ICARDA-117 En

METHODS TO SIMULATE DISTRIBUTIONS OF CROP YIELDS BASED ON FARMER INTERVIEWS

Ulrich Maerz

METHODS TO SIMULATE DISTRIBUTIONS OF CROP-YIELDS BASED ON FARMER INTERVIEWS

International Center for Agricultural Research in the Dry Areas

P.O. Box 5466 Aleppo, Syria

August 1987

METHODS TO SIMULATE DISTRIBUTIONS OF RAINFED CROP YIELDS BASED ON FARMER INTERVIEWS

by Ulrich Maerz*

ABSTRACT

Methods to derive and simulate crop yield distributions for use in stochastic analysis of farming systems are explained with examples from a rainfed farming area of NW Syria. MBASIC code, for implementation of the methods on micro-computers, is documented in the Appendix.

Farmers were asked about their yields per hectare in good, normal and bad years, and the frequency of such years in the past ten years. These estimates are combined to form an array of ten yield values for each farmer. The results for individual farmers are aggregated in a linear hierarchial model which allows calculation of a grand mean, a grand sum of squares, and sums of squares due to variation within farms over time and between farms. The mean sum of squares due to within-farm variation can be regarded as a true estimator for the year-to-year variability of yields per hectare in the study area: variance for the average farmer.

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. This approach also assumes, for the sake of simplicity, that crop yields follow statistically normal distributions, fully described by their means and variances.

A random series of normally distributed yield values are simulated with the Box-Muller approximation, using empirical estimates of the mean and standard deviation. This model is extended to the multivariate case of simulating correlated random series of yield values for n crops, based on a vector of empiric mean yields and an empiric variance-covariance matrix. Derivation of the latter from farmer interviews requires the additional asumption that, for each farmer, a "good year" for one crop is a "good year" for the other crops, a "normal year" for one crop is a "normal year" for the other crops, and so on.

It is shown that crop yield distributions can be reproduced in the sense that (in the parameters) the simulated distributions are not significantly different from the empiric distributions. Such simulated yields are appropriate for driving stochastic whole-farm models. Where long time-series of yield data are not available, empiric estimates of crop yield distributions may be derived from interviews of farmers with long experience in the area.

---

* Institute of Agricultural Economics and Social Sciences, Center for Tropical Agriculture, University of Hohenheim, Stuttgart, West Germany; presently with the Farm Resource Management Program, ICARDA, Aleppo, Syria.

ACKNOWLEDGEMENTS

The author is indebted to George Arab, who assissted as a guide and interpreter in the field survey, and to Thomas Nordblom, Wolfgang Goebel, Michael Jones and Peter Cooper, who provided critical comments on earlier drafts of this document. Thanks go also to Abdelkader Summakieh for collation and translation of barley yield records from the Tel Dhaman Agricultural Office in Breda.

The final form of this document has benefited from editing by Thomas Nordblom. Responsibility for the methods described, and for any errors which remain, are the author's alone.

TABLE OF CONTENTS

	page
Title	i
Abstract	iii
Acknowledgements	iv
Table of Contents	v
List of Tables and Figures	vi
A. Introduction	1
B. Estimation of Yield Distributions from Farmer Interviews	2
C. Simulating Random Yields of a Single Crop	4
D. Validation of the Single Crop Model	5
E. Extension of the Model to the Multivariate Case	11
F. A Model for the Simulation of "n" Correlated Products	12
G. Validation of the "n" Crop Model	16
H. Limitations and Conclusions	18
REFERENCES	23
APPENDICES	
Appendix 1. Single Crop Model, MBASIC program	27
Appendix 2. Cholesky's decomposition, MBASIC program	29
Appendix 3. Matrix Multiplication, MBASIC program	32
Appendix 4. n-Crop Simulation Model, MBASIC program	34
Appendix 5. Corroborating Data	38

LIST OF TABLES AND FIGURES

		Page
TABLE 1.	Characteristics of the simulated barley grain yield distribution for the average farmer in the study area, and test values of the model after 100 runs	7
TABLE 2.	Empirically estimated vector of n mean yields, and the variance-covariance and correlation matrices, based on farmer interviews	13
TABLE 3.	Vector of n mean yields, and the variance-covariance and correlation matrices, based on 100 runs of the model	19
FIGURE 1.	Flowchart of the Single-Crop Simulation Model	6
FIGURE 2.	Results of 50 runs of the Barley Yield Simulation	9
FIGURE 3.	Results of 50 Simulation runs for Random Correlated Yields of Wheat Grain, Barley Grain, Wheat Straw and Barley Straw	17

A. INTRODUCTION

Information on the distribution of yields of field crops is required for stochastic* analyses of farming systems. The usual approach of analyzing long time-series of yield data leads to good approximations of yield distributions, but demands an extensive data base. Finding an appropriate sequence of observations over a time-span of sufficient length is often problematic or impossible in some cases, and contributes to the reluctance of scientists to analyze farming systems in dynamic or stochastic terms.

In this document, an approach to define yield distributions, and simulate random yield values which conform to the underlying parameters, is presented with examples from NW Syria. First, a method is shown for the estimation of yield distribution parameters from farmer interviews. Next, a method is shown for using the empirical parameters to simulate the yields of a single crop. These methods are then extended to the multivariate case for simulation of correlated yields of "n" crops. Finally, a documented listing of programs, written in MBASIC, is given in the Appendix to facilitate implementation of the above methods on micro-computers.

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

---
* "stochastic" is a mathematical term designating a process having a progression of jointly distributed random variables.

survey of farmers. The approach is explained with respect to crop yields in a dry farming area of Aleppo province, in North-west Syria, which receives 200-275 mm annual precipitation and where barley is the most widely cultivated crop.

The total study area, extending mainly from the districts east and south of Breda village to a point about 36 km south of Breda, has approximately 3,500 farms from which a sample of 100 farms, randomly selected, provided the base for interviews. The 100 farmers kindly gave their estimates of crop yields among a larger set of questions for the author's study of farm resource management in the area.

B. ESTIMATION OF YIELD DISTRIBUTIONS FROM FARMER-INTERVIEWS

Assuming that farmers have reliable knowledge about their crop yields and can express this as estimates of yield in "good," "normal" and "bad" years as well as estimates of how often such years have occurred over time, the following two questions can be asked:

1. "How many years out of the last ten do you regard as 'good,' 'normal' and 'bad' with respect to your barley crop?"

   An example of a farmer's answer might be: 1 "good" year, 4 "normal" and 5 "bad" years.

and,

2. "How many bags of barley grain do you get per hectare when years are 'good,' 'normal' and 'bad'?"

   An example of a farmer's answer might be: 10 bags in a "good" year, 5 bags in a "normal" year and 2 bags in a "bad" year.

No special definition of good, normal and bad was imposed; this was left open to the judgement of each farmer. Combining the answers to these questions, a gross picture of the farmer's yield distribution, over time, can be constructed. Assuming the weight of a standard bag of barley grain is 110 kg, the yields of the past ten years may be expressed as a ranked array of ten values for each farmer (i.e., 1100, 550, 550, 550, 550, 220, 220, 220, 220, 200 kg/ha, using the above examples).

The ten-value arrays of each farmer are aggregated across all farmers to calculate a grand mean yield and a 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 hierarchial model described by Hartung (1985, p. 630), 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 for a 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.

For the study area, the average barley grain yield (grand mean) is estimated at about 396 kg/ha, with a standard deviation of 260 kg/ha for the average farmer, representing year-to-year variation only (the coefficient of variation is about 66%)*.

---

* for comparative estimates from independent sources, see Appendix 5.

Since the intention is to simulate a crop yield distribution, over time, the underlying statistical parameters have to be defined. Log-normal or a truncated normal distributions could possibly give better fits to the empiric data (Day, 1965). However, in this document, crop yield distributions over time are assumed to be statistically normal because the simplest methods to simulate and test random series are based on this assumption, especially in the multivariate cases discussed later.

C. SIMULATING RANDOM YIELDS OF A SINGLE CROP

The empirical yield estimates from farm interviews provided a basis for specifying the parameters ($\bar{x}$ = 396 and s = 260) of a normal distribution of barley grain yields over time for the average farmer in the study area. These parameters can be used in a simulation model to generate random "observations" which follow the same distribution. Parameters derived by other methods (i.e., from a long time-series of measured yields in the study area, if available) could also be used at this point.

