

An Investigation of the Factors Affecting Paddy yields from Two Districts

ANN LEAKER*

Statistical Unit, University of Colombo, Colombo 3, Sri Lanka

(Date of receipt: 08 July 1982)

(Date of acceptance: 21 May 1984)

Abstract: Factors affecting paddy yield in the Anuradhapura and Ratnapura districts are investigated. Regression analyses are performed using data from 1962 - 1978, and particular attention is given to the growth stages of the Paddy plant and its water requirements at each stage. The results indicate that if the total extents cultivated has been increasing then this has a large influence on yield per acre. This was seen at Ratnapura for both seasons and in the Maha season at Anuradhapura. The study also shows that the non-climatic factors influenced by technological research, (called technological factors in this paper), such as transplanting, type of variety, fertilisers and irrigation have played a major role in paddy production, particularly at Anuradhapura. Variations in climatic variables have not influenced variation in yield, as greatly as changes in technological factors. Also the results indicate that the climatic variables which are important depend on the type of technology employed. For example at Anuradhapura, the Yala season has increasingly relied on major irrigation schemes and temperature was shown to be important, but in the Maha season some farmers still rely, to a certain extent, on rainfall, and consequently rainfall was indicated as an important climatic factor. The study has highlighted many areas of further work and the direction further research should take to enable a comprehensive knowledge of the effects of factors and their interactions to be developed.

1. Introduction

Self sufficiency in rice has been a target that Sri Lanka has been striving to achieve for many years. This can be achieved basically by

- (a) increasing acreage under paddy.
- (b) increasing paddy plant yields.
- (c) reducing risk of crop failure.

There have been several measures taken to achieve this goal of which the development of the Dry Zone for irrigated paddy cultivation has been a major one.⁷ In the mid 1960's the emphasis for self sufficiency was on increasing paddy plant production as opposed to area expansion.¹² For the paddy farmer this meant improved credit facilities, subsidised fertilizers and an improvement of cultural practices, through better extension. Consequently paddy production increased

* Present address: *Department of Applied Statistics, University of Reading, Reading, RG6 2AN, U.K.*

by 55% during 1966 - 1970.⁷ This was then followed by the introduction of the new high yielding varieties¹ which were capable of yielding 100 bushels per acre compared to the 60 bushels per acre of the traditional varieties.

In 1972 the Ministry of Planning developed "The five year plan"¹³ which aimed at self sufficiency in 1977. The plan recognised the scope for more extensive use of land and labour, for increasing output with high yielding strains and for better use of water sources. Large scale irrigation schemes were planned and developed. Also increased supplies of fertilizer, weedicide and pesticide were made available in conjunction with programmes for rural credit. However, in 1974 the government withdrew the fertilizer subsidies which subsequently meant that most farmers were using fertilizers at levels of efficiency much lower than recommended. Chambers, in 1975⁴ pointed out that both the severe shortage of the necessary inputs, on which the green revolution is based, and the economic difficulties of Sri Lanka at the same time, made the immediate prospect of self-sufficiency less feasible. Chambers stressed that a greater emphasis ought to be placed on water management. The above can be summarised by a quote from Seneviratne and Appadurai.¹⁴ "Self-sufficiency in rice has been declared the objective of agricultural policy in this country for decades and numerous measures have been adopted in an attempt to realise this goal with varying degrees of success."

This paper is not intended, to propose methods designed to achieve self-sufficiency. If a clearer understanding of the role played by the various factors affecting paddy yield can be formed, then paddy farmers can make efficient use of available resources to obtain maximum results. If this could be achieved then self-sufficiency would not be such a distant target. There has been much speculation and discussion of the factors affecting paddy plant growth but very few real statistical analyses with concrete results have emerged. This paper sets out the results of a preliminary study carried out on the factors affecting paddy. The results obtained are primarily intended only to give an indication of areas into which further research regarding crop-climatic relationships could be pursued more deeply.

2. Material and Methods

To simplify our initial study we analysed paddy yields for Maha and Yala seasons from two districts; one in the dry zone (Anuradhapura) and one in the wet zone (Ratnapura). This enabled us to obtain a picture of the regional differentiation in productivity potentials and also to provide an indication as to why these differences exist. Factors which affect paddy yields can be categorised as those related to:

- (i) **methodology** eg. variety of seed, method of irrigation, fertilizer.
- (ii) **climatic** eg. rainfall, temperature, number of rainy days.
- (iii) **economic** eg. fertilizer subsidy schemes, rent acts, insurance.
- (iv) **physical location** eg. soil, drainage, vegetation cover.

The factors covered by this study are to be found in categories (i), (ii) & (iii).

The paddy statistics and data related to category (i) were obtained from the Agricultural Division of the Department of Census & Statistics. This data is available at district level from 1962. The paddy yields per acre are estimated from crop cutting surveys.

