Elicitation of subjective crop yield PDF for DSS implementation

Clop-Gallart, M. Mercè & Juárez-Rubio, Francisco
University of Lleida. Av. Rovira Roure, 191. 25198 Lleida (Catalonia, Spain).
Tel.: + 34 973 702610; e-mail: mclop@aegern.udl.es; fjuarez@aegern.udl.es

Paper prepared for presentation at the 11th seminar of the EAAE
(European Association of Agricultural Economists),
'The Future of Rural Europe in the Global Agri-Food System', Copenhagen, Denmark,
August 24-27, 2005

ELICITATION OF SUBJECTIVE CROP YIELD PDF FOR DSS IMPLEMENTATION

Abstract

The aim of this research is to establish the persistence of annual crop yield point values subjective estimates, and the coherence and reliability of subjective crop yield probability density functions (PDF) elicited from a series of interviews carried out on a wide group of farmers, and then to determine whether they should be included or not in a decision support system (DSS). Three different elicitation techniques were used:

a) The Two Step PDF estimation method
b) Triangular distribution
c) Beta distribution

Although the results are noteworthy, further studies should be carried out to perfect the aforementioned techniques before crop yield PDF's are used in decision making processes.

Keywords: Decision support systems, Subjective crop yield PDF elicitation, Two Step PDF estimation method, Triangular and Beta distributions.

JEL classification: Q10, L20, L23, D81

1. Introduction

The attempts to develop a Decision Support System (DSS) to enhance the outcome of farmers' decision making activities and other subjects such insurance premiums, have been complicated by the need to assess Probability Density Functions (PDF) for key variables such as crop yield. Pease (1992) has highlighted certain difficulties related to the information available on the four main sources of data on crop yield: country yield series, series from agronomic experiments, farm level historical yield series, and elicited subjective yield forecasts.

The use of data from country yield series and agronomic experiments in the elaboration of individual farm yield probability distributions faces serious methodological problems. Namely, they:

(a) Need data from 20 or more years for statistical purposes;
(b) Depend on atmospheric and environmental conditions and management of the specific farm (microclimate, soil quality, input utilization, crop variety, tillage practices, etc.);
(c) Need the recorded yield of mean production from many other dispersed fields, and
(d) Must take into account the data registered throughout several years regarding the environmental, economic and technological conditions; these data influence the production plans of individual farmers, and they can change drastically from year to year.

In this work, we will focus on subjective yield forecasts. There is little empirical evidence on the efficacy of these estimates, and agricultural economists have little confidence on the possibility of obtaining reliable results from the measurement of subjective probability distributions in decision support models. Nevertheless, Anderson (1997) has pointed out that "risk perception is an art form that is quintessentially subjective". Miranda (1991), in his discussion on the experience of the Federal Crop Insurance Program, recognizes that "producers are better informed about the distribution of their own yields and thus are better able to assess the actuarial fairness of their premium than the insurer…" The right or wrong decisions which farmers make when designing their production plans imply, on their part, an estimation of crop yield PDF.

Several authors have carried out experiments to elicit farmers' subjective PDF. These studies have discussed the efficiency of the elicitation techniques used (accuracy, reliability, acceptability and predictive accuracy; see Norris and Kramer, 1990). Our research follows these guidelines.

Subjective (personalistic) PDF estimates generally result in skewed probability density functions, and acceptable intervals for the mean value estimation (Bessler, 1980; Grisley and Kellogg, 1983; Pease, 1992; Smith and Mandac, 1995); an additional research line, started in psychology by

Winkler (1967), tries to experimentally evaluate the properties of the techniques used in data collection from surveys. Nevertheless, variance is systematically underestimated. Variance underestimation is conjectured in the literature as a consequence of cognitive biases (e.g. Smith and Mandac, 1995; Tversky and Kahneman, 1974).

A program for the systematic collection of data has been set up in order to accumulate experience on crop yield PDF. Agronomy students with a background in traditional statistics, interviewed farmers to elicit their crop yield PDF, in order to create a methodology which would be readily accessible to potential users of DSS.

2. Research method

In 1999, fifty-two farmers were interviewed for the first time. Two students interviewed each farmer. Student number one carried out the interview with what we called the "first day questionnaire". Approximately two weeks later, another student interviewed the same farmer with the "second day questionnaire", which was organized in a different way (questionnaires available from the authors). A total of 104 interviews were carried out with 52 farmers.

The main objectives of the research were:

(1) To determine the point and PDF functions reliability or persistence of the answers provided by the farmers and,

(2) To establish the coherence between:

   (a) The calculated mean values of the studied distributions and the mean values explicitly declared,

   (b) The calculated mean values from the three estimated PDFs and,

   (c) The calculated variances from the three estimated PDFs.

Persistence (one of the requirements of coherence) was evaluated by the responses given to the two different students and time spans with the use of different subjective probability elicitation methods. Farmers come from a wide range of geographical areas with very different environmental and technological conditions from farm to farm. Although Pease (1992) has pointed that "Geographical location plays a larger role than crop in comparison of relative variability of yields", our aim did not lie in determining a given operational PDF for each of the crops analyzed, but to verify the persistence of the responses.

Based upon the experience acquired from the interviews carried out during 1999, the 2000 questionnaire (second year of the interviews) was changed slightly. The second year interviews followed the same methodology described above. Forty-four different farmers were questioned, providing a total of 88 new interviews.

Each farmer indicated the annual crop he would provide information for, depending on his own experience. From the answers obtained, only those with the most number of responses (annual crops with more than five responses) were taken into account (Table 1).

