UNITED STATES DEPARTMENT OF AGRICULTURE
Bureau of Agricultural Economics

16

USE OF CONDITION REPORTS AND WEATHER DATA IN FORECASTING THE YIELD PER ACRE OF WHEAT

* * *

by

Fred H. Sanderson

* * *

Washington, D. C.
October 1942

YRB 42-01

UNITED STATES DEPARTMENT OF AGRICULTURE
Bureau of Agricultural Economics

October 1942

USE OF CONDITION REPORTS AND WEATHER DATA IN FORECASTING THE YIELD PER ACRE OF WHEAT

By Fred H. Sanderson, Research Assistant 1/

PURPOSE OF THE STUDY

The importance of forecasts of crop production is increased during critical times. With the Nation at war and committed to the task of supplying food and other agricultural products to its armed services and its Allies, adequate forecasts are needed as early as possible in the growing season. The method of improving crop forecasts developed in this study points the way to increasing precision in the monthly crop reports of the Department. The study was originally undertaken as a part of the general project of research in crop-estimating methodology under Bankhead-Jones Funds. The author was assigned that part of the project which had to do with the estimating of spring-wheat yields from condition reports and weather data, in the four principal spring wheat States of the United States (Minnesota, North Dakota, South Dakota, and Montana) and the contiguous Prairie Provinces of Canada.

Whether reasonably reliable forecasts of wheat yields could be made on the basis of the condition figures based upon crop reporters' returns in conjunction with one or more weather factors which crop correspondents perhaps do not consider sufficiently when judging the condition of the growing crop, was a fundamental question. The portion of the study relating to the Canadian figures was completed first and the results of this study are set forth in this report. Similar analyses are being completed for the Dakotas, Minnesota, and Montana. These will be used by the Crop Reporting Board as additional indications of wheat yields in those States. The method developed here is applicable to other crops.

BACKGROUND OF THE PROBLEM

Use of Crop Condition Figures as Indicators of Prospective Wheat Yields: Up to 1918, the Canadian Government issued official forecasts of yield and production of wheat, using the "par" system adopted by the United States Department of Agriculture in 1912 2/. The condition as reported by crop correspondents was expressed as a percentage of a "full" or "standard" crop. Forecasts of production were issued early in June, July, and August.

In 1919, at the suggestion of the International Institute of Agriculture, the Dominion Bureau of Statistics shifted to a presumably more exact method whereby the correspondents were asked to report condition as a percentage of a 10-year average.

1/ This study was made while the author was a part-time employee of the Bureau and a research associate at Harvard University.

2/ In the United States, the par system was in general use in a somewhat modified form until the end of the 1929 season. See U.S.D.A. Misc. Publ. No. 171, The Crop and Livestock Reporting Service of the United States, Nov. 1933, pp. [illegible]

STATEMENT OF THE HYPOTHESES THAT ARE TO BE TESTED STATISTICALLY

The aim of the present study was to determine whether or not reasonably reliable forecasts can be made on the basis of the condition figures published by the Dominion Bureau of Statistics, 7/ and whether such forecasts can be improved significantly by the inclusion of certain weather factors in the analysis.

Condition as an Indicator of Prospective Yield: In general, the reported condition or appearance of the growing crop is highly useful in forecasting crop yields, especially late in the season, as harvest approaches. It is obvious that the earlier in the season a condition figure is taken, the less dependable it is as an indicator of prospective yield because of the greater opportunity for extremes in subsequent weather to influence the crop outturn either beneficially or adversely.

The condition figure is usually a much better indicator of prospective yield after the reproductive stage of the plant's growth has advanced to the point where the fruit is readily visible---heads of wheat, ears of corn, bolls of cotton, or actual fruit in the case of apples, peaches, grapes, etc.

It is to be expected that the condition of wheat at or after heading time would be a much better indicator of probable yield than would be the case either with cotton, where fruiting is indeterminate or continuous, or with root and tuber crops, where the product is underground. The condition of wheat in the Prairie Provinces should become, therefore, an increasingly reliable indicator of prospective yield per acre as the season advances.

However, condition reports can hardly be interpreted in a literal sense. When 100-percent condition is defined by the Dominion Bureau of Statistics as a "condition promising a yield equal to the long-time average," it does not follow that in practice, a 100-percent condition is associated, on the average, with the mean yield, a 50-percent condition with one half of the mean yield, and so forth;

6/ (Continued) Hopkins, J. W.: Weather and Wheat Yield in Western Canada. Canadian Journal of Research, 1935, June and Sept. 1936.

Wilson, C. F.: The influence of rainfall and temperature on wheat yields in Saskatchewan, 1921-1937. Dominion Bureau of Statistics, Agr. Branch, Ottawa, Dec. 23, 1937.

Wilson, C. F.: Relations between weather factors and wheat yields in Western Canada. Paper presented at the joint meeting of the Canadian Society of Technical Agriculturalists and the American Statistical Association, Ottawa, June 29, 1938. (Mimeographed) Cf. Proceedings of the Tenth Annual Meeting of the Canadian Agricultural Economics Society. June 1938, pp. 73-86.

The following study appeared when the present paper was nearing completion: The Influence of Precipitation and Temperature on Wheat Yields in the Prairie Provinces, 1921-1940. Canada, Dominion Bureau of Statistics, Quart. Bull. of Agr. Statistics, Vol. 34, #391, July-Sept. 1941, pp. 167-187.

Several studies of the effect of weather on Canadian wheat yields, most of them unpublished, have been made by the U. S. Bureau of Agricultural Economics. For a summary of these studies, cf.: Swenson, W. E.: A Survey of Wheat-Crop Estimation Studies for Canada, Australia, and Argentina. U.S.D.A., Statistics and Agriculture, #3, July 1941, pp. 2-6.

7/ The publication of condition figures based on the returns of crop correspondents was discontinued by the Bureau in 1941. Canada, Dominion Bureau of Statistics, Agricultural Branch; Quar. Bul. of Agr. Stat., Vol. 34, No. 390, April-June 1941, pp. 116-119.

or that a crop condition defined by 100 is equal four times a crop condition defined by 25. Crop correspondents, in reporting conditions, as a rule, show a tendency to underestimate yield prospects, to underestimate the intensity of the variation in yield, and to shy away from reporting more than a "fair" yield at any time.

Crop reporters underestimate yield prospects in that a reported condition of 100-defined as promising a yield equal to the long-time average--would actually indicate a probable yield considerably exceeding the long-time average. If final yield is expressed in percent of the long-time average and plotted along the y-axis of a system of coordinates, and condition along the x-axis, the line of average relationship would not be the diagonal defined by the points 0:0; 50:50; 100:100 etc., the 1-1 line, but would actually lie above this line. This downward bias of the condition reports would increase as the season advances. (See table 1) 8/

If the condition figures would correctly reflect the intensity of variations of yield, an increase in condition by a given percentage would indicate, on the average, an equal relative increase in yield. Hence, if condition is plotted along the horizontal and yield along the vertical axis, both expressed as percentages of their respective means, and using the same unit along both axes, the regression coefficient would be 1. In other words, the regression line or line of average relation would have a slope of 45 degrees. Actually an increase of ten percent in condition in most cases indicates an increase in yield of more than ten percent, and the line of regression would meet the x-axis at an angle of more than 45 degrees.

In addition, this bias is not constant over the whole range of prospective yields. Crop reporters apparently hesitate to report more than a "slightly better than average" yield at any time. In other words, reported condition would not increase proportionally with prospective yield. Condition tends to be a sensitive indicator of low yield prospects, but becomes less sensitive as prospects improve 9/

Curvilinear relations between reported condition and final yield have been observed in many countries where crop forecasts are based on numerical indications regarding the appearance of the growing crop as gathered from crop correspondents. The lack of proportionality in the relation of reported condition to final yield as early as 1927 led the United States Department of Agriculture to abandon the "par" method of interpreting condition figures and to adopt the more flexible correlation method.

However, straight line relations between "condition" and yield may be found

---

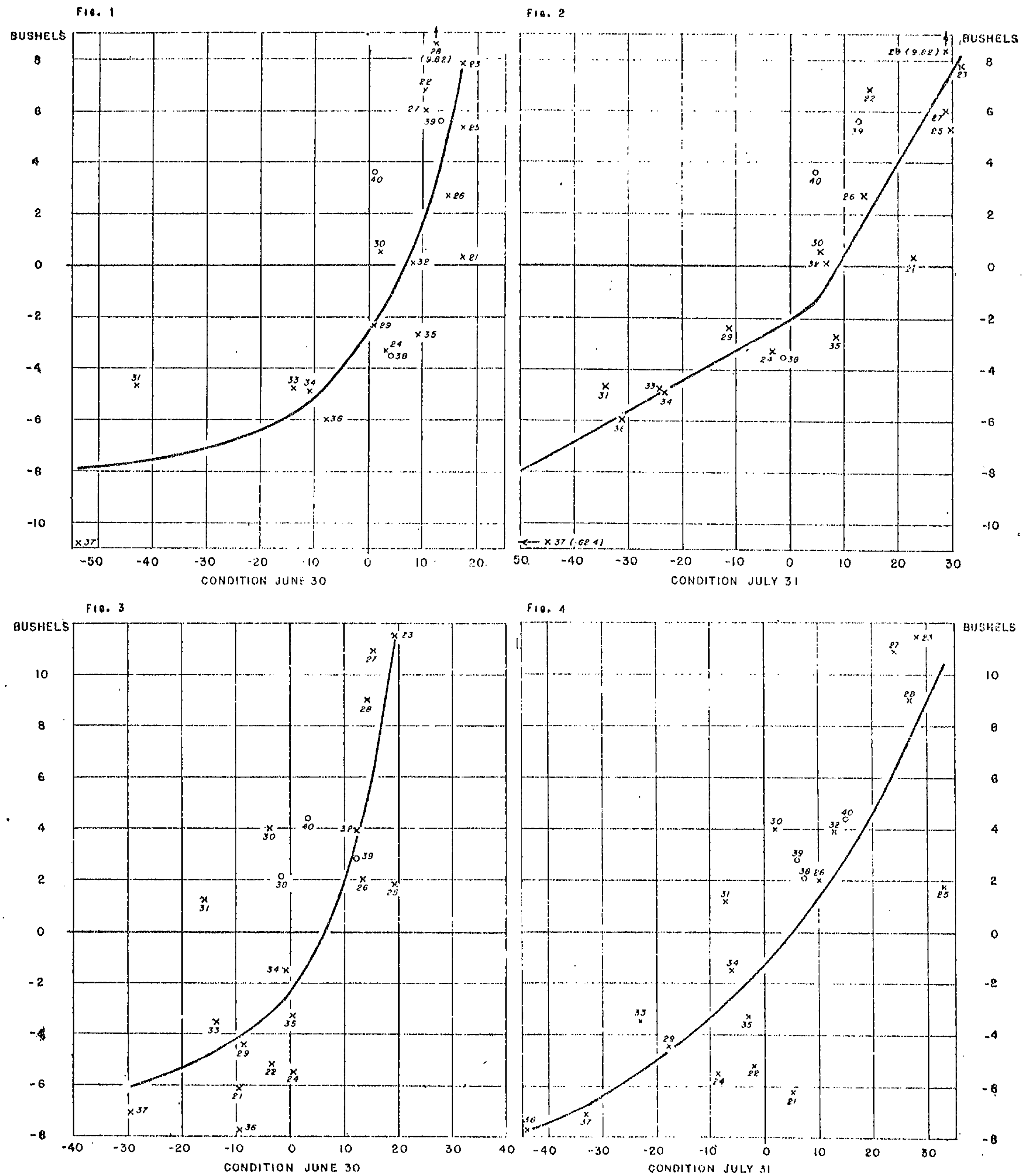
8/ In the provinces of Saskatchewan and Alberta, it was found that a condition of "100" reported at June 30 would correspond, on the average, to a yield ranging between 100 and 124 percent of the long-time average (1921-37), whereas the same condition reported on July 31 would indicate a probable yield ranging between 120 and 140 percent of the average. Consequently, in interpreting condition reports allowance must be made for this tendency to decline as the season advances. The average condition as of May 31 during the period 1921-37 was 94 for Saskatchewan and 98 for Alberta. The corresponding figures for June 30 are 87 and 93. A reported decline in the condition figure has real significance only when considered in relation to the average change during the period.

9/ Take, for example, the case of Saskatchewan, June 30 (fig. 1). When yield prospects were poor, an increase from 8 bushels below average to 6 bushels below average was accompanied, on the average, by an increase in "condition" as reported by crop correspondents from -53 to -16, that is, 37 points. The same increase in yield when prospective yields were high (+5 to +7 bushels) was accompanied by an increase in reported "condition" from +14.5 to +16.8, that is, 2.3 points.

TABLE I. CONDITION, PRESEASONAL PRECIPITATION, MAXIMUM TEMPERATURES, YIELD AND PRODUCTION OF WHEAT, 1921-1940

Year	CONDITION			Pres.Pr. (Sept.-Nov.)	10 days' (June 21-30)		20 days' (June 21-July 10)		Acreage	Yield	Prod.
	May 31	June 30	July 31		Prec.	Max.Temp.	Prec.	Max.Temp.			
					SASKATCHEWAN						
1921	102	105	99	3.69	0.40	85.9	1.20	83.0	13,557	13.8	187
1922	101	98	91	6.64	1.22	74.0	1.88	74.0	12,332	20.3	250
1923	98	105	108	2.53	3.81	68.2	4.93	72.6	12,791	21.3	[illegible]
1924	96	91	73	2.60	0.11	73.0	0.44	78.8	13,033	10.2	133
1925	100	105	106	3.50	1.23	77.1	2.18	77.4	12,509	18.8	235
1926	104	102	90	3.64	0.08	79.5	0.55	80.0	13,558	16.2	220
1927	94	98	105	3.67	0.39	77.5	2.30	76.0	12,978	19.5	253
1928	98	100	105	5.15	0.54	72.3	1.24	73.0	13,791	23.3	321
1929	99	89	65	1.06	0.57	69.7	0.92	72.7	14,445	11.1	160
1930	95	90	82	2.93	1.76	74.0	2.14	77.8	14,764	14.0	207
1931	77	45	42	3.55	1.79	81.4	2.23	77.1	15,026	8.8	132
1932	92	96	83	3.01	0.78	78.8	3.03	73.8	15,543	13.6	211
1933	99	74	52	2.34	1.59	74.8	1.98	78.2	14,743	8.7	128
1934	73	77	53	3.65	1.31	74.0	1.72	75.0	13,232	8.6	114
1935	97	97	85	1.90	0.82	76.0	2.50	77.7	13,206	10.8	143
1936	95	80	45	2.62	0.44	82.0	0.67	84.8	14,744	7.5	111
1937	78	34	14	1.72	0.19	86.0	0.26	88.8	13,893	2.7	37
Aver. 1921-37	94.0	87.4	76.4	3.19	1.00	76.7	1.77	77.7	13,775	13.5	186
1938	99	92	75	2.63	0.97	75.6	0.90	77.2	13,793	10.0	138
1939	92	101	89	3.48	1.77	70.6	1.00	74.6	14,233	19.1	271
1940	94	89	81	1.52	0.55	72.2	1.19	76.1	15,571	17.1	267
					ALBERTA						
1921	102	83	89	1.68	0.42	78.4	1.52	75.1	5,038	10.3	52
1922	102	89	82	2.16	0.93	72.1	2.32	71.8	5,701	11.3	64
1923	100	112	112	1.73	3.28	66.8	4.90	68.8	5,088	23.0	142
1924	97	93	75	1.28	0.93	73.8	1.31	76.6	5,537	11.0	61
1925	107	112	117	2.85	0.87	76.6	1.56	75.8	5,334	18.3	98
1926	105	106	94	5.02	1.06	74.4	2.04	76.2	6,139	18.5	114
1927	100	108	108	6.50	1.43	70.0	3.27	69.2	6,165	27.4	169
1928	103	107	111	5.54	1.92	69.7	3.48	69.8	6,597	25.5	168
1929	102	84	66	1.11	0.51	72.0	1.89	73.8	7,423	12.1	90
1930	97	89	86	2.88	1.03	67.9	1.53	72.4	7,164	20.5	147
1931	84	77	77	3.39	1.67	69.8	2.41	69.8	7,943	17.7	141
1932	102	105	97	2.75	0.77	74.8	1.89	72.2	8,201	20.4	167
1933	98	79	61	3.67	1.31	71.0	2.07	73.8	7,898	13.0	103
1934	88	92	78	3.68	1.09	70.3	2.32	70.9	7,501	15.0	113
1935	96	93	81	3.41	0.59	72.5	1.00	72.6	7,500	13.2	99
1936	96	83	40	3.07	0.36	81.0	0.65	81.1	7,537	8.8	66
1937	93	63	51	2.69	0.10	77.7	0.42	79.0	7,834	9.4	74
Aver. 1921-37	98.4	92.6	83.8	3.14	1.07	72.9	2.03	73.5	6,741	16.5	111
1938	99	91	91	3.75	1.30	74.0	1.48	73.6	7,969	18.6	148
1939	96	105	90	2.69	0.63	67.4	0.52	71.0	8,379	19.3	161
1940	98	[illegible]	99	3.81	0.68	70.6	1.69	72.6	8,667	20.9	181

Units of measurement: Precipitation, Total inches, average per station; Temperature, Average daily maximum temperature, in degrees Fahrenheit; Acreage, 1,000 acres; Yield, Bushels per acre; Production, Million bushels.



U. S. DEPARTMENT OF AGRICULTURE BUREAU OF AGRICULTURAL ECONOMICS

Fig. 1. Saskatchewan, June 30 (Formula 3), Regression of Yield on Condition at June 30.

Fig. 2. Saskatchewan, July 31 (Formula 7), Regression of Yield on Condition at July 31.

Fig. 3. Alberta, June 30 (Formula 13), Regression of Yield on Condition at June 30.

