The Pakistan Development Review
32:4 Part II (Winter 1993) pp. 799–808

Simulation Model for Distributions of Crop Yields Based on Farmer Perception–A Case Study from Highland Balochistan, Pakistan

USMAN MUSTAFA

1. INTRODUCTION

Balochistan is a vast, arid rugged, sunbaked stretch of land with multi-coloured mountain ranges and green juniper valleys. This is the largest province of Pakistan in terms of area and the smallest in terms of population. Seventy-five percent of its population earn their livelihood from agriculture and livestock. The farmers are very poor and the literacy rate in Balochistan is the lowest, only 10.06 percent and in the case of the rural areas it is further decreased to 5.61 percent. Only 35 percent of the cultivated area is irrigated [Government of Balochistan (1991)]. Balochistan has an arid or semi-arid climate with annual precipitation varying from 50 mm in the West and over 400 mm in the East. The rainfall distribution pattern is erratic with extremely high and low temperatures [Kidd *et al.* (1988)].

In such a harsh and unpredictable environment and with a very low literacy rate amongst the farmers, it is very difficult and expensive to get and generate time-series information from farmers on the distribution of field crops. Information on yield of field crops of a specific area is vital for planning and development of that area/location. The usual approach of analysing yield distributions demand an extensive data base. There is a need to provide a way to generate estimates of crop yield distributions (over time) based on farmer interviews at any particular location or area. There should be a rapid, inexpensive and objective way to measure yield gaps between the potentials (from growth model) and farmers experience. Area with large gaps can, therefore, be identified without long running trials.

In the present paper, an approach of Maerz' is discussed and tested [Maerz (1987)]. This approach defines yield distributions, and simulates random yield values which confirm to the underlying parameters, is presented for highland Balochistan. A random series of normally distributed n yield values are simulated with the Box-Muller approximation, using empirical estimates of the mean and standard deviation.

Usman Mustafa is an Agriculture Economist at Arid Zone Research Institute (AZRI), Pakistan Agriculture Research Council (PARC) at Quetta.

Author's Note: I am grateful to Dr Abelardo Rodriguez of ICARDA for his helpful comments and supplying of "MULTISIM" package.

Although these methods can only lead to approximations of reality, they have the advantage of simplicity and need only a minimum of data, obtainable for a particular area in a rapid survey of farmers [Maerz (1987)]. They also may indicate performance improvements possible from different crop cultivars the farmers are using. The alternative, of establishing agronomy and yield trials at a number of locations for several years is very expensive which could never be considered seriously. In this paper Maerz's approach is explained with respect to crop yields in a dry farming area of highland Balochistan.

2. METHODS AND ESTIMATION OF YIELD DISTRIBUTIONS FROM FARMER INTERVIEWS

Farmers were interviewed at their premises in Southern parts of Khuzdar and Kalat and Northern parts of Loralai in highland Balochistan. Annual rainfall in highland Balochistan averages 200 mm in the southern districts and 300 mm in the northern districts [Maerz (1987)]. In each area 15 farmers were randomly selected keeping in view that they followed the common farming systems, same soil types and also grow both wheat and barley on their farm. Only experienced farmers were interviewed. The data used in this analysis is a part of an extensive survey on "Farmers' Perception of Water Harvesting in Highland Balochistan" conducted by the Agricultural Economics and Farming Systems Section of AZRI during early 1992.

From each farmer two questions were asked. First how many years in ten, he would consider to be "good", "normal" or "bad" years for wheat and barley crops on his soil, according to his experience in that location (not simply the last ten years). There were no specific definitions for "good", "normal" and "bad" years and it solely depended upon farmers' perceptions. Second question was how much yields/bags per hectare of wheat and barley he associated with good, normal and poor years. The yield data were recorded from farmers even though such low yields would not be harvested as fodder/straw or grazing value of the crop.

A total picture of farmer's yield distribution, over time, can be constructed by combining their answers to the questions asked. The yields of wheat and barley for the past ten years may be expressed as a ranked array of ten values for each crop for each farmer accordingly. This is further aggregated across all farmers to calculate grand mean yield and grand sum of squares; the latter includes both the within farm variation over time and the between farm variation. Decomposing the grand sum of squares into the between farm and within farm sums of squares with a simple hierarchical, the mean sum of squares due to within farm variation can be regarded as a true estimator for the average variance resulting from year to year variability of crop yields in the area. The calculations are similar to those of one way analysis of variance except that here we are not seeking a test of significance; only the grand mean and the partitioned sum of squares are of interest [Maerz (1987)].