The climatic data was obtained from the Colombo Meteorology Department. The Statistical Unit, Colombo University has been collecting and processing climatic variables from most of the districts in Sri Lanka. An efficient data entry system has been developed for entering, checking and editing the data, thus the climatic variables are thought to be reliable. Although daily weather data is available, monthly values were used.

For each season and each district the data set thus consisted of paddy yield per acre, total extent cultivated, percentage of area cultivated under major, minor, or rain fed irrigation schemes; per cent acreage under transplanted methodology; per cent acreage under pure-line seeds; fertilizer issues; monthly rainfall; monthly maximum temperature; monthly minimum temperature; monthly average temperature; total monthly rainfall and number of rainy days observed in each season.

3. Method of Analysis

In crop climatic studies there are several different approaches to understanding the degree of influence that various climatic factors have on crop yields. These can be roughly classified as follows:³

- (a) Models which consider the impact of meteorological variables on specific processes e.g. transpiration, photosynthesis. These models then simulate the processes by mathematical equations.
- (b) Models which use meteorological variables on a daily basis and which relate to the development periods of the plant.
- (c) Models which use a sample of yield and meteorological data and produce estimates of the coefficients of a regression equation.

The approach used in this paper is influenced by methods (b) and (c). Since this study is a preliminary investigation then interest is in using a method which will indicate the important factors rather than in developing a model. Consequently regression analysis is used to highlight the relative importance of the variables under consideration. It is intended that the work done in this study will lead to a more detailed investigation where a model will be developed. This is discussed later.

Before any statistical analyses were performed, histograms and time series plots of each variable were constructed to check the data for anomalies, possible errors and to obtain an idea of how each of the factors have been changing over the time period considered. The data was then summarised in the form of means and standard deviations to enable one or two simple comparisons between districts to be made.

For the climatic data, to consider monthly figures meant that for each season a maximum of 8 monthly figures for each variable is possible, however not all 8 monthly values for all the 10 weather variables can be incorporated simultaneously into a regression analysis. We have only 18 yield observations so it is not possible to conduct a multiple regression with 80 explanatory variables. This means that the explanatory variables need to be condensed or reduced into a manageable set before it is possible to include them in a multiple regression analysis. There are several methods of doing this. Examining the work of other authors in this area showed that some authors used multivariable techniques such as principal components⁹ and others have used techniques which screen the meteorological variables,² i.e. reject or select variables depending on their correlation with yield. Kendall has indicated that there are disadvantages and advantages implicit with both of these techniques.¹⁰ Also it is known that some of the weather variables have a considerable effect on the paddy plant during specific growth periods and no effect during others. Therefore some of the explanatory variables could well be unnecessary, consequently the latter method of screening is employed in this study, although it is realised that it has its limitations, and a variable was only discarded if it was indicated statistically and known agronomically as not being important.

The aim of the paper is to discover which factors play a significant role in explaining the variation in paddy yields. This is done through a multiple regression analysis¹⁶ which is conducted in two steps. The first step, as outlined in the above paragraph, identifies the factors which appear to be related to yield. The second step uses forward selection to choose a subset, from the explanatory variables identified at the first step. There are several other methods of selecting an appropriate subset of explanatory variables, apart from forward selection. The drawbacks of most of these techniques are discussed in Draper and Smith and although forward selection may not be the most appropriate technique,⁶ it was the only one available at the time of this study.

STEP I: The correlation structure of the factors was examined as it was necessary to understand the interrelationships which exist between the variables. Simple linear regressions between the yield per acre and those factors believed to affect

paddy yields were performed. In order to decide which regressions were appropriate the percentage of variation accounted for by each regression and the corresponding F-values were noted. From these simple regressions and supported by agronomic knowledge the factors could be classified into 3 groups:

- (A) Those which affected yields directly.
- (B) Those which did not appear to have marked effect on yield but whose rejection was not obvious.
- (C) Those which had no effect on yield at all.

The factors categorised into A are those which appear to be the most effective in changing the yield per acre. However, most of these factors are highly correlated so when they are all incorporated into a multiple regression the total amount of variation explained by them all may not be as large as expected and usually only one or two need to be included in the final subset. Also, multicollinearity amongst the explanatory variables can have serious consequences on a multiple regression analysis.

The factors grouped in B are those variables which seem ineffective on their own but may be operationally important when considered with other variables in the regression model. A non-significant term in the simple regressions analysis cannot always be ignored especially if it is close to the tabulated F-value. In this case it is preferable to include those factors with a large but non-significant F-ratio and to test them with other variables rather than reject them at this stage.