Fig. 4. Alberta, July 31 (Formula 17), Regression of Yield on Condition at July 31.

frequently, especially early and late in the season. In the early part of the growing season, the total correlation between "condition" and yield may be poor, and the curvilinearity of the relationship may be blurred by errors. As the season advances, the departure from linearity would become more apparent. Shortly before harvest, however, prospects of bumper crops begin to become recognized, and the relationship between "condition" and yield would approach a straight line. (See figures 2 and 4)

Where condition is plotted against yield per acre harvested (a practice resorted to when data for acreage sown are unavailable), it can be argued that the departure from linearity might possibly be due to abandonment of acreage subsequent to the condition report. Indeed, it is possible that the condition reports made by crop correspondents, particularly the earlier reports, refer more closely to acreage sown than to acreage harvested. Widespread abandonment of acreage subsequent to the report will, therefore, introduce an apparent downward bias in the condition figures when the latter are compared with final yield per harvested acre. As abandonment of acreage tends to be associated with low yields rather than with high yields, this apparent downward bias will be particularly noticeable in years of poor yields, while in years of above-average yields and correspondingly small extent of abandonment, it would scarcely affect condition figures at all.

If this view were more than a partial explanation of the curvilinear bias present in condition figures, the relationship between reported condition and yield per acre sown would be a straight line. But it was found that where yield data were available per seeded acre as well as per harvested acre, the curvature was somewhat attenuated in the former case, but did not disappear. A further confirmation of this is provided by the case at hand, for the Canadian yield data refer to acreage sown.

In using condition figures as a basis of forecasting yields, it must be kept in mind, of course, that these are not crop data, but opinions or judgment appraisals of crop reporters. There is a danger of instability of the psychological factors on which the estimates are based. The systematic bias may be modified by a slight modification of the questionnaire, or of the definition of "normal" or "average," or by changes in the underlying conditions. Any change in the procedure constitutes a break in homogeneity which tends to introduce entirely unpredictable changes in the relationship between reported condition and yield. To eliminate bias from the reported condition figures, a long series of observations under homogeneous conditions is needed, and if an innovation is introduced and is sufficiently radical, it may take some years before the condition reports can again be used for forecasting purposes.

Use of Weather Factors to Supplement Condition Reports: It is reasonable to suppose that reported condition--a subjective estimate of the appearance of the growing crop--may not include all the important factors that may be in existence at the time the forecast is made. For example, it is possible that in reporting the current condition of the growing crop, crop correspondents do not take into full consideration the amount of soil moisture available to the plant.

Reserve of Subsoil Moisture: Unfortunately, no direct measurements concerning subsoil moisture were available for this study. It was assumed that under the climatic conditions prevailing in the Prairie Provinces of Canada, the moisture content of the soil at the beginning of the growing season might be determined to a considerable extent by the amount of rainfall received during the 3 months of September, October, and November of the year preceding harvest. There is little precipitation during the winter months in these provinces, and that which occurs is probably mostly lost by evaporation before it has a chance to enter the ground when the spring thaws come.

Lack of subsoil moisture would not necessarily affect the appearance of the crop early in the season, provided spring rains had been sufficient to germinate the crop and give it a good start. In fact, lack of subsoil moisture might not affect the appearance of the crop until heading time when the plant requirements for moisture are at or near the maximum; and the moisture due to precipitation during the growing season is nearly or completely used up by the plant 10/.

One might expect, therefore, that the condition of the crop in the early part of the season is influenced largely by current precipitation, and the reserve of moisture in the lower levels of the soil might well be neglected by the crop correspondents when estimating condition. Hence, the hypothesis to be tested is whether or not preseasonal rainfall, when used in conjunction with condition, will give a better basis of forecasting the yield of wheat than condition alone. In years when the preseasonal precipitation is above average, do crop correspondents tend to underestimate yields more than in years of below-average preseasonal rainfall? However small the total influence of preseasonal rainfall may be in comparison with other meteorological factors, it may be expected to be an influential factor among those that are not sufficiently taken account of by crop correspondents. It is a factor that might be expected to have less relative importance in comparison with condition as the season advances, for in the later stages of growth, lack of soil moisture reserves is more and more reflected by the appearance of the crop.

<u>Weather Conditions During the Period Immediately Preceding the Date of the June 30 Condition Report</u>: It is possible that crop correspondents would fail to allow for rainfall immediately preceding the report. Barnes and Hopkins have pointed out that a heavy downpour does not lead to a considerable increase in soil moisture at the depth available for plant use until 10 days or more after its occurrence. That is, the appearance of the crop may not be affected immediately, and it is on the basis of the appearance of the crop that correspondents are supposed to estimate condition.

It may be expected that this "lag-effect" of rainfall will be particularly important during the period preceding heading time (about the end of June and beginning of July), which is considered a critical period with regard to wheat yields in the Prairie region 11/. Obviously, the failure of crop correspondents to take sufficiently into account the probable effect of rainfall received during the period immediately preceding the report may considerably reduce the value and timeliness of their reports particularly during the periods of maximum sensitivity of the plant with regard to weather influences. On the other hand, if the reports obtained from the correspondents could be supplemented by information regarding weather conditions during the period immediately preceding the mailing of the correspondents' reports and, possibly, during the time elapsing between the date of the individual reports of the correspondents and the date of the release of the official crop report of the Canadian Government, the reliability of the forecast might be considerably increased.

10/ Compare Barnes, S. and Hopkins, E. S.: Soil Moisture and Crop Production. Dominion of Canada, Department of Agriculture, New Series Bulletin 130: Ottawa, 1930.

11/ A careful analysis, by means of a regression integral, of spring wheat yields at Dickinson, N.D. indicates that the crop reaches its maximum susceptibility to rainfall about June 20. At that time, an additional inch of rain would lead, on the average, to an increase in yield of four bushels per acre. After heading (which occurs about July 10), the beneficial effect of an additional inch of rain in terms of yield per acre decreases rapidly and becomes zero about July 20. (See Davis, F.E. and Pallesen, J.E.: "Effect of the Amount and Distribution of Rainfall and Evaporation during the Growing Season on Yields of Corn and Spring Wheat," Journal of Agricultural Research, 60, 1, Jan. 1, 1940, pp. 1-23; particularly pp. 12-23.) Consequently, no improvement of the forecast should follow from the inclusion of weather conditions during the 10 days preceding the July 31 report.

Crop correspondents' estimates of condition were obtained by the Canadian Government at three different times during the growing season--May 31, June 30, and July 31. The official crop report of the Canadian Government was issued about 10 days later. It would therefore be possible to supplement the June 30 condition figures as reported by the correspondents with meteorological data up to the day prior to the publication of the official crop report. Accordingly, two sets of meteorological variables were computed, one representing the weather prevailing during the 10 days before the date of the crop correspondents' condition report, (that is, the period June 21-30), the other representing weather conditions during the 20 days preceding the date of release of the official crop report of the Canadian Government (that is, the period June 21-July 10).

Slow Changes In the Relationship Between Reported Condition and Yield, and Between Weather Factors and Yield: Psychological factors, such as the effect of the previous year's or previous years' level of yield, may influence the crop reporter's concept of an average or normal. Shifts of acreage, changing methods of cultivation, introduction of new varieties may cause slow changes in the general level of yields, as well as in the nature of the relationship between weather factors and yields.

Since the early twenties, the continuous westward shift of acreage has come to a standstill. During the period covered by the study, the slight westward movement during the earlier years has been followed by a slight eastward trend during the later years. A small part of the wheat is grown on irrigated land.

A new rust-resisting variety of wheat, Thatcher, was introduced commercially in Canada in 1935. In 1939, Thatcher wheat was grown on about nine million acres, mainly in Manitoba and in the eastern part of Saskatchewan. With respect to yield and drought resistance, the new variety seems to be about at par with the standard varieties. However, Thatcher does not appear as leafy as the standard varieties and therefore does not look as well, and it is entirely possible that crop correspondents would at first tend to underestimate both condition and drought resistance.

METHOD OF ANALYSIS

General Remarks on the Testing of Hypotheses: In the present investigation, the principle of testing a rather rigid hypothesis has been adhered to more strictly than is customarily the case in the study of weather-yield relationships. Many analysts start out by plotting final yields against a large number of monthly means of temperature, rainfall, sunshine, humidity, etc. Those meteorological series which show an apparent correlation with yield are then used to construct forecasting formulas. In other words, the procedure consists in selecting, out of a large number of series originally considered, that set of variables which yields the highest multiple correlation coefficient. The tests of significance which are applicable to a hypothesis consisting of a number of linearly related variables selected a priori or drawn at random do not apply to such cases. The greatly increased probability of obtaining stated high correlations from random data tends to reduce drastically the significance of correlation coefficients obtained in such a manner 12/. The same objection applies to the selection of curve types and to the combination of the predictor series with reference to the sample. The rejection of this sort of procedure does not imply, of course, that "one cannot learn from the data." It simply means that the correlation coefficients thus obtained do not have the significance which is usually attributed to them.

12/ No general solution of this problem has been put forward. See, however, the discussion by R. A. Fisher in "The Influence of Rainfall on the Yield of Wheat at Rothamsted," Phil. Trans. of the Royal Society of London, Series B, Vol. 213, 1924, pp. 94-95.

FIG. 5

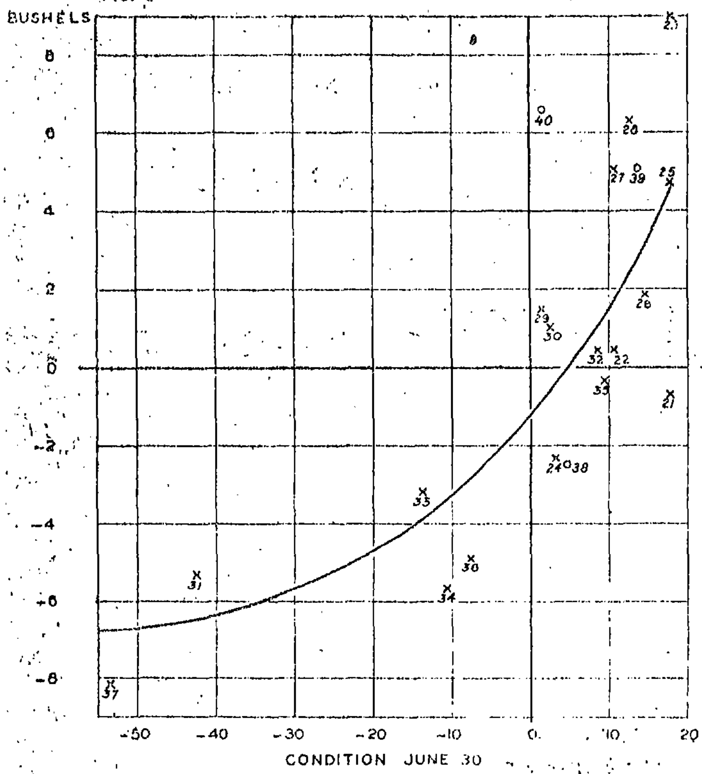


FIG. 6

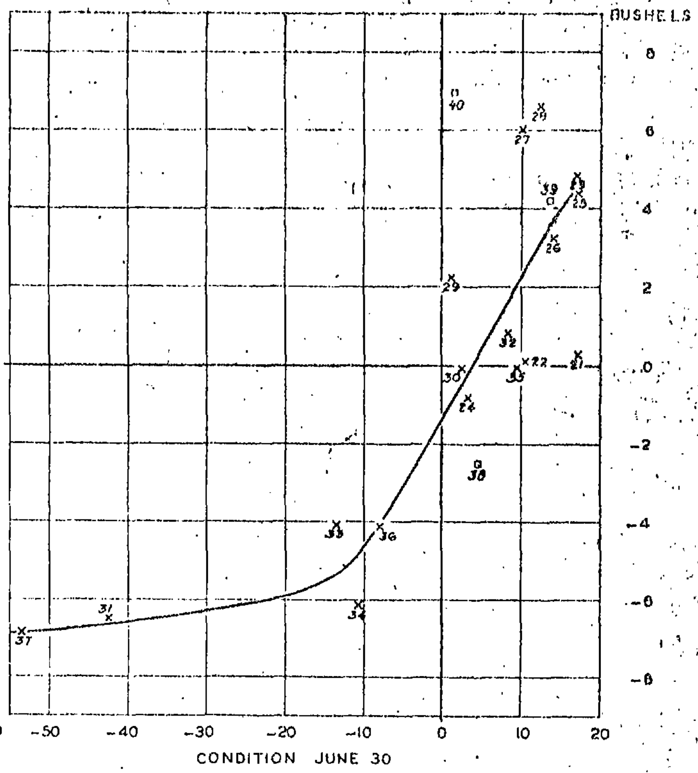


FIG. 7

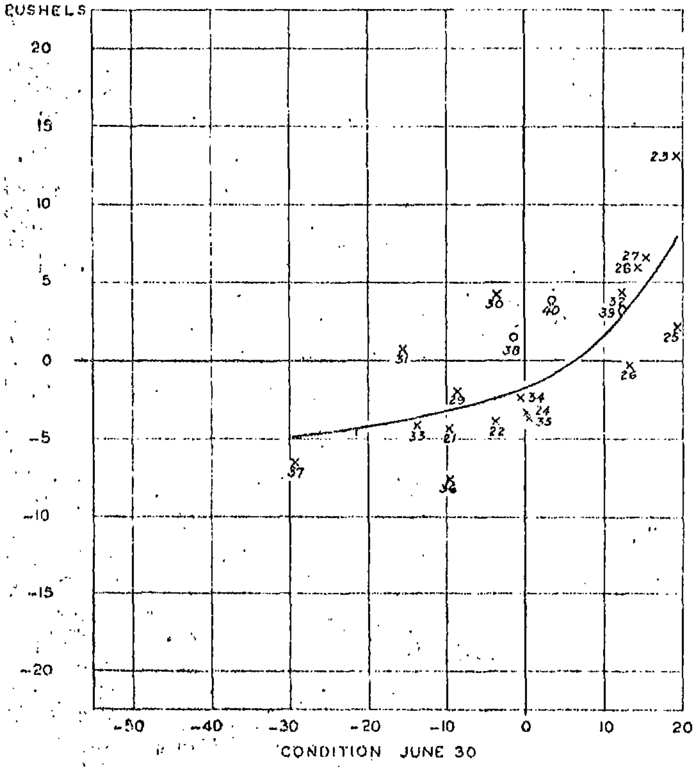
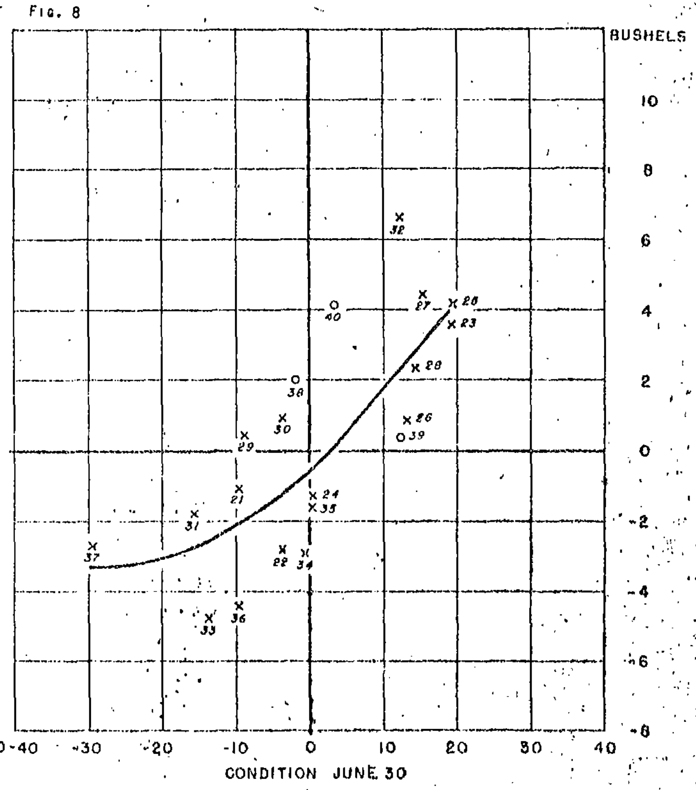


FIG. 8



U. S. DEPARTMENT OF AGRICULTURE   NEG. 43003   BUREAU OF AGRICULTURAL ECONOMICS

Fig. 5. Saskatchewan, June 30 (Formula 3), Net Regression of Yield on Condition at June 30, After Removing Effect of Preseasonal Precipitation.

Fig. 7. Alberta, June 30 (Formula 14), Net Regression of Yield on Condition at June 30; Preseasonal Precipitation Held Constant.

Fig. 6. Saskatchewan, June 30 (Formula 5), Net Regression of Yield on Condition at June 30, After Removing Effect of Preseasonal Precipitation and Precipitation During the 10 Days Prior to the Report.

Fig. 8. Alberta, June 30 (Formula 16), Net Regression of Yield on Condition at June 30; Preseasonal Precipitation, Precipitation and Temperature During the 10 Days Prior to the Report Held Constant.

None of the major hypotheses used in the present study was derived from the data. The choice of the weather variables was made *a priori*. Unfortunately, no reasonable well-founded hypotheses were available as to the exact shape of some of the regressions. Curvilinear regressions of yield on condition were expected on the strength of independent evidence. The general shape of the regressions could be specified as an increasing function of condition, bent concave upward. It was anticipated, however, that for the earlier reports the curvilinearity of the condition-yield relationships might be blurred by errors; it has been pointed out, furthermore, that there is usually a tendency for the curves to straighten out as harvest time approaches and prospects of bumper crops become apparent. Apart from these considerations, there was no theoretical reason, or empirical evidence given *a priori*, which would have enabled one to specify the exact mathematical form of the relationship.

In a case like this, it is usually advisable to divide the available data into two independent halves, to derive a hypothesis from the first half and to test it on the basis of the second half. But this is possible only if enough observations are available.

If a sufficient number is not available, a trial hypothesis must be formulated on the basis of whatever *a priori* knowledge there exists regarding the nature of the relationship which is to be investigated. If the general shape of the relationship can be specified only in vague terms such as "concave upward," it is customary to fit a simple type of curve to the data, such as a second-degree parabola, with the further restriction of a positive coefficient in one or both terms. However, the *type of curve* must be specified *a priori* and without reference to the sample, whereas the values of the parameters are estimated on the basis of the sample.