Table 1. Response from Selected Crops

	Cases	
Annual crop	**1999**	**2000**
Non irrigated barley	46	22
Non irrigated wheat	25	12
Irrigated wheat	13	5
Irrigated maize	20	10

Three elicitation techniques were used: Two Step PDF estimation, Triangular and Beta distributions.

In the **Two Step PDF estimation technique**, farmers first assess the frequency (in percentage) for each of the five crop yield year classes: very poor yield years, poor yield years, normal yield years, good yield years and very good yield years. In this first step, farmers assess frequencies at five ordinal interval classes. In the second step, farmers indicate the yield interval that they consider appropriate to describe a very poor yield year, a poor yield year, etc. The final product is a crop yield five-interval histogram. Mean and variance values of the elicited distribution are then calculated. If x_i is the midpoint value of the estimated frequencies, interval i, and f_i is the corresponding frequency, the mean (E) and variance (VE) are given by the expressions:

$$E = \Sigma_i x_i f_i \quad (1)$$
$$VE = \Sigma_i (x_i - E)^2 f_i \quad (2)$$

As far as we know, this technique has not been used in previous studies.

In the **Triangular distribution method**, farmers specify the crop's highest possible yield (H), lowest possible yield (L), and most frequent yield (M) (Sonka and Patrick, 1984). The mean (T) and variance (VT) of the Triangular distribution are:

$$T = (1/3)\,(L+M+H) \quad (3)$$
$$VT = (1/18)\,[(H-L)^2 + (M-L)(M-H)] \quad (4)$$

In the **Beta distribution method**, the three previously declared values are fitted to a Beta function with a mean value and variance given by the known PERT approach expressions (Moskowitz and Bullers, 1979). The mean (B) and variance (VB) of the Beta – PERT distribution are:

$$B = (1/6)\,(L + 4\,M + H) \quad (5)$$
$$VB = (1/36)\,(H - L)^2 \quad (6)$$

3. Crop yield persistence analysis

Ideally, crop yield data given by a decision maker at two different time spans, should be the same, and thus a null difference would be interpreted as an argument for the persistence. Nevertheless, in case of discrepancy between the declared crop yield values of the first and the second day by a decision maker we should not infer a violation of the persistence hypothesis, for there might be certain cognitive mechanisms taking part which would account for that discrepancy in the voiced information. Some explanations for that discrepancy might be: the tendency to give rounded figures (and, with some intermediate figures, it would be possible to round both up or down); the existence of a range of values which farmers believe to be equivalent; or the voicing of uncertainty through an interval out of which a point figure is declared.

Farmers who were interviewed gave both estimates for point crop yield (mean, H, M and L) and for the PDF based on interval estimations. To evaluate persistence, a concept of persistence (which we may call "time persistence") was used, based on measuring the difference between the estimates declared at two different time spans. Therefore, if a subject declares a point figure on the first day interview and a different one two weeks later, a big difference between the two is not expected as a first criterion to judge the coherence of the estimates. In the same way, in making a PDF estimation, persistence will be measured by its mean and variance, so that, if the farmer declares a PDF on the first interview and a different one two weeks later, both shall be considered as similar if the values they show are close in their respective means and variances, verifying the (time) persistence criterion.

A second meaning for the concept of persistence –"methodological persistence"- would be related to the estimation technique. In this way, if a declared value is maintained, not just in time, but also in the techniques it has been declared through, it will be supposed that the criterion of persistence will be verified. “Methodological” persistence is analysed by:

(1) The comparison between the direct point estimation of the mean crop yield, made by the farmers, and the estimation calculated from the elicited PDFs; (methodological) persistence is verified if the declared mean crop yield and mean yield implicit in PDFs are similar.
(2) The comparison between the mean crop yield values and their variances inferred from the different functional estimations, and the relationship between the results and the accessibility of extreme values.

To avoid the biases pointed out by Bland and Altman (1995, 1999), if d_1 and d_2 are the values to be compared, as in, for example, the estimations made by a famer on the first day and second day, or two values corresponding to different PDFs; relative differences throughout this research will be expressed thus: $(d_1 - d_2) / [(d_1 + d_2) / 2]$.

The rounding hypothesis needs no further explanation. In the case of the existence of a range of equivalent values, it can be argued that there is sufficient evidence in many areas for the fact that people do not distinguish certain values in a continuous way, but their perception of “similar” values is represented by one value (perception or segregation thresholds).

Farmers tend to answer using rounded values, which may lead to a discrepancy between the amounts declared on one day or another.

4. Time persistence of point estimates

Estimates of point values for both interview days are available. What follows is a study of the persistence of data, taken from the declared values of the mean (year 2000), lowest, highest and most frequent crop yields.

4.1. Declared Mean Crop Yield Differences

Generally, farmers talk about mean crop yields when they compare results from a single crop or from different crops. It seems that farmers are familiar with mean crop yields, which are easily and cognitively available for them. Mean yield could be an anchorage value in the estimation of other crop yield parameters (availability and anchorage are used in the way that Tversky and Kahneman (1974) discussed). When farmers were asked for a mean crop yield value, their response was identified as “declared mean crop yield” (m).

In the 1999 survey, the “declared mean crop yield” value was not included in the first interview, as the availability of this value might condition the answers for other values. This question was included in the second interview, where the farmer was asked to declare the “mean crop yield” for each crop.

When mean yields from the responses obtained were compared in the three PDF elicitation techniques used, no evidence was found to indicate that there were significant differences in the responses between the first and second day interviews that could be attributed to the question formulated in the second day questionnaire regarding “declared crop yield”. For this reason, in the 2000 survey, the farmer was asked about “declared mean crop yield” in both questionnaires.

The declared mean crop yields given by each subject were taken into account in the year 2000 interviews. Statistics have been reproduced in Table 2. Answers for the first and second days, both for non irrigated and irrigated crops, seem to belong to the same population (Wilcoxon test, $\alpha = 0.05$).

