應用有限混合分佈理論評估田間作物減產的因子、時期與面積

Application of Finite Mixture Distributions to Assess Factor, Timing, and Area of Yield Decrease in the Field Survey

國立台灣大學農業工程學系副教授

張　文　亮

Wen-Lian Chang

摘　要

本文以差分法與貝爾方法，推估有限混合分佈曲線的機率特徵值，再由此設立判決參數將混合分佈區分為獨立分佈。應用此方法，分析一個以高鹽分與低鹽分混合再灌溉的高粱田，發現有40%的高粱減產，在種苗成活期間的缺水，其他有6%的減產是在孕穗期的灌溉不足。鹽分在此沒有造成玉米的減產，甚至有利於旱作灌溉，可能的原因是減低入滲率，增加土壤有效水分。

關鍵詞：混合分佈，圖形辨識、差分法、貝爾理論、高粱產量、直線噴灌。

ABSTRACT

The differences and Bayes methods were used to estimate probability features of mixture distributions. The results of estimation provide decision criteria to classify observed objects and to investigate relations of two mixture variables. Using saline and non-saline water to irrigate sorghum in growth period, 40% of the yields declined mainly by water-deficit in seedling period, and another 6% in booting period were found. Salt blended water used in irrigation can benefit yield by increasing water availability from decreasing infiltration.

Additional Index Words: mixture distribution, pattern recognition, differences method, Bayes theory, sorghum yield, line sprinkler irrigation.

INTRODUCTION

Crop damaged by environmental stress has became a serious worldwide problem to limit agricultural development. For a crop damage assesment in a field, we need to identify damage factor, time, and area.

Identification is based on classification. Recently, pattern recognition methods (e.g., principal components, Fisher's discriminate, and cluster analysis) have been widely used in classification. Some researchers applied these techniques to identify plant damage by outside environmental stress. For example, Harding et al. (1985) identified slope and soil nutrient are two main factors which effect growth condition of white spruce. Simmleit and Schulten (1989) identified spruce damaged from air pollution through effects on leave and nutrient deficient in soil. However, before application of pattern recognition methods, one need to determine the number of classified categories, to assure normal distribution of variables, and to know the learning samples to establish discriminant function. Generally speaking, they do not have strong theoretical support but leave to prior knowledge to decide. However, most of the field surveys of plant damage may not have those prior knowledge available.

The purpose of this paper is to offer a quantitative method to estimate those prior knowledge before classification and identification. Because the fundamental unsolved question over simplified characteristics of probability density function of variables; the basic hypothesis of this study is that as objects affected significantly by outside effects, the probability density function does not show a single but a mixture distribution. Then separate a mixture distribution to several components may serve as a required process in classification. In the experiment, two line-source irrigation systems were used to apply saline and non-saline water, respectively, to sorghum [Sorghum *bicolor* (L.) Moench] and found some yield decreased after harvesting. This paper attempts to apply this method to answer the following questions: 1) What was the cause of sorghum yield decline? 2) When did it occur? 3) How much of the yield was affected?

THEORY

A probability density function of a finite mixture distribution is defined by Titterington et al. (1985) as

$$P(X/\theta) = \sum_{j=1}^{k} \pi_j \ f(X/\theta_j) \quad \ldots\ldots\ldots\ldots\ (1)$$

where X is a finite mixture variable, $P(X/\theta)$ is a finite mixture conditional density function, π is the mixing weight, k is number of component, θ_j is parameters of the jth component, and $f(X/\theta_j)$ is the jth conditional component density. Unlike a single component distribution, a mixture distribution does not have a straightforward method to estimate various parameters.

Harding (1949) plotted the cumulative distribution function in normal probability paper to see whether it has evidence of mixture. Since a single normal distribution shows plot as linear, derivation from linearity may have certain mixture. He suggested the inflection point of non-linear may serve as separation indicator of mixture component. Extending Harding's method, Tanner (1959) used the first difference of probability to estimate mean and the second difference of probability density to estimate standard derivation of a finite mixture normal distribution. The zero point of the first and the second difference may approximate the mean value and the standard deviation. However, Tanner (1962) considered his method is depend on the histogram interval selected to generate the probability density. Without an optimum histogram interval to present a density function, parameters recognized from Tanner's method are still questionable.