If free-hand curves are used, a similar degree of rigidity of the form of the original hypothesis can be approximated only by an extremely conservative use of this method. For tests of significance to be applicable, the assumption must be that in the process of successive approximation, the curves have been adapted to the data only to an extent which would be equivalent to the change of the coefficients in a mathematically specified relationship, while the "form" of the relationship has not been permitted to be influenced by the sample.

In view of the paucity of certain data, however, it is often necessary to violate the principle of a rigid hypothesis to a certain extent. Minor modifications of an otherwise given hypothesis may have to be introduced without benefit of a further independent test. To the extent to which such modifications are made, the significance of the correlations obtained is reduced. Strictly speaking, if the significance of a hypothesis involving several variables can be established only with reference to some of the variables, it does not seem sufficient to reject those variables which appeared not to be significant while retaining the others. In order to fit the available methods of statistical generalization, the larger hypothesis as a whole must be thrown out, and the more restricted hypothesis must be tested on the basis of independent material. Very often, however, data are not available in sufficient quantity to permit this.

In the present study care was taken that modifications of the original hypothesis were confined to minor aspects of it. The fitting of *curves* instead of straight lines in the case of average maximum temperature constitutes the only outright deviation from the original hypothesis. However, the curvilinearity in this case is barely significant, and substantially the same correlation coefficients, corrected for the number of constants used, would have been obtained with linear regressions. It is also recognized that the use of free-hand curves in the analysis of the condition yield relationships presents certain dangers with respect to the required rigidity of the hypothesis.

Fig. 9

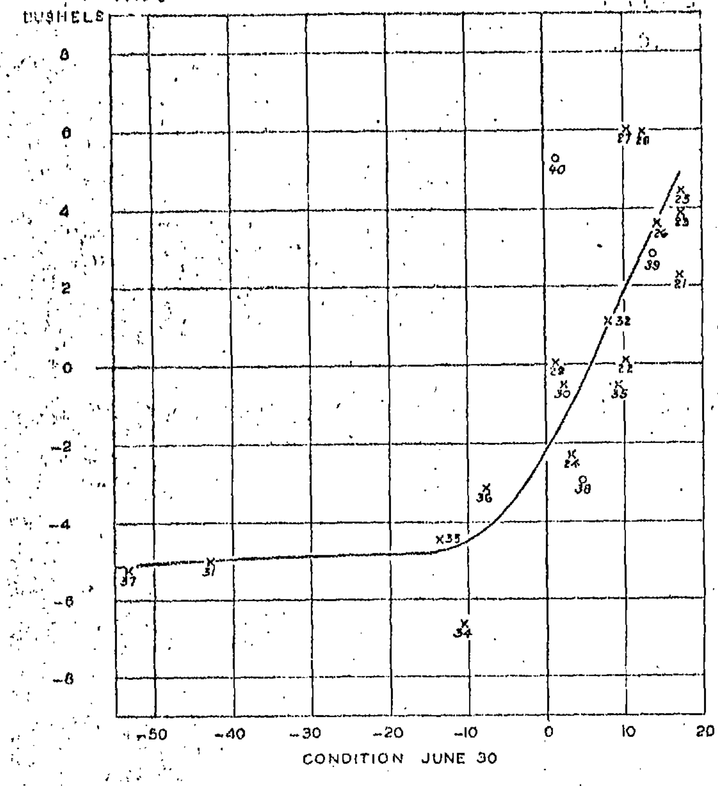


Fig. 10

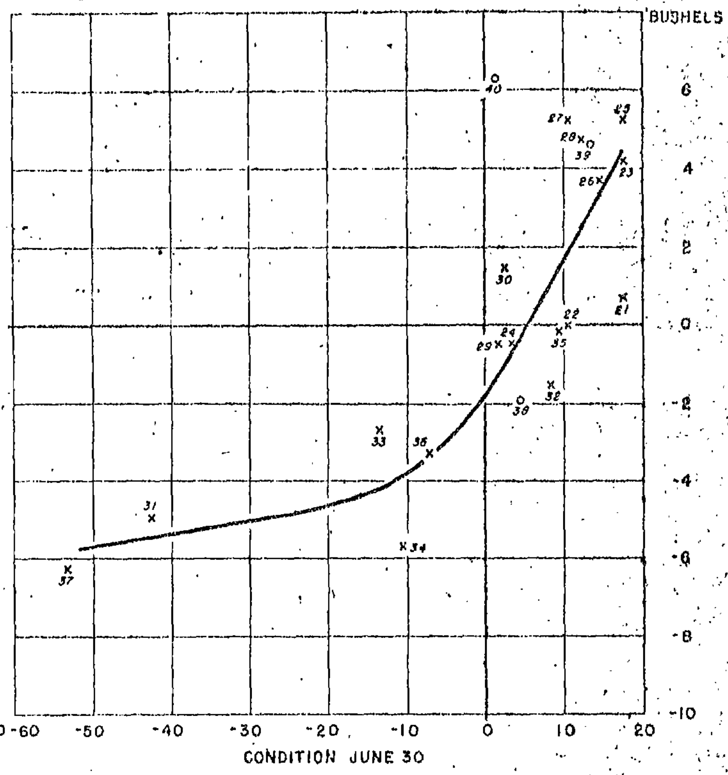


Fig. 11

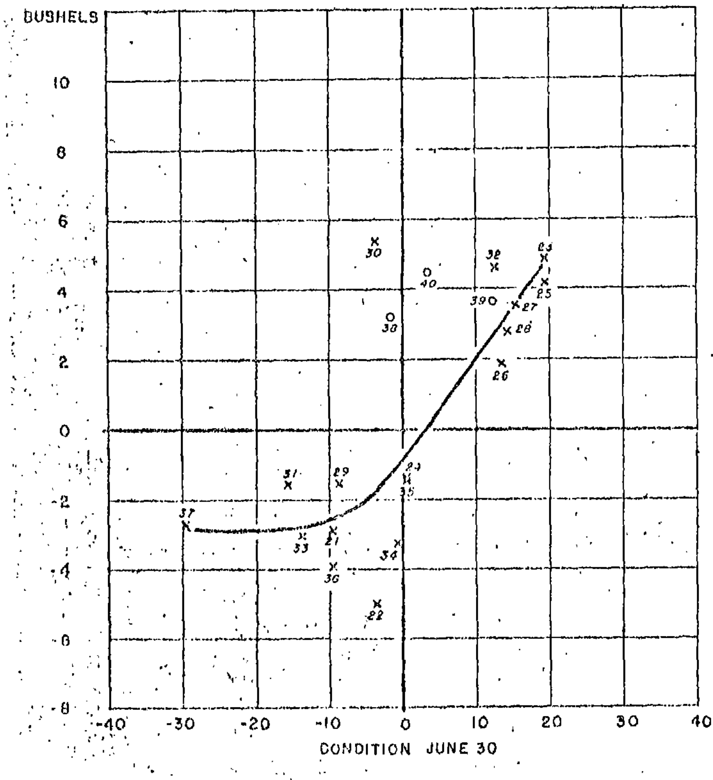
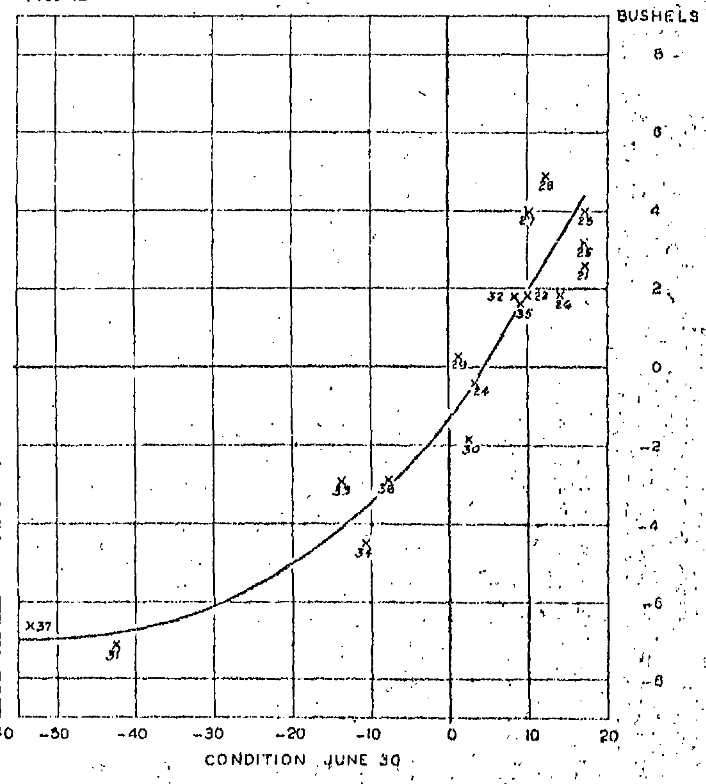


Fig. 12



U. S. DEPARTMENT OF AGRICULTURE   NEG. 43004   BUREAU OF AGRICULTURAL ECONOMICS

Fig. 9. Saskatchewan, June 30 (Formula 6), Net Regression of Yield on Condition at June 30; Preseasonal Precipitation, and Precipitation and Temperature During the 10 Days Prior to the Report Held Constant.

Fig. 10. Saskatchewan, June 30 (Formula 6a), Net Regression of Yield on Condition at June 30; Preseasonal Precipitation, and Precipitation and Temperature During the 20 Days Prior to the Release of the Official Crop Report Held Constant.

Fig. 11. Alberta, June 30 (Formula 16a), Net Regression of Yield on Condition at June 30; Preseasonal Precipitation, and Precipitation and Temperature During the 20 Days Prior to the Release of the Official Crop Report Held Constant.

Fig. 12. Saskatchewan, June 30 (Formula [illegible]), Net Regression of Yield on Condition at June 30; Preseasonal Precipitation During the 10 Days Prior to the Report, and Supposed Influence of [illegible] Held Constant.

[illegible] lowing procedure was adopted. A multiple linear regression equation was fitted to the observations. For those of the determining variables which were expected to show a curvilinear relationship with the dependent variable, net regression curves were determined from the residuals by a series of graphic approximations. These approximations were guided by a priori expectations as to the form of the relationships, while the steepness of the curves and the degree of curvature were determined from the group averages.

<u>Degree of Total Determination</u>: The square of the multiple correlation coefficient (or index) is used rather than the coefficient or index itself. The squared correlation coefficient or index measures the percentage of the total year-to-year variability in the yield data that is associated with the determining factors--condition and one or more weather factors.

The distribution of the multiple correlation coefficient is very skew in small samples, and "there is a tendency for the multiple correlation coefficient shown by the sample to be in excess of the correlation existing in the universe from which the sample was drawn, especially when the number of observations is small, or the number of variables large." 13/

An estimate of the multiple correlation coefficient corrected for bias was computed according to the formula.

$$\bar{R}^2 = 1 - \left[(1-R^2)\frac{N-1}{N-m} \right]$$

(where N is the total number of observations, and m the number of constants used) 13/

<u>Tests of Significance</u>: Appropriate tests of significance were applied at each step of the analysis, appreciating the fact that the time series used may not be strictly random series.

In testing the significance of correlation coefficients or indexes, Snedecor's convenient table was used 14/. The significance of additional variables or constants introduced in the analysis was determined by means of the following test:

$$z = \log_e \sqrt{\frac{(R_1^2 - R_2^2)(N-m_1-1)}{(1-R_1^2)(m_1-m_2)}} \quad (\text{where } m_1 > m_2) \quad 15/.$$

N represents the total number of observations, m_1 the number of constants involved in the larger set of variables, and m_2 the number of constants in the smaller set of variables. By this test, it is possible to determine whether the increase in R^2 (or P^2) caused by the inclusion of an additional variable (or constant) is significant, or whether it is probably due to random effects.

13/ Ezekiel, M., Methods of Correlation Analysis.

14/ Snedecor, G. W.: Statistical Methods, 1938 ed., p. 286

15/ Tippett, L.H.C.: The Methods of Statistics, p. 208 (m_1-m_2) and $(N-m_1-1)$ are entered under n_1 and n_2, respectively, in Table IV, pp. 250-253 of R. A. Fisher's "Statistical Methods for Research Workers," 7th edition, 1938.

FIG. 13

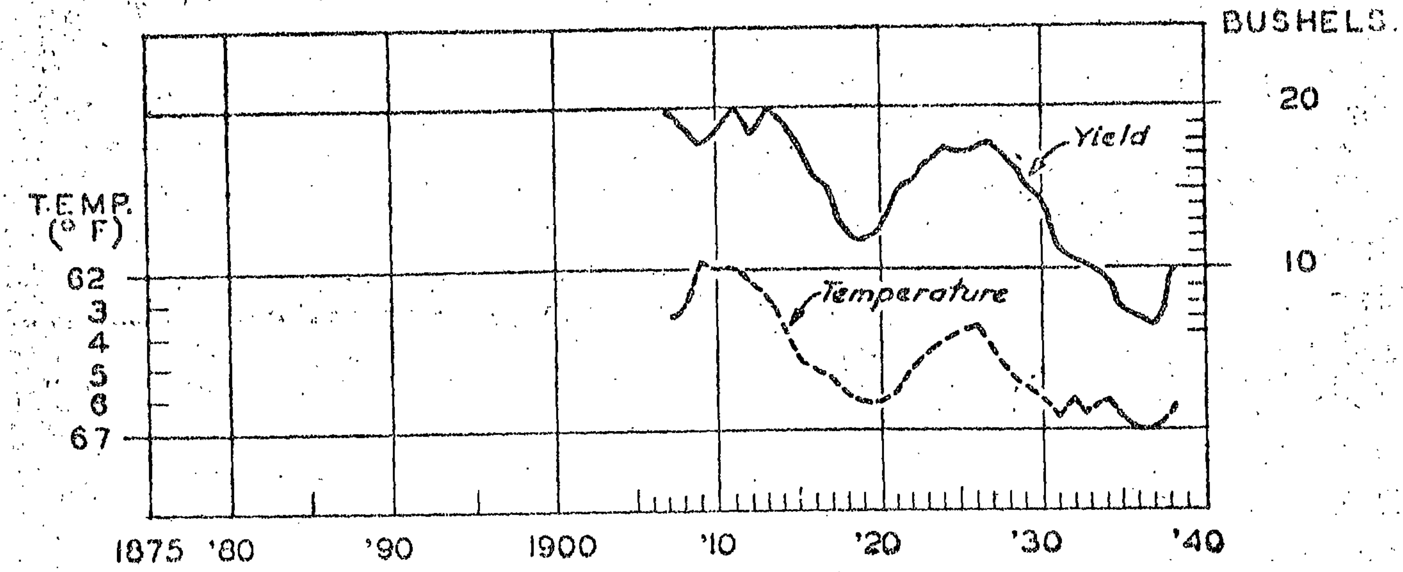


FIG. 14

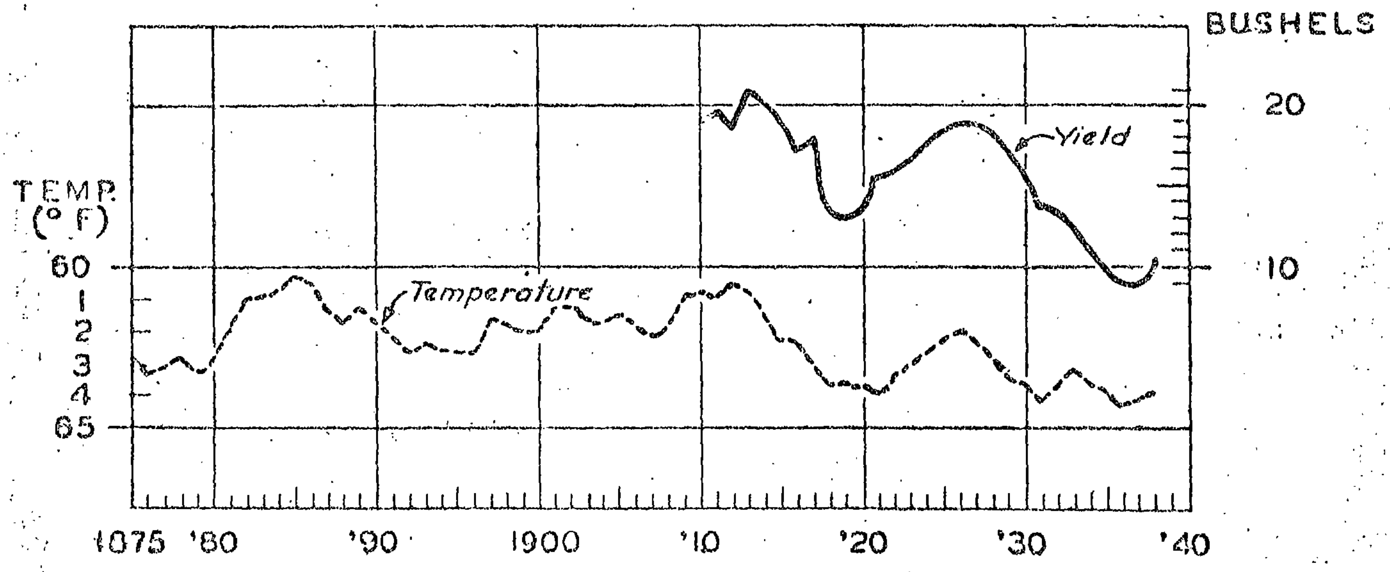


Fig. 13. Saskatchewan, Comparison of General Tendency in Wheat Yield and July-August Average Temperature, 1907-1938. Five-Year Moving Average (19[illegible] extrapolated).

Temperature, (Inverted Scale) at Battleford, Swift Current, Qu'Appelle, Moose Jaw (weight 2).

Fig. 14. Prairie Provinces, Comparison of General Tendency in Wheat Yield and July-August Average Temperature. Five-Year Moving Average (19[illegible] extrapolated).

Temperature (Inverted Scale) at Winnipeg, Qu'Appelle, Edmonton, Calgary. (Wheat Yields are Averages for the Dominion.)