Differences are normally distributed (Kolmogorov – Smirnov, $\alpha = 0.01$) for both non irrigated and irrigated crops.

Extreme differences were found, though not often, and raised the question: which would be the width of the range or interval compatible with a reasonable degree of approximation and, therefore, also compatible with the persistence hypothesis. It would seem logical for that criterion to be expressed as a percentage.

Table 2. “Declared Mean Crop Yield” variable statistics, kg / ha, year 2000

Declared Mean Crop Yields	n	Mean	S.D.	Median	Min	Max	K-S+	Wilcoxon*
Non irrigated crops								
day 1	33	3148.2	775.6	3000	1400	4375	0.927	
day 2	33	3204.7	849.2	3000	1400	5400	0.289	0.503
Relative Difference	33	-0.0176	0.115	0	-0.27	0.29	0.319	
Irrigated crops								
day 1	14	8225	3338.5	7950	4000	14000	0.663	
day 2	14	8542.8	3657.4	7750	4000	14000	0.723	0.236
Relative Difference	14	-0.031	0.093	0	-0.2	0.15	0.251	

+ Kolgomorov – Smirnov Test for a normal distribution.
* Wilcoxon Test for independent populations (first day vs. second day).

Also in Table 2, we might observe how, for mean crop yield in non irrigated crops, the mean for the relative difference is 1.8%, with a standard deviation of 11.5%, and with extreme differences ranging between 27% and 29%. For irrigated crops, the mean relative difference is 3.1%, with a standard deviation of 9.3% and with extreme differences ranging between 15% and 20%. The median for relative differences is null for both non irrigated and irrigated crops.

Establishing the threshold, from which the persistence of values may be admitted, possibly ought to be judged by the economic impact of relative differences, and future research is needed. In this preliminary investigation – with no reference to that economic impact- it is suggested that the estimates that have been researched verify the persistence criterion, in the case of the declared mean crop yields or, at least, in a high percentage of the surveyed population.

4.2. Observed differences in lowest possible yields, highest possible yields and most frequent yields

Both in 1999 and in 2000, farmers were asked about the estimation of lowest possible yields, most frequent yields and highest possible yields in order to fit the Triangular and Beta functions. Pairs of data generated on both survey days are available. The results of the analysis have been summarised in Table 3. Generally, previous studies suggest that farmers have difficulty in understanding the meaning of subjective estimates of lowest and highest possible values (it is not about obtaining “possible” values –and, in this case, declared crop yields could range between 0 and infinite-, but about extreme values that may have been observed with a frequency of at least 5%).

Table 3. Relative differences in declared crop yields on survey days (%)

Declared Yields and year	n	Mean	S.D.	Median	Min	Max
Non irrigated crops						
Lowest possible yield (1999)	70	-4.7%	36.8%	0%	-125%	86%
Lowest possible yield (2000)	33	-1.15%	41.3%	0%	-113%	100%
Highest possible yield (1999)	70	-0.8%	9.6%	0%	-40%	18%
Highest possible yield (2000)	33	1.1%	12%	0%	-18%	46%
Most frequent yield (1999)	70	-0.37%	11.4%	0%	-31%	50%
Most frequent yield (2000)	33	-4.2%	12.7%	0%	-51%	13%
Irrigated crops						
Lowest possible yield (1999)	33	8%	20.3%	0%	-57%	46%
Lowest possible yield (2000)	14	1.3%	35.1%	0%	-67%	67%
Highest possible yield (1999)	33	-1.9%	7.9%	0%	-29%	18%
Highest possible yield (2000)	14	-3.6%	7.9%	0%	-15%	6%
Most frequent yield (1999)	33	-2.8%	9.9%	0%	-34%	15%
Most frequent yield (2000)	14	-5.2%	8.1%	-3.5%	-22%	8%

Table 4 shows the figures obtained in Kolgomorov - Smirnov’s normality test and in Wilcoxon’s test for independent populations (first day vs. second day), with a significance level of 5%, for the relative differences of the different point estimates of lowest possible, highest possible and most frequent yields.

Table 4. Normality Kolgomorov – Smirnov (K-S) and Wilcoxon tests results for relative differences (1999, 2000). Probability values.

Declared Yields	**K-S**		**Wilcoxon**	
	1999	**2000**	**1999**	**2000**
Non irrigated crops				
Lowest possible yield	0	0.005	0.522	0.879
Highest possible yield	0	0.007	0.423	0.827
Most frequent yield	0	0.053	0.961	0.129
Irrigated crops				
Lowest possible yield	0.005	0.764	0.550	0.593
Highest possible yield	0.002	0.117	0.233	0.121
Most frequent yield	0.004	0.696	0.054	0.042

Most of the Kolgomorov – Smirnov test seems to suggest that distributions do not comply with a normal distribution (5%), except for in some values corresponding to the year 2000. On the other hand, Wilcoxon's test makes it impossible to reject the similarity hypothesis between the means in all cases, except for the ones referring to the most frequent yield in irrigated crops for the year 2000.

The high value for the standard deviation in lowest possible crop yields is due to the existence of some extreme differences, in some cases, of 75% or more. Although the means and medians of the relative differences for the lowest possible crop yields are small, standard deviations show a probable misunderstanding or indetermination surrounding this concept. It is conjectured that that indetermination is due to the fact that lowest possible crop yields are not readily accessible, as they do not play a major part in farmers's estimations or goals.