Based on the Bayes theory (Melsa and Cohn, 1978), Chang (1986) extend Tanner's method by assuming a criteria (Xo) exists to separate a normal-normal mixture distribution by two normal components. Xo may be obtained from

$$\ell n\frac{\sigma_2}{\sigma_1} + \frac{1}{2}\left(\frac{X_0 - \mu_2}{\sigma_2}\right)^2 - \frac{1}{2}\left(\frac{X_0 - \mu_1}{\sigma_1}\right)^2 - \ell n\frac{\pi_2}{\pi_1} = 0 \quad \ldots\ldots\ldots\ldots\ (2)$$

where σ_1, μ_1, and π_1 are standard deviation, mean and the weight of the first component σ_2, μ_2, and π_2 are standard deviation, mean and the weight of the second component. For a three-component finite mixture distribution, two decision criteria will be required.

Similar to Eq. (2), for an exponential-normal distribution, Xo may be estimated by (Chang, 1986)

$$\ell n \frac{\sigma_2}{\mu_1} + \frac{1}{2}\left(\frac{X_0 - \mu_2}{\sigma_2}\right)^2 - \frac{X_0}{\mu_1} - \ell n \frac{\pi_2}{\pi_1} = 0 \ldots . (3)$$

In this study, two types of component distributions are considered, normal and exponential. Since Eq. (2) and Eq. (3) depend on knowing parameters μ, σ, and π, Tanner's method may serve as initial step to estimate those parameters. After knowing Xo, the first compoment is a group of sample smaller than Xo and the second component is a group of sample larger than Xo. Mean and standard deviation in each component is easy to be calculated from classified samples by classical equations. Then using those parameters put into Eq. (2) or Eq. (3) obtain Xo again. Repeat this procedure until criteria is approximated to a constant. In this condition, the determination of the parameters is more accurate than the Tanner's method. After this, the probability of classified error, P(e), may be expressed as

$$P(e) = \pi_1 f(Xo/\mu_1, \pi_1) + \pi_2 f(Xo/\mu_2, \pi_2) \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots (4)$$

and the probability of corrected classification, P(c), may be expressed as

$$P(c) = \pi_1 [1-f(Xo/\mu_1, \pi_1)] + \pi_2 [1-f(Xo/\mu_2, \pi_2)] \ldots \ldots \ldots \ldots \ldots \ldots (5)$$

Thus either correct or error probability is equal to

$$P(e) + P(c) = \pi_1 + \pi_2 \ldots \ldots \ldots \ldots \ldots \ldots (6)$$

which shall be equal to one. In this study, the optimal histogram have a minimum P(e) or a minimum P(c). The decision criteria from the optimal historgram was used to classify investigated variable. After this, samples may be classified to several components.

Assuming each variable has equal number of components, the probability of samples allocated to same classified component may be expressed by a joint probability. For example, for i classified component Wi, X variables have weight π_{xi}, and Y variables have weight π_{yi}, then the joint probability of x and y, P (x, y), can be expressed as

$$P(x, y) = \sum_{i=1}^{k} \left(1 - \frac{\pi_{xi} \quad \pi_{yi}}{\sum_{i=1}^{k} \pi_{xi} \quad \pi_{yii}}\right) f(x \in w_i \text{ and } y \in w_i) \ldots \ldots \ldots \ldots \ldots \ldots (7)$$

Eq. (7) shows that a component, which has less weight in a mixture variable, may have more weight in the joint probability, because it has a smaller probability to match correctly with another mixture variable. As both variables match correctly in a less weight component, degree of similarity, or the joint probability will be increased. As the joint probability close to one, variables X and Y have a stronger relationship.

EXPERIMENTAL DESIGN

Site Conditions:

A field study conducted in 1982 at the West Side Field Station of the University of California, located 64 km southwest of Fresno, is the source of the experimental data to be analyzed here. The soil at the site is a Panoche clay loam (Typic Torriorthents). The experimental plot has dimensions of 45 m by 50 m.