In certain cases, it seemed desirable to establish the significance of an increase in the correlation coefficient (or index) obtained when substituting one determining variable for another. By extreme good fortune, an unpublished paper by Professor Harold Hotelling was available which gives the solution for the linear case. 16/

Let x_2' and x_2'' be two alternative variables that might be used to predict a variable, x_1. r_1 is the correlation between x_2' and x_1, and r_2 is the correlation between x_2'' and x_1. The problem is whether the difference between r_1 and r_2 is significant.

The test, derived by Hotelling, using Fisher's t, is as follows:

$$t = (r_1 - r_2)\sqrt{\frac{n(1+r_0)}{2D}}$$

where n is the number of degrees of freedom, r_0 is the correlation between x_2' and x_2'', and D is the determinant

$$\begin{vmatrix} 1 & r_1 & r_2 \\ r_1 & 1 & r_0 \\ r_2 & r_0 & 1 \end{vmatrix}$$

The absolute values of r_1 and r_2 are used in computing $(r_1 - r_2)$.

Suppose that x_2' and x_2'' are alternative variables in a multiple set comprising, in addition, a second determinant variable, x_3. Instead of comparing two multiple correlation coefficients, one can compare the simple correlation coefficients obtained by correlating the two alternative variables separately with the residuals from the dependent variable after eliminating the effect of the determinant variable that is common to the two alternative regression equations.

Let

$$z_2' = x_2' + Ax_3$$

$$z_2'' = x_2'' + Bx_3$$

$$z_3 = x_3$$

where

$$A = \frac{S(x_2'x_3)}{S(x_3^2)}$$

and

$$B = \frac{S(x_2''x_3)}{S(x_3^2)}$$

A and B are determined so that z_2' and z_3, and z_2'' and z_3 are orthogonal ($S(z_2'z_3) = 0$ and $S(z_2''z_3) = 0$); that is, we add to x_2' and x_2'' that part of the variation of x_3 that is associated with x_2', resp. x_2''.

Then, the line $x_1 = a - cx_3$ is fitted to the data and the residuals (v) computed.

16/ Professor Hotelling's study, entitled "The Selection of Variates for Use in Prediction With Some Comments on the General Problem of Nuisance Parameters," appeared in the Annals of Mathematical Statistics, Vol. XI, No. 3, September 1940.

FIG. 15

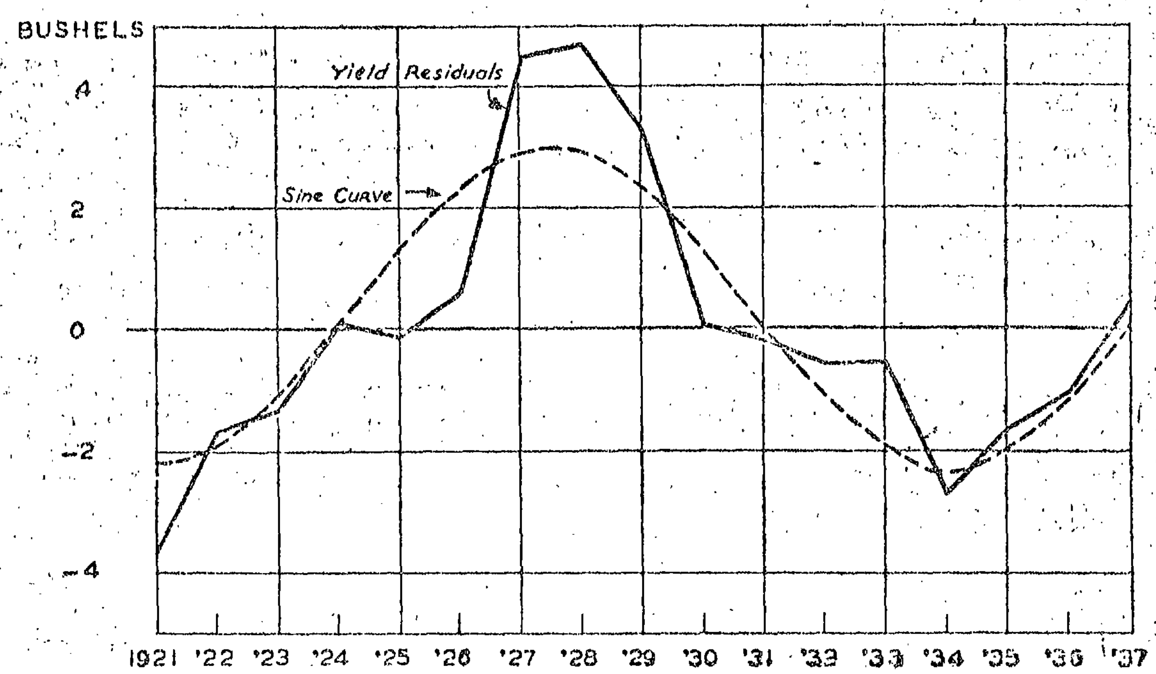


FIG. 16

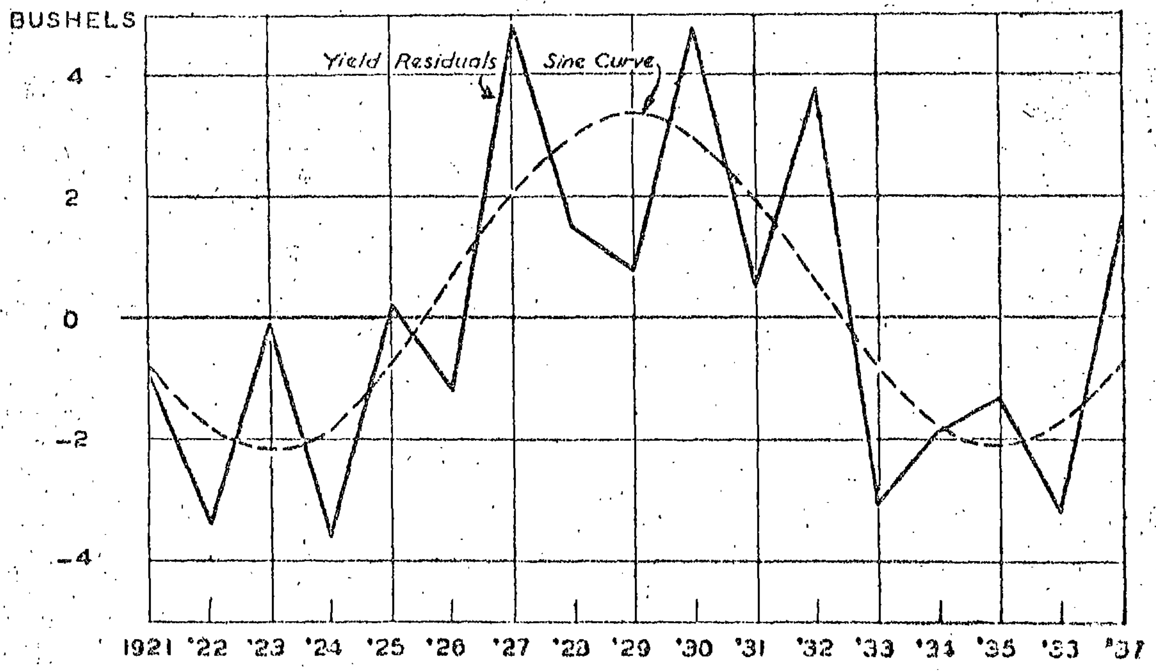


Fig. 15. Saskatchewan, June 30 (Formula 8a), Net Regression of Yield on "Time," Condition at June 30, Preseasonal Precipitation, and Precipitation During the 10 Days Prior to the Report Held Constant.

Fig. 16. Alberta, June 30 (Formula 16a), Net Regression of Yield on "Time," Condition at June 30, Preseasonal Precipitation, and Precipitation during the 10 Days Prior to the Report Held Constant.

Origin, Halfway Between 1925 and 1926; Period, 12 years.

Let

r_0 be the correlation between z_2' and z_2'',
r_1 " " " " z_2' and v,
r_2 " " " " z_2'' and v,

and compare r_1 and r_2 by the t-test as before, except that $n = N-4$ in this case, instead of N-3 in the case of the simple correlation.

DESCRIPTION OF THE DATA

In table 1 are shown the basic data utilized in obtaining the forecasting formula in this study. For Saskatchewan, the yield series was taken from C.F. Wilson's study; the revised figure for 1937 was obtained from the Monthly Bulletin of Agricultural Statistics. For Alberta and Manitoba, the last published figures from the Monthly Bulletin of Agricultural Statistics were used. The revised figures for 1938 and 1939 were taken from the official crop reports, September 1939 and 1940. The data for 1940 are the preliminary estimates issued in September. The Canadian yield data refer to acreage sown.

Until 1930, the 100-percent condition of wheat was defined as the condition promising the average yield per acre of the previous ten years. Since 1931, the crop condition of 100 percent is defined more vaguely as equal to the "long-time average yield per acre."

A statement by the Dominion Bureau of Statistics regarding the probable effect of this change in definition is available only for the year 1930 17/. According to this statement, the actual condition given on the basis of the old system would have been equivalent to the following figures on the basis of the new system:

Comparison of Condition Expressed in Terms of (a) the Previous Ten-Year Average Yield and (b) the Long-Time Average Yield Per Acre

Province	Date of Report	Actual (a) (Old System)	Hypothetical (b) (New System)	Diff.
Saskatchewan	May 31	95	97	+ 2
	June 30	90	91	+ 1
	July 31	82	83	+ 1
Alberta	May 31	97	99	+ 2
	June 30	89	91	+ 2
	July 31	86	88	+ 2
Manitoba	May 31	99	93	- 6
	June 30	102	96	- 6
	July 31	100	94	- 6

Only in Manitoba is the difference appreciable. Moreover, it is uncertain whether crop correspondents paid much attention to this change in the definition of the 100-percent condition. Experience in many countries shows that whatever the basic definition adopted by the crop reporting service, the base or 100-percent condition in the mind of crop correspondents is always a fairly stable concept, corresponding approximately, not to the long-time average, but to a long-time

17/ Monthly Bulletin of Agricultural Statistics, Vol. [illegible], No. [illegible], April [illegible]

Fig. 17

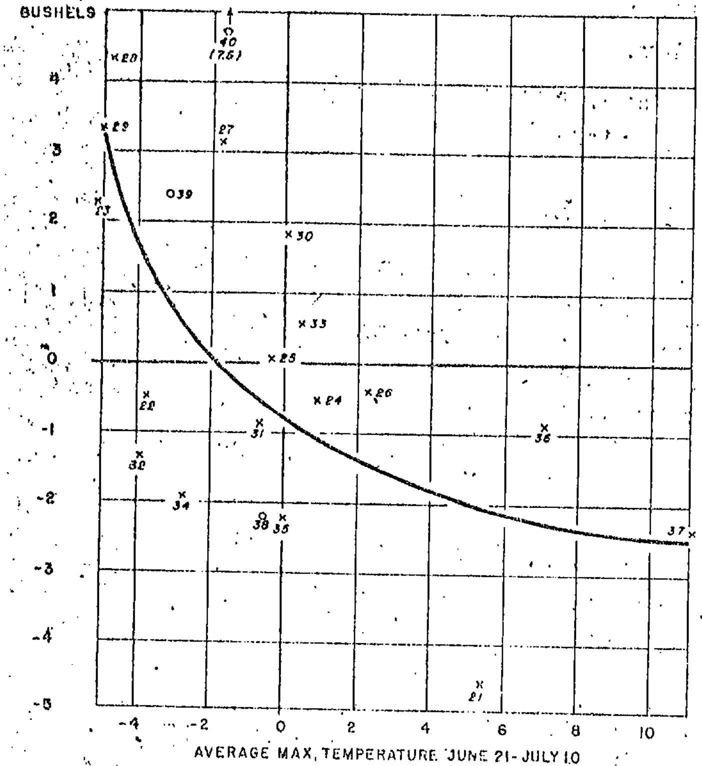


Fig. 18

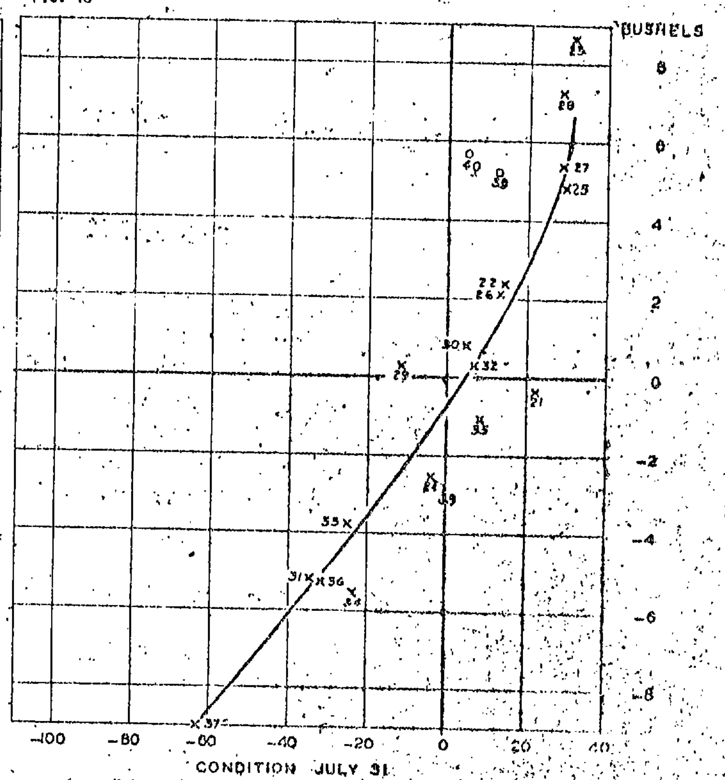


Fig. 19

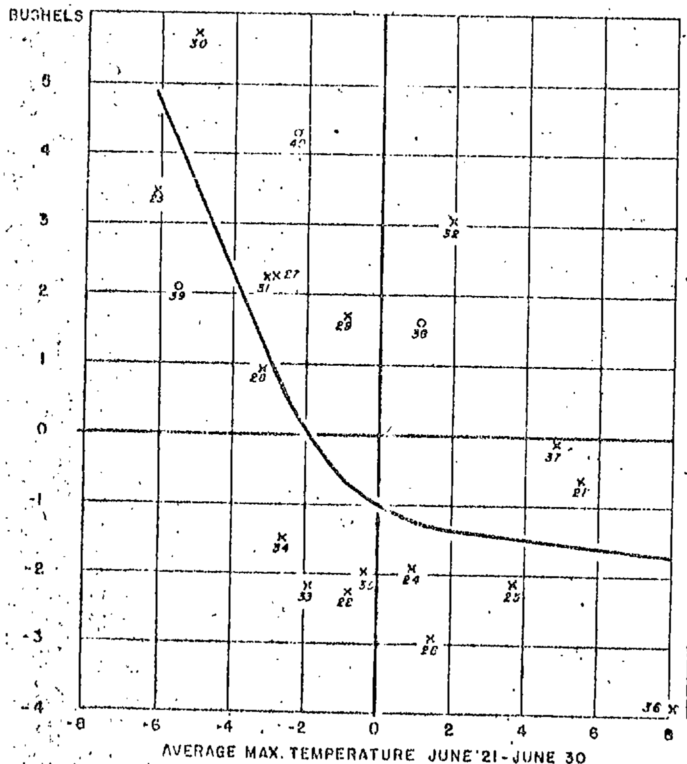
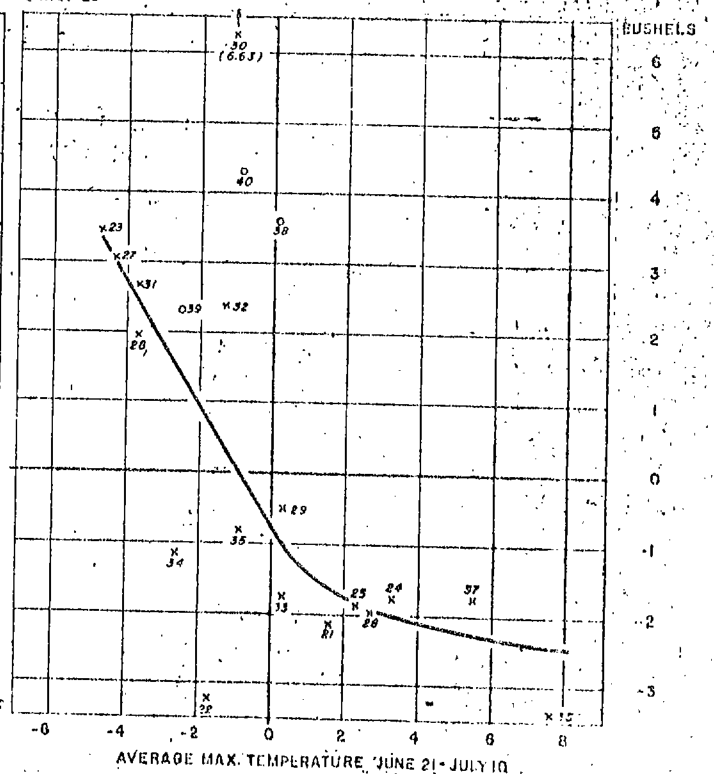


Fig. 20



[illegible]S DEPARTMENT OF AGRICULTURE NEG 43005 BUREAU OF AGRICULTURAL ECONOMICS

Fig. 17. [illegible], June 30 (Formula [illegible]), Net Regression of Yield on Average Maximum Temperature During the 20 Days Prior to the [illegible] of the Official Crop Report (July 10); Condition at June [illegible], [illegible] Precipitation, and 20 Days' Precipitation Held Constant.

Fig. 18. Saskatchewan, July 31 (Formula 8), Net Regression of Yield on Condition at July 31; [illegible] Precipitation Held Constant.

Fig. 19. Alberta, June 30 (Formula 16), Net Regression of Yield on [illegible] Temperature During the 10 Days Prior to the [illegible]; Condition [illegible], [illegible] Precipitation, and [illegible] Precipitation Held Constant.