A significative rank correlation between lowest possible and highest possible crop yields is yet to be found. It is unlikely that the same people would make poor estimations for both types of crop yield. The results aforementioned indicate a higher persistence in the estimated highest possible crop yield values than in the lowest possible ones. The greater differences found in the estimation of the second variable may be due to the fact that the margin for the highest possible crop yields may be subjectively more restricted than the margin for the lowest possible crop yields. In the same way, it is possible that this variable may act as a goal for farmers, therefore being more accessible than the estimated lowest possible crop yields.

5. Time persistence of PDF estimates

PDF estimates for both survey days, every year, are available. The following analyses the persistence of data, evaluated by the value of the mean crop yield and the variance of the crop yields calculated through Equations (1) to (6).

5.1. Mean crop yields calculated from estimated PDFs

The differences between the mean crop yields calculated from the three already analysed distributions (Equations (1), (3) and (5)) between the first and second days for the years 1999 and 2000 were established. The results can be found in Tables 5 and 6.

Generally, mean values calculated from the Two Step PDF for both days are very similar. The mean for the relative differences ranges between –0.7% and –0.1% in non irrigated crops and between –2.1% and –1.4% in irrigated crops, with medians being null in non irrigated crops and practically null in irrigated crops. The standard deviations for that variable are 8.1% and 11.7% in non irrigated crops and 6.6% and 6.9% in irrigated crops. The similarity of both distributions is obvious. There are no differences between the different populations (Wilcoxon, $\alpha = 0.05$).

Similar results were found for the Triangular and Beta–PERT functions.

Table 5. Relative differences in the calculated mean crop yields (non irrigated crops)

PDF and year	n	Mean	S.D.	Median	Min	Max	K-S+
Two Step, 1999	68	-0.001	0.081	0	-0.33	0.18	0.180
Two Step, 2000	33	-0.007	0.117	0	-0.47	0.16	0.205
Triangular, 1999	68	-0.011	0.089	0	-0.31	0.17	0.009
Triangular, 2000	33	-0.015	0.074	0	-0.21	0.23	0.254
Beta, 1999	68	-0.011	0.085	0	-0.30	0.16	0.001
Beta, 2000	33	-0.020	0.078	0	-0.19	0.16	0.159

+ Kolgomorov – Smirnov Test for a normal distribution.

Table 6. Relative differences in the calculated mean crop yields (irrigated crops)

PDF and year	n	Mean	S.D.	Median	Min	Max	K-S+
Two Step, 1999	33	-0.014	0.066	-0.004	-0.20	0.15	0.322
Two Step, 2000	14	-0.021	0.069	0	-0.24	0.07	0.191
Triangular, 1999	33	-0.018	0.096	0	-0.36	0.22	0.101
Triangular, 2000	14	-0.029	0.077	-0.030	-0.17	0.16	0.553
Beta, 1999	33	-0.023	0.095	0	-0.34	0.16	0.054
Beta, 2000	14	-0.040	0.066	-0.037	-0.12	0.12	0.976

+ Kolgomorov – Smirnov Test for a normal distribution.

5.2. Variance calculated from estimated PDFs

Relative differences between variances calculated using the different analysed distributions, have been compared using Equations (2), (4) and (6) on the first and second days for years 1999 and 2000. Results are summarised in Tables 7 and 8.

Table 7. Relative differences between the variances of calculated crop yields (non irrigated crops)

PDF and year	n	Mean	S.D.	Median	Min	Max	K-S+
Two Step, 1999	68	0.001	0.596	0	-1.93	1.41	0.038
Two Step, 2000	33	0.031	0.267	0	-0.39	0.50	0.295
Triangular, 1999	68	-0.003	0.516	0	-1.20	1.78	0.012
Triangular, 2000	33	0.008	0.518	-0.008	-1.15	1.30	0.047
Beta, 1999	68	-0.008	0.515	0	-1.20	1.78	0.009
Beta, 2000	33	0.009	0.514	0	-1.15	1.34	0.006

+ Kolgomorov – Smirnov Test for a normal distribution.

Table 8. Relative differences between the variances of calculated crop yields (irrigated crops)

PDF and year	n	Mean	S.D.	Median	Min	Max	K-S+
Two Step, 1999	33	0.057	0.466	0	-0.61	1.73	0.011
Two Step, 2000	14	0.267	0.597	0	-0.63	1.62	0.175
Triangular, 1999	33	-0.057	0.369	0	-0.92	1.42	0.026
Triangular, 2000	14	-0.233	0.819	-0.135	-1.26	1.22	0.816
Beta, 1999	33	-0.057	0.363	0	-0.92	1.42	0.009
Beta, 2000	14	-0.281	0.891	-0.153	-1.89	1.20	0.902

+ Kolgomorov – Smirnov Test for a normal distribution.

In general, PDF estimations in the three aforementioned distributions for non irrigated crops provide very similar values to the variances calculated from the data obtained on both survey days. Nevertheless, there are some cases that show great differences between the estimates of one day and the other, measured using variance. These cases probably show evidence in some farmers to be biased and anchor interval values with the mean, having little regard for extreme values.

In the case of relative differences for irrigated crops, we have obtained unfavourable results in the year 2000 for the persistence hypothesis; these results were probably magnified by the sample's limited size, but the general trend is conceptually similar to the one discussed for non irrigated crops.

Samples appear to belong to the same population in all analysed cases (Wilcoxon, $\alpha = 0.05$).

6. Methodological persistence

Mean declared crop yield values on the second day of 1999 and the first and second days of 2000 are available. In the following section, the declared mean crop yield values will be compared to the mean values calculated for the Two Step estimation method -using E from Equation (1)-, and the mean values calculated for the Triangular and Beta estimations –using T from Equation (3) and B from Equation (5), respectively.

If m1 is the declared mean crop yield value, and m2 is the value calculated from the Two Step estimation method or the Triangular and Beta approximations (m2 = E or T or B), the relative difference will be defined as (m1-m2) / [(m1+m2)/2].