Experimental Design:

A line source sprinkler irrigation system was used to apply measured quantities of water. Two parallel sprinkler irrigation lines were located 15 meters apart in a 45 meter wide area yielding three zones of 15 meters width lying parallel to the direction of prevailing winds. One line applied low salt water over a region 30 meters wide. The low-salt irrigation water, pumped from a deep well, had an average electrical conductivity of 1.5 dS/m. The second line applied high-salt water, again 30 meters wide but overlapping low salt water on one side by 15 m. The high-salt irrigation water pump from a shallow well had an average electrical conductivity of 4.3 dS/m. The depth and timing of irrigation were based on experience. Rainfall during the growing season was insignificant.

Irrigation treatments were started on July 8, 1982. Sorghum (Sorghum *bicolor* (L.) Moench) seeds were planted on July 19, 1982 with two rows per meter equally spaced. Seeds were cultivated in rows parallel to the line source pipeline and extended outward from the pipeline to beyond the range of the sprinkler wetted zone.

From before seedling until harvest, soil water distribution was measured with a neutron meter. Soil water was measured on ten dates: August 6, August 13, August 19, August 27, September 2, September 13, September 28, October 6, and October 13 at five different depths: 0-15 cm, 15-30 cm, 30-60 cm, 60-90 cm, and 90-120 cm. There are three transects of 30 neutron access tubes normal to the sprinkler lines providing 90 sample sites.

Sorghum grain was harvested on November 11, 1982, as 5 of 10 m by 1 m plots. Since the area is 45 m wide, it provides 225 sorghum samples (45 rows x 5 replication). After harvest, the sorghum heads were thrashed and the grain was oven dried at 70°C.

Soil samples were collected before seeding by using a power sampler. Soil samples were taken at five different depths: 0-15 cm, 15-30 cm, 30-60 cm, 60-90 cm, and 90-120 cm. Each depth was sampled at 225 locations (45 rows x 5 transects). The five transects were spaced 10 meters apart normal to the irrigation line. Within each transect 45 samples were taken 1 meter apart. Electrical conductivity of the soil solutions were measured from the saturated paste extract.

For comparing more or less equal numbers of soil water, EC and sorghum measurements, the number of sorghum and EC samples were reduced by arbitrarily selecting values from transects 1, 3, and 5. Hence, the number of sorghum and EC samples statistically analyzed was reduced from 225 (45 rows x 5 transects) to 135 (45 rows x 3 transects). The number of soil water content was increased by changing the sample spacing in each transect from 1.5 m to 1 m. This increase was achieved by interpolating the values of the soil water content linearly with distance.

RESULTS AND DISCUSSION

1. Classified results of the sorhum yield

After classification, the sorghum yield is recognized as mixture distributions with two optimum histograms: two components in the six class intervals and three components in the ten class intervals. The first component located at the left boundary of the mixture density is exponential, following by one or two normals (Fig. 1, 2), regardless of whether two or three components are recognized in the sorghum yield variable.

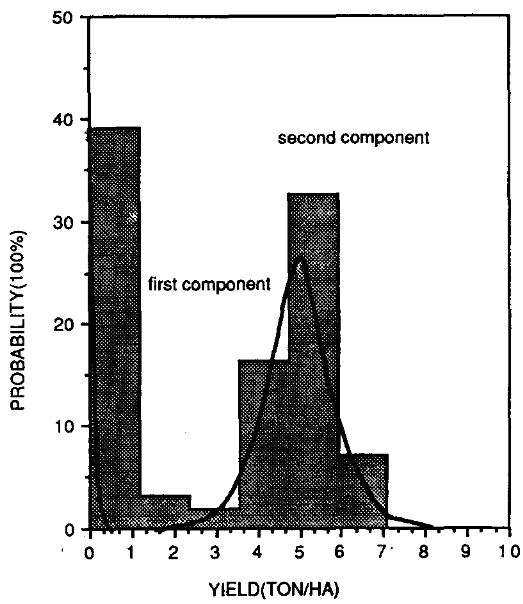

Figure 1. Mixture Distribution of Two Components of Sorghum Yield

a) Two components

As regards to the two components, a harvest higher than .730 ton/ha is classified as the "normal yield". Most of the yield assigned to this component appeared in the overlapping irrigation zone or near the two sprinkler lines (Figure 3).