A pseudo-random number generator* gives independent, uniformly distributed numbers $r_i \ldots r_{i+1}$ in the interval of 0 to 1. Two such independent uniformly distributed random numbers are required to produce a single random number (u_i) from a normal distribution with a mean of zero and a standard deviation of 1,

---

* see Fruehwirth and Regler (1983, p. 92f.) for several of the standard algorithms. Here, the MBASIC command RND is used.

N(Ø,1), using the Box-Muller approximation (Fruehwirth and Regler, 1983, p. 1Ø3-1Ø6):

$$u_i = \sqrt{-2 \ln r_i} \cdot \cos(2\pi r_{i+1})$$

Multiplying each u_i by the empiric standard deviation (26Ø), then adding the product to the empiric mean (396), normally $N(\bar{x}', s')$ distributed values, x_i, can be generated. A program, written in MBASIC computer language, for the single-crop simulation model is given in Appendix 1. A flowchart of the model is given in Figure 1.

D. VALIDATION OF THE SINGLE CROP MODEL

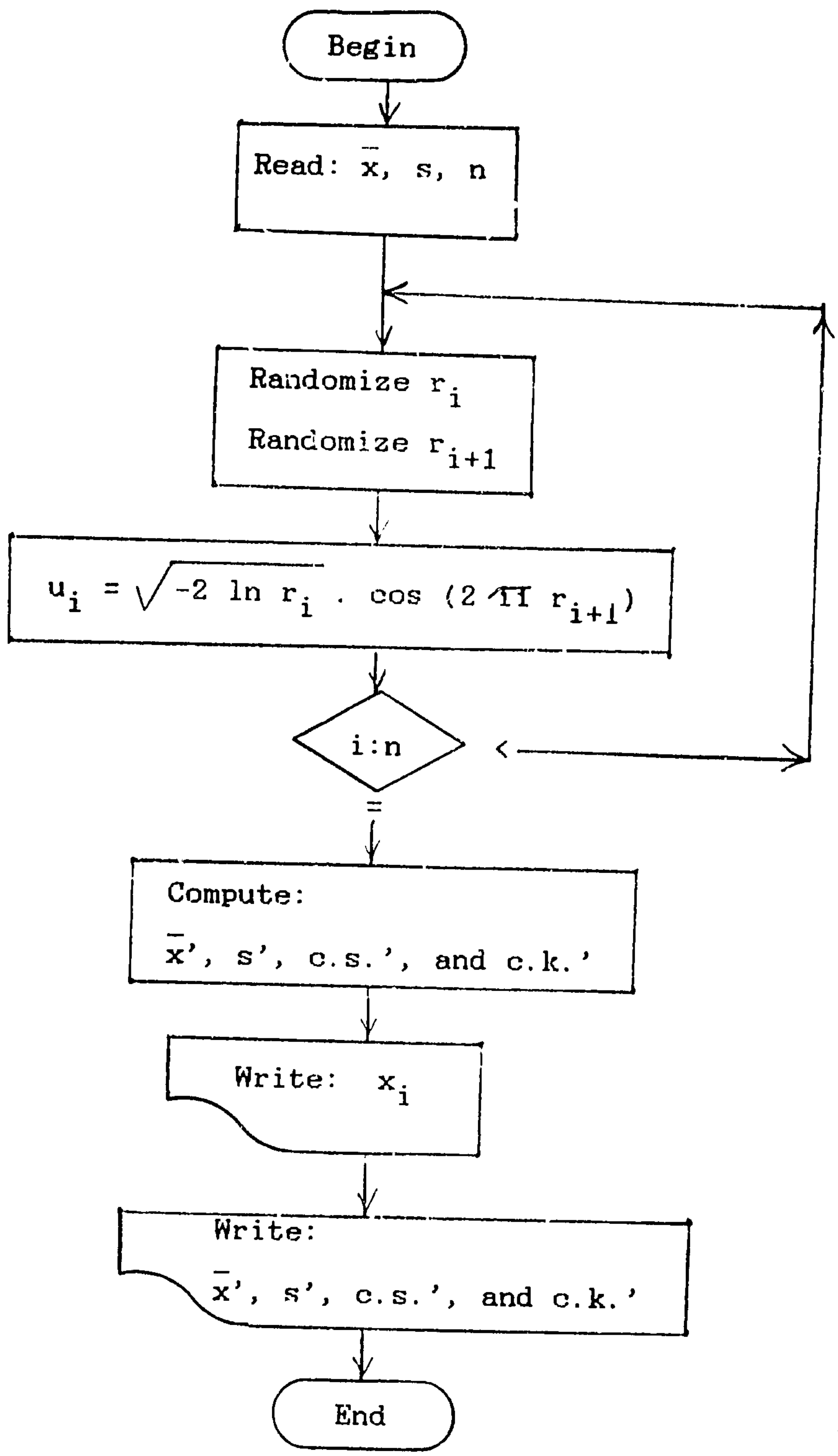
To test the model, two tests are proposed where an error level of p = Ø.1Ø is regarded as acceptable:

1. For the test on randomness of the pseudo-random numbers, Mann's test statistic C can be used (Hartung, 1985, p. 249). Such a test is necessary since a bad "seed number" can cause the pseudo-random number generator to perform poorly, especially if the number of runs is small.

2. For testing the series of pseudo-random numbers for uniform distribution, Fruehwirth and Regler (1983, p. 93) propose the Chi-square test for equal expected frequencies;

The basic characteristics of the distribution of simulated values from 1ØØ runs of the model, as well as the results of the two proposed tests, are given in Table 1. The Chi-square

FIGURE 1. Flowchart of the Single-Crop Simulation Model[1/]



1/ see Appendix 1 for MBASIC program code

TABLE 1. Characteristics of the simulated barley grain yield distribution for the average farmer in the study area, and test values of the model after 100 runs

$\bar{x}$	s 1/	c.k. 2/	c.s. 3/
363	313	3.44	0.51

Chi-square,	$r_i \ldots r_m$:	6.0	($\text{Chi-square}_{9,90}$: 14.7)
Chi-square,	$r_{i+1} \ldots r_m$:	12.8	
Mann's C-value,	$r_i \ldots r_m$:	-1.0	(u_α 4/ : 1.65)
Mann's C-value,	$r_{i+1} \ldots r_m$:	0.16	

1/ the generated standard deviation (s') differs from the empiri standard deviation (s) by 17 % here, with only 100 runs. Re-running the model 300 times reduced the difference to 0.01 %

2/ coefficient of kurtosis

3/ coefficient of skewness

4/ u_α is the quantile of the standard normal distribution N(0,1) at the significance level p = 0.90

and C values confirm (at p = .90) that the basic randomness conditions for the model are met.

The mean and standard deviation of the simultion results can be expected to converge on the underlying empiric values (396,260) as the number of runs is increased. Likewise, the coefficients of kurtosis and skewness will converge on those for a normal distribution: 3 and zero, respectively. See Table 1 for the results after only 100 runs.

Simulated barley yields from the first 50 runs are plotted in Figure 2 to illustrate the sort of variation which may be expected by the average farmer in the study area over time.

According to this model, the probability that an average farmer gets a barley grain yield less than 100 kg/ha, is about 12 %. This means in 12 out of 100 years the hypothetical farmer could not harvest more than 100 kg/ha. The lower tail of the normal distribution puts a probability of 0.065 on yields of zero "or less." Since negative yields are impossible, random values less than zero are set to zero, and the resulting error is neglected.