3. SIMULATING RANDOM YIELDS

For the simulation of random yield, it is assumed that yields in a "good" year for one crop are associated with the "good year" yields of the others grown by the farmer, and so on for "normal" and "bad year". If there were n crops, the data for each farmer would be given as n arrays of ten values, providing a framework for calculation of yield mean, variance and correlations.

A user friendly version of Maerz's (1987) multivariate simulation programme was developed at ICARDA, Aleppo [Nordblom and Shomo (1991)]. This programme "STAT1" and "MULTISIM" was used for simulation purposes. This programme finds out the vector and matrix summary statistics of the data and makes available the results on the correlation matrix of the simulation data, the variance-covariance matrix of the simulated data, and the vector of mean values. Results of the first 50 runs out of a 100 runs of wheat and barley yield simulations are presented in Figures 1 to 3.

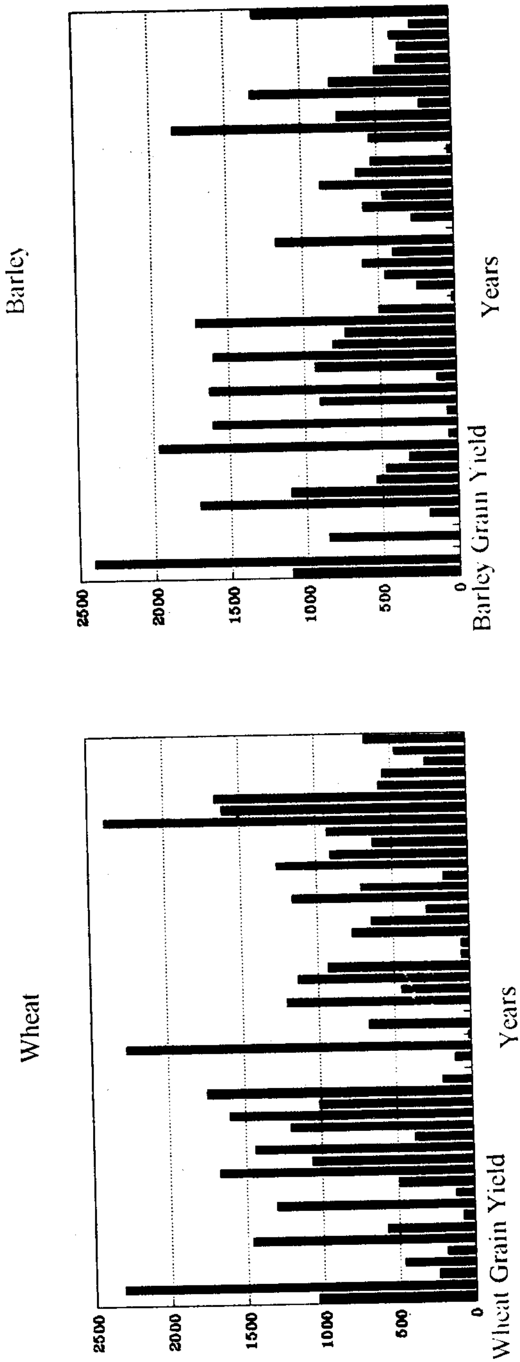
Two tables (Table 1 and Table 2) of summary statistics are presented. Table 1, reports the crop mean yields and variances-covariance matrix taken from farmer interview data and for comparison, the same kind of the summary of the simulation runs. Table 2 contains measures of errors comparing the original "farmers data" and the results of the simulation runs.