Fig. 20. Alberta, June 30 (Formula [illegible]), Net Regression of Yield on Average Maximum Temperature During the [illegible] Days Prior to the [illegible] of the Official Crop Report; [illegible] June 30, Seasonal Precipitation, and [illegible] Held Constant.

"normal" condition which would indicate above-average yield prospects. Years of widespread or complete crop failure appear to be largely forgotten when the current crop is appraised in terms of the "normal" condition. On this assumption it was decided not to convert the figures prior to 1931 to the new basis, or vice versa.

As no data concerning subsoil moisture were available, recourse was taken to the indirect representation of variations in soil moisture by one of its determinant factors, preseasonal precipitation. Under the climatic conditions prevailing in the Prairie Provinces, the moisture content of the subsoil appears to be strongly affected by the amount of rainfall received during the months of September, October, and November of the year preceding harvest 18/. The weighted average of four stations (Swift Current, Battleford, Qu'Appelle and Moose Jaw, the latter used twice) was used in Saskatchewan. This weighting corresponds approximately to the distribution of acreage under wheat in the province of Saskatchewan in 1929 19/.

In Alberta, the simple average of four stations (Edmonton, Stettler, Calgary, Lethbridge) was used. This sample, though admittedly too small, may be considered as fairly representative of the spring wheat region of Alberta. It is true that part of the acreage that is scattered in the northern part of the province may not be adequately represented by weather conditions at Edmonton.

The rainfall records at Stettler are not complete. It was necessary, therefore, to substitute the average of two neighboring stations in the following instances:

November 1928:	Lacombe and Hanna
November 1933:	Lacombe and Coronation
November 1934:	Red Deer and Hanna
November 1935:	Lacombe and Alix
November 1936:	Lacombe and Hanna

In Manitoba, the average of two stations was used (Morden, Experimental Farm, and Minnedosa). Each of these stations is located in a center of wheat production, and was considered representative of about one-half of the wheat acreage of the province. The wheat acreage of Manitoba is much smaller than in the other two provinces, and the area devoted to wheat does not cover a large territory 20/.

18/ Obviously, factors other than September to November rainfall must also have a part in determining the soil moisture available at the start of the growing season. Unfortunately, there seemed to be no practicable way of estimating the carry-over of moisture left in the soil after the harvest of the preceding year's crop. This carry-over may be expected to depend principally on the amount of rainfall received during the preceding crop year, minus the amount lost by evaporation or removed by the preceding crop. The extent of summer-fallowing and new breaking in the previous year and other variations in farming practices are of importance in this connection (See Thysell, J. C.: "Conservation and Use of Soil Moisture at Mandan, N.D.," USDA Tech. Bull. 617, 1938.) It may be assumed, however, that the year-to-year variations in farming practices are negligible and that such changes as have taken place are in the nature of trends. In accordance with expectations no significant correlation was found to exist between yield and the current winter and early spring precipitation.

19/ Sources (acreage):
Manitoba: Province of Manitoba, Dep. of Agr., Report on Crops, 1929
Saskatchewan: Annual Report of the Dep. of Agr. of the Province of Saskatchewan, 1929
Alberta: Census returns, 1931

20/ Sources:
1920-37: Monthly Record of Meteorological Observations
1938-39: Monthly Weather Map

TABLE 2.–SUMMARY OF RESULTS – SASKATCHEWAN

Date of Report and Variables Used	Number of Formula	Number of Constants	Degrees of Freedom	Regression Coeff. or Function	Correlation Coeff. or Index			β^2
					Symbol	Uncorr.	Corr.	
May 31 x_2 = condition	1	2	15	+ .35808	r^2	.329	.284	
x_2 = condition x_3 = presea. precip.	2	3	14	+ .30973 +2.38985	R^2	.621	.567	$\beta^2_{12.3}$ = .25 $\beta^2_{13.2}$ = .20
June 30 x_2 = condition	3	3	14	$f(x_2)$	ρ^2	.624	.570	
x_2 = condition x_3 = presea. precip.	4	4	13	$f(x_2)$ +1.80257	ρ^2	.798	.751	$\beta^2_{12.3}$ = .41 $\beta^2_{13.2}$ = .18
x_2 = condition x_3 = presea. precip. x_4 = 10 days precip.	5	5	12	$f(x_2)$ +1.64862 +1.30089	ρ^2	.860	.800	$\beta^2_{12.34}$ = .53 $\beta^2_{13.24}$ = .14 $\beta^2_{14.23}$ = .04
x_2 = condition x_3 = presea. precip. x_4 = 10 days precip. x_5 = time	5a	8	9	$f(x_2)$ +1.71450 +1.98290 +.40 +2.5964 cos.$(\frac{2\pi t}{13} - \pi)$	ρ^2	.966	.940	$\beta^2_{12.345}$ = .43 $\beta^2_{13.245}$ = .15 $\beta^2_{14.235}$ = .10 $\beta^2_{15.234}$ = .11
x_2 = condition x_3 = presea. precip. x_4 = 10 days precip. x_5 = 10 days temp.	6	6	11	$f(x_2)$ +1.62578 + .68276 − .29060	ρ^2	.907	.865	$\beta^2_{12.345}$ = .43 $\beta^2_{13.245}$ = .14 $\beta^2_{14.235}$ = .01 $\beta^2_{15.234}$ = .07
x_2 = condition x_3 = presea. precip. x_4 = 10 days precip. x_5 = 10 days temp.	6a	7	10	$f(x_2)$ +1.47275 + .25432 + $f(x_5)$	ρ^2	.904	.846	$\beta^2_{12.345}$ = [illegible] $\beta^2_{13.245}$ = .11 $\beta^2_{14.235}$ = .00 $\beta^2_{15.234}$ = .11
July 31 x_2 = condition	7	3	14	$f(x_2)$	ρ^2	.859	.829	
x_2 = condition x_3 = presea. precip.	8	4	13	$f(x_2)$ +1.27024	ρ^2	.920	.914	$\beta^2_{12.3}$ = [illegible] $\beta^2_{13.2}$ = .03

Total precipitation and average maximum temperature serve to describe weather conditions during the period immediately preceding the June 30 condition report.

The meteorological stations are the same as for preseasonal precipitation. For reasons explained above (p.9), two alternative periods were used: the ten-day period from June 21 to June 30, and the 20-day period from June 21 to July 10. 21/ Precipitation was measured in total inches. In the case of temperature, average daily maximum in degrees Fahrenheit was used.

No complete records regarding temperature during the period June 21 to July 10 were available for the years 1936 and 1937. In each case, the temperature of about 2 days out of 10 had to be interpolated from neighboring days. For these same years, the observations at only two stations (Calgary and Edmonton) were available for Alberta. The average of these two stations was adjusted on the basis of the normal average maximum temperature during the period for the four stations. In the case of the precipitation data, no such adjustment was made, as the average interlocal variation of rainfall is insignificant in comparison with year-to-year variation and local variation in each year. 22/

Evaluation of the Data: There may be a question why both total precipitation and average maximum temperature should have been used to describe weather conditions during the period immediately preceding the June 30 report. These two variables are highly intercorrelated, and therefore no great importance should be attributed to the net regressions. The additive correlation method cannot always satisfactorily account for the complex interrelation of the effects of these two factors. Experiments under artificial conditions have shown that even the highest temperatures experienced in the spring wheat region do not materially affect the crop provided the moisture supply is ample. Moderately high temperatures, however, may cause considerable damage if the moisture stored in the soil is insufficient.

In spite of these considerations, it was thought that the inclusion of both precipitation and average maximum temperature in the analysis would offer certain advantages. Continuous meteorological observations covering the period 1921-37 were available for this study for a limited number of stations only. Obviously, an average of precipitation of four stations is a rather poor indicator of the amount of rainfall received over so vast a territory as the spring wheat area of Alberta or Saskatchewan. The shorter the period of time, the higher the variability among stations.

Furthermore, the rainfall received is much more variable in mid-summer than in either the spring or the fall months. The characteristic type of rainstorm in mid-summer is thermal in character; that is, the result of convexional processes that produce what are commonly known as local thunderstorms. In the other months of the year a large part of the precipitation occurs in connection with either warm or cold frontal storms and is much more uniformly distributed over an area than is the case with the convexional storms of the summertime.

21/ It will be seen below (p. 27) that, in addition to increasing the timeliness of the forecast, the inclusion of the first 10 days of July may somewhat attenuate the standard error of the computed averages of precipitation and temperature.

22/ Sources:
- 1921-35: Monthly Record of Meteorological Observations
- 1936-37: Daily Weather Map
- 1938-39: Data furnished by Meteorological Office, Dept. of Transport, Toronto, Ont.
- 1940: Daily Weather Map and Data furnished by Meteorological Office, Winnipeg, Man.

FIG. 21

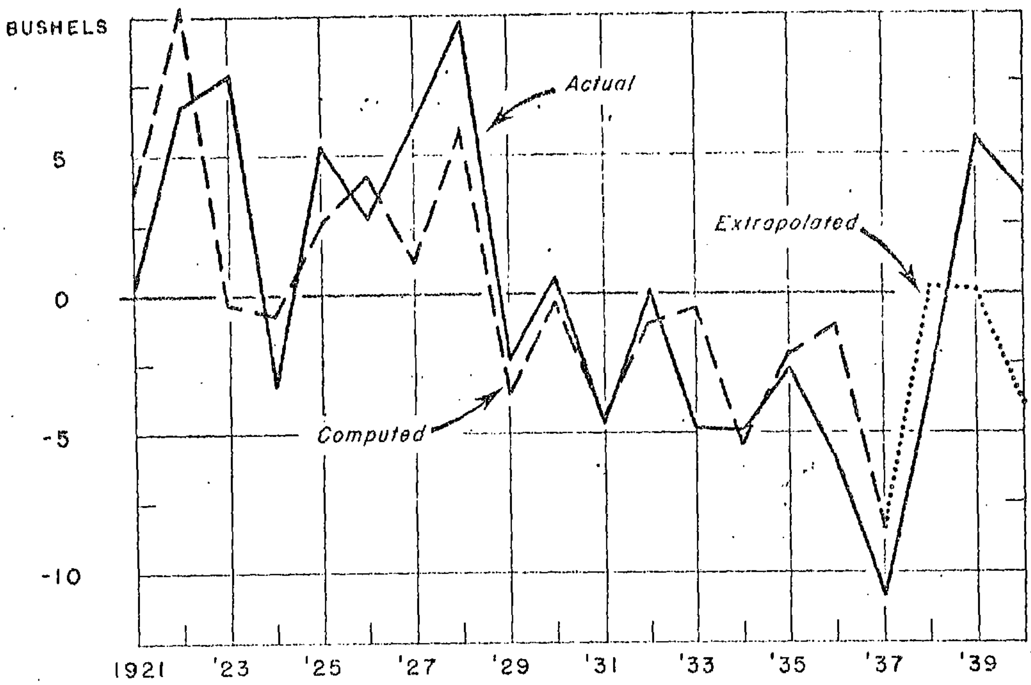


FIG. 22

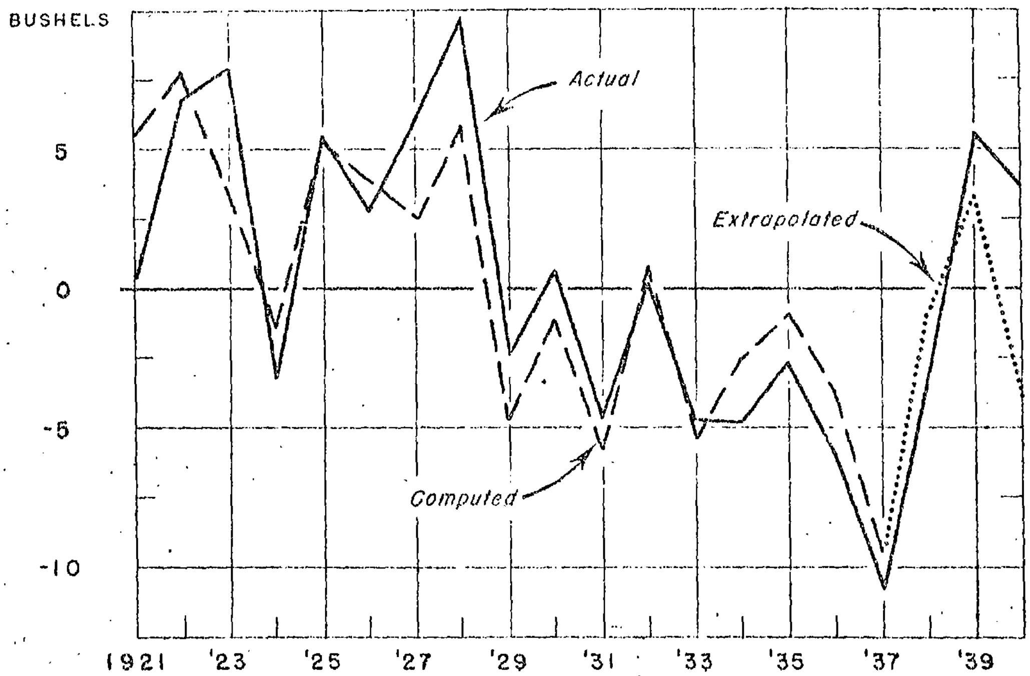


Fig. 21. Saskatchewan, May 31 (Formula 2), Actual and Computed Yield.
x_1, Yield; x_2, Condition at May 31; x_3, Presseasonal Precipitation; R^2, .631

Fig. 22. Saskatchewan, June 30 (Formula 4), Actual and Computed Yield.
x_1, Yield; x_2, Condition at June 30; x_3, Presseasonal Precipitation; R^2, .723

In view of these meteorological principles, "preseasonal rainfall" is undoubtedly a much more reliable indication of the rainfall received in the wheat areas of the three provinces than is the case with precipitation during the period June 21-30, or June 21-July 10. Rainfall data of four stations for as short a period would give an average of only very slight precision, and there is a question whether the average rainfall calculated for these short periods is at all indicative of the actual precipitation received.

On the other hand, temperature data are much less variable over a region than precipitation. In fact, maximum temperature may be a better measure of variation in the conditions that give rise to a period of localized showers than the actual measurement of precipitation at a limited number of stations. Maximum temperature is quite sensitive to cloudiness. A condition of general cloudiness and low maximum temperature would be indicative of the prevalence of localized rainstorms. In addition to being correlated with actual precipitation received, average maximum temperature is also affected by other factors which in turn influence the condition of the crop. It is very sensitive to the degree of cloudiness and thereby reflects the rate of evaporation. It seemed advisable, therefore, to include average daily maximum temperature along with precipitation during the period immediately preceding the June 30 report.

Much more satisfactory results could have been expected if direct and extensive measurements of soil moisture reserves at the start of the growing season, and complete data describing weather conditions during the growing season had been available. The inadequacy of existing meteorological data cannot be overemphasized.

However, not all the existing possibilities of representing the amount of rainfall in the Canadian spring wheat belt were exhausted in this study. The choice of a limited number of meteorological stations was governed primarily by considerations of time and the availability of current data for forecasting purposes. In the "Monthly Record of Meteorological Observations," a complete collection of meteorological data for all important stations in the Prairie Provinces is now available for the period 1921-37. For subsequent years, one must refer to the summary published in the Monthly Weather Map and the Daily Weather Map in which only a limited number of stations are carried.

The present study should be considered merely as a preliminary survey. It would be desirable to verify whether the relations found on the basis of the small number of meteorological stations used in this study will be confirmed by a similar study using the observations at a greater number of stations.

In the case of preseasonal precipitation, such a check was made possible through the kindness of Dr. Wilson of the Dominion Bureau of Statistics, who made available to us series covering the period 1921-37, based on a much larger sample of stations. Not only did no appreciable difference appear as to the size and shape of the computed regressions, but contrary to expectations, there was no significant improvement of the fit. Consequently, the average of four stations was preferred, mainly because of the accessibility of current data.

Unfortunately, a similar check was not available regarding precipitation and temperature during the period preceding the June 30 report. It is highly probable that in this case, an improvement would result from the use of a greater number of meteorological stations. When this study was nearly completed, it was learned that weekly averages of precipitation and temperature for each province are now being computed by the Meteorological Service of Canada and published currently during the growing season in the series of fifteen weekly "Telegraphic Crop Reports" issued by the Agricultural Branch of the Dominion Bureau of Statistics. It would be highly desirable to extend these series of averages to cover a sufficient number of years, and to derive new formulas on this basis.