It is worth noting that poor barley crops have value for sheep grazing as an alternative to their harvest value. Mazid and Hallajian (1983, p. 20) estimated economic thresholds, for harvesting vs. grazing of mature barley crops, as high as 321 kg/ha in NW Syria, and 235 kg/ha for all of Northern Syria. Such high thresholds would imply the average farmer in the study area may choose to graze his barley crop in a third or

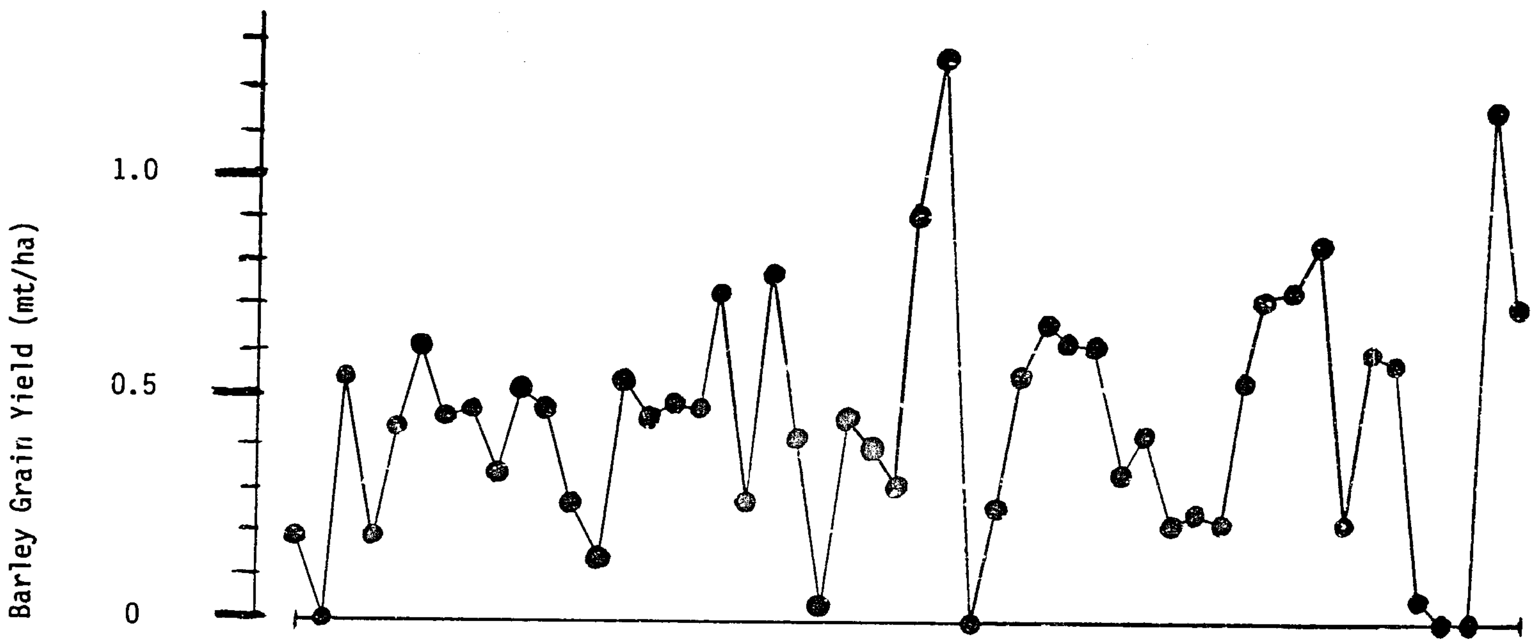


FIGURE 2. Results of 50 runs of the Barley Yield Simulation, in mt/ha.

more of the years.

Based on the model, the probability for the average farmer to get a barley grain yield of 1000 kg/ha or more is only 1.3 %. Thus, the average farmer in the study area could expect such a good year in only one or two years out of a hundred.

Individual farms in the study area will be found with much higher barley grain yields in their good years than indicated for the average farmer (see Nour and Nygaard, 1986 p.131, for an example of this in Hawaz village, near the study area).

In addition to the questions on "good", "normal" and "bad" years, the 100 farmers were asked to provide estimates on their barley yield in the harvest of 1986. The distribution of yields among the 100 farmers in 1986 can be described as follows:

- With a mean of 490 and a standard deviation of 309 kg/ha, the yields are comparable with earlier survey estimates by Somel, et al (1984, p. 59) for the 1981 and 1982 harvests in Zone 4 of NW Syria: means of 473 and 376, with standard deviations of 275 and 335 kg/ha, respectively. However, the mean of 490 kg/ha for the study area in 1986 was about 24 % higher than the average yield of the past ten years; this is consistent with opinions expressed by the majority of sampled farmers that the harvest of 1986 was "not bad."

- With a high positive coefficient of skewness of 1.00 and a coefficient of kurtosis of 3.53, the 1986 distribution cannot be regarded as normal (Kolmogorov-Smirnov D-value: 0.149).

- Yields less than 100 kg/ha were reported by about 10 % of the farmers for 1986, while grain yields above 1000 kg/ha were reported by about 5 %.

Sampling true crop yields, over a number of years, could provide a basis for testing the accuracy of the model. Unfortunately, such records are not currently available in the study area, nor in many areas in the ICARDA region where plant breeding and agronomy research are needed in the future. It is encouraging, therefore, that a rapid and inexpensive method is available, for estimating crop yield variation over time, when farmers with long experience in an area can be interviewed.

E. EXTENSION OF THE MODEL TO THE MULTIVARIATE CASE

Farmers in many places grow more than one crop each year, the yields of which are more or less correlated due to the simultaneous influence of several factors, weather in particular. A method to simulate normally distributed, correlated, random, yield values for several (n) crops is explained below.

The method is exemplified with empiric data from 18 farmers, among the original 100, who grow both wheat and barley and who reported harvest values for wheat grain (WG), barley grain (BG), wheat straw* (WST) and barley straw* (BST) for "good,"

---

* Amounts of straw collected from fields vary with method of grain harvest, with the season and with anticipated need for winter stocks of feed. Since the amounts of straw collected by farmers are not linked in exact proportion to grain yields, straw could be used in this example.

"normal" and "bad" years. An additional assumption, needed for use of interview data in the multivariate case, is 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" years. In the present case, with four crop products, the data for each farmer would be given as four arrays of ten values, providing a framework for calculation of yield means, variances (as described above) and correlations.

If "n" crop yields are correlated and should be analyzed simultaneously, multivariate statistical methods have to be applied. Analogous to the arithmetic mean and standard deviation in the single-crop model, the multivariate model requires a vector of empiric estimates of the mean values and empiric estimates for the respective variance-covariance matrix. In the present case of the four crop products, these data and the correlation matrix are given in Table 2.

The elements in the main diagonal of the variance-covariance matrix are the estimated average within-farm crop variances. The other elements are the respective covariances ($Cov_{x,y}$):

$$Cov_{x,y} = r_{x,y} \cdot \sqrt{Var_x \; Var_y}$$

where $r_{x,y}$ is the correlation coefficient of yields X and Y.

F. A MODEL FOR THE SIMULATION OF "n" CORRELATED PRODUCTS

A simulation model was written to generate random values for n correlated yield distributions whose characteristics, the mean

TABLE 2. Empirically estimated vector of n mean yields, and the variance-covariance and correlation matrices, based on farmer interviews

vector of mean yields

$$M = \begin{bmatrix} 513 \\ 46Ø \\ 5Ø2 \\ 458 \end{bmatrix} = \begin{bmatrix} \text{mean Wheat Grain (WG), in kg/ha} \\ \text{mean Barley Grain (BG), in kg/ha} \\ \text{mean Wheat Straw (WST), in kg/ha} \\ \text{mean Barley Straw (BST), in kg/ha} \end{bmatrix}$$

variance-covariance matrix E =

	WG	BG	WST	BST
WG	94771	63366	585ØØ	38591
BG		79273	68924	45213
WST			1Ø6Ø66	4Ø596
BST				4Ø647

correl-ation matrix R =

	WG	BG	WST	BST
WG	1	Ø.774	Ø.618	Ø.628
BG		1	Ø.793	Ø.8Ø5
WST			1	Ø.625
BST				1

vector (M') and variance-covariance matrix (E'), should not be significently different from the respective empiric mean vector (M) and the empiric variance-covariance matrix (E). This model generally follows the same calculation steps as outlined in Figure 1. However, several important modifications, departing from the single crop model, are applied:

- For "n" crop yields, the n-element empiric mean vector (M), has to be read in, in place of a single empiric mean, $\bar{x}$;

- in place of the single empiric standard deviation, s, an n x n dimensional triangular matrix (A) must be read in. The matrix A, however, must first be derived by decomposition of the n x n dimensional empiric variance-covariance matrix (E), by the method of Cholesky. By Cholesky's method, the matrix A is found such that $E = A^T A$, where A^T is the transpose of A (Engeln-Muellges and Reuter, 1985, p. 51, and Fruehwirth and Regler, 1983, p. 104f.).