Hence the above classification and the inter-relationships highlighted by the correlation analysis produced a final subset of variables to be considered in the final multiple regression analysis. For the simple regressions we have used

$$y_i = \beta_0 + \beta_1 x_i + \varepsilon_i \text{ where } E(\varepsilon_i) = 0, \text{ Cov}(\varepsilon_i, \varepsilon_j) = 0, i \neq j$$

$$V(\varepsilon_i) = \sigma^2 \text{ and } \varepsilon_i \sim N(0, \sigma^2)$$

y_i = Yield/acre (the dependant variable),

x_i = the factor being considered (the independent variable).

STEP 2: The final subset of variables comprised of groups A and B for both of the seasons, Maha and Yala in each of the districts. A forward selection, multiple regression analysis was performed on the above subset, by using a statistical computer package available at Colombo University. This is a sequential procedure in which explanatory variables are added into the regression analysis one by one. At each stage, the explanatory variable included is the one which contributes most, out of those not yet selected, to the regression sum of squares. The selection stops

when the contribution is judged to be statistically insignificant. The final regression analysis then consists of a set of explanatory variables whose combination cannot be improved on in terms of explaining the variation in yield (i.e. any other variable added to the final combination does not significantly improve the proportion of yield variance explained by climatic factors).

It is the variables included in this set which are of interest for helping to understand which factors play an important role in the production of paddy.

When the selected independent variables are x_1, x_2, \dots, x_t then the final multiple regression equation takes the form

$$y_i = \beta_0 + \beta_1 \times x_{1j} + \beta_2 \times x_{2j} + \beta_3 \times x_{3j} \dots + \beta_t \times x_{tj} + \xi_i \quad i = 1 \text{ (1) } n.$$

where ξ_i are assumed to be independently and identically distributed random variables with an $N(0, \sigma^2)$ distribution. However the regression equation is not a sum of factors "making" a yield but a combination of factors which explain the variation in yield, and as it was not intended in this paper to develop a model, then, apart from the sign, the coefficients are of little interest. The estimation of effects is discussed later in areas of further work. The relative importance of the explanatory variables finally chosen and the implications these have on areas of further work are the main concern of this paper.

4. Analysis

4.1 Descriptive Analysis

(a) Paddy yield per acre

The variable of most interest is the yield per acre of paddy whose variation is what we are trying to explain. Table 1 presents the yield per acre at each location and each season. It shows paddy yields at Anuradhapura exhibiting greater variation when compared to yields at Ratnapura.

TABLE 1. Statistics related to yield per acre at Anuradhapura and Ratnapura

<i>District</i>	<i>Season</i>	<i>Mean</i>	<i>Std. Deviation</i>
Anuradhapura	Maha	51.54	10.71
	Yala	46.07	10.21
Ratnapura	Maha	37.75	06.30
	Yala	38.23	04.47

Table 2 presents differences between the means at the two locations (1 & 2) and the difference between the two seasons at Anuradhapura. A significant difference in the yields between the two districts for both Maha and Yala is indicated. In each case Anuradhapura reveals a significantly higher yield per acre. Hence paddy plant production at Anuradhapura is higher than that at Ratnapura. There is no significant difference in the mean yields per acre between the two seasons at Anuradhapura, hence the paddy plants perform equally well in the two seasons.

TABLE 2. t-test for differences in mean between locations and between seasons at one location

	<i>Differences between Means</i>	<i>Std. Error of Difference</i>	<i>Observed t-statistic</i>
1. Maha	13.79	02.93	04.71* } based
2. Yala	7.84	02.63	02.98* } on 17
3. Anuradhapura	05.47	03.48	01.57 } d.f. based on 34 d.f.

(* - Significant at the 5% level)

The differences between the locations and seasons, over the period of interest, can also be seen from Figure 1. In studying the factors affecting paddy production, we should be able to explain why these differences in the means and standard deviation exist. (The yields for Maha 74/75 at Anuradhapura was considerably lower than others. This was due to a severe drought in the area at that time).

(b) *Extent Cultivated*

For Anuradhapura there is a large difference, in the extent cultivated, between the two seasons. The extents cultivated in Maha have a very definite increasing trend over the period 1960 - 1980, whereas in Yala acreage fluctuates about a constant value. This implies that at Anuradhapura in the Maha season conditions are (or becoming more) conducive to supporting a much larger acreage for paddy production than in Yala. One obvious factor which causes this difference is the difference in the rainfall patterns. Extents cultivated in both seasons at Ratnapura, are almost the same and show a similar pattern over years of a small steady increase which levels off eventually.

(c) *Irrigation*

Anuradhapura

A larger percentage of the acreage in Yala than in Maha is under the major irrigation scheme. This is to be expected as rainfall received in Yala is a lot less than in Maha. The percentage of paddy acreage under rainfed in Yala is almost nil.