FIG. 23

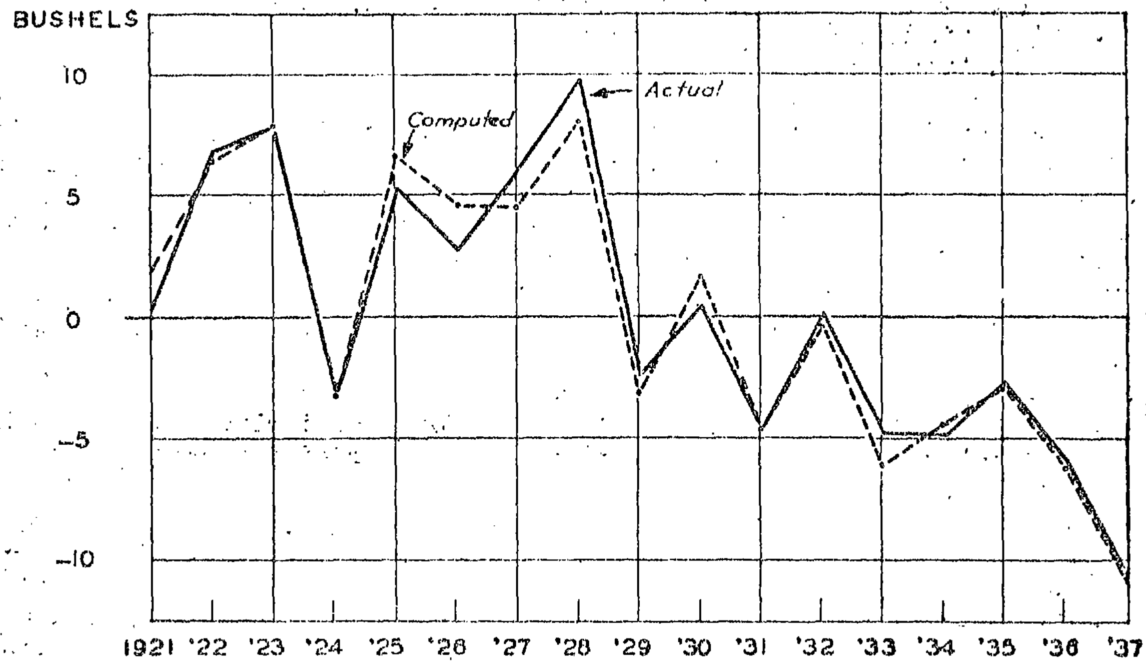


FIG. 24

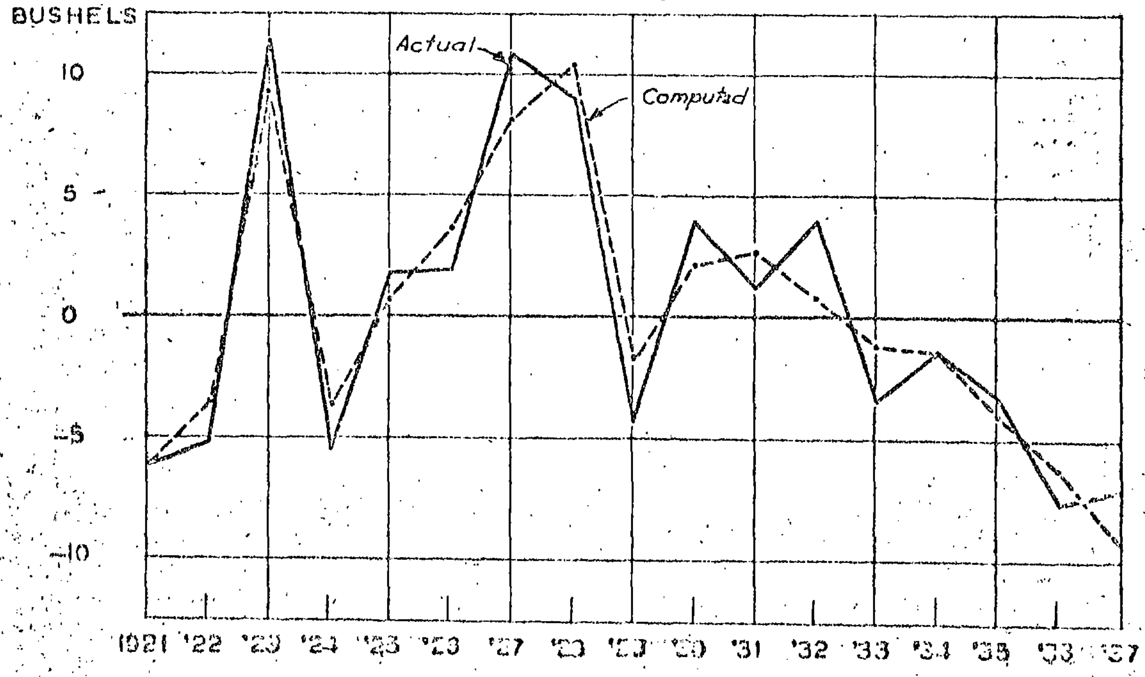


Fig. 23. Saskatchewan, June 30 (Formula 8a). Actual and Computed Yield.

x_1, Yield; x_2, Condition at June 30; x_3, Preseasonal Precipitation; x_4, Precipitation During 10 Days Prior to the Report; x_5, Sine Function of Time; R^2, .968

Fig. 24. Alberta, June 30 (Formula 16a). Actual and Computed Yield.

x_1, Yield; x_2, Condition at June 30; x_3, Preseasonal Precipitation; x_4, Precipitation During the 10 Days Prior to the Report; x_5, Sine Function of Time; R^2, [illegible]

THE STUDY

An analysis was made for each of the three prairie provinces separately: Saskatchewan, with about 55 percent of the wheat acreage in 1938; Alberta, with 32 percent; and Manitoba, with 13 percent of the acreage. The total wheat acreage in the three provinces was estimated to be about 25 million acres in 1938, compared with 26 million acres for all of Canada.

The study covers the 17-year period from 1921 to 1937, inclusive. The formulas developed on this basis were used to forecast wheat yields for 1938-40 by a process of extrapolation.

The variables used in the study (see table 1) are as follows:

(1) The dependent variable, final revised estimate of yield per acre.

(2) The reported condition at May 31, June 30 and July 31.

(3) Preseasonal precipitation—total precipitation during September to November of the previous fall.

(4) Total precipitation during the 10 days preceding the June 30 condition report—or, alternatively, during the 20 days preceding the release of the June 30 official crop report (June 21-July 10).

(5) Average of daily maximum temperatures during the 10 days preceding the June 30 condition report—or, alternatively, during the 20 days preceding the release of the June 30 official crop report (June 21-July 10).

(6) Sine function of time.

Results for Saskatchewan: The results of the study for Saskatchewan are summarized in table 2 (page 20) in which are shown the regression equations for all formulas. Thus the equation for formula 1 is: $x_1' = -+.35808x_2$. The province is by far the most important of the wheat-producing provinces of Canada. In 1938, it accounted for 55 percent of the acreage sown in spring wheat.

Condition at May 31: The reported condition at May 31, as might be expected, shows a low degree of correlation with final yield per acre ($r^2 = .329$). It is surprising to find, therefore, that when used with preseasonal precipitation, the May 31 condition gives a multiple linear correlation coefficient of $R^2 = .621$. (See fig. 21 on page 22.) The difference between the correlation obtained with condition alone ($r^2 = .329$), and that obtained by adding preseasonal precipitation as a second determining factor ($R^2 = .621$) is highly significant 23/. It is interesting to note that the factor "preseasonal rainfall," considered alone, is a better indicator of final yield than the May 31 condition report when that variable is used alone.

As the season progresses, the relative influence of preseasonal precipitation declines, while the relative importance of "condition" as an explanatory factor of wheat yields increases. In other words, the inclusion of this meteorological factor does improve the earlier forecasts more than those made later in the season.

Condition at June 30: At June 30, a rather reliable forecast of yield appears to be possible on the basis of a curvilinear interpretation of condition ([illegible]). This relationship is formula 3 in table 2. The curve is shown in fig. 1 on page [illegible]

23/ Table 2, formula 1. z = 1.1813 which would occur by chance less than once in 100 cases (the 1 percent point is z = 1.1027).

FIG. 25

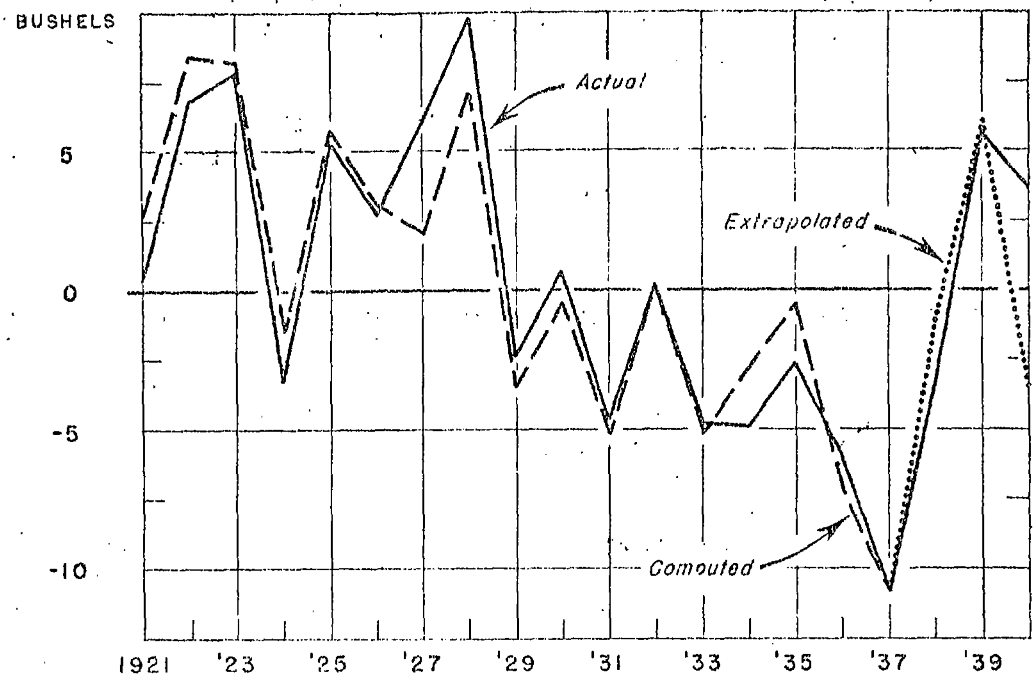


FIG. 26

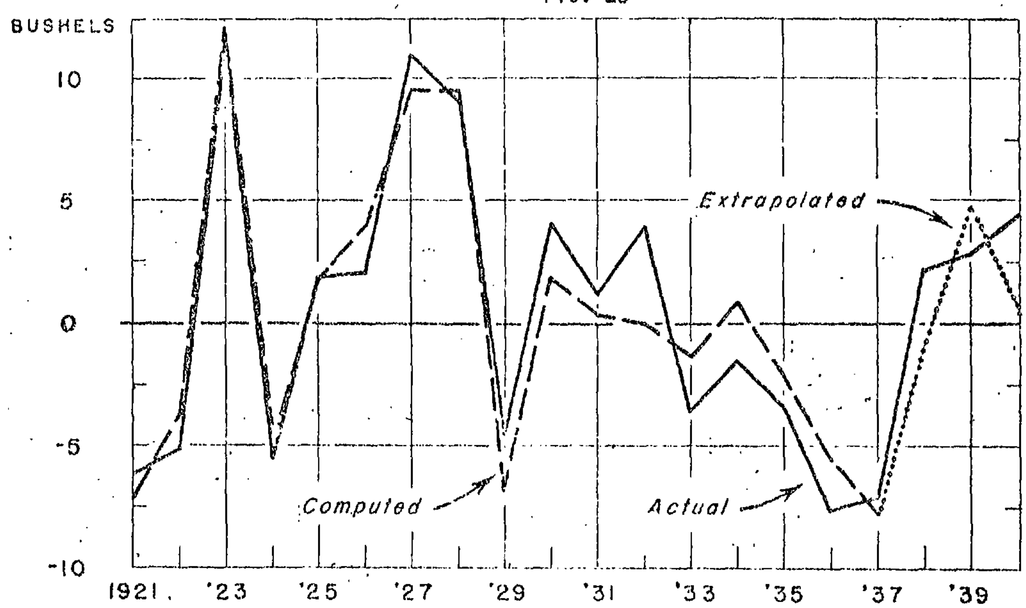


Fig. 25. Saskatchewan, June 30 (Formula [illegible]), Actual and Computed Yield.

x_1, Yield; x_2, Condition at June 30; x_3, Preseasonal Precipitation; x_4, Precipitation During 10 Days Prior to the Report; x_5, Average Maximum Temperature During the 10 Days Prior to the Report; R^2, .907

Fig. 26. Alberta, June 30 (Formula 16), Actual and Computed Yield.

x_1, Yield; x_2, Condition at June 30; x_3, Preseasonal Precipitation; x_4, Precipitation During the 10 Days Prior to the Report; x_5, Average Maximum Temperature During the 10 Days Prior to the Report; R^2, .916

In formula 3 the $f(x_2)$ represents the regression of yield on condition determined according to the "Ezekiel" method of successive approximations. This procedure is adapted to the principle of a rigid "form of hypothesis." The inclusion of preseasonal precipitation in the regression equation leads to a significant increase of the total determination (P^2 = .798). In this case z = 1.1675, which exceeds the 1 percent point (1.1160). The corrected multiple correlation index is $\bar{P}^2$ = .75; that is, about 75 percent of the variation in yields during the period covered by this analysis is associated with the variation in "condition" and preseasonal rainfall. The linear net regression coefficients is given in table 2 as formula 4; the curvilinear net regression together with the scatter of the actual observations around the regression curve, is shown in fig. 5 on page 10. In fig. 22 on page 22 the computed yields are compared to actual yields.

The inclusion of the amount of precipitation received during the 10 days prior to the condition report leads to a further increase of the squared multiple correlation index from .798 to .850. On the basis of the short series available, this increase is not clearly significant. In this case z = .6378. With 12 degrees of freedom, the .05 level of significance is .7889. The regressions are shown in table 2 as formula 5, and depicted graphically in fig. 6 on page 10.

The introduction of a fifth factor, average maximum temperature during the 10 days prior to the report (the period ending June 30) leads to an increase of the squared multiple correlation index to P^2 = .907. The regression is formula 6 and is shown in fig. 9 on page 12. The computed yields are compared with the actual yields in fig. 25 on page 26. The increase in the degree of total determination over formula 5 is significant, z being equal to .9065, which is between the .05 and .01 points (.8012 and 1.1535, respectively). It would appear, therefore, that the four determining variables--condition as of June 30, preseasonal precipitation, and rainfall and average maximum temperature during the 10 days immediately preceding the report--might be used as a basis of forecasting wheat yields as soon after June 30 as the current condition and the meteorological data can be obtained.

As previously mentioned, the official June 30 crop report of the Canadian Government is not released until about July 10. To secure a timely forecast of production on the basis of June 30 condition, it would be possible, therefore, to supplement the condition figure with weather data for the period beginning June 21 up to the day prior to the release of the official crop report. However, the correlation index obtained on this basis (formula 6a) is not significantly different from that obtained on the basis of the 10-day period. In this case P^2 = .904. The increase of the correlation index over formula 4 (based on condition and preseasonal precipitation) is not clearly significant, since z = .5988 and the 5 percent point is .6757. The net regression of yield on condition is shown in fig. 10 on page 12. In the analysis of this set of variables, the net relationship between yield and average maximum temperature during the period June 21-July 10 appeared to be curvilinear. The net regression of yield on temperature is shown in fig. 17 on page 16. The departure from linearity was significant, with z = .9266. With 9 degrees of freedom, the 5-percent point is .8163.

One may wonder why this should be so. Very likely, the explanation lies in the peculiar form of the relationship between maximum temperature and precipitation. It has been pointed out that average maximum temperature may be a better indicator of actual rainfall received over a large area than is the case of average precipitation as reported at only a few stations in that area. But the relationship between average maximum temperature and precipitation is curvilinear. With increasing maximum temperature, precipitation decreases first rapidly, then more slowly. If we assume, therefore, that actual precipitation received in the wheat area is [illegible] approached by average maximum temperature than by our "rainfall" variable, it is not surprising that this particular curvilinear relationship between [illegible] and yield should appear. In fact, it tends to corroborate this hypothesis.

TABLE 3.--SUMMARY OF RESULTS - ALBERTA

Date of Report and Variables Used	Number of Formula	Number of Constants	Degrees of Freedom	Regression Coeff. or Function	Correlation Coeff. or Index			ρ^2
					Symbol	Uncorr.	Corr.	
May 31 x_2 = condition	11	2	15		r^2	.049	0	
x_2 = condition x_3 = prescas. precip.	12	3	14	+ .22663 +2.20709	R^2	.318	.231	$\rho^2_{12.3}$ = .05 $\rho^2_{13.2}$ = .27
June 30 x_2 = condition	13	3	14	$f(x_2)$	P^2	.584	.525	
x_2 = condition x_3 = prescas. precip.	14	4	13	$f(x_2)$ + .91696	P^2	.702	.633	$\rho^2_{12.3}$ = .45 $\rho^2_{13.2}$ = .07
x_2 = condition x_3 = prescas. precip. x_4 = 10 days precip.	15	4	13	+ .12895 +1.20661 +4.42222	R^2	.770	.717	$\rho^2_{12.34}$ = .08 $\rho^2_{13.24}$ = .03 $\rho^2_{14.23}$ = .33
x_2 = condition x_3 = prescas. precip. x_4 = 10 days precip. x_5 = time	15a	7	10	+ .13846 + .70938 +4.47000 $+.22 + 2.7738 \cos\left(\frac{2\pi t}{12} - \frac{4}{3}\pi\right)$	R^2	.858	.789	$\rho^2_{12.345}$ = .10 $\rho^2_{13.245}$ = .03 $\rho^2_{14.235}$ = .23 $\rho^2_{15.234}$ = .10
x_2 = condition x_3 = prescas. precip. x_4 = 10 days precip. x_5 = 10 days temp.	16	7	10	$f(x_2)$ +1.38576 +2.34211 + $f(x_5)$	P^2	.916	.866	$\rho^2_{12.345}$ = .17 $\rho^2_{13.245}$ = .11 $\rho^2_{14.235}$ = .08 $\rho^2_{15.234}$ = .09
x_2 = condition x_3 = prescas. precip. x_4 = 20 days precip. x_5 = 20 days temp.	16a	7	10	$f(x_2)$ + .67008 +1.27451 + $f(x_5)$	P^2	.887	.820	$\rho^2_{12.345}$ = .[illegible] $\rho^2_{13.245}$ = .02 $\rho^2_{14.235}$ = .05 $\rho^2_{15.234}$ = .10
July 31 x_2 = condition	17	3	14	$f(x_2)$	P^2	.662	.614	
x_2 = condition x_3 = prescas. precip.	18	3	14	+ .20264 +1.28892	R^2	.727	.6[illegible]	$\rho^2_{12.3}$ = .[illegible] $\rho^2_{13.2}$ = .[illegible]

Similar curvilinear temperature regressions were also found in Alberta. The computed yields from formula 6a are compared with actual yields in fig. 27 on page 30.

At one stage of the analysis it appeared that the residuals of yield exhibited a pronounced systematic cyclical co-variation with time. A simple mathematical curve---a sine curve with a period of 13 years and minima in 1921 and 1934---was fitted to these residuals. This comparison is made graphically in fig. 15 on page 16. Using formula 5a a multiple correlation index of P^2 = .966 was obtained when time was included as an additional variable, along with condition at June 30, preseasonal precipitation, and precipitation for the 10 days before June 30. The net regression is shown in fig. 12 on page 12; the computed yields compared with the actual yields in fig. 23 on page 24.