- For n crops, 2n independent and uniformly distributed random numbers, r_i and r_{i+1}, in the interval I(0,1) have to generated for each run.

- To transform the n independent N(0,1) - normally distributed numbers, obtained by n applications of the Box-Muller approximation each run, into n $N(\bar{x}', s')$ - normally distributed, correlated random values, the single crop formula $x_i = (u_i\ s) + \bar{x}$ has to be replaced by the matrix operations:

$$x_i = u_i\ A + M$$

where A is the triangular matrix derived from the empiric variance-covariance matrix, and M is the empiric mean vector. The result, x_i, is a vector of n correlated random yield values from each run of the model.

The process can be summarized in a paragraph for the case where one wishes to generate a vector denoting four crop yields: eight independent and uniformly distributed random numbers have to be generated; these are transformed into a vector of four N(Ø,1) - normally distributed random numbers (u_i); this vector is multiplied by the triangular matrix, A, and; the resulting vector is added to the mean vector. The vector of resulting sums comprises the four random crop yields from the run. With a sufficiently large number of runs, the four random crop yield values should prove to be correlated as specified and, individually, to be following $N(\bar{x},s)$ - normal distributions.

For the reader's convenience, Cholesky's algorithm for decomposition of the variance-covariance matrix is given in the form of an MBASIC program in Appendix 2, with an example using the empiric matrix of Table 2; an MBASIC program for matrix multiplication is given in Appendix 3, using $E = A^T A$ as an example; an MBASIC program for simulation of random yield values for n crops, which are normally distributed and correlated, is given in Appendix 4. The statistical algorithms were adapted from Jacob and Jacar (1985, p. 73), Kahmann (1985, p. 1), Hartung and Elpelt (1984, p. 51) and Lillifors (1967).

G. VALIDATION OF THE "n" CROP MODEL

At this point, one could use estimates of means, variances and covariances from long-run time-series of measured yields (if such data were available), as the basis for simulation and tests of the model. Since they are not available for the study area, the author's survey estimates (in Table 2) are used.

Two tests are proposed to compare the results of the model with the empiric data:

1. For comparison of the variance-covariance matrices, to test the hypothesis

$$H_o: E = E' \quad \text{against} \quad H_1: E \neq E'$$

a modified Chi-square test, suggested by Hartung and Elpelt (1984, p.236), is applied.

2. For comparison of the mean vectors, to test the hypothesis

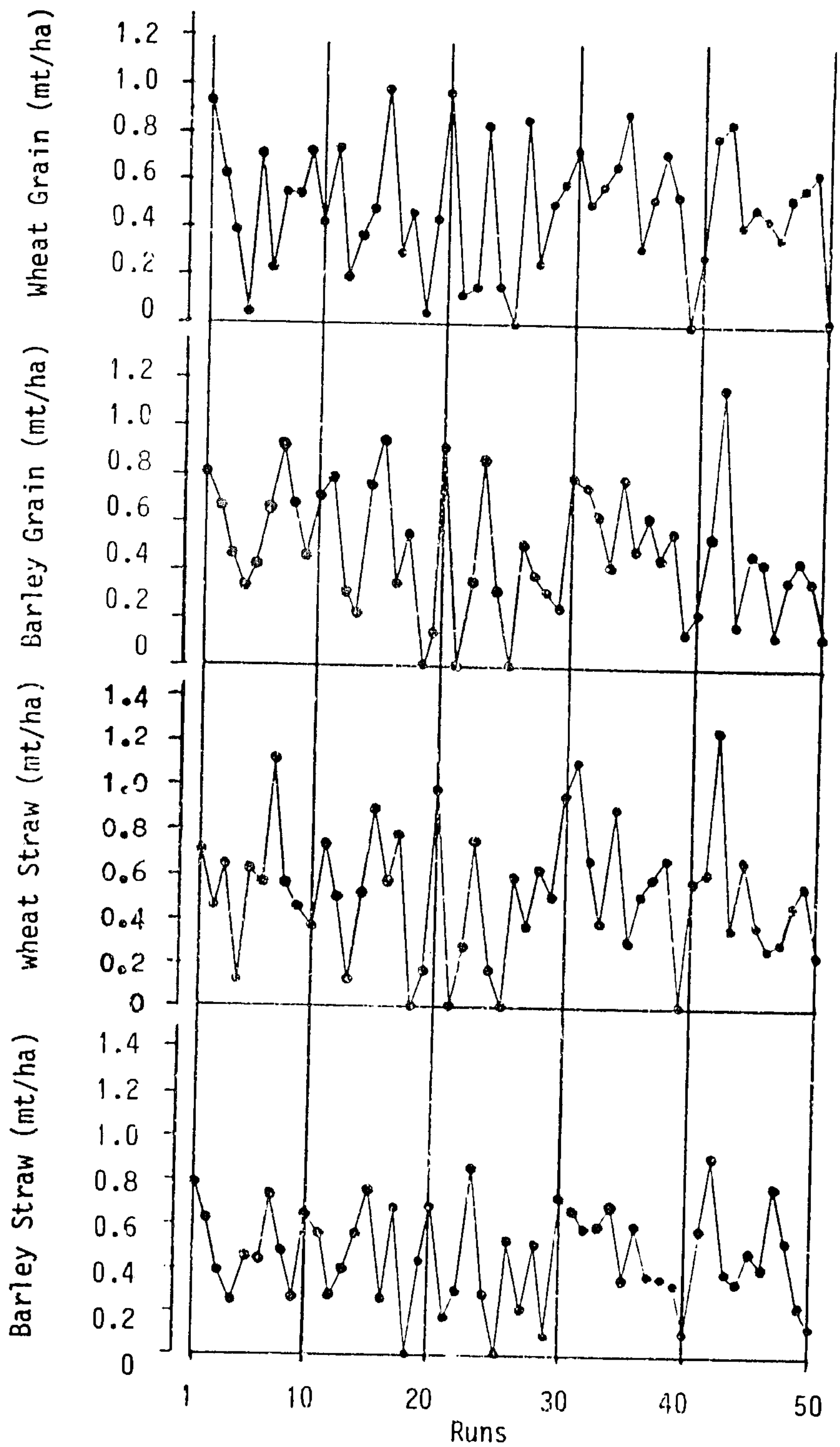
$$H_o: M = M' \quad \text{against} \quad H_1: M \neq M'$$

Hotelling's T^2 - Statistic for unequal data length can be applied (Hartung and Elpelt, 1984, p. 230).

To test the generated pseudo-random numbers for equal distribution and for randomness, the tests proposed in Section D, above, can be applied.

After 100 runs of the model, from which the results of the first 50 are shown in Figure 3, the simulated values were used to calculate a vector of mean yields (M'), a variance-covariance matrix (E') and a correlation matrix (R'), which are

FIGURE 3. Results of 50 Simulation runs for Random Correlated Yields of Wheat Grain, Barley Grain, Wheat Straw and Barley Straw, in mt/ha.



given in Table 3. These summaries of the simulation runs may be compared with the empiric parameters (in Table 2) on which the simulation was based.

The calculated Chi-square value of 1.73 is smaller than the critical value of 10.64, showing that the variance-covariance matrices E and E' are not significantly different at p = .90; Hotelling's T^2, used to compare the mean vectors M and M', shows with a calculated value of 8.2, smaller than the critical value of 12.73, that the mean vectors are not significantly different at p = .90 Thus, results of the simulation model can be regarded as reliable estimates of the four crop yield distributions whose characteristics were described simply by the mean vector of yields and a 4 x 4 dimensional variance-covariance matrix.