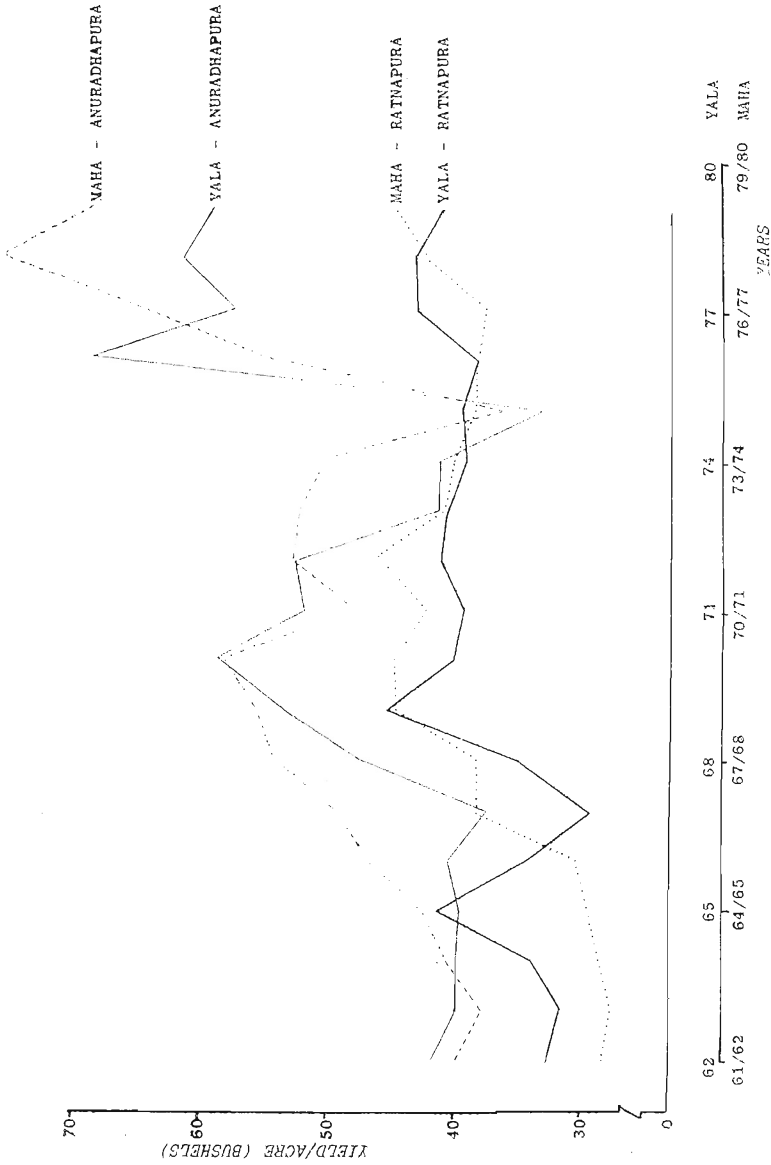


FIGURE 1. Yield per acre at Anuradhapura and Ratnapura (1962 - 1979).

Also within Yala there is an increasing trend in percentages under major irrigation schemes. In Maha there has been only a slight increase in percentages under major irrigation over the past 10 years but still a greater percentage are dependent on minor irrigation with a small almost constant percentage dependent on rainfall alone. Hence in Yala, farmers at Anuradhapura are totally dependent on irrigation schemes and thus have almost completely changed over from minor irrigation schemes to supplement the water received from the Maha monsoon.

Ratnapura

There is very little difference in the patterns of the irrigation schemes between the two seasons. This is to be expected, as the difference in the rainfall experienced in the two monsoons is not as great as that in Anuradhapura. In Ratnapura the farmers rely mainly on minor irrigation and rainfall. The percentage under rainfed has remained fairly constant over the period of study. The percentage under rainfed in Ratnapura is much higher than percentage under rainfed in Anuradhapura during the Maha season.

(d) Weather variables

Paddy production in Sri Lanka is dependent on two monsoon seasons. An idea of the pattern of the climatic variations experienced in the two districts is necessary before correlating yields with climatic factors.

For a representative picture of the rainfall, data was collected from meteorology stations within each district. The daily rainfall figures from 5 stations scattered throughout the Anuradhapura district were averaged to give the pattern of daily rainfall expected for Anuradhapura district. Similarly, for Ratnapura 3 meteorology stations were used. Figures 2 and 3 show the differences between the two seasons at each place.

Anuradhapura

The total precipitation experienced in the Maha season is almost twice the amount in the Yala season. The graph also shows there is a decreasing trend in rainfall amounts recorded in the Maha season, over the period of study.

Ratnapura

The difference between the rainfall in the two seasons is not as large as at Anuradhapura. Generally the Yala season receives almost the same amount of rainfall as the Maha season.

