, 2nd Ed., McGraw-Hill Book Company, New York, 1972.
9. Mather, John R., Climatology: Fundamentals and Applications, McGraw-Hill Book Company, New York, 1974.
10. Office of the Agricultural Attache, Brief on Indian Agriculture 1975, American Embassy, New Delhi, India, March 1975.
11. Parker, John B. Jr., New Technology in India's Agriculture and Outlook for Grain Production, ERS-F 341, USDA-ERS, Washington, D.C. June 1972.
12. Randhawa, M. S., Green Revolution, John Wiley & Sons, New York, 1974.