H. LIMITATIONS AND CONCLUSIONS

Three major limitations of the proposed methods should be kept in mind by anyone considering their use:

1. The methods of interview, and analysis of interview data, are based on the assumption that farmers, with long experience at farming in a given area, have reliable knowledge about their own crop yields and can express this information accurately in terms of "good," "normal" and "bad" crop years. This assumption cannot be verified in areas where no records of measured yields exist; therefore, over- or under-estimations of the frequency of year-types may happen. Also, a ten-year

TABLE 3. Vector of n mean yields, and the variance-covariance and correlation matrices, based on 100 runs of the model.

vector of mean yields M'=

491	=	mean Wheat Grain (WG), in kg/ha
457		mean Barley Grain (BG), in kg/ha
500		mean Wheat Straw (WST), in kg/ha
442		mean Barley Straw (BST), in kg/ha

variance-covariance matrix E'=

	WG	BG	WST	BST
WG	90698	68726	69351	47638
BG		96578	90064	57235
WST			124926	55015
BST				54578

correl-ation matrix R'=

	WG	BG	WST	BST
WG	1	0.742	0.658	0.684
BG		1	0.828	0.796
WST			1	0.673
BST				1

---

* The probabilities to get "negative yields," where the lower tails of the normal distributions cross zero, are between 1% and 6% in this example. In the summary of simulation results here, the raw data were used In Figures 2 and 3, simulated negative values were set to zero.

period may be too long for accurate recall by farmers on one hand, and on the other hand too short to adequately characterize statistical distributions of variable yields.

2. Crop yields, in general, are not normally distributed, but positively skewed. Therefore, by assuming normality, the true means and respective measures of dispersion are not precisely estimated for, or reproduced by, the simulation model.

3. By calculating yield means, and variances over time, for the average farmer, an analysis of crop yield variability for an identifiable single farmer is impossible. This dificulty may be relieved somewhat by a priori sampling from a highly homogeneous class of farmers and farm conditions, or by ex post clustering which allows the analysis to be carried out within homogeneous clusters of conditions to which identifiable farmers are associated.

Keeping in mind the above limitations, the approach of estimating yield distribution parameters for the average farmer in an area, and to simulate random yield values which follow the empirical distribution, has some important advantages for time-bounded research on farm resource management:

1. The data base needed for the simulation model is easily collected in a one-visit survey; since only a few questions have to be asked, the information is quickly collected and the costs of field work remain low. This can be very favorably contrasted, in the dimensions of research cost and timeliness, with the approach of waiting while multiple field measurements

are taken over a long series of years in a new study area. If not a replacement for the latter approach, the survey and simulation approach can provide valuable early and complementary information.

2. The structures of the single-crop and n-crop models are simple and readily implemented on micro-computers. Therefore, the models should be considered as accessable tools for most national research programs in the ICARDA region today and, in the future, increasingly so.

3. Combined with whole-farm economic models, the yield simulation models can help determine the over-time-stability of alternative organizations of single farm plans. This will be a key to the prediction of the adoption and impact of new farm technologies.

4. The yield simulation models based on farmer interviews will provide a measure of "ground truth" for the validation of crop-growth simulation models, particularly for areas where no extensive field measurements have been made and where the driving weather variables have to be generated by interpolation from other sites.

5. Simple yield simulation models may provide a link between the results of crop-growth models and yields of the other crops for which growth models are not yet available. For example, wheat crop growth may be simulated now with detailed models of the physical processes but, in some areas, wheat is

grown in rotation with lentils and water melons, two crops for which growth models do not exist. The simple yield simulation models should provide a basis for estimating yield distributions for these two crops with respect to wheat yields generated by the growth model; this would enable economic analyses in a stochastic whole-farm context.

Where wild annual fluctuations in crop yields are a very major feature of a farming system, as they are in the lower rainfall dry-farming areas of main interest to ICARDA, methods are needed for rapid and cost-effective characterization of agr-ecological conditions for research and development planning. The present survey and simulation approach is a step in that direction.

REFERENCES

Day, R. 1965. Probability distributions of field crops. Journal of Farm Economics. 47, pp. 713-741.

Engeln-Muellges, G. and R. Reuter 1985. Formelsammlung zur numerischen Mathematik mit Basic Programmen. BI-Wissenschaftsverlag, Mannheim.

Fruehwirth, R., and M. Regler 1983. Monte-Carlo-Methoden. BI-Wissenschaftsverlag, Mannheim.

Hartung, J., and B. Elpelt 1984. Multivariate Statistik. Oldenbourg Verlag, Munich.

Hartung, J. 1985. Statistik. Oldenbourg Verlag, Munich.

Jacob, P., and S. Jacar 1985. Basic Gleichungssysteme - Eigenwerte. Vieweg Verlag, Braunschweig.

Kahmann, J. 1985. Basic Programme zur numerischen Mathematik. Vieweg Verlag, Braunschweig.

Lillifors, H. 1967. On the Kolmogorov-Smirnov test on normality with mean and variance unknown. Journal of the American Statistical Association. 62, pp. 399-402.

Mazid, A., and M. Hallajian 1983. Crop-livestock interactions, information from a barley survey in Syria. Report No. 10, Farming Systems Program, ICARDA, Aleppo.

Nour, M.A., and D.F. Nygaard 1986. "Perspectives for Food Production - the Middle East and North Africa." in Nygaard, D.F. and P.L. Pellett (eds.) Dry Area Agriculture, Food Science and Human Nutrition. Proceedings of a workshop held at the University of Aleppo, Aleppo, Syria. Pergamon Press, New York.

Somel, K., A. Mazid and M. Hallajian 1984. Survey of Barley Producers, Northern Syria, 1981/1982, Volume III, Descriptive Statistics. Research Report 12-III, Farming Systems Program, ICARDA, Aleppo.

Weber, E. 1980. Grundriss der biologischen Statistik. Fischer Verlag, Stuttgart.

APPENDICES

APPENDIX 1. Simulation of a random series of normally distributed yield values for a single crop

1.1 Algorithms and formulae: see text pages 4 to 6.

1.2 Listing of an MBASIC program for the simulation of a random series of normally distributed yield values for a single crop:

```
1 HOME
2 CLEAR
3 DIM X1(5ØØ), X2(5ØØ), X3(5ØØ)
4 DIM Y(5ØØ), A(5ØØ), B(5ØØ), X(5ØØ)
5 PRINT"THIS PROGRAM SIMULATES A CROP YIELD DISTRIBUTION"
6 PRINT"BASED ON THE ASSUMPTION OF NORMALITY"
7 PRINT"THIS SIMULATION PROGRAM IS APPROPRIATE WHEN ONLY ONE"
8 PRINT"INDEPENDANT CROP YIELD IS TO BE SIMULATED"
9 PRINT"IF MORE THAN 5ØØ RUNS ARE NEEDED, THE DIMENSION"
1Ø PRINT"COMMAND IN LINE 3 HAS TO BE CHANGED"
15 PRINT:PRINT"DATA ENTRY SECTION"
2Ø PRINT:INPUT"NUMBER OF RUNS?";N
25 PRINT:INPUT"ARITHMETIC MEAN OF THE SAMPLE";M1
3Ø PRINT:INPUT"STANDARD DEVIATION OF THE SAMPLE";SS1
35 PRINT:INPUT"ENTER RANDOM SEED NUMBER TO START ALGORITHM";Z
4Ø PRINT:PRINT"CALCULATION OF THE DISTRIBUTION"
8Ø FOR I=1 TO N
9Ø A(I)=RND(I+Z):B(I)=RND(I+Z+1)
1ØØ X(I)=SQR(-2*LOG(A(I)))*COS(2*3.14159*B(I))
13Ø X(I)=M1+SS1*X(I)
17Ø NEXT I
18Ø PRINT:PRINT"CALCULATION OF THE DISTRIBUTION PARAMETERS"
19Ø FOR I=1 TO N
2ØØ AA=AA+X(I)
21Ø NEXT I
215 S1=AA/N
22Ø FOR I=1 TO N
23Ø S2=S2+(X(I)-S1)^2
24Ø S3=S3+(X(I)-S1)^3
245 S4=S4+(X(I)-S1)^4
246 NEXT I
247 ST=(S2/(N-1))^.5
25Ø CS=(S3/(N-1))/ST^3
26Ø CK=(S4/(N-1))/ST^4
```