During the Yala period and the inter-monsoonal periods, the amount of rainfall and the number of rainy days is much less for Anuradhapura than Ratnapura.

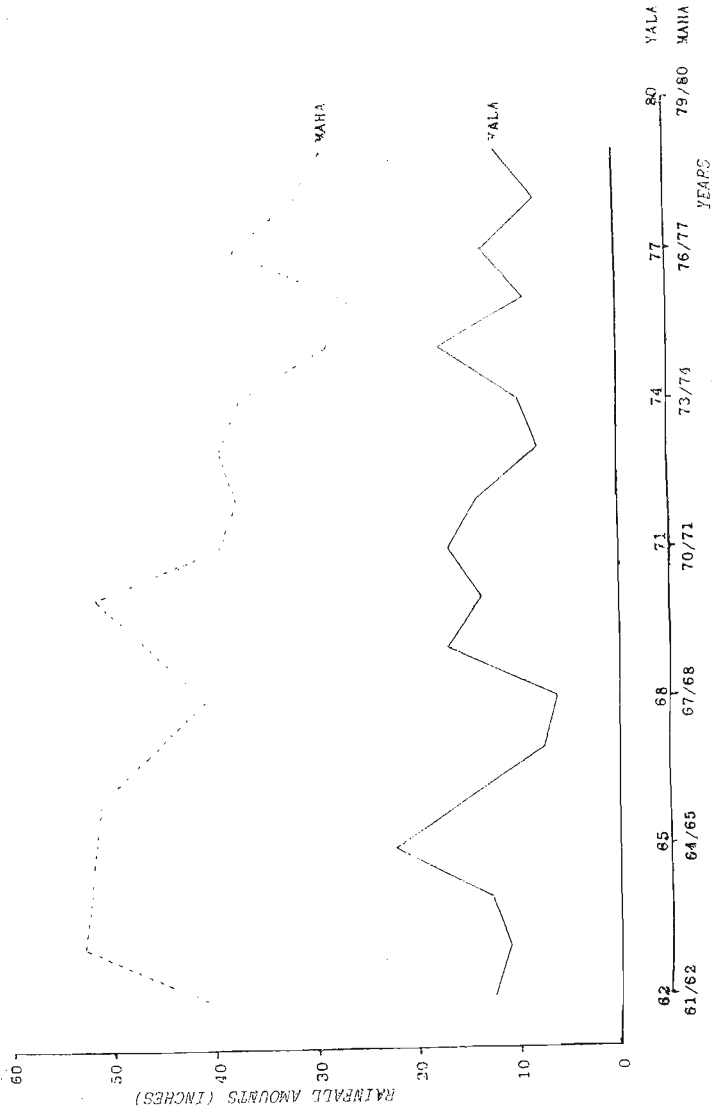


FIGURE 2. Rainfall amounts at Anuradhapura (1962 - 1979).