Table 1

Statistical Comparison of Original Data from Farmer Interviews and Simulation Results for Three Areas/Locations from Highland Balochistan

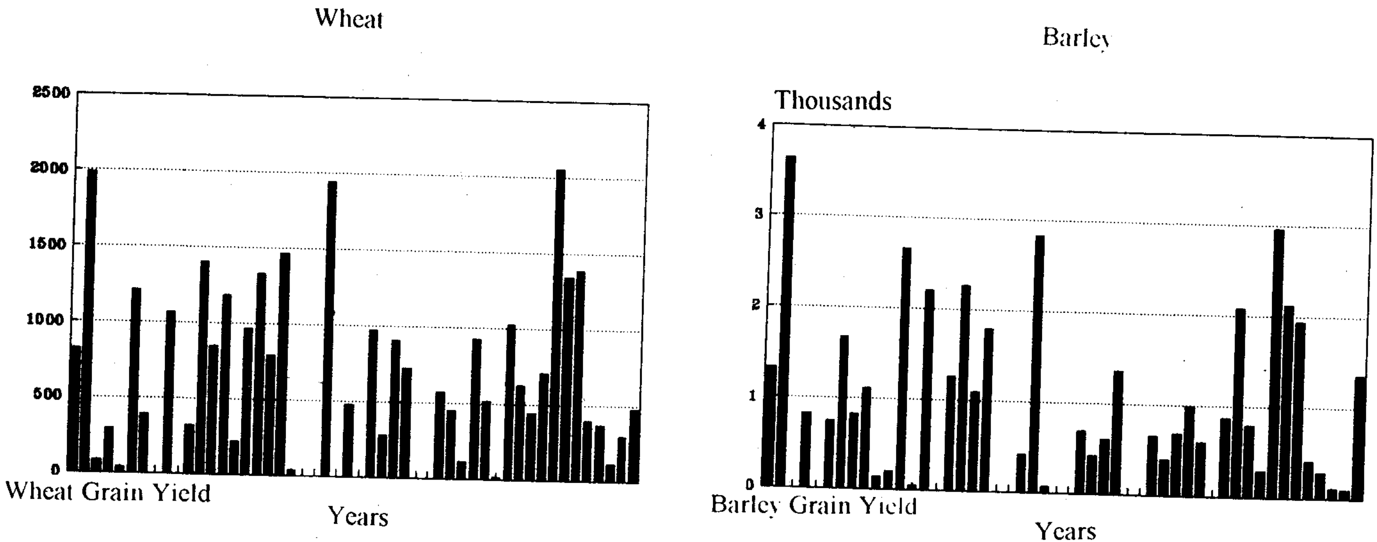
	Original Data from Farmers 100 – Run Simulation Results					
Areas/ Locations	Mean Vector	Variance-Covariance Matrix		Mean Vector	Variance-Covariance Matrix	
Loralai	Qx/ha	Wheat	Barley	Qx/ha	Wheat	Barley
Wheat	841.45	399685.5	174696.8	846.76	388870.3	172002.5
Barley	736.91		392698.4	825.41		359890.0
Khuzdar						
Wheat	635.42	335391.4	398713.5	656.76	302984.2	364233.3
Barley	837.50		869343.8	967.43		752746.8
Kalat						
Wheat	506.67	354830.5	3399026	560.52	281425.6	285025.8
Barley	562.13		439027.8	649.33		371328.6

Fig. 1.



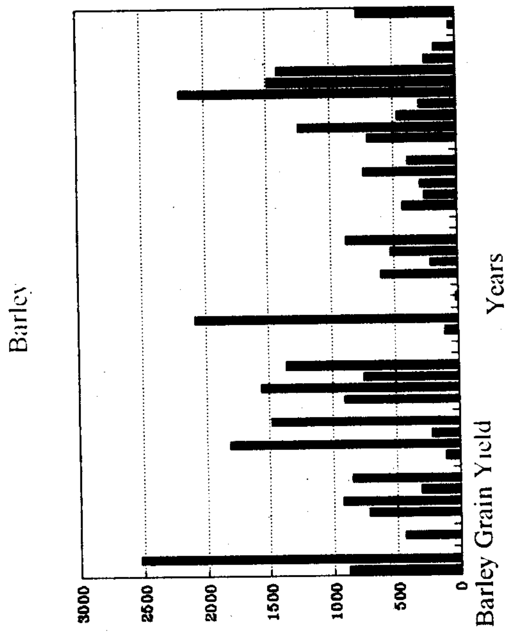
Results of 50 Runs of the Wheat and Barley Yields Simulation in Kg/Ha. (Loralai Area).