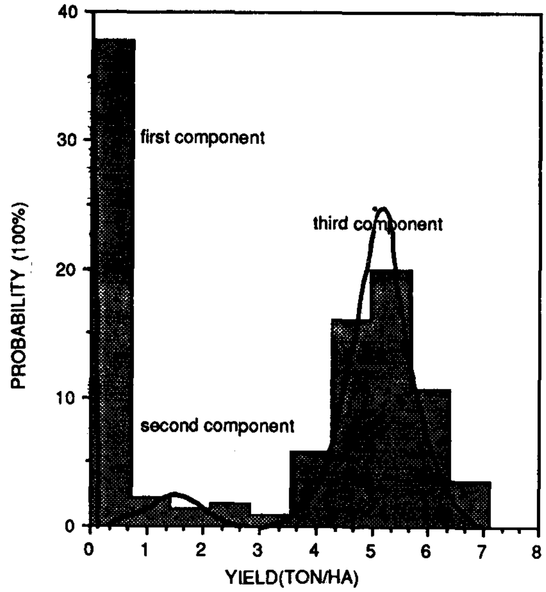

Figure 2. Mixture Distribution of Three Components of Sorghum Yield.

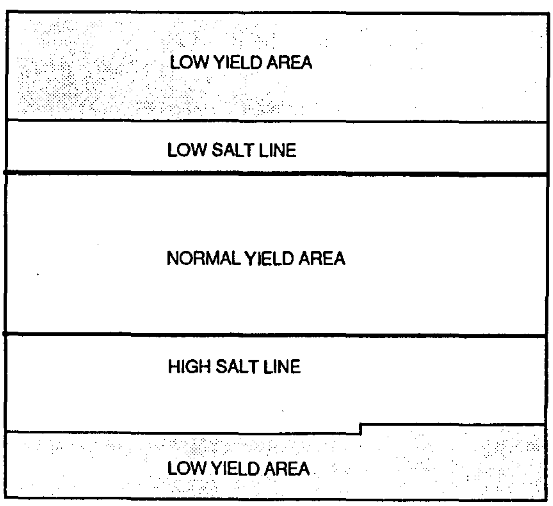

Figure 3. Classified Map of the Two-Component Sorghum Yield.

On the other hand, a harvest lower than .730 ton/ha is classified as the "low yield". Samples belonging to this component are mainly located close to the boundary outside the two sprinkle lines.

b) Three components

Consider for three components, a harvest higher than 3.028 tons/ha is classified as "normal yield". Samples assigned to this component are located in the overlapping irrigation zone or near the two sprinkle lines (Figure 4). A harvest smaller than 3.028 tons/ha but larger than 0.116 ton/ha is classified as "middle yield", which is located outside the two sprinkle lines. A harvest smaller than 0.116 ton/ha is classified as "low yield". Samples assigned to this component are mainly close to the boundary of the field.

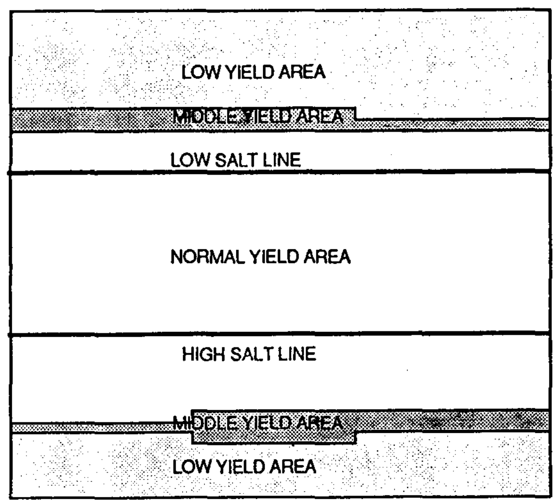

Figure 4. Classified Map of the Three-Component Sorghum Yield

Under either two or three-component classification, yield outside of the two sprinkler lines have different spatial patterns.