APPENDIX 1. continued

```
27Ø REM OUTPUT SECTION
28Ø PRINT:PRINT"CROP YIELD DISTRIBUTION"
29Ø PRINT:PRINT
3ØØ PRINT TAB(1) "NR.";TAB(1Ø) "YIELD"
31Ø PRINT"------------------------------"
32Ø FOR I=1 TO N
33Ø PRINT TAB(1) I;TAB(1Ø) X(I);
34Ø NEXT I
35Ø PRINT:PRINT
36Ø PRINT"PARAMETERS OF THE DISTRIBUTION"
37Ø PRINT"MEAN:";TAB(2Ø) S1
38Ø PRINT"STD:";TAB(2Ø) ST
39Ø PRINT"C.S.:";TAB(2Ø) CS
4ØØ PRINT"C.K.:";TAB(2Ø) CK
41Ø PRINT"NR. RUNS:";TAB(2Ø) N
43Ø PRINT"DO YOU WANT TO PRINT THE OUTPUT? (Y/N):";:GET A$
44Ø IF A$="Y" THEN 45Ø ELSE END
45Ø PRINT:PRINT"DO YOU WANT TO SKIP DETAIL? (Y/N):";:GET A$
46Ø IF A$="Y" THEN GOTO 54Ø ELSE 47Ø
47Ø LPRINT"ONE-DIMENSIONAL CROP YIELD DISTRIBUTION":LPRINT
48Ø LPRINT TAB(1) "NR.";TAB(1Ø) "YIELD"
49Ø LPRINT"------------------------------"
5ØØ FOR I=1 TO N
51Ø LPRINT TAB(1) I; TAB(1Ø) X(I)
52Ø NEXT I
53Ø LPRINT
54Ø LPRINT:LPRINT"PARAMETERS OF THE DISTRIBUTION"
55Ø LPRINT"MEAN: ";S1
56Ø LPRINT"STD:  ";ST
57Ø LPRINT"C.S.: ";CS
58Ø LPRINT"C.K.: ";CK
59Ø LPRINT"NUMBER OF RUNS: ";N
6ØØ END
```

1.3 Example: see text, pages 5 - 11.

APPENIX 2. Decomposition of a matrix by the method of Cholesky

1.1 Algorithms and formulae:

The variance-covariance matrix S

$$S = \begin{bmatrix} s_{11} & s_{12} & \cdots\cdots & s_{1n} \\ s_{21} & s_{22} & \cdots\cdots & s_{2n} \\ \cdot & \cdot & & \cdot \\ \cdot & \cdot & & \cdot \\ \cdot & \cdot & & \cdot \\ s_{n1} & s_{n2} & \cdots\cdots & s_{nn} \end{bmatrix}$$

is to be decomposed into the triangular matrix A, such that $S = A^T A$ (where A^T is the transpose of A)

$$A = \begin{bmatrix} a_{11} & \cdots\cdots\cdots\cdots & a_{1n} \\ & \cdot & \cdot \\ & \quad\cdot & \cdot \\ & \qquad\cdot & \cdot \\ & & a_{nn} \end{bmatrix}$$

by the recursive formulae (Fruehwirth and Regler, 1983, p.104):

$$a_{i1} = s_{i1}/\sqrt{s_{11}} \qquad 1 \le i \le n$$

$$a_{ii} = \sqrt{(s_{ii} - \sum_{k=1}^{i-1} a_{ik}^2)} \qquad 2 \le i \le n$$

$$a_{ij} = (s_{ij} - \sum_{k=1}^{i-1} a_{ik}*a_{jk})/a_{jj} \qquad 2 \le j \le i-1$$

$$a_{ji} = 0 \qquad i+1 \le j \le n$$

APPENDIX 2. continued

2.2. Listing of an MBASIC program for the decomposition of a matrix by the method of Cholesky:

```
1 HOME
2 CLEAR
3 PRINT"THIS PROGRAM CALCULATES THE CHOLESKY DECOMPO-":PRINT
4 PRINT"SITION OF MATRIX S INTO TRIANGULAR MATRIX A.":PRINT
5 PRINT"MATRIX S MUST BE SYMETRIC AND POSITIVE DEFINIT":PRINT
6 PRINT"IF DIMENSION OF MATRIX S IS GREATER THAN":PRINT
7 PRINT"20 X 20, THEN THE DIMENSION COMMANDS IN LINES":PRINT
8 PRINT"20 AND 25 NEED TO BE CHANGED"
9 PRINT:
25 DIM S(20,20),A(20,20)
27 PRINT"YOU CAN SELECT THE DATA ENTRY OPTION OR, IF YOU":PRINT
28 PRINT"ALREADY ENTERED YOUR DATA, YOU CAN SELECT THE":PRINT
29 PRINT"CHOLESKY ALGORITHM":PRINT
30 PRINT"1)DATA ENTRY"
35 PRINT"2)CHOLESKY DECOMPOSITION"
40 PRINT"3)END"
50 INPUT"YOUR OPTION:";OP:PRINT
55 ON OP GOSUB 100,400,65
60 IF OP <> 0 THEN 30
65 END
100 REM DATA ENTRY SECTION
180 HOME:PRINT:PRINT"DATA ENTRY SECTION"
190 PRINT:INPUT"DIMENSION OF THE MATRIX=";N
200 FOR J=1 TO N
210 FOR I=1 TO N
220 PRINT"S(";I;",";J;")="
230 INPUT S(I,J):S(J,I)=S(I,J)
240 NEXT I: NEXT J
250 PRINT:RETURN
400 HOME:PRINT"CHOLESKYS TRIANGULAR MATRIX IS NOW CALCULATED"
410 GOSUB 1000
420 PRINT:IF D>0 THEN 470
430 PRINT"MATRIX NOT POSITIVE DEFINIT, THEREFORE,"
440 PRINT"CHOLESKY DECOMPOSITION NOT POSSIBLE"
450 PRINT:RETURN
470 FOR J=1 TO N
480 PRINT:PRINT"COLUMN";J
490 FOR I=1 TO N
500 PRINT"A(";I;",";J;")=";A(I,J)
510 NEXT I:NEXT J
530 LINE INPUT"OUTPUT TO PRINTER(P)?";I$
540 IF I$="P" THEN 550 ELSE RETURN
550 PRINT"MAKE PRINTER READY":LINE INPUT"PAGE HEADER:";Z$
560 LPRINT:LPRINT Z$:LPRINT
570 FOR J=1 TO N
580 LPRINT:LPRINT"COLUMN";J
```

APPENDIX 2. continued

```
590 FOR I=1 TO N
600 LPRINT "A(";I;",";J;")=";A(I,J)
610 NEXT I:NEXT J
620 PRINT:RETURN
1000 REM SUBPROGRAM CHOLESKY
1020 FOR I=1 TO N
1030 FOR J=1 TO I
1040 A(I,J)=0:A(J,I)=A(J,I)
1050 NEXT J:NEXT I
1060 D=A(1,1)
1070 IF D≤0 THEN RETURN
1080 A(1,1)=SQR(A(1,1))
1090 FOR J=2 TO N
1100 A(1,J)=A(1,J)/A(1,1)
1110 FOR K=2 TO J
1120 FOR I=2 TO K
1130 A(K,J)=A(K,J)-A(I-1,J)*A(I-1,K)
1140 NEXT I
1150 IF K=J THEN 1180
1160 A(K,J)=A(K,J)/A(K,K)
1170 GOTO 1210
1180 D=D*A(J,J)
1190 IF D≤0 THEN RETURN
1200 A(J,J)=SQR(A(J,J))
1210 NEXT K:NEXT J: RETURN
```

2.3. Example of Cholesky's decomposition:

The variance covariance matrix S

$$S = \begin{bmatrix} 94771 & 63366 & 58500 & 38591 \\ 63366 & 79273 & 68924 & 45213 \\ 58500 & 68924 & 106066 & 40596 \\ 38591 & 45213 & 40596 & 40647 \end{bmatrix}$$

is decomposed into the triangular matrix A:

$$A = \begin{bmatrix} 307.849 & 205.835 & 190.028 & 125.357 \\ 0 & 192.107 & 155.172 & 101.038 \\ 0 & 0 & 214.189 & 5.11844 \\ 0 & 0 & 0 & 121.234 \end{bmatrix}$$

such that $S = A^T A$, where A^T is the transpose of A.