6.1. Differences between Declared and Two Step estimation method calculated mean crop yields

Declared mean crop yields in 1999 (relating to the second day) and 2000 (first and second days) by each farmer were compared to the calculated mean crop yields, using Equation (1). Statistics relating to all crops can be seen in Table 9.

Table 9. Relative differences between declared mean crop yields and calculated using the Two Step estimation method

Crops	**n**	**Mean**	**S.D.**	**Median**	**Min**	**Max**	**K-S⁺**	**Wilcoxon***
1999 crop yield (second day)								
Non irrigated crops	68	0.040	0.125	0.042	-0.31	0.45	0.600	0.008
Irrigated crops	33	-0.008	0.132	0	-0.24	0.50	0.631	0.675
2000 crop yield (first day)								
Non irrigated crops	33	0.048	0.149	0.022	-0.31	0.47	0.940	0.102
Irrigated crops	14	-0.029	0.171	-0.011	-0.45	0.23	0.876	0.422
2000 crop yield (second day)								
Non irrigated crops	33	0.059	0.13	0.059	-0.23	0.46	0.994	0.007
Irrigated crops	14	-0.020	0.189	-0.008	-0.50	0.23	0.981	0.683

[+] Kolgomorov – Smirnov Test for a normal distribution.
[*] Wilcoxon Test for independent populations.

The mean of relative difference for non irrigated crops tends to be small, around 4% to 6% (with acceptable standard deviations); the median ranges between 2% and 6%. This factor indicates a slight underestimation of mean crop yield when expressed through the Two Step estimation method, in relation to the directly declared mean crop yield value.

For irrigated crops, the mean of relative differences ranges between -2.9% and –0.8%; the median ranges between –1.1% and 0%. These negative values probably lead to a slight overestimation of the mean crop yield when expressed through the Two Step estimation method. Standard deviation ranges between 13.2% and 18.9%.

6.2. Differences between Declared mean crop yields and calculated using the Triangular and Beta approximations

Mean crop yields declared in the 1999 (for the second day) and 2000 (for the first and second days) surveys by each farmer have been compared to mean crop yields calculated by the Triangular approximation –using Equation (3); the results are summarised in Table 10. Results obtained by the Beta approximation –using Equation (5)- are shown in Table 11.

In the case of non irrigated crop yields, the mean for the relative diference reaches values between -2% and –1.7%; the median has values between –2% and 0%. Relative difference follows a similar pattern to the Two Step estimation for irrigated crops. In irrigated crops, relative difference of the estimated mean and the mean calculated by the Triangular method lies between –1% and –0.4%, with the median lying between –2.3% and 0%. Standard deviation ranges from 7.5% to 11.8%, for non irrigated crops, and 5.8% to 9.9% for irrigated crops.

Table 10. Relative differences between declared mean crop yields and calculated by the Triangular approximation

Crops	**n**	**Mean**	**S.D.**	**Median**	**Min**	**Max**	**K-S**[+]	**Wilcoxon***
1999 crop yield (second day)								
Non irrigated crops	70	-0.02	0.118	0	-0.33	0.18	0.020	0.635
Irrigated crops	33	-0.004	0.099	0	-0.29	0.20	0.261	0.981
2000 crop yield (first day)								
Non irrigated crops	33	-0.019	0.092	-0.004	-0.29	0.18	0.283	0.344
Irrigated crops	14	-0.01	0.07	0	-0.17	0.08	0.605	0.859
2000 crop yield (second day)								
Non irrigated crops	33	-0.017	0.075	-0.02	-0.24	0.18	0.388	0.303
Irrigated crops	14	-0.009	0.058	-0.023	-0.09	0.10	0.765	0.729

[+] Kolgomorov – Smirnov Test for a normal distribution.
[*] Wilcoxon Test for independent populations.

Table 11. Relative differences between declared mean crop yields and calculated by the Beta – PERT approximation

Crops	**n**	**Mean**	**S.D.**	**Median**	**Min**	**Max**	**K-S**[+]	**Wilcoxon***
1999 crop yield (second day)								
Non irrigated crops	70	-0.016	0.103	0	-0.32	0.40	0.022	0.664
Irrigated crops	33	-0.015	0.087	0	-0.31	0.19	0.042	0.683
2000 crop yield (first day)								
Non irrigated crops	33	-0.015	0.099	-0.013	-0.22	0.18	0.309	0.274
Irrigated crops	14	-0.019	0.054	0	-0.13	0.05	0.578	0.328
2000 crop yield (second day)								
Non irrigated crops	33	-0.026	0.086	-0.023	-0.31	0.18	0.156	0.041
Irrigated crops	14	-0.029	0.061	-0.023	-0.12	0.09	0.990	0.177

[+] Kolgomorov – Smirnov Test for a normal distribution.
[*] Wilcoxon Test for independent populations.

With regard to the Beta approximation on non irrigated crops, the mean for relative differences ranges from –2.6% to –1.5%; on irrigated crops, it ranges from –2.9% to –1.5%. The median for both non irrigated and irrigated crops lies between –2.3% and 0%. Standard deviations vary from 8.6% to 10.3% (for non irrigated crops) and 5.4% to 8.7% (for irrigated crops).

Negative values found for the mean of relative differences – as a result of both the Triangular and Beta approximations-, seem to indicate a slight overestimation of the mean crop yield (regarding the directly declared mean).

6.3. Differences between mean crop yields calculated by various PDFs

An interesting practical issue is whether the results obtained by the Triangular and Beta approximations predict values that are very different to those predicted by the Two Step estimation method, as the first two methods would, initially, appear to be easier ways of interviewing farmers.