2. **Classified results of soil water EC**

a) Two components

Two components are identified above 60 cm soil depth. Criteria to assign samples to different components is shown in Table (1). The "high EC level" are mostly clustered around the high-salt sprinkler line, e.g. in 0-15 cm depth (Figure 5). Apparently, "high EC level" is caused by high-salt sprinkler water, because soil closest to the irrigation line received the largest amount of salt.

EC decision criteria decreased with increasing soil depth. But percent of total samples assigned

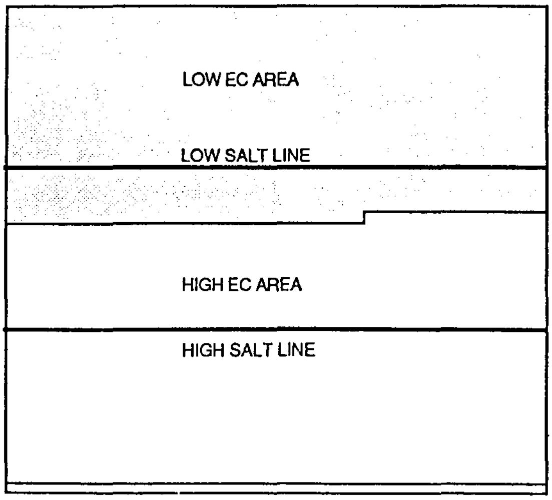

Figure 5. Map of the Soil Water EC for the two-component on 15-30 cm depth.

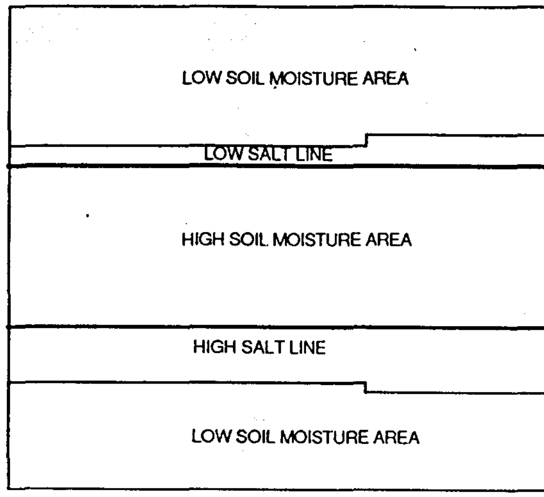

Figure 7. Map of the Soil Water for Two-Component on 9/13/1982 at 0-15 cm depth.

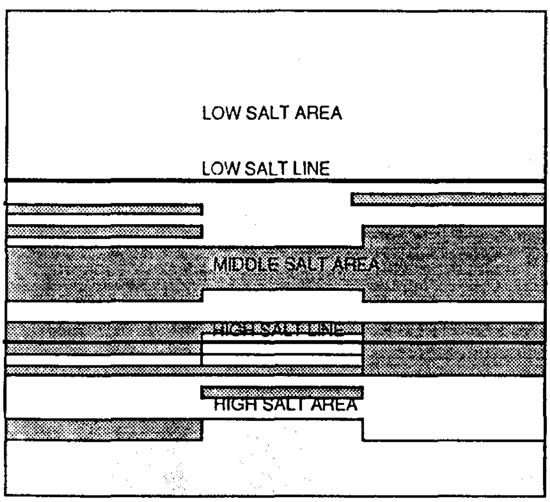

Figure 6. Map of the Soil Water EC for the three-component on 15-30 cm depth.

to the "high EC level" decreased with increasing soil depth. Such a result is reasonable, because soil acts as a filter to retain a certain amount of salt on the soil top. Samples assigned to the "high EC level" decreased more significantly outside the high-salt sprinkler line than overlapping irrigation area. This infers that more salt found in the deeper zones in the overlapping zone was responding to a larger amount of infiltration.

b) Three components

Three components are shown below 15 cm depth. Criteria are presented in Table (1). All high EC samples are located evenly in the vicinity of the high salt sprinkle line, e.g. in 15-30 cm depth (Figure 6). The "middle EC level" mainly located in the region of the high salt sprinkle lines.