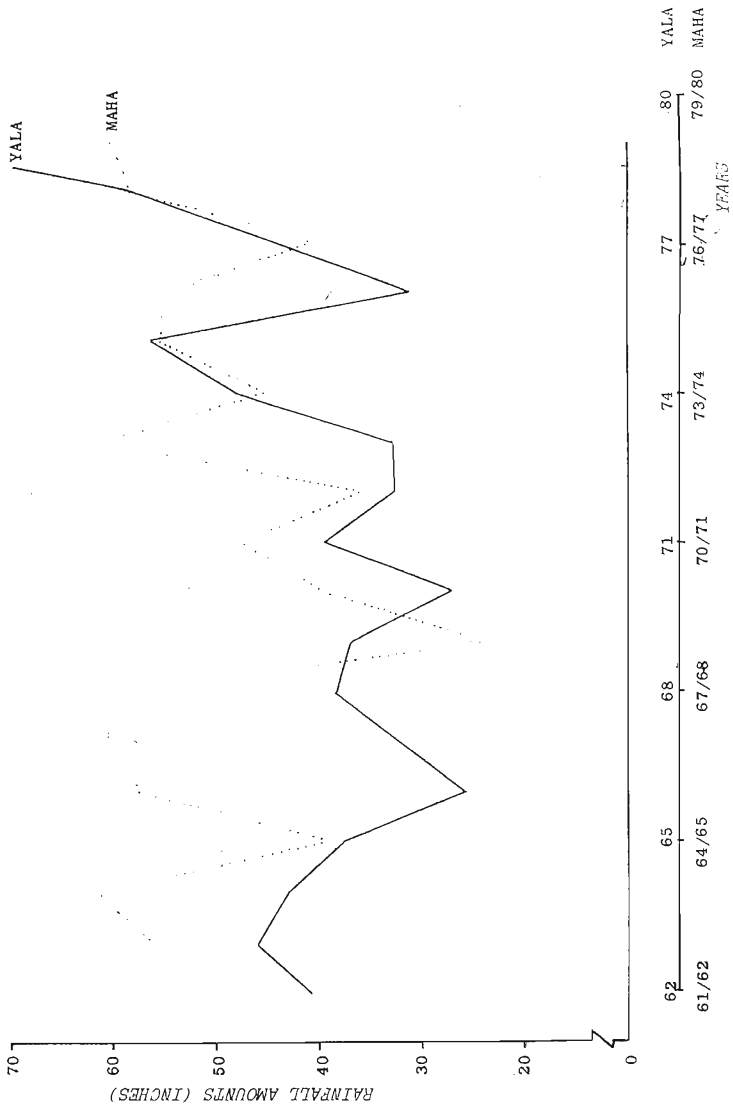


FIGURE 3. Rainfall amounts at Rainapura (1962 - 1979).

The other climatic variables considered were maximum and minimum temperatures. Both of these temperature variables were more variable in Anuradhapura than in Ratnapura. Although the temperatures experienced were not identical for the two places, differences were not as marked as those found in the rainfall figures.

4.2 Use of weather variables in the regression

In considering the effect of weather variables on paddy production, the crop weather analysis approach discussed in section 3 was adopted. Here the climatic variables were related to the different development periods of the crops.

The total growing period of paddy can vary in length between 95 and 150 days and this can be classified by growth stages as follows (see 7).

- (i) Establishment stage (with a length of upto 10-15 days)
- (ii) Vegetative stage (with a length of upto 40-60 days)
- (iii) Flowering stage (with a length of upto 10-15 days)
- (iv) Yield formation stage (with a length of upto 25-35 days)
- (v) Ripening stage (with a length of upto 10-25 days).

The length of, and the requirements at, each stage depend on the paddy variety, its sensitivity to sunlight, temperature and other climatic conditions.

For example, Grist⁸ has shown that Japonicarice in certain environments can yield up to twice that of Indica rice. Kung¹¹ states that, in general, germination does not occur below 12° C and Domros⁵ states that optimum day time, air and water temperatures for paddy yields are between 28° C and 32° C, which agrees with temperatures quoted by Kung, who also reports that small differences between day and night temperatures (i.e. maximum and minimum) during flowering and yield formation are required for good yields.

The total water requirements of the paddy plant depend on climate and length of growing period. However, the water requirements of the plant are different at the different growth stages. Domros⁵ gives the corresponding details for rice cultivation in Sri Lanka. Evapotranspiration increases with vegetative growth and is highest just before flowering. The most sensitive periods of water deficit are flowering and the second half of the vegetative period. Also prolonged submersion of paddy rice adversely affects plant growth, particularly during the yield formation stages.¹¹ So the time at which water deficits or water surpluses occur is an important factor. Thus some weather variables have a considerable effect on the paddy plant during specific growth periods and no effect during others.

In order to investigate the effect on yield, of water deficits and other climatic factors at certain crop development periods of the plant, these periods had to be interpreted in terms of particular months of the year. The total growth period for a Maha season crop is between September/October and February/March and for Yala between April and August. These monsoon periods coincide with the paddy cropping seasons and can be divided into the different crop development stages. The breakdown of the total growing period into the five stages in terms of the number of days was shown above. This was used to obtain a corresponding breakdown for the Maha and Yala cropping seasons. By calculating what proportion of the total growing period each of the five development stages takes, the time periods in each season, which correspond to each stage were found. Table 3 gives an idea of the length of each of the growth stages, the corresponding months and an indication of the water requirements.¹¹ This breakdown is not exact as it would be difficult to obtain a precise correspondence between the months and the growth stages since these will vary over the years, depending on planting dates. This aspect will be discussed later. The amount of rain falling in September/October and April was considered important as this is the time when the rains should have started and consequently affects planting dates. Water requirements are also large in the flowering period. i.e. January and June. The effect of the amounts of rain in the ripening stage, i.e. February/March and August was also examined.

TABLE 3. Length of each growth stage of the paddy plant and their water requirements

Stage	No. of days in period	Months corresponding to growth stage		Water requirements
		Maha	Yala	
1. Establishment	10 - 15	Sept.-Oct.	April	Deep
2. Vegetation	40 - 60	Oct.-Dec.	April-May	Medium
3. Flowering	10 - 15	January	June	Deep
4. Yield formation	25 - 35	February	July	Shallow
5. Ripening	10 - 25	Feb.-March	August	No water

To determine the adequacy of the total amount of rainfall available during the growth period

- (a) average monthly rainfall totals for Maha and Yala;
- (b) number of rainy days within each season;

were examined. For the effect of differences in day and night temperatures during the flowering and yield formation, the difference of maximum and minimum temperatures in January and February for Maha and in June and July for Yala were considered. Average temperatures for these periods and the yield formation stage were also considered.