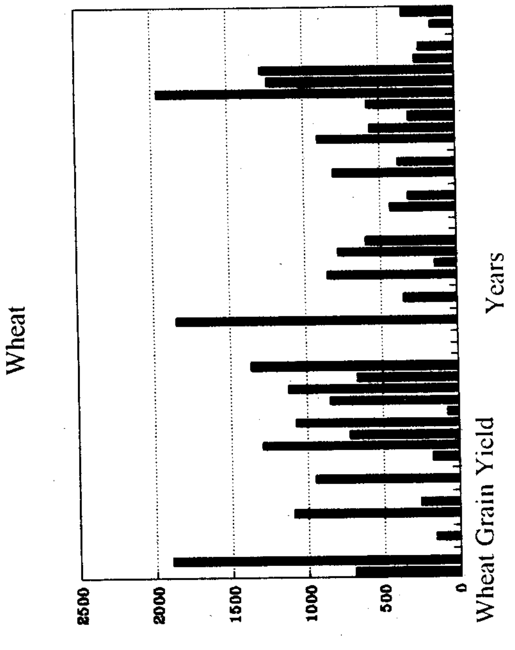
Fig. 2.



Results of 50 Runs of the Wheat and Barley Yields Simulation in Kg/Ha. (Khuzdar Area).

Fig. 3.





Results of 50 Runs of the Wheat and Barley Yields Simulation in Kg/Ha. (Kalat Area).

Table 2

Degrees of Error from Assigning Zero to Simulated Negative Values: The Mean (Ms) and Standard Deviation (Ss) of a Simulated Series, Expressed as Ratios to those of the Original Data Set Mo, (So), with the Original Coefficients of Variation (CV = So/Mo). Simulation Results for Three Area/Locations in Highland Balochistan

Area/ Locations	Original CV (= So/Mo)	Ms/Mo	Ss/So
Loralai			
Wheat	0.7352	1.006	0.991
Barley	0.6961	1.120	1.042
Khuzdar			
Wheat	0.7984	1.033	1.002
Barley	1.0014	1.155	1.221
Kalat			
Wheat	0.9090	1.106	1.064
Barley	0.9023	1.155	0.998

4. SUMMARY AND CONCLUSIONS

Maerz's method to derive and simulate crop yield distributions for use in stochastic analysis of farming systems was explained with an example from the rainfed farming areas of highland Balochistan. A user friendly version of Maerz's method developed by Nordblom and Shomo was used. The approach, of course, assumes that farmers have reliable knowledge about their own crop yields, can express these in terms of good, normal and bad yields, and that the past ten years are representative of a longer series of years. The yield of one crop is associated with the yield of the other crop. There should be a highly homogeneous class of farmers and farm conditions. This method is only applicable where there is a joint normal distribution of the data set.

Although these are very strong assumptions, the approach of estimating yield distribution parameters for the average farmer in an area, and to simulate random yield values, has some important advantages [Nordblom *et al.* (1991)]. This model is applicable where data is normally distributed, where no extensive field measurements have been made and the driving weather variables have to be generated by interpolation from other sites.

There are very few questions to be asked to the farmers, so it required very simple and short questionnaires. The data are easily collected with minimum cost and also can be easily analysed with a friendly computer package "MULTISIM".

It is a rapid, inexpensive and objective way to measure yield gaps between the potentials (from growth models) and farmers experience [Nordblom *et al.* (1991)]. Areas with large yield gaps can, therefore, be identified without long-running trials. Simulated yields are appropriate for driving stochastic whole-farm models. Where long time series of yield data are not available, empirical estimates of crop-yield distributions may be derived by interviews of farmers with long experience in that particular locality. Further improvement can be made by incorporating meteorological and other crop-yield data in the model. There is a need of developing a model which can simulate joint non-normal distribution.

REFERENCES

Balochistan, Government of (1991) Development Statistics of Balochistan 1989-90. Quetta: Bureau of Statistics. Planning and Development Department.

Kidd, C. H. R., D. J. Rees, J. D. H. Keatinge, F. Rehman, A. Samiullah and S. H. Raza (1988) Meteorological Data Analysis of Balochistan. Quetta: ICARDA, AZRI. (Research Report No. 19.)

Maerz, Ulrich (1987) Methods to Simulate Distributions of Crops Yields Based on Farmer Interviews. International Centre for Agricultural Research in the Dry Areas, ICARDA-117 En. Aleppo, Syria.

Nordblom, T., and Farouk Shomo (1991) "STAT1" and "MULTISIM" Programme. International Centre for Agricultural Research in the Dry Areas, Aleppo, Syria.