Comparing the Triangular and Beta methods holds no empirical interest because differences between the mean and variance values, which are estimated from the same set of values L, M and H (the lowest possible, most frequent and highest possible crop yields), are mathematically determined by functional forms. In estimating the mean crop yield, the difference given by both functions is the result of using Equation (7), and Equation (8) for the variance.

$$T - B = (1/6)\,(H + L - 2M) \qquad (7)$$
$$VT - VB = (1/36)\,[(H-M)^2 + (M-L)^2] \qquad (8)$$

The difference between the means reaches its lowest value with M = (1/2) (H+L), as the variance. This result shows that, when the most frequent value M coincides with the midpoint of L and H, mean and variance values of both approximations also coincide. The difference between variances (8) is always positive, therefore: VT ≥ VB.

Tables 12 and 13 summarise the differences found between the mean crop yield from the Two Step estimation method (E) and the Triangular (T) and Beta (B) approximations.

Table 12. Relative differences in calculated mean crop yields (non irrigated crops)

Year and day	**n**	**Mean**	**S.D.**	**Median**	**Min**	**Max**	**K-S[+]**
T – E							
1999 day 1	68	0.053	0.146	0.043	-0.35	0.38	0.612
1999 day 2	68	0.063	0.133	0.035	-0.31	0.41	0.453
2000 day 1	33	0.067	0.132	0.040	-0.13	0.38	0.665
2000 day 2	33	0.075	0.133	0.077	-0.13	0.51	0.835
B – E							
1999 day 1	68	0.053	0.139	0.052	-0.30	0.45	0.731
1999 day 2	68	0.063	0.126	0.048	-0.31	0.44	0.744
2000 day 1	33	0.063	0.13	0.041	-0.13	0.37	0.673
2000 day 2	33	0.076	0.148	0.068	-0.17	0.58	0.810

[+] Kolgomorov – Smirnov Test for a normal distribution.

Table 13. Relative differences in calculated mean crop yields (irrigated crops)

Year and day	**n**	**Mean**	**S.D.**	**Median**	**Min**	**Max**	**K-S[+]**
T – E							
1999 day 1	33	-0.008	0.123	-0.012	-0.25	0.33	0.423
1999 day 2	33	-0.003	0.115	-0.014	-0.18	0.32	0.960
2000 day 1	14	-0.018	0.193	0	-0.46	0.26	0.878
2000 day 2	14	-0.010	0.181	0.005	-0.46	0.25	0.963
B – E							
1999 day 1	33	-0.002	0.131	-0.019	-0.27	0.34	0.801
1999 day 2	33	0.007	0.117	-0.014	-0.21	0.32	0.867
2000 day 1	14	-0.011	0.176	-0.019	-0.40	0.26	0.977
2000 day 2	14	0.008	0.171	0.019	-0.40	0.26	0.990

[+] Kolgomorov – Smirnov Test for a normal distribution.

With regards to non irrigated crops, both the Triangular and Beta methods tend to show estimations for the mean crop yield which are slightly higher than the estimates resulting from the Two Step estimation method, though it is not a significative difference. With regards to irrigated crops, the mean crop yield inferred from the Two Step estimation method is higher than the Triangular estimation. There is no clear trend for its comparison with the Beta method. This suggests very similar values in the mean crop yield estimates.

6.4. Differences between crop yield variances resulting from various PDFs

Variances calculated by the Two Step estimation method and the Triangular and Beta – PERT approximations differ a great deal. In the case of the difference between the crop yield variances evaluated from the Two Step method and the Triangular approximation (VE-VT), the relative differences are higher than 34.4% in all cases, with standard deviations of up to 101%. In the comparison between the Two Step estimation and the Beta approximation (VE-VB), the relative differences are less favourable than in the previous example; the minimum mean for the relative differences being 48.8%, with a maximum standard deviation of 100.3%. Results for irrigated crops were slightly more favourable than for non irrigated crops.

6.5. Differences between extreme crop yield values

The aforementioned result may be a consequence of both Triangular and Beta functional specifications, and possible difficulties in the persistence of extreme values, especially L. In order to examine this last issue, the extremes in the density function estimated in Two Steps were compared to the point estimations that serve to define the Triangular and Beta approximations. In order to achieve this, two relative differences indexes were created (I_m and I_M). If *r* is the lower extreme in the interval defined as "very poor crop yields" and *s* is the upper extreme in the interval defined as "very good crop yields", we have:

$$I_m = (r - L) / [(r + L)/2] \quad (9)$$
$$I_M = (s - H) / [(s + H)/2] \quad (10)$$

Results obtained in the relative differences of indexes (9) and (10) are shown in Tables 14 and 15.

Table 14. Value of index I_m (relative differences between "very poor crop yield" r and the lowest L)

Year and day	n	Mean	S.D.	Median	Min	Max
Non irrigated crops						
1999 day 1	68	-0.51	0.72	-0.20	-2	0.67
1999 day 2	68	-0.57	0.69	-0.38	-2	0.86
2000 day 1	33	-0.57	0.61	-0.4	-2	0
2000 day 2	33	-0.58	0.65	-0.54	-2	0.46
Irrigated crops						
1999 day 1	33	-0.54	0.71	-0.29	-2	0
1999 day 2	33	-0.32	0.49	-0.19	-2	0.18
2000 day 1	14	-0.66	0.72	-0.46	-2	0.4
2000 day 2	14	-0.49	0.71	-0.25	-2	0.29

Table 15. Value of index I_M (relative differences between "very good crop yield" s and the highest H)

Year and day	n	Mean	S.D.	Median	Min	Max
Non irrigated crops						
1999 day 1	68	0.052	0.12	0.024	-0.18	0.46
1999 day 2	68	0.034	0.13	0.038	-0.48	0.31
2000 day 1	33	0.12	0.18	0.08	-0.15	0.71
2000 day 2	33	0.09	0.18	0	-0.15	0.71
Irrigated crops						
1999 day 1	33	0.12	0.11	0.09	0	0.4
1999 day 2	33	0.09	0.11	0.07	-0.13	0.4
2000 day 1	14	0.09	0.13	0.07	-0.06	0.45
2000 day 2	14	0.05	0.08	0.03	-0.07	0.24

It can be observed that farmers tend to give relatively stable estimations for the highest possible crop yield and/or the higher extremes of the range. On the other hand, very different estimations for lowest possible values and/or the lower extremes of the range were given.