5. Results and Discussion

5.1 Results of Step I

It was noted in section 4.1a that the yields for Anuradhapura Maha '74/'75 were very low. For the simple linear regressions performed the analysis was repeated, omitting this point and the corresponding data. If the effect of this produced a significantly different regression to that obtained earlier, then the '74/'75 data would be regarded as outliers of an extreme nature. However, this was not found to be the case and so it was not necessary to reject the '74/'75 data points.

The large number of factors which were expected to affect yields were categorised into the three groups by the method outlined in section 3, i.e. by looking at the correlation structure of the independent variables and by performing simple linear regressions between yields and each independent factor. The corresponding results are set out in Table 4. Only the factors in categories A and B are given since category C contains factors which were considered to have no effect on yields.

So a set of twenty possible factors which were believed to effect paddy yields was reduced to a smaller subset of 13 and 9 variables for Anuradhapura Maha and Yala respectively and 11 variables for both Ratnapura Maha and Yala. Each of these smaller subsets comprises of the factors listed in categories A and B, which have been shown statistically to be associated with the yield per acre.

5.2 Discussion of Results

The following section discusses the relative importance of each of the variables chosen in the final subset. This is done by examining the changes in R^2 and \bar{R}^2 on the introduction of each variable.

$$R^2 = \frac{\text{variation of yield explained by all regressors}}{\text{total variation of yield}}$$

$$\bar{R}^2 = \left(R^2 - \frac{k}{n-1}\right) \left(\frac{n-1}{n-k-1}\right) \quad (\text{called the adjusted or corrected } R^2)$$

where k = number of regressors in the (analysis) model

n = number of observations

R^2 provides an overall index of how well y can be explained by the regressors. Since the inclusion of an irrelevant regressor will increase R^2 , then it is desirable to correct for this, as shown above by \bar{R}^2 ,¹⁶ Table 6 shows the values of R^2 and \bar{R}^2 at each stage of the analyses.

TABLE 5. Variables selected as having a significant effect on paddy yields.

<i>Anuradhapura</i>		<i>Ratnapura</i>	
<i>Maha</i>	<i>Yala</i>	<i>Maha</i>	<i>Yala</i>
1. Transplanted extents.	1. Extents under major irrigation	1. Total extents cultivated.	1. Total extents cultivated.
2. Total extent cultivated.	2. Temperature range in June.	2. Rainfall in March.	2. Rainfall in August.
3. Extents under pureline seed.	3. Extents under pureline seed.	3. Extents under pureline seed.	3. Extents under major irrigation
4. Rainfall in September.	4. Temperature range in June and July.	4. Average temperature in March	4. Extents under minor irrigation.
5. Number of rainy days.	5. Number of rainy days.	5. Fertilizer.	5. Transplanted extents.
6. Rainfall in March.	6. Average temperature in August.	6. Extents under minor irrigation.	6. Temperature range in June.
7. Rainfall in January.	7. Fertilizer.	7. Transplanted extents.	
8. Extents under minor irrigation.			
9. Average temperature in February.			
10. Fertilizer.			
$R^2 = 98.0\%$	$R^2 = 81.4\%$	$R^2 = 95.0\%$	$R^2 = 73.9\%$

TABLE 6. Values of R^2 and \bar{R}^2

	<i>Anuradhapura</i>				<i>Ratnapura</i>			
	<i>Maha</i>		<i>Yala</i>		<i>Maha</i>		<i>Yala</i>	
	R^2	\bar{R}^2	R^2	\bar{R}^2	R^2	\bar{R}^2	R^2	\bar{R}^2
1.	0.7	0.68	0.522	0.489	0.793	0.779	0.43	0.392
2.	0.836	0.812	0.569	0.508	0.89	0.874	0.56	0.507
3.	0.886	0.859	0.632	0.547	0.904	0.882	0.64	0.558
4.	0.91	0.879	0.686	0.581	0.916	0.888	0.682	0.576
5.	0.927	0.894	0.712	0.582	0.923	0.889	0.699	0.562
6.	0.937	0.9	0.776	0.64	0.951	0.922	0.737	0.579
7.	0.955	0.919	0.814	0.669	0.958	0.925		
8.	0.964	0.927						
9.	0.974	0.941						
10.	0.981	0.946						

5.2.1 *Anuradhapura Maha*

The forward selection stopped after 10 variables had been included in the analysis. Including all 10 variables explains 98% of the total variation and would appear to explain a substantial amount of the variation. However a closer look at the changes in \bar{R}^2 gives an indication of the relative importance of the explanatory variables. The 10th variable, fertilizer, gives a negligible increase in \bar{R}^2 and cannot be regarded as really contributing to the explanation of variation in yields. The first three variables which produce quite sizeable changes in \bar{R}^2 , are non-climatic factors more associated with technology and account for 89% of the variation. The separate plots of these three factors over time showed a steady increase and the multiple regression analysis showed that all three were positively related to yield. Thus an increase in the use of new technology has helped improve paddy production. The other variables only gradually increase \bar{R}^2 . Variables 4-7 are concerned with rainfall and their combined total change in \bar{R}^2 is = 0.05. Therefore they seem less important in explaining changes in yields as compared to the first three factors which are more associated with technology.