As usual in similar cases, a number of possible interpretations regarding the causes of such an undulatory movement occur to the analyst, but causes determined <u>a posteriori</u> should be viewed with great skepticism.

It could be reasoned that the concept of "average" or "normal" yield or condition in the "mass" mind of crop correspondents might be less stable than is usually supposed and perhaps might be subject to cululative effects of a series of good or bad years.

In fact, the yield series itself seems to follow some kind of cyclical movement. If, then, the idea which the correspondents have of a "100-percent condition" varies, it would also follow an undulatory movement, though with a certain time lag. If these variations are not allowed for, the condition series would therefore include an underlying cyclical movement.

To investigate this problem, condition figures were expressed as percentages of the moving average of the yields of the preceding 5 (or 10) years. It was observed, then, that the undulatory movement present in the residuals was by no means syncnronous with the adjusted condition series. An economic interpretation, on the basis of farm income of the preceding years, was dismissed on the same grounds.

When the analyst has had time to overcome at least a part of the shock sustained when the large size of the correlation becomes apparent, he realizes that a squared correlation index as high as P^2 = .966 on June 30 would seem reasonable only if the weather factors influencing yield after the heading and throughout the filling stage can be expressed in terms of a sine curve similar to the one fitted to the residuals.

In fig. 13 on page 14 is shown a comparison of the (smoothed) series of wheat yields in the province of Saskatchewan with the (smoothed and inverted) series of average July-August temperatures. Both series show, besides a slow downward trend, a synchronous, undulatory movement since about 1910. Naturally, the temperature series are less regular than the yield series, for the average wheat yield is a better integrator of weather conditions prevailing in regions where wheat is grown practically everywhere, than can be the case with the average of only four thermometers.

The same undulatory movement and downward trend is present in the wheat yields for all of Canada and July-August average temperature series for the Prairie Provinces as a whole. There seems to be a strong indication---but by no means a certainty---that the undulatory movement (see fig. 14) in the residuals may be due to a similar trend in summer weather.

It is obvious, however, that no value whatever for forecasting purposes can be claimed for our formula unless it is proved that, over a considerably longer

FIG. 27

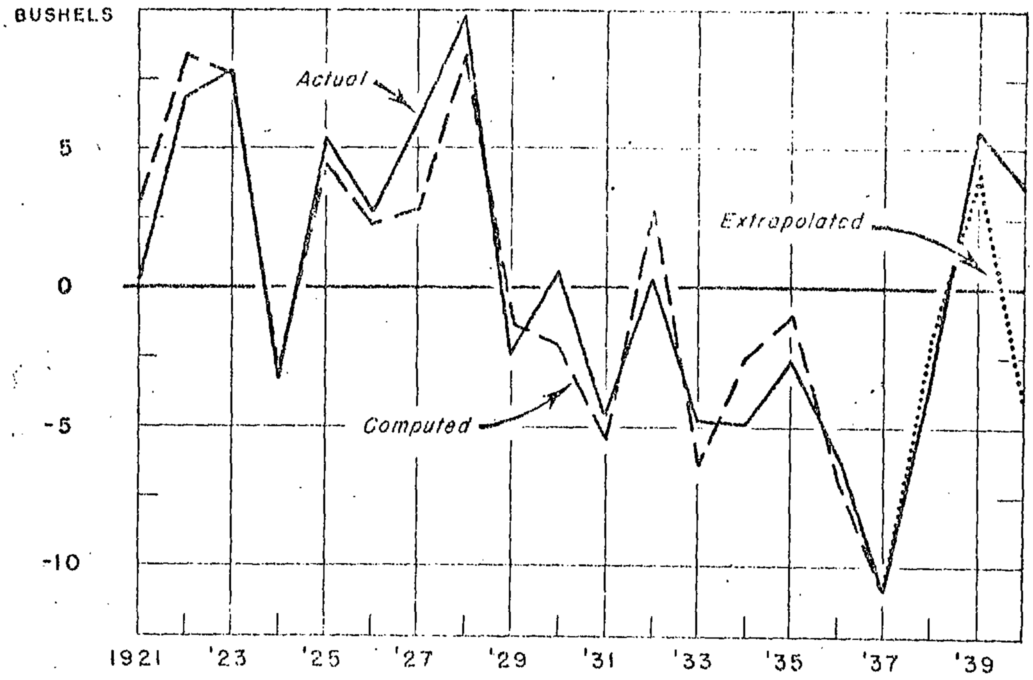


FIG. 28

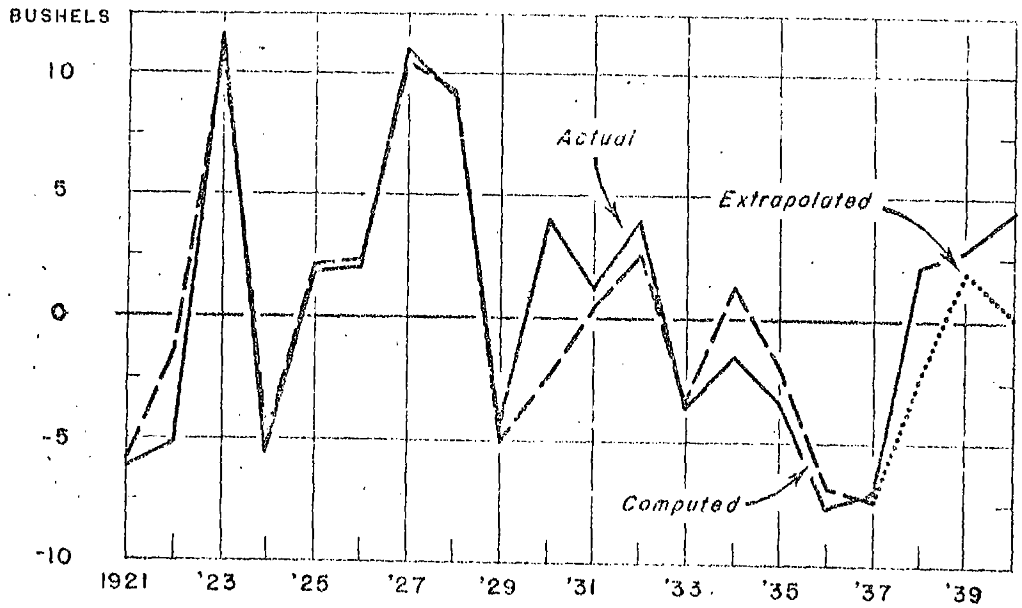


Fig. 27. Saskatchewan, June 30 (Formula 6a), Actual and Computed Yield.

x_1, Yield; x_2, Condition at June 30; x_3, Preseasonal Precipitation; x_4, Precipitation During the 20 Days Prior to the Release of the Official Crop Report; x_5, Average Maximum Temperature During the 20 Days Prior to the Release of the Official Crop Report; R^2, .904

Fig. 28. Alberta, June 30 (Formula 10a), Actual and Computed Yield.

x_1, Yield; x_2, Condition at June 30; x_3, Preseasonal Precipitation; x_4, Precipitation During the 20 Days Preceding the Release of the Official Crop Report; x_5, Average Maximum Temperature During the 20 Days Preceding the Release of the Official Crop Report; R^2, .687

series of years, yields and summer temperatures follow indeed a fairly regular cyclical movement. Unfortunately, yield series for the Prairie Provinces are not available prior to 1908. If we extrapolate the series of average July-August temperatures of the four oldest stations in western Canada (Winnipeg, Qu'Appelle, Edmonton and Calgary) beyond 1910 backward to 1873 24/, there does not remain much evidence of a regular cyclical movement of a period of 11, 12 or 13 years. We may conclude that, if the apparent periodic movement in yields is due to the effect of July-August temperature, it is no more likely to continue in the future, and the computed "prediction" formula, as well as the squared correlation index of $P^2 = .966$, are spurious and cannot be recommended for forecasting.

Moreover, although the same quasi-cyclical movement seems to be present in both series, no conclusive evidence was found of correlation between year-to-year fluctuations of yield residuals with summer temperature. Therefore, all we can say is that the two series show a certain parrallelism in their long-run movements since 1910; but there is no evidence that one is the cause of the other.

<u>Condition at July 31</u>: Condition at July 31 as reported by crop correspondents shows a curvilinear correlation of $p^2 = .859$ with final yield. The regression is shown in fig. 2 on page 6. This is formula 7, in which the $f(x_2)$ represents the regression of yield on condition only. If preseasonal precipitation is used together with condition, formula 8 is derived. The multiple correlation index increases to $P^2 = .930$. The increase is highly significant, with $z = 1.2499$, exceeding the 1 percent point (1.1166). The corrected multiple correlation index is high, ($\bar{P}^2 = .914$). The regression is shown in figure 18 on page 18 and computed yields are compared with actual yields in fig. 29 on page 32.

<u>Results for Alberta - Condition at May 31</u>: The results for Alberta are summarized in table 3. In that Province 32 percent of the wheat acreage was grown in 1938. The May 31 condition, when taken alone, did not show any correlation with final yield (formula 11). When taken along with September-November precipitation, however, it gave formula 12 with a squared correlation coefficient of $R^2 = .318$, which is at the 5-percent level of significance.

Thus, the degree of total determination is much less satisfactory than was the case in Saskatchewan.

<u>Condition at June 30</u>: Condition at June 30 provides a considerably better basis of forecasting yield than is the case a month earlier. For formula 13 the squared simple correlation index is $p^2 = .584$. The regression is shown in fig. 3 on page 6. When preseasonal precipitation is used along with June 30 condition, the resulting multiple correlation index is $P^2 = .702$. This is formula 14. The regression is shown in fig. 7 on page 10. The improvement of the fit over formula 13 (where only condition is used) is fairly significant, with $z = .7789$, which is at the 5-percent level of significance.

This prediction formula may be improved further by the introduction of the amount of precipitation received during the 10 days prior to the report. The squared correlation coefficient, which is based on the same number of constants as formula 14, is $R^2 = .770$ 25/. The new variable, considered alone, is a better indicator of final yield than reported June 30 condition when that variable is used alone.

24/ Sources: World Weather Records. Smithsonian Misc. Coll., Vols. 79 and 90; Monthly Record and Monthly Weather Map. At Winnipeg, observations are available since 1873, at Edmonton and Qu'Appelle since 1883, at Calgary since 1884. 1873-83: Adjusted to average of the four stations.

25/ Table 2. Formula 15. If a mathematical curve had been fitted to the net regression of yield on condition in formula 14, the significance of the increase of the correlation coefficient could have been established by means of Hotelling's test,

FIG. 29

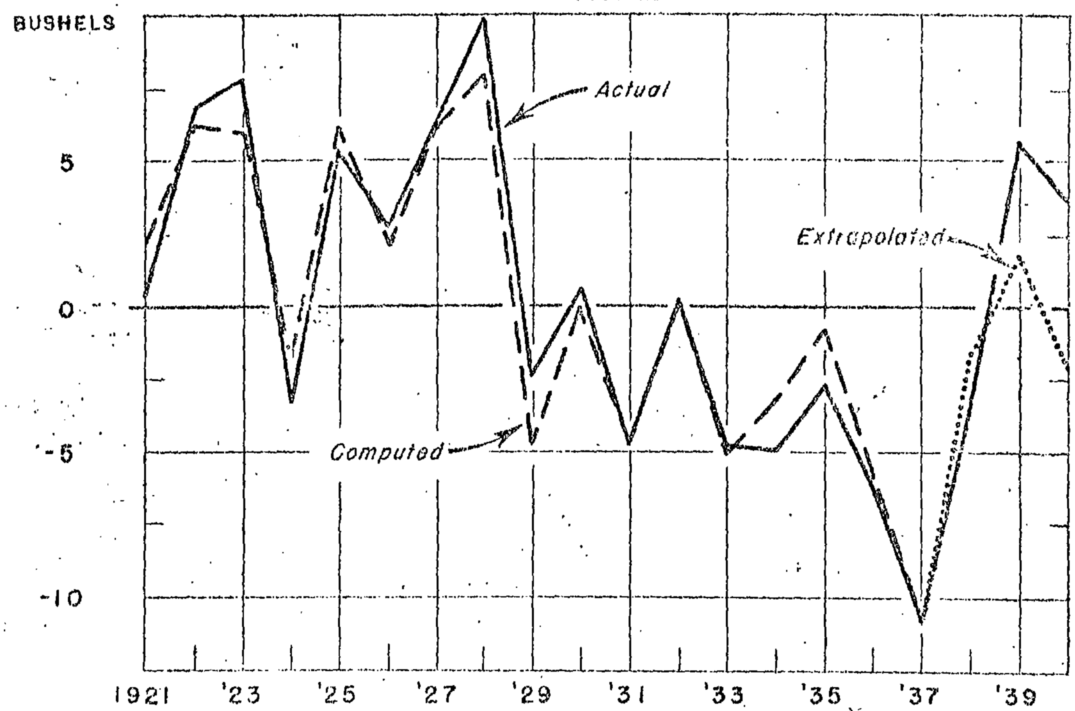


FIG. 30

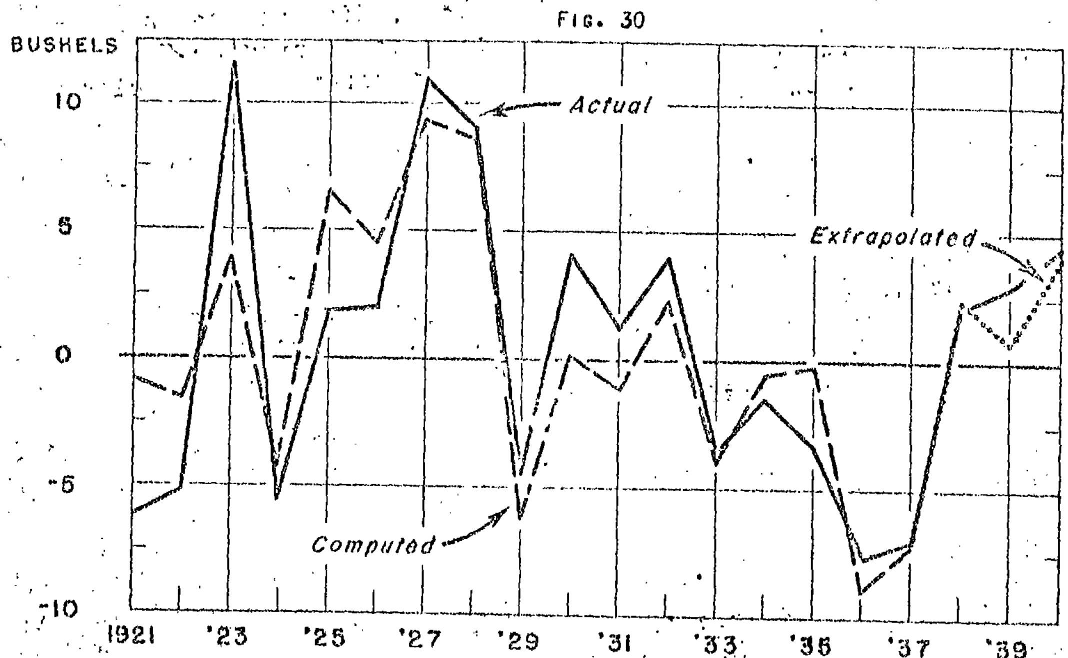


Fig. 29. Saskatchewan, July 31 (Formula 6), Actual and Computed Yield.

x_1, Yield; x_2, Condition at July 31; x_3, Preseasonal Precipitation; R^2, .930

Fig. 30. Alberta, July 31 (Formula 16), Actual and Computed Yield.

x_1, Yield; x_2, Condition at July 31; x_3, Preseasonal Precipitation; R^2, .737

The regression coefficient for the 10 days' precipitation is possibly somewhat too high, on account of one exceptionally large positive deviation in 1923. One may be inclined, therefore, to put less reliance in formulas 15 and 16 than in formula 16a, derived on the basis of 20 days' precipitation.

The introduction of average maximum temperature prevailing during the 10 days preceding the report yields formula 16 with a multiple correlation index of $P^2 = .916$. This correlation index is actually greater than the highest that could be obtained on the basis of condition as of July 31. It remains to be seen whether the June 30 formula, when extrapolated, will actually prove to be a more reliable basis of forecasting yield than that based on July 31 condition. The net regression of yield on June 30 condition is shown in fig. 8 on page 10; the net regression of yield on temperature in fig. 19 on page 18.

The increase over formula 15 is significant, with z = .8256 which is between the 5 percent (.6757) and the 1 percent (.9724) points of significance. Computed yields from formula 16 are compared with actual yields in fig. 26 on page 26.

As in Saskatchewan, a curve was fitted to the temperature regression. Although the significance of the departure from linearity was not clearly established, this relationship deserves consideration as a reasonable hypothesis for future studies as data for subsequent years become available. In this case z = .7220, which is below the 5 percent point (.8163). The fact that a curvilinear regression (fig. 20) was found in Saskatchewan as well as in Alberta is in itself a strong indication of the validity of the hypothesis. The correlation between wheat yields in Saskatchewan and in Alberta is low.

If rainfall and average maximum temperature during the 20 days preceding July 10 are substituted for weather conditions during the last 10 days of June, equation 16a with a somewhat smaller multiple correlation index is obtained ($P^2 = .887$). The net regression is shown in fig. 11 on page 12. Computed yields are compared with actual yields in fig. 28 on page 30. However, in view of the undoubtedly greater accuracy of the 20-day averages, one might be inclined to prefer this formula over formula 16. If formula 16a is compared to formula 14, it appears that the inclusion of weather conditions during the 20 days prior to July 10 has resulted in a significant increase of the total determination since z = .7953, which is between the 5 percent (.6757) and 1 percent (.9724) points of significance. The fitting of a curvilinear regression for the temperature variable gives a slightly higher correlation index than would be the case on the basis of a straight regression line, but the difference is not statistically significant.