7. Declared Mean and Most Frequent yields relative differences

The difference between the most frequent value and the declared mean crop yield for 1999 (second day) and 2000 (first and second days) has been analysed; the results are summarised in Table 16. It can be seen that the differences between the variables are similar to the ones for different days. Wilcoxon's test (α=0.05) for mean crop yield estimations and the most frequent estimations does not allow us to discard the hypothesis of mean equality for all cases. Rank correlation between these two variables is positive and significant in all cases.

Table 16. Relative differences between declared mean and most frequent crop yields

Crops, year and day	n	Mean	S.D.	Median	Min	Max
Non irrigated crops, 1999	70	0.01	0.11	0	-0.67	0.32
Irrigated crops, 1999	33	0.02	0.08	0	-0.18	0.34
Non irrigated 2000 (day 1)	33	0.009	0.12	0	-0.29	0.22
Non irrigated 2000 (day 2)	33	0.03	0.12	0	-0.18	0.51
Irrigated 2000 (day 1)	14	0.02	0.05	0	-0.07	0.13
Irrigated 2000 (day 2)	14	0.05	0.08	0.02	-0.07	0.2

The previous results lead us to conjecture that declared mean crop yield and most frequent crop yield figures appear to belong to the same population, an argument which seems to work in favour of the idea that one of the values anchors itself with the other.

8. Concluding remarks

The time persistence of the mean, lowest possible, highest possible and most frequent physical crop yield values declared by farmers on surveys on annual crops, carried out in 1999 and 2000 has been analysed. Limited relative differences have been found for all variables, except for the lowest possible crop yield estimates, which have a broad dispersion. There is no correlation between relative differences of the highest possible and the lowest possible crop yields. This indicates that there is no personal bias in the differences observed in the estimations.

It is conjetured that:

(1) The relative differences in the highest possible and lowest possible crop yield estimates are due to the fact that the highest possible crop yields are more accessible than the lowest possible crop yields.
(2) The lowest possible crop yields do not play an important role in the calculations made by farmers, whereas the highest possible crop yields may be acting as a goal, according to which annual results are measured against.
(3) Farmers truncate the left tail of the crop yields distribution curve.

The similarity between the mean and the variance of estimated functions for each of the days that the survey was conducted on, was used as the criterion to examine the time persistence of PDF estimations given by farmers. It has been found that differences were relatively small, except for the variance on irrigated crops in 2000 (with a smaller number of observations). These results seem to suggest a great stability in the mental image that farmers have of PDFs.

In order to calibrate methodological persistence, the direct estimation of mean crop yields and the calculated from estimated PDFs was compared. The differences were found to be small. In the case of the Two Step estimation method, PDF in non irrigated crops tends to give a slight underestimation of the mean in relation to the declared; whereas the opposite occurs in irrigated crops. Using the Triangular and Beta approximations, a slight overestimation of the mean crop yield, as opposed to the declared crop yield, was given.

This overestimation of the mean crop yield (shown in the negative figures of the relative differences in Tables 9 to 11) could be interpreted as the result of a broader range on the frequency histogram (relating to the Two Step estimation on irrigated crops, and the Triangular and Beta approximations on all crops), contrary to what is cited in the literature regarding variance underestimations. Values are so small, that it is neither possible nor useful to confirm that they may be anything other than zero (meaning that the declared and the calculated crop yields coincide). Nevertheless, it is significant that no general trend of positive relative differences had been detected, which could endorse the tendency to subestimate the variance.

When mean crop yields (calculated by the Two Step estimation method) are compared to mean crop yields (obtained by the Triangular and Beta approximation methods), similar results are obtained; this shows a great methodological persistence in measuring this variable. Crop yield variances are very different, possibly due to the functional specifications and problems regarding the estimation of extremes, especially lower ones.

The methods used for the PDF elicitation were clearly acceptable by both the farmers interviewed and the students who carried them out. They are very simple elicitation methods that can be incorporated into DSS with relative ease.

From the data obtained from the interviews, it is possible to conclude that, in general, we cannot assume a normal PDF to describe crop yield distributions. However in some cases, a near normal PDF distribution can be found. This result ties in with the results obtained in the research done from observed series yields. In general, the literature has pointed out that crop yield PDFs are skewed and non-normal (e.g. Day (1965), Teigen and Thomas (1995), Kaufmann and Snell (1997), Gallagher (1986, 1987), Nelson and Preckel (1989), Nelson (1990), Taylor (1990), Kaylen and Koroma (1991), Moss and Shonkwiler (1993), Ramírez (1997), Goodwin and Ker (1998)). However, Just and Weninger (1999) have argued in favor of not rejecting normal distributions of crop yields. The possible non-normality of the PDF has important implications on the methods which should be used for optimization, as is well-known in the decision making theory.

There is a large number of farmers who show a great accuracy and reliability in their responses for the first and second day interviews. This circumstance is interpreted as an indication of a good knowledge of PDF, although we are aware that this hypothesis still has to be investigated. There is also a great variability between farmers regarding coherence and reliability of the responses.