APPENDIX 3. Matrix Multiplication

3.1. Algorithms and formulae (i.e., for check of results of Cholesky's decomposition in Appendix 2). The product of two matrices, A and B

$$A\,B = \begin{bmatrix} a_{11} & a_{12} & \cdots & a_{1h} \\ . & . & & . \\ . & . & & . \\ a_{n1} & a_{n2} & \cdots & a_{nh} \end{bmatrix} \begin{bmatrix} b_{11} & b_{12} & \cdots & b_{1m} \\ . & . & & . \\ . & . & & . \\ b_{h1} & b_{h2} & \cdots & b_{hm} \end{bmatrix}$$

is calculated as C (Hartung and Elpelt, 1984, p.51), where

$$C = \begin{bmatrix} \sum_{i=1}^{h} a_{1i}b_{i1} & \sum_{i=1}^{h} a_{1i}b_{i2} & \cdots\cdots & \sum_{i=1}^{h} a_{1i}b_{im} \\ \vdots & \vdots & & \vdots \\ \sum_{i=1}^{h} a_{ni}b_{i1} & \sum_{i=1}^{h} a_{ni}b_{i2} & \cdots\cdots & \sum_{i=1}^{h} a_{ni}b_{im} \end{bmatrix}$$

3.2. Listing of an MBASIC program for calculating the product of two matrices:

```
1 HOME:
2 CLEAR
4 DIM A(20,20), B(20,20), C(20,20)
5 PRINT"THIS PROGRAM CALCULATES THE PRODUCT OF 2 MATRICES,"
6 PRINT"A AND B.  IF MATRIX A OR MATRIX B HAVE MORE THAN 20"
7 PRINT"ROWS OR COLUMNS THEN THE DIMENSION COMMANDS IN"
8 PRINT"LINE 4 MUST BE CHANGED"
10 PRINT:
20 PRINT
30 PRINT"DATA ENTRY SECTION"
35 PRINT:INPUT"NUMBER OF ROWS IN MATRIX A=?";N
40 PRINT:INPUT"NUMBER OF COLUMNS (H) IN MATRIX A=?";L
45 FOR J=1 TO L
50 FOR I=1 TO N
55 PRINT"A(";I;",";J;")=";
60 INPUT A(I,J)
65 NEXT I:NEXT J
75 PRINT:INPUT"NUMBER OF COLUMNS IN MATRIX B=?";M
80 FOR K=1 TO M
85 FOR J=1 TO L
90 PRINT"B(";J;",";K;")=";
95 INPUT B(J,K)
100 NEXT J:NEXT K
101 PRINT:PRINT
```

APPENDIX 3. continued

```
102 PRINT"THE PROGRAM IS NOW MULTIPLYING THE TWO MATRICES"
107 PRINT:
11Ø FOR K=1 TO M
115 FOR I=1 TO N
12Ø S=Ø
125 FOR J=1 TO L
13Ø S=S+A(I,J)*(J,K)
135 NEXT J
14Ø C(I,K)=S
145 NEXT I
15Ø NEXT K
155 PRINT"OUTPUT SECTION"
16Ø PRINT:PRINT"THE MATRIX C, THE PRODUCT OF A X B, IS:"
161 FOR K=1 TO M
162 FOR I=1 TO N
165 PRINT"C(";I;",";K;")=";C(I,K)
17Ø NEXT I:NEXT K
175 LINE INPUT"OUTPUT TO PRINTER(P)?";I$
18Ø IF I$="P" THEN 2ØØ ELSE END
2ØØ PRINT"MAKE PRINTER READY":LINE INPUT"PAGE HEADER:";Z$
2Ø5 LPRINT:LPRINT Z$:LPRINT
2Ø6 LPRINT"PRODUCT MATRIX C"
21Ø FOR K=1 TO M
22Ø FOR I=1 TO N
23Ø LPRINT"C(";I;",";K;")=";C(I,K)
24Ø NEXT I:NEXT K
25Ø END
```

3.3 **Example:** Multiplying the matrices A^T and A from the example in Appendix 2,

$$A^T = \begin{bmatrix} 307.849 & 0 & 0 & 0 \\ 205.835 & 192.107 & 0 & 0 \\ 190.028 & 155.172 & 214.189 & 0 \\ 125.357 & 101.038 & 5.11844 & 121.234 \end{bmatrix}$$

$$A = \begin{bmatrix} 307.849 & 205.835 & 190.028 & 125.357 \\ 0 & 192.107 & 155.172 & 101.038 \\ 0 & 0 & 214.189 & 5.11844 \\ 0 & 0 & 0 & 121.234 \end{bmatrix}$$

the original variance-covariance matrix is re-calculated:

$$A^T A = C = \begin{bmatrix} 94771 & 63366.1 & 58499.9 & 38591 \\ 63366.1 & 79273.2 & 68924 & 45213 \\ 58499.9 & 68924 & 106066 & 40595.9 \\ 38591 & 45213 & 40595.9 & 40646.9 \end{bmatrix}$$

APPENDIX 4. Simulation of a random series of n-correlated crop yields

4.1. Algorithms and formulae: see text, pages 12 - 16.

4.2. Listing of an MBASIC program for the simulation of n random series correlated yield values:

```
1 CLEAR:
2 HOME:
10 DIM A(5), D(200), E(200), F(5,200), L(5,200), C(5,5)
11 DIM X(200,5), P(5,200), PP(5,5), R(5,5), PP1(5), U(5,5)
20 PRINT"THIS PROGRAM CALCULATES A RANDOM SERIES OF N CORREL-"
21 PRINT"ATED CROP YIELDS BASED ON A MEAN VECTOR AND THE"
22 PRINT"DECOMPOSED VARIANCE-COVARIANCE MATRIX."
23 PRINT"IF MORE THAN 5 CORRELATED CROPS, OR A SERIES OF MORE"
24 PRINT"THAN 200 RUNS SHOULD BE GENERATED, THE DIMENSION"
25 PRINT"COMMANDS IN LINES 10 AND 11 SHOULD BE MODIFIED"
26 PRINT:PRINT"DATA ENTRY SECTION"
27 PRINT:INPUT"NUMBER OF RUNS:";M
30 PRINT:INPUT"NUMBER OF CROPS:";N
40 FOR I=1 TO N
50 PRINT"ARITHMETIC MEAN OF CROP";I;:INPUT":";A(I)
70 NEXT I
80 PRINT:PRINT"ENTER CHOLESKY'S TRIANGULAR MATRIX"
90 FOR I=1 TO N
100 FOR J=1 TO N
110 PRINT "C(";I;",";J;")=";
120 INPUT C(I,J)
130 NEXT J:NEXT I
150 PRINT:PRINT:INPUT"ENTER SEED NUMBER";Z
151 PRINT:PRINT"THE PROGRAM GENERATES CORRELATED CROP YIELDS"
155 RANDOMIZE (Z)
160 FOR S=1 TO M
170 FOR J=1 TO N
180 D(S)=RND(Z+S+J)
190 E(S)=RND(Z+S+J+1)
200 X(S,J)=SQR(-2*LOG(D(S)))*COS(2*3.14159*E(S))
220 NEXT J:NEXT S
260 REM MATRIX MULTIPLICATION
267 FOR S=1 TO M
280 FOR I=1 TO N
285 U=0
290 FOR J=1 TO N
300 U=U+C(I,J)*X(S,J)
310 NEXT J
320 F(I,S)=U
330 NEXT I
340 NEXT S
410 FOR S=1 TO M
420 FOR I=1 TO N
430 L(I,S)=F(I.S)+A(I)
```

APPENDIX 4. continued

```
440 NEXT I
445 NEXT S
1000 REM CALCULATION OF PARAMETERS FOR N DISTRIBUTIONS
1010 FOR I=1 TO N
1015 AA(I)=0
1020 FOR S=1 TO M
1030 AA(I)=AA(I)+L(I,S)
1040 NEXT S
1050 B(I)=AA(I)/M
1070 NEXT I
1080 FOR I=1 TO N
1090 FOR S=1 TO M
1100 P(I,S)-L(I,S)-B(I)
1200 NEXT S
1220 NEXT I
1230 FOR I=1 TO N
1240 FOR S=1 TO M
1260 PP1(I)=PP1(I)+P(I,S)^2
1270 NEXT S
1276 NEXT I
1279 PP(I,J)=0
1280 FOR I=1 TO N
1290 FOR J=I+1 TO N
1300 FOR S=1 TO M
1310 PP(I,J)=PP(I,J)+(P(I,S)*P(J,S))
1325 NEXT S
1330 NEXT J
1340 NEXT I
1400 FOR I=1 TO N
1410 R(I,I)=1
1420 FOR J=I+1 TO N
1430 R(I,J)=PP(I,J)/(SQR(PP1(I)*PP1(J)))
1435 R(J,I)=R(I,J)
1440 NEXT J:NEXT I
1450 PRINT:PRINT"CALCULATION FINISHED"
1460 PRINT"SKIP DETAILED RUN RESULTS(Y/N):";:GET A$
1470 IF A$="Y" THEN 1800 ELSE I500
1500 PRINT:PRINT"N-CROP SIMULATION"
1510 PRINT"----------------"
1520 FOR I=1 TO N
1530 Y=Y+10
1540 PRINT TAB(Y)"CROP" I;
1550 NEXT I
1560 FOR S=1 TO M
1570 Y=0
1580 FOR I=1 TO N
1590 Y=Y+10
1600 PRINT TAB (Y) L(I,S);
1610 NEXT I
1620 NEXT S
1630 PRINT
```