The low joint probability of yield and soil water EC for two and three components (Table 1) shows that soil salinity did not cause yield stress.

3. Classified results of soil water content

Most mixture distributions of soil water variables at different class intervals, soil depths and measured periods consisted of two or three normal components.

a) Two components

Based on Table (2), the decision criteria are similar for different measured periods. The only exception occurred on September 2, 1982 (45 days after seedling), when the dicision criteria decreased significantly above 30 cm depth. That may cause by the deficiency of the irrigation applied

Table 1. Classified Results of Soil Water EC

Soil Depth (CM)	Optimum Class Interval	Number of Components	Name of Component	Weight	Bayes Decision Criteria (dS/cm)	Similar Factor With Sorghum
0-15 cm	6	2	Low	0.50	5.41	0.36
			High	0.50		
15-30 cm	7	2	Low	0.55	3.59	0.28
			High	0.45		
	11	3	Low	0.59	3.87	0.23
			Middle	0.24	5.07	
			High	0.17		
30-60 cm	6	2	Low	0.65	2.69	0.27
			High	0.35		
			Low	0.50	0.14	0.23
			Middle	0.19	3.01	
			High	0.31		
60-90 cm	12	3	Low	0.59	0.82	0.16
			Middle	0.34	1.15	
			High	0.07		
90-120 cm	11	3	Low	0.26	0.79	0.31
			Middle	0.44	1.03	
			High	0.30		

Table 2. Decision Criteria of Two Components of Soil Water At different Depths for Different Times

	TIME									
DEPTH (CM)	8/6/82	8/13/82	8/19/82	8/27/82	9/2/82	9/13/82	9/21/82	9/28/82	10/6/82	10/13/82
0-15	.199	.192	.180	.197	.149	.192	.223	.210	.186	.162
15-30	.209	.220	.213	.215	.176	.213	.234	.242	.213	.186
30-60	.204	.212	.215	.211	.199	.210	.202	.207	.225	.200
60-90	.170	.181	1.76	.177	.174	.201	.182	.182	.195	.186
90-120	.194	1.91	—	—	—	.229	.212	—	.219	.212

—: unrecognized decision criteria

from two sprinkle lines or severe evaportranspiration loses during this period, so that the amount of soil moisture decreased up to 30 cm depth.

All spatial maps show samples assigned to the "high soil moisture" are surrounding the two sprinkle lines and distributed unevenly outside the two sprinkle lines above 0-90 cm depth, e.g. 0-15 cm, 9/13/1982 (Figure 7). More high soil moisture samples are located close to the high salt line. However, at 90-120 cm soil depth, more "high soil moisture" is found near the low-salt sprinkle line, e.g. 90-120 cm, 9/13/1982 (Figure 8).

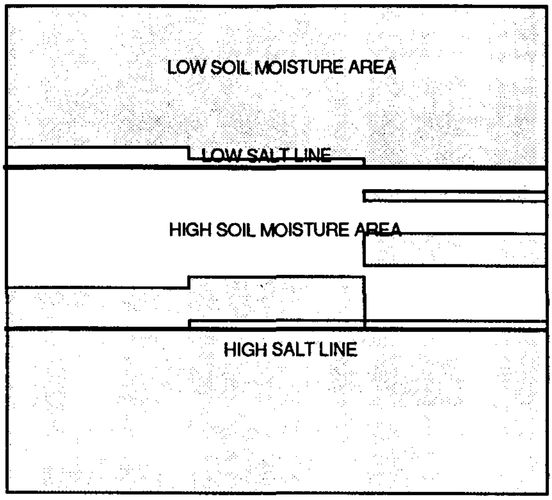

Figure 8. Map of the Soil Water for Two-Component on 9/13/1982 at 90-120 cm depth.