Nordblom, T., F. Shomo, H. Farihane, M. el Mourid, M. Boughlala and H. Harris (1991) Estimating the Frequency Distribution of Farmers, Crop Yields. Farm Resource Management Programme. International Centre for Agricultural Research in the Dry Areas, Aleppo, Syria. (Annual Report for 1991.)

Comments on
"Simulation Model for Distributions of Crop Yields Based on Farmer Perception–A Case Study from Highland Balochistan, Pakistan"

The authors have presented a simulation model drawing a case-study from highland Balochistan, Pakistan. The purpose of this paper is to test the Ulrich Maerz approach to simulate distributions of crop yields based on farmers' perception, especially in areas where data are not readily available. The authors have selected four locations namely Khuzdar, Kalat, parts of Loralai and Zhob in highland Balochistan. They have interviewed 68 experienced farmers and asked only two simple questions i.e. in the past ten years how a respondent farmer would consider as "good", "normal" and "bad" years for wheat and barley yields. The second question was the yield per hectare associated with these attributes. A random series of normally distributed n yield values were simulated with Box-Muller approximation, using the mean and standard deviation of farmer's perception of yields. The results did not show significant variation in mean and variance of farmers's response and simulation results except the wheat yield in Zhob.

The authors made an effort to simulate yield distribution of wheat and barley from an area of Pakistan where data are not easily available. But the assumptions of this paper are heroic. They have banked upon the memory of a few selected farmers who are knowledgeable about their farming conditions. They have not even defined good/normal and bad years. No doubt farmers know much about agro-economic conditions of their area but they are illiterate and can hardly recollect their memories for the past ten years. It is also not reasonable to assume that good/normal or bad years for wheat and barley will be the same for other crops. The authors claim of several advantages like cost-effectiveness, simple and easily to get a timely answer through a short questionnaire, surely has merit for this methodology.

For the information of the learned authors, time-series data are available for Balochistan on micro-computer discs in the Economic Wing of the Ministry of Food, Agriculture and Cooperatives which can easily be accessed for such kind of analysis. Furthermore the Government of Pakistan has introduced Area Sampling Frame (ASF) methodology which is being used by the U.S. Department of Agriculture. With this methodology the whole of the province is mapped up. This technique will provide reliable, cost-effective and timely annual data on area, production of major crops.

I ventured to use the time-series data of both crops for the past ten years (1982–91) for the districts studied by the authors and calculated the basic statistics in the attached table. If one compares these results with the one calculated by the authors, there is significant difference in both estimates. The coefficient of variation (CV) varies from (Loralai 0.16 and 0.27; Zhob 0.18 and 0.25; Kalat 0.17 and 0.27; Khuzdar 0.17 and 0.21) for wheat and barley crops respectively. The CV's estimated by the learned authors are quite high showing greater variation in time-series data generated through farmers' perceptions or simulation analysis. This is even shown in authors's graphical presentation. The learned authors's estimates are overly pessimistic for barley crop. For wheat, they have not identified whether this crop is irrigated or un-irrigated and variety is local or high-yielding or both. I have taken the average yield of wheat both for irrigated and un-irrigated areas. The mean yield of the authors is quite low, perhaps representing the un-irrigated land of a local wheat variety.

Thus data generated through such an exercise, will not be stable time-series and hence not useful for planning and policy purposes. This kind of analysis is a good academic exercise which can be used as a pedagogic device. This is a highly subjective approach which cannot be used for objective analysis.

Descriptive Statistics for Barley Yield

Districts↓	Mean	Std Dev	Variance	Min	Max	N
Loralai	1108.30	306.03	93652.01	714.0	1500.0	10
Zhob	820.60	208.44	43448.04	500.0	1000.0	10
Kalat	881.70	245.79	60413.12	500.0	1200.0	10
Khuzdar	847.00	181.06	32782.44	520.0	1160.0	10

Descriptive Statistics for Wheat Yield

Districts↓	Mean	Std Dev	Variance	Min	Max	N
Loralai	1339.20	226.76	51421.29	1022.0	1764.0	10
Zhob	1363.10	257.72	66417.66	1052.0	1773.0	10
Kalat	1378.80	246.83	60923.29	936.0	1741.0	10
Khuzdar	1324.90	234.21	54855.66	1057.0	1867.0	10

Zakir Hussain

USAID,
Islamabad.