As in Saskatchewan, the residuals left by formula 15 show a distinct oscillatory movement which was dealt with by fitting a sine curve, defined by a period of 12 years and origin halfway between 1925 and 1926. A multiple linear regression equation was fitted to the data on the basis of condition, preseasonal precipitation, precipitation during the last 10 days of June, and a sine function of time. This set of avriables yields a squared correlation coefficient $R^2 = .868$. This is formula 15a. The net regression of yield on "time" is shown in fig. 16 on page 16. Actual and computed yields are shown in fig. 24 on page 24. Nevertheless, the same reasoning as for Saskatchewan leads us to discard the formula as probably spurious and useless for forecasting purposes.

<u>Condition at July 31</u>: The squared simple correlation index between July 31 condition and final yield (formula 17) is $p^2 = .662$. The regression of yield on condition is shown in fig. 4 on page 6. When preseasonal precipitation is used together with a linear interpretation of condition, the squared correlation

---

15/Cont'd - regarding the square term of the curved regression, and the new variable, 10 days' precipitation, as "alternative" variables.

coefficient, based on the same number of constants as the index in formula 17, increases to $R^2 = .727$. This is formula 18 and the computed yields are compared with the actual yields in fig. 30 on page 32. However, this coefficient scarcely exceeds the index obtained a month earlier when the same factors are used, and it is actually smaller than the indexes computed for June 30 when weather conditions prior to the date of the report were included. No significant improvement is obtained by using July 31 condition rather than a linear interpretation of condition at June 30. ($t = 1.1395$ which, with 13 degrees of freedom is at the .30 level of significance.) As the curvilinear interpretation of June 30 condition yields a significantly higher correlation than the linear regression, the improvement due to July 31 condition becomes even less significant. The significance of the increase of the multiple correlation coefficient in formula 18 over that of formula 17 could have been established by means of Hotelling's test, if a mathematical curve had been fitted to the regression of yield on July 31 condition.

Manitoba: Whereas it appeared that for Saskatchewan and Alberta (which account for almost nine-tenths of the Canadian wheat acreage) fairly reliable forecasts of the yield of wheat might be possible more than two months prior to the harvest, on the basis of condition reports and one to three meteorological variables, no satisfactory basis has been discovered for forecasting wheat yields in Manitoba in connection either with the May 31 or June 30 condition reports.

Fortunately, it is less important to have reliable prediction formulas for Manitoba, for the following two reasons: 1. Only 13 percent of the Canadian wheat acreage is in this province; 2. the year-to year variability of yields is much smaller in Manitoba than in Saskatchewan or in Alberta. The variability of wheat yields, as measured by the mean square of the deviations from the average, is only 45 percent of that prevailing in Saskatchewan, and only 37 percent of Alberta.

Condition as of May 31 and June 30 do not show any significant relation to the final yield. Nor does the scatter diagram reveal any relation between yield and preseasonal rainfall. However, the simple correlation of yield with average maximum temperature prevailing during the 10 days prior to the June 30 report gives a corrected squared correlation coefficient of $\bar{r}^2 = .221$, which is near the .02 level of significance. Inclusion of condition as a third variable does not improve this formula significantly.

Thus, for Manitoba, no reliable forecast seems to be possible until July 31.

Several suggestions may be put forward as to why this should be so. Experience has shown that early condition reports would furnish a satisfactory basis of forecasting yields where the crop is subject to the variations of one predominant factor--as, for example, the moisture supply. Where wheat is grown under conditions approaching the meteorological optimum, yield variations become less intensive, it is true, but also less predictable, as they depend on a number of relatively minor influences--meteorological or non-meteorological. Weather conditions during the later part of the growing season and at harvest time become relatively more important.

Condition at July 31 is correlated with final yield with a squared correlation coefficient of $r^2 = .551$ ($\bar{r}^2 = .521$). The regression equation is $x_1' = +.15810\, x_2$. The inclusion of preseasonal precipitation as a second determinant variable does not lead to a significant improvement of the formula: $R^2 = .601$ ($\bar{R}^2 = .543$) and $x_1' = +.18179\, x_2 - .52373\, x_3$.

FORECASTS FOR 1938, 1939 AND 1940

Considering only the results obtained for the period covered by the study,

(1) The official condition reports, based on crop reporters' returns, which were issued by the Canadian Government as of June 30 and July 31, may be used as a basis of forecasting yield and production of wheat in the Prairie Provinces, except for Manitoba, where only the July 31 condition exhibits a sufficiently high correlation with final yield.

(2) Forecasts based on these official condition reports may be significantly improved by allowing for variations in subsoil moisture (as represented by one of its determinant factors, preseasonal precipitation) and for weather conditions during the period immediately preceding the report.

It would seem, in fact, that in Alberta, condition taken in conjunction with the above weather factors would provide a better basis of forecasting wheat yields early in July than this was possible a full month later on the basis of condition alone. In Saskatchewan, forecasts made early in July by means of the condition-and-weather formulas might be expected to be about as reliable as forecasts made early in August based on condition alone.

But in evaluating the reliability for forecasting purposes of the formulas developed in this study, it should be kept in mind that all extrapolation is based on the assumption that the relations studied will remain essentially the same as in the past, and that no new relevant factors will appear in the future.

Furthermore, it must be noted that the formulas are based mainly on the condition figures issued by the Canadian Government. Except in the case of the May 31 formulas, "condition" is the predominant factor, while the meteorological factors are of relatively minor importance. Consequently, our results cannot be expected to be accurate whenever the collective opinion of the crop correspondents falls far off the mark as an indicator of yield.

Finally, it must be remembered that the formulas derived in this study are average relationships, liable to breaking down if special circumstances prevail. Sometimes, it will be possible to detect such special circumstances in advance, if the premises on which the study is based are kept in mind. It is conceivable, for instance, that in years in which the precipitation during the growing season is so plentiful that the subsoil moisture reserves are not tapped by the plant, the influence of the variable "preseasonal precipitation" will be negligible.

The formulas derived in the present study have been used as a basis of forecasting wheat yields in Saskatchewan and Alberta in 1938, 1939 and 1940. The results of the extrapolation for Saskatchewan are shown in table 4 and for Alberta in table 5.

If one compares the charts showing the time series of actual and computed yields, it appears that, with the exception of the 1940 figures for Saskatchewan, the errors made in forecasting wheat yields are well within the range of the errors of estimate during the period covered by the analysis. In Alberta, of 22 forecasts 11 do not exceed the standard error of estimate; none exceeds twice the standard error. In Saskatchewan, of 18 forecasts made for 1938-1939, 11 do not exceed the standard error of estimate; only one is in excess of twice the standard error. Of the 9 forecasts made for 1940, however, 6 are in excess of twice the standard error.

More relevant for an evaluation of the study is the question whether the inclusion of one or more weather variables along with reported condition has actually led to an improvement of the forecasts as compared with the formulas based on "condition" alone.

TABLE 4.—FORECASTS FOR 1938, 1939, AND 1940 – SASKATCHEWAN

Date of Report, Formula, and Year		Condition		Preseasonal precipitation		10 days precipitation 1/		10 days temperature 2/	
		x_2	bx_2 or $f(x_2)$	x_3	cx_3	x_4	dx_4	x_5	ex_5 or $f(x_5)$
May 31									
1	1938	+ 5.0	+1.79						
	1939	- 2.0	-0.72						
	1940	0.0	0.00						
2	1938	+ 5.0	+1.55	- .56	-1.34				
	1939	- 2.0	-0.62	+ .29	+0.69				
	1940	0.0	0.00	-1.67	-3.99				
June 30									
3	1938	+ 4.6	-1.10						
	1939	+13.6	+4.20						
	1940	+ 1.6	-2.20						
4	1938	+ 4.6	0.00	- .56	-1.01				
	1939	+13.6	+2.70	+ .29	+0.62				
	1940	+ 1.6	-0.85	-1.67	-3.01				
5	1938	+ 4.6	+0.20	- .56	-0.92	-0.03	-0.04		
	1939	+13.6	+3.40	+ .29	+0.48	+0.77	+1.00		
	1940	+ 1.6	-0.90	-1.67	-2.75	-0.45	-0.59		
6	1938	+ 4.6	-0.60	- .56	-0.91	-0.03	-0.02	-1.1	+0.52
	1939	+13.6	+3.20	+ .29	+0.47	+0.77	+0.53	-6.1	+1.77
	1940	+ 1.6	-1.75	-1.67	-2.72	-0.45	-0.31	-4.5	+1.31
6a	1938	+ 4.6	-0.30	- .56	-0.82	-0.87	-0.22	-0.5	-0.60
	1939	+13.6	+3.00	+ .29	+0.43	-0.77	-0.20	-3.1	+0.82
	1940	+ 1.6	-1.30	-1.67	-2.46	-0.58	-0.15	-1.6	-0.10
July 31									
7	1938	- 1.4	-2.25						
	1939	+12.6	+1.40						
	1940	+ 4.6	-1.30						
8	1938	- 1.4	-0.85	- .56	-0.71				
	1939	+12.6	+1.20	+ .29	+0.37				
	1940	+ 4.6	0.00	-1.67	-2.13				

1/ For formula 6a, 20 days precipitation. 2/ For formula 6a, 20 days temperature

TABLE 4 (Cont.)--FORECASTS FOR 1938, 1939, and 1940 - SASKATCHEWAN

Date of Report, Formula, and Year		x_1^1	Computed Yield 1/	Final Official Yield	Acreage	Computed Production	Final Official Production
May 31							
1	1938	+1.8	15.3	10.0	13793	211	138
	1939	-0.7	12.8	19.1	14233	182	271
	1940	0.0	13.5	17.1	15571	210	267
2	1938	+0.2	13.7	10.0	13793	189	138
	1939	+0.1	13.6	19.1	14233	194	271
	1940	-4.0	9.5	17.1	15571	148	267
June 30							
3	1938	-1.1	12.4	10.0	13793	171	138
	1939	+4.2	17.7	19.1	14233	252	271
	1940	-2.2	11.3	17.1	15571	176	267
4	1938	-1.0	12.5	10.0	13793	172	138
	1939	+3.2	16.7	19.1	14233	238	271
	1940	-3.9	9.6	17.1	15571	149	267
5	1938	-0.8	12.7	10.0	13793	175	138
	1939	+4.9	18.4	19.1	14233	262	271
	1940	-4.2	9.3	17.1	15571	145	267
6	1938	-1.2	12.3	10.0	13793	170	138
	1939	+6.0	19.5	19.1	14233	278	271
	1940	-3.5	10.0	17.1	15571	156	267
6a	1938	-1.9	11.6	10.0	13793	160	138
	1939	+4.0	17.5	19.1	14233	249	271
	1940	-4.0	9.5	17.1	15571	148	267
July 31							
7	1938	-2.2	11.3	10.0	13793	156	138
	1939	+1.4	14.9	19.1	14233	212	271
	1940	-1.3	12.2	17.1	15571	190	267
8	1938	-1.6	11.9	10.0	13793	164	138
	1939	+1.7	15.2	19.1	14233	216	271
	1940	-2.1	11.4	17.1	15571	178	267

1/ Computed yield = $x_1^1 + \bar{x}_1$ $(\bar{x}_1 = 13.5)$

TABLE 5. - FORECASTS FOR 1938, 1939, AND 1940 - ALBERTA

Date of Report, Formula, and Year		Condition		Preseasonal precipitation		10 days precipitation 1/		10 days temperature 2/	
		x_2	bx_2 or $f(x_2)$	x_3	cx_3	x_4	dx_4	x_5	ex_5 or $f(x_5)$
May 31									
12	1938	+ 0.6	+0.14	+0.61	+1.35				
	1939	- 2.4	-0.54	-0.45	-0.99				
	1940	- 0.4	-0.09	+0.67	+1.48				
June 30									
13	1938	- 1.6	-2.85						
	1939	+12.4	+3.40						
	1940	+ 3.4	-1.40						
14	1938	- 1.6	-1.90	+0.61	+0.56				
	1939	+12.4	+3.00	-0.45	-0.41				
	1940	+ 3.4	-0.90	+0.67	+0.61				
15	1938	- 1.6	-0.21	+0.61	+0.74	+0.23	+1.02		
	1939	+12.4	+1.60	-0.45	-0.54	-0.44	-1.95		
	1940	+ 3.4	+0.44	+0.67	+0.81	-0.39	-1.72		
16	1938	- 1.6	-0.85	+0.61	+0.85	+0.23	+0.54	+1.1	-1.26
	1939	+12.4	+2.20	-0.45	-0.62	-0.44	-1.03	-5.5	+4.10
	1940	+ 3.4	+0.10	+0.67	+0.93	-0.39	-0.91	-2.3	+0.32
16a	1938	- 1.6	-1.25	+0.61	+0.41	-0.55	-0.70	+0.1	-0.85
	1939	+12.4	+2.70	-0.45	-0.30	-1.51	-1.92	-2.5	+1.40
	1940	+ 3.4	+0.10	+0.67	+0.45	-0.34	-0.43	-0.9	-0.02
July 31									
17	1938	+ 7.2	+0.55						
	1939	+ 6.2	+0.26						
	1940	+15.2	+2.95						
18	1938	+ 7.2	+1.47	+0.61	+0.79				
	1939	+ 6.2	+1.26	-0.45	-0.58				
	1940	+15.2	+3.10	+0.67	+0.66				

1/ For formula 16a, 20 days precipitation. 2/ For formula 16a, 20 days temperature.

TABLE 5. (Cont.)-- FORECASTS FOR 1938, 1939, AND 1940 - ALBERTA

Date of Report, Formula, and Year		x_1^1	Computed Yield 1/	Final Official Yield	Acreage	Computed Production	Final Official Production
May 31							
12	1938	+1.5	18.0	18.6	7969	143	148
	1939	-1.5	15.0	19.3	8367	126	161
	1940	+1.4	17.9	20.9	8667	155	181
June 30							
13	1938	-2.8	13.7	18.6	7969	109	148
	1939	+3.4	19.9	19.3	8367	166	161
	1940	-1.4	15.1	20.9	8667	131	181
14	1938	-1.3	15.2	18.6	7969	121	148
	1939	+2.6	19.1	19.3	8367	160	161
	1940	-0.3	16.2	20.9	8667	140	181
15	1938	+1.6	18.1	18.6	7969	144	148
	1939	-0.9	15.6	19.3	8367	131	161
	1940	-0.5	16.0	20.9	8667	139	181
16	1938	-0.7	15.8	18.6	7969	125	148
	1939	+4.7	21.2	19.3	8367	177	161
	1940	+0.4	16.9	20.9	8667	146	181
16a	1938	-2.4	14.1	18.6	7969	112	148
	1939	+1.9	18.4	19.3	8367	154	161
	1940	+0.1	16.6	20.9	8667	144	181
July 31							
17	1938	+0.6	17.1	18.6	7969	136	148
	1939	+0.2	16.7	19.3	8367	140	161
	1940	+3.0	19.5	20.9	8667	169	181
18	1938	+2.3	18.8	18.6	7969	150	148
	1939	+0.7	17.2	19.3	8367	144	161
	1940	+4.0	20.5	20.9	8667	178	181

1/ Computed yield = $x_1^1 + \bar{X}_1$ $(\bar{x}_1 = 16.5)$

In Alberta, the inclusion of meteorological variables led to a reduction of the error of prediction in 15 cases out of 18. 26/ For Saskatchewan, the record was not as impressive. In 1938 and 1939, the inclusion of one or more weather factors led to an improvement of the forecast in seven cases out of twelve. In 1940, however, the forecast based on condition alone---although very poor indeed---is actually better than the forecasts based on condition and one or more weather variables.

26/ In the absence of a significant correlation between May 31 condition and yield, the average yield for 1921-1937 was used to forecast 1938-1940 yields.

APPENDIX

TABLE 1 on page 5 sets forth the data series used in this study.

TABLES 2 and 3 on pages 20 and 28 give a summary of the results of this study. Column (1) shows the date of the condition report on which the various formulas are based. All data other than condition which are used in the respective equations would be available by the time the official crop report was released by the Canadian Government. As the official report was released about a week to 10 days after the date of the correspondents' reports, a forecast on the basis of the prediction formulas would not be possible before June 8-10, July 8-10, or August 8-10, respectively.

The variables used in each analysis are shown in the stub; the serial numbers of the formulas in column 1, number of constants involved in column 2, where all free-hand regression curves which were fitted to the data were assumed to absorb 2 degrees of freedom. Three constants are required to describe the sine curve in formulas 5a and 15a. The remaining degrees of freedom (number of observations minus number of constants used) are shown in column 3. All linear regression coefficients and the sine function in formulas 5a and 15a are found in column 4. The symbol f(x) refers to the semi-mathematical free-hand regressions which are shown in diagrams. The degree of total determination (square of correlation coefficient or index, uncorrected and corrected) is also shown. The symbol r^2 refers to simple linear correlation; p^2 measures simple curvilinear correlation; R^2 is used in the case of multiple linear correlation, and P^2 in the case of multiple curvilinear correlation.

TABLES 4 and 5 on pages 36, 37, 38, and 39 show the operations involved in the extrapolation of the formulas for 1938, 1939 and 1940. In the odd columns the determinant variables x_2, x_3, x_4 and x_5 are expressed in deviations from their 1921-37 average; in the even columns, their effect is expressed in terms of yield per acre, in deviations from the average. These deviations are summed up to obtain the computed yield, x_1' (in deviations from the average). The latter is added to the average yield in 1921-1937 ($\bar{x}$) to obtain the computed yield in bushels per acre. The computed yield is multiplied by the acreage (in 1000 acres) to obtain the computed production (in million bushels).

The curvilinear net regressions are illustrated in the figures. For each year, the determinant variable is plotted along the x-axis, the dependent variable (either actual yields, or yields corrected for the effect of other determinant variables, as specified in the caption) is plotted along the y-axis. Each point thus obtained is indicated by means of a small cross, except those which are the result of extrapolation (1938-40). The latter are indicated by small circles.

For formulas involving one or more weather factors, comparisons of computed and actual yields in time are illustrated in the figures. Actual yields are connected by full lines; computed yields by a dash line, except those obtained by extrapolation (1938-40) which are indicated by a dotted line.