The empirical evidence gathered on the efficacy of crop yield PDF estimates with the simple methods discussed in this paper will most surely facilitate their inclusion into DSS in their interfaces for the compilation of information. Such a simple approach as the Two Step PDF estimation method is most probably at the same time easily comprehensible and acceptable for farmers and sufficiently efficient as a means of suitably describing PDFs.

References

Anderson, J.R. (1997). An 'ABC' of risk management in agriculture: overview of procedures and perspectives. In Huirne, R.B.M. et al. (eds), Risk Management Strategies in Agriculture: State of the Art and Future Perspectives, Ch. 1. Wageningen, The Netherlands: Mansholt Institute, 1-13.

Bessler, D.A., (1980). Aggregated Personalistic Beliefs on Yields of Selected Crops Estimated Using ARIMA Processes. American Journal of Agricultural Economics 62 (4): 666-674.

Bland, J. M. and Altman, D.G. (1995). Comparing methods of measurement: why plotting difference against standard method is misleading. Lancet 346: 1085-87.

Bland, J. M. and Altman, D.G. (1999). Measuring agreement in method comparison studies. Statistical Methods in Medical Research 8: 135-160.

Day, R.H. (1965). Probability Distributions of Field Crop Yields. Journal of Farm Economics 47 (3): 713-741.

Gallagher, P. (1986). U.S. Corn Yield Capacity and Probability: Estimation and Forecasting with Nonsymmetric Disturbances. North Central Journal of Agricultural Economics 8 (1): 109-122.

Gallagher, P. (1987). U.S. Soybean Yields: Estimation and Forecasting with Nonsymmetric Distributions. American Journal of Agricultural Economics 69 (4): 796-803.

Goodwin, B.K. and Ker, A.P. (1998). Nonparametric Estimation of Crop Yield Distributions: Implications for Rating Group-Risk Crop Insurance Contracts. American Journal of Agricultural Economics 80 (1): 139-153.

Grisley, W. and Kellogg, E.D. (1983). Farmers' Subjective Probabilities in Northern Thailand: An Elicitation Analysis. American Journal of Agricultural Economics 65 (1): 74-82.

Just, R.E. and Weninger, Q. (1999). Are Crop Yields Normally Distributed?. American Journal of Agricultural Economics 81 (2): 287-304.

Kahneman, D., Slovic, P., Tversky, A. (eds.) (1974). Judgment Under Uncertainty: Heuristics and Biases. Cambridge: Cambridge University Press.

Kaufmann, R.K. and Snell, S.E. (1997). A Biophysical Model of Corn Yield: Integrating Climatic and Social Determinants. American Journal of Agricultural Economics 79 (1): 178-190.

Kaylen, M.S. and Koroma, S.S. (1991). Trend, Weather Variables, and the Distribution of U.S. Corn Yields. Review of Agricultural Economics 13 (2): 249-258.

Miranda, M.J. (1991). Area-Yield Crop Insurance Reconsidered. American Journal of Agricultural Economics 73 (2): 233-242.

Moskowitz, H. and Bullers, W.I. (1979). Modified PERT versus Fractile Assessment of Subjective Probability Distributions. Organizational Behavior and Human Performance 24 (2): 167-194.

Moss, C.B. and Shonkwiler, J.S. (1993). Estimating Yield Distributions with a Stochastic Trend and Nonnormal Errors. American Journal of Agricultural Economics 75 (4): 1056-1062.

Nelson, C.H. and Preckel, P.V. (1989). The Conditional Beta Distribution as a Stochastic Production Function. American Journal of Agricultural Economics 71 (2): 370-378.

Nelson, C.H. (1990). The Influence of Distributional Assumptions of the Calculation of Crop Insurance Premia. North Central Journal of Agricultural Economics 12: 71-78.

Norris, P.E. and Kramer, R.A. (1990). The Elicitation of Subjective Probabilities with Applications in Agricultural Economics. Review of Marketing and Agricultural Economics 58 (2-3): 127-147.

Pease, J.W. (1992). A Comparison of Subjective and Historical Crop Yield Probability Distributions. Southern Journal of Agricultural Economics 24 (2): 23-32.

Ramírez, O.A. (1997). Estimation and Use of a Multivariate Parametric Model for Simulating Heteroskedastic, Correlated, Nonnormal Random Variables: The Case of Corn Belt Corn, Soybean, and Wheat Yields. American Journal of Agricultural Economics 79 (1): 191-205.

Smith, J. and Mandac, A.M. (1995). Subjective versus Objective Yield Distributions as Measures of Production Risk. American Journal of Agricultural Economics 77 (1): 152-161.

Sonka, S.T. and Patrick, G.F. (1984). Risk Management and Decision Making in Agricultural Firms. In Barry, P.J. (ed.), Risk Management in Agriculture. Ames, Iowa: Iowa State University Press, 95-115.

Taylor, C.R. (1990). Two Practical Procedures for Estimating Multivariate Nonnormal Probability Density Functions. American Journal of Agricultural Economics 72 (1): 210-218.

Teigen, L.H. and Thomas Jr., M. (1995). Weather and Yield, 1950-94: Relationships, Distributions and Data. U.S.D.A. Commercial Agriculture Division, Economic Research Service, Washington D.C.

Tversky, A. and Kahneman, D. (1974). Judgment under Uncertainty: Heuristics and Biases. In Kahneman, D., Slovic, P., Tversky, A. (eds.), Judgment Under Uncertainty: Heuristics and Biases. Cambridge: Cambridge University Press.

Winkler, R.L. (1967). The assessments of prior distributions in bayesian analysis. Journal.of American Statististics Association 62.