Infiltration affted by soil salinity may be the cause for uneven spatial distribution of "high soil moisture" at different depths. High sodium may decrease infiltration rate. Consequently, closer to the high-salt sprinkle line, more water is found in the shallow depth, and less infiltration to deeper depths because of decreasing infiltration rates. This salt effect may benefit sorghum growth under water stress.

b) Three components

In three components, most of the high soil moisture located in the overlapping irrigation

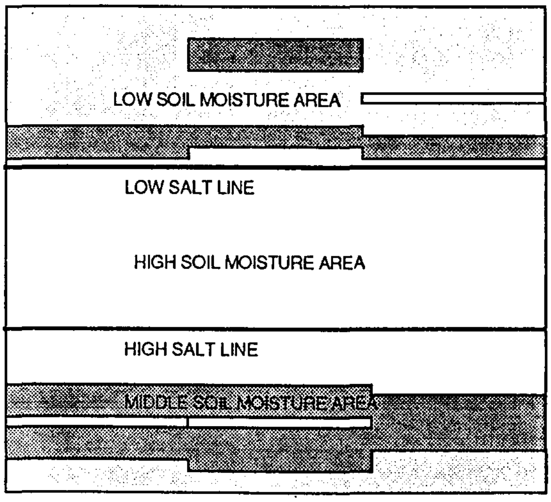

Figure 9. Map of the Soil Water for Three-Component on 9/13/1982, at 0-15 cm depth.

zone. Middle soil moisture level more appear around the high salt irrigation, e.g. 0-15, 9/13/-1982 (Figure 9).

4. Assessing soil moisture effect on sorghum yield

a) Two components

Table (3) shows that the high joint probability ($p > .8$) between soil moisture and yield during the sorghum growth period. This result indicates that soil moisture had a significant effect on sorghum yield at least to a depth of 90 cm. Since the high joint probability of sorghum yield and soil moisture started to occur on August 6, 1982, which is 18 days after seeding emergence, 40% of the area, the low yield (<0.73 ton/ha), was affected by soil moisture stress from this early stage.

b) Three components

Different from two components, Table (4) shows that the high joint probability ($p > .8$) of yield and moisture appeared 56 days (9/13/1982) after seedling, corresponding to the booting period to harvest in the soil surface 0-15 cm depth. Therefore, the middle yield component (<3.208 ton/ha), 6% of the area, responds to water defi-

Table 3. Joint Probability of Two Components of Soil Water and Sorghum Yield at Different Depths for Different Times

	TIME									
DEPTH (CM)	8/6/82	8/13/82	8/19/82	8/27/82	9/2/82	9/13/82	9/21/82	9/28/82	10/6/82	10/13/82
0-15	.937	.867	.900	.879	.857	.895	*.91	*.963	.791	.860
15-30	*.958	*.958	.924	.900	.850	*.91	.903	.902	.904	*.905
30-60	.858	.931	*.931	*.945	*.932	.888	.903	.903	*.918	.903
60-90	.731	.819	.887	.910	.900	.770	.816	.828	.871	.846
90-120	.617	.668	–	–	–	.558	.583	–	.671	.672

*: HIGHEST JOINT PROBABILITY (%) WITH YIELD AT EACH DATE

–: unrecognized decision criteria

Table 4. Joint Probability of Three Components of Soil Water and Sorghum Yield at Different Depths for Different Times

	TIME									
DEPTH (CM)	8/6/82	8/13/82	8/19/82	8/27/82	9/2/82	9/13/82	9/21/82	9/28/82	10/6/82	10/13/82
0-15	0.155	–	–	*.438	0.37	*.817	*.813	*.818	0.28	*.779
15-30	0.123	–	–	–	*.468	0.144	0.2	0.186	*.947	0.681
30-60	0.13	0.14	–	0.317	–	0.284	0.337	0.14	0.142	0.165
60-90	*.259	*.309	*.309	0.339	–	0.347	0.3	0.28	–	–
90-120	–	0.109	0.113	–	–	0.12	–	–	–	–

*: HIGHEST JOINT PROBABILITY (%) WITH YIELD AT EACH DATE

–: unrecognized decision criteria

ciency in the soil surface during the booting period.

CONCLUSIONS

Theoretical Algorithm:

Contrasted to most "typical" classification techniques, the technique described here is to establish decision criteria from a finite mixture distribution. The sequence of this technique is the following.