APPENDIX 4. continued

```
1640 PRINT"OUTPUT TO BE PRINTED? (Y/N):";:GET A$
1650 IF A$="Y" THEN 1660 ELSE 1800
1660 LPRINT"SIMULATION FOR ";N;" CROPS"
1670 LPRINT
1680 Y=10
1690 LPRINT TAB(1) "NR.";
1700 FOR I=1 TO N
1710 LPRINT TAB(Y) "CROP" I;
1720 Y=Y+10
1730 NEXT I
1740 FOR S=1 TO M
1750 Y=10
1760 LPRINT TAB(1) S;
1780 FOR I=1 TO N
1790 LPRINT TAB(Y) L(I,S)
1792 Y=Y+10
1794 NEXT I
1796 NEXT S
1800 PRINT:PRINT"DISPLAY OF CORRELATION MATRIX (Y/N):";:GET A$
1810 IF A$="Y" THEN 1820 ELSE 2000
1820 HOME:
1830 PRINT:PRINT"CORRELATION MATRIX OF THE SIMULATION MODEL"
1840 FOR I=1 TO N
1850 PRINT STR$(I)+":";
1860 FOR J=1 TO N
1870 PRINT USING"###.####";R(J,I);
1880 NEXT J
1890 PRINT:NEXT I
1900 PRINT:PRINT"PRINT THE CORRELATION MATRIX? (Y/N):";: GET A$
1910 IF A$="Y" THEN 1915 ELSE 2000
1915 LPRINT:LPRINT
1920 LPRINT"CORRELATION MATRIX OF THE SIMULATION MODEL RESULTS"
1930 FOR I=1 TO N
1940 LPRINT STR$(I)+":";
1950 FOR J=1 TO N
1960 LPRINT USING "###.####";R(J,I);
1970 NEXT J
1980 LPRINT
1990 NEXT I
2000 PRINT:PRINT"DISPLAY COVARIANCE MATRIX? (Y/N):";:GET A$
2010 IF A$="Y" THEN 2020 ELSE 2200
2020 HOME:
2025 PRINT:PRINT"COVARIANCE MATRIX OF SIMULATION RESULTS"
2030 FOR I=1 TO N
2040 FOR J=1 TO N
2050 U(I,J)=R(I,J)*(SQR(PP1(I)*PP1(J)))
2060 NEXT J
2070 NEXT I
2075 FOR I=1 TO N
2080 PRINT STR$(I)+":";
2090 FOR J=1 TO N
```

APPENDIX 4. continued

```
2100 PRINT USING"########.###";U(I,J)/M;
2110 NEXT J:PRINT:NEXT I
2120 PRINT:PRINT"PRINT COVARIANCE MATRIX? (Y/N):";:GET A$
2130 IF A$="Y" THEN 2135 ELSE 2200
2135 LPRINT:LPRINT
2140 LPRINT"VARIANCE-COVARIANCE OF SIMULATION MODEL RESULTS"
2150 FOR I=1 TO N
2155 FOR J=1 TO N
2160 U(I,J)=R(I,J)*(SQR(PP1(I)*PP1(J)))
2170 NEXT J:NEXT I
2180 FOR I=1 TO N
2185 LPRINT STR$(I)+":";
2190 FOR J=1 TO N
2192 LPRINT USING"########.###";U(I,J)/M;
2194 NEXT J:LPRINT:NEXT I
2200 PRINT:PRINT"DISPLAY MEAN VECTOR OF RESULTS?(Y/N):";:GET A$
2210 IF A$="Y" THEN 2220 ELSE END
2220 HOME:
2230 PRINT:PRINT TAB(1) "NR.";TAC(10) "MEAN"
2240 PRINT
2250 FOR I=1 TO N
2260 PRINT TAB(I) I;TAB(10) B(I)
2270 NEXT I
2280 PRINT:PRINT"PRINT OUT OF THE MEAN VECTOR? (Y/N):";:GET A$
2290 IF A$="Y" THEN 2295 ELSE END
2295 LPRINT:LPRINT
2300 LPRINT"MEAN VECTOR OF THE SIMULATION RESULTS"
2320 LPRINT:LPRINT TAB(1) "NR.";TAB(10) "MEAN"
2340 LPRINT:FOR I=1 TO N
2350 LPRINT TAB(1) I;TAB(10) B(I)
2360 NEXT I
2370 END
```

4.3. Example: see text, pages 12 - 19

APPENDIX 5. Corroborating Data.

The author is satisfied that barley yield estimates, over a series of years from measured samples in a random selection of farmers' fields in the study area, which would be appropriate for testing the present survey estimates, do not exist.

Nevertheless, other independent estimates may be compared with the present results. First is a ten-year series of annual estimates of barley grain yields by the local office of the Syrian Ministry of Agriculture and Agrarian Reform (in Breda village) for villages in the study area. The second set of data are results of a three-year ICARDA study in Hawaz village, nearby and with similar rainfall conditions.

1. In the Study Area:

The Tel Dhaman District Agricultural Office, in Breda, produces estimates each year for barley areas harvested, yield per hectare and total production. Unpublished records of these estimates were kindly made available for several villages in Zone 3 and Zone 4 of the study area, for the ten year period 1977-86. The aggregate zone yield estimates are presented below, allowing an independent comparison with the present survey estimates of average barley grain yields and their variance over time.

	Barley Grain (kg/ha)	
YEAR	Zone 3	Zone 4
1977	400	250
1978	600	600
1979	270	250
1980	920	900
1981	800	500
1982	550	550
1983	700	400
1984	0	0
1985	400	200
1986	500	400

	Zone 3	Zone 4
Mean	514	405
Standard Deviation	267	251
Coefficient of Variation	52%	62%
Coefficient of Skewness	-0.35	0.37
Coefficient of Kurtosis	2.66	2.85

Correlation Coefficient (Zone 3, Zone 4): 0.89

APPENDIX 5. continued

The mean and standard deviation calculated for the Zone 4 estimates (including the zero-yield value for 1984) are close to those derived with the present survey method: the mean being 2.3 percent higher and the standard deviation 3.5 percent lower than the author's estimates.

However, half of the study area is classified as Zone 3 farmland where average yields for the ten-year period are calculated at 514 kg/ha: about 30 percent higher than the author's aggregate estimate. Thus, if the above data are accurate, the author's estimates have understated aggregate mean yields in the area by about 14 percent while overstating the coefficient of variation by about 18 percent.

2. In Hawaz village:

Nour and Nygaard (1986, pp. 115 and 131) reported on barley yields at Hawaz village, in a Zone 4 farming area about 40 km east of Breda, for three seasons (1977/78, 78/79 and 79/80). In that period, the lowest measured seasonal rainfall was 150mm, the next highest was 240mm and the highest was 270mm. The corresponding mean harvested barley grain yields from village fields (unweighted for field size) were 170, 336 and 681 kg/ha, respectively. These data were derived from a multiple-visit survey by ICARDA staff and were based on farmer interviews, not actual measures of yield samples in farmers' fields.

When one considers only the three mean yield values to calculate a grand mean and standard deviation <u>over time</u>, the results ($\bar{x}$ = 396 kg/ha, and s = 261 kg/ha) are virtually identical to the author's estimates. However, this remarkable coincidence should not be given too much importance since it is based on data specific only to three years, and the location was outside the study area.

المركز الدولي للبحوث الزراعية في المناطق الجافة
ايكاردا
ص. ب. 5466 ، حلب ، سورية

INTERNATIONAL CENTER FOR AGRICULTURAL
RESEARCH IN THE DRY AREAS
Box 5466, Aleppo, Syria

1987