1). Assume normal and exponential distributions for decomposition of a finite mixture distribution. Results show that this hypothesis is workable to express yield, soil salinity and soil water phenomena in this field investigation.

2). The mean is derived from a root of the first derivative of the probability density function.

3). The standard derivation is obtained from one half of the difference of two roots of the second derivative of the probability density function. Results show no mixture component can really be derived by a normal distribution. So, the mean of a "proximated" normal distribution is bounded by { mean + standard deviation } and { mean - standard deviation }.

4). Weight of a component is determined by Bayes method with iteration.

5). Bayes decision criteria is estimated from the known probability density function by Bayes method with iteration.

6). Based on Bayes decision criteria, each sample can be assigned to a classification category.

7). An optimum class interval expresses a mixture distribution with a minimum mis-classification error and mis-match error.

8). The relationship of two mixture variables is expressed by a joint probability.

Conclusion of Classification:

The results of the classification can answer the five following important soil-plant managerial problems.

1). What is the cause of the yield decline?

Water deficiency, not salinity stress, causes yield reduction in this experiment. On the other hand, salt water seems to improve sorghum yield due to a higher moisture retention resulting from a decreasing transmission rate. Therefore, the mixing of saline and non-saline irrigation water is a possible alternative to increase water volume in a water deficiency area.

2). When did the cause occur?

Yield reduction by water deficiency occurred during the germination and booting periods. Forty percent of the yield was affected by water deficieny during the germination period, and 5.9% of the yield was affected by water deficiency during the booting period. Both periods may be considered as highly sensitive to water deficiency and thus to yield production.

3). How did it develop?

Water deficiency occurred distance from the sprinkle line increased. As a result, the amount of evaportranspiration exceeded the irrigation supply.

4). What depth in the root zone responded to irrigation?

It was recognized that the root zone responded to irrigation up to 90 cm.

5). How deep in the soil profile is the soil salinity affected by irrigation?

The high-salt irrigation water caused soil water EC to increase up to the 90 cm depth, and the low-salt irrigation also caused soil water EC to increase up to 120 cm depth. The difference between the two is caused by the decreasing water conductivity in the high-salt treatment, probably sodium.

BIBLIOGRAPHY

Chang, W. L. 1986. Fitting of Mixture Model for Evaluation of Water and Salinity Stress on Sorghum Yield. Master Thesis in the U. of California at Davis.

Hanks, R. J., J. Keller, V. P. Rasmussen, and G. D. Wilson. 1976. Line source sprinkler for continuous variable irrigation-crop production studies. Soil Sci. Soc. Am. Proc. 40: 426-429.

Harding, J. P. 1949. The use of probability paper for the graphical analysis of polymodal frequency distributions. J. Marin Biol. Assoc., 28:141-153.

Harding, R. B., D. F. Grigal, and E. H. White. 1985. Site quality evaluation for white spruce plantations using discriminant analysis. Soil Sci. Soc. Am. J. 49:229-232.

Melsa, J. L. and D. L. Cohn. 1978. Decision and estimation theory. McGraw-Hill Inc. New York.

Simmleit, N. and H. Schulten. 1989. Pattern recognition of spruce trees. An integrated, analytical approach to forest damage. Environ. Sci. Technol. 23:1000-1006.

Tanner, W. F. 1959. Sample components obtained by the method of differences. J. Sed. Pet., 29:408-411.

Tanner, W. F. 1962. Components of the hypsometric curve of the earth. J. Geophys. Res., 67:2841-2843.

Titterington, D. M., Smith, A. F. M., and U. E. Markov. 1985. Statistical analysis of finite mixture distributions. John Wiley & Sons Ltd.

ACKNOWLEDGEMENTS

I would like to express my sincere thanks to Professor D. R. Nielsen who gives constructive criticism to this study, also thanks to Professor J. W. Biggar for his inspiration on the concept to solve the difficult problem of this study. In addition, I thank Kearny Foundation for its financial support, which made this study possible.

收稿日期：民國81年 8月 8日
修正日期：民國81年 8月27日
接受日期：民國81年11月10日