However the indication is that once certain aspects of technology have been accounted for then rainfall is the important factor which affects yields. The monthly rainfall variables chosen were those for September, January, March which indicates the growth periods where rainfall can affect yields. The yields could be adjusted for technology to allow the effects of rainfall to be studied, however technology-weather interactions may drastically affect this approach. This will be discussed later.

5.2.2 *Anuradhapura Yala*

The forward selection stopped after 7 variables had been included in the analysis. Extents under major irrigation explains 52% of the variation in yields. Extents under major irrigation have increased over the period of study from 40% to 90% and since the total extent cultivated has remained constant then this implies that farmers have turned to using major irrigation schemes to cultivate their crops and have become less dependent on rainfed systems. The third variable included is extents under pureline seed which produces a change in \bar{R}^2 of 0.04 and fertilizer produces a change of 0.03, although it was the seventh variable to be included. (Transplanted extents were not included in the final set. In the simple linear regressions the correlations between yield and each of the technology factors such as transplanted extents, extents under pureline, and fertilizer was quite high, thus transplanted extents may not have been included due to the nature of its correlation with other technological factors already included in the final set). So again the technological factors seem important, have been steadily increasing and are positively related to yield. The other variables included are mainly based on temperatures

and these produce increases in \bar{R}^2 which indicate that they are important in explaining variation in yield. The growth periods which seem to be affected by temperatures are flowering and ripening. It is interesting to note that in this analysis the climatic variables included are based on temperature, presumably because rainfall is not such a limiting factor when irrigation is used. Again it would be interesting to remove the effect of technology, if no interaction with climate existed to study the effect of temperature.

5.3 Ratnapura Maha

The forward selection stopped after 7 variables had been included in the analysis. Total extents cultivated explains 79% of the variation in yield. Increases in yield per acre are related here, to increases in use of lands previously uncultivated for paddy. A time plot of total extents cultivated at Ratnapura Maha showed an increase over the period of study. If the nutrient content of soil is higher in these uncultivated lands, then this may explain this relationship. The technological factors included in this set were extents under pureline seed, transplanted extents, fertilizer and extents under minor irrigation. These do not produce increments in \bar{R}^2 like those in the Anuradhapura Maha analysis. Also they do not account for a large proportion of the yield variation as in the Anuradhapura analysis. Here the effects of technology are present but not so pronounced, either because they are not suited to the environment, or because farmers do not practice them extensively. Certainly there appears to be less of a need for major irrigation than in Anuradhapura. However plots of extents under pureline seed against time showed that at the end of the study, 40% of the farmers in Ratnapura use pureline seed as opposed to 70% in Anuradhapura. The climatic variables were mainly concerned with the weather in March, namely rainfall and temperature. Rainfall in March produced a change of 0.095 in \bar{R}^2 and was the second variable included into the regression analysis. The analysis showed that rainfall in March had a detrimental effect on yield.

5.4 Ratnapura Yala

The forward selection stopped after 6 variables had been included in the analysis. Total extents cultivated explains 43% of the variation in yield. The discussion about total extents cultivated in Ratnapura Maha also applies here. The technological factors included were extents under major and minor irrigation and transplanted extents, altogether these produce an increase in \bar{R}^2 of 0.07. Again the discussion in Ratnapura Maha with respect to the technological factors applies here. The decrease in \bar{R}^2 , on including the 5th variable, is due to the increase in \bar{R}^2 being less than $\frac{1}{n-1} \cdot 16$. The weather variables included are rainfall in August and temperature range in June. Although neither of these produce large changes in \bar{R}_2 .

5.5 Summary

		<i>Std. Deviation</i>	\bar{R}^2	$\hat{\sigma}$
Anuradhapura	Maha	10.71	0.946	2.19
	Yala	10.21	0.669	5.75
Ratnapura	Maha	6.30	0.925	1.7
	Yala	4.47	0.579	2.85

The estimate, $\hat{\sigma}$, is based on the residual sums of squares from the final regression analysis in each case. Thus it can be seen that more of the variation in the Maha yields can be explained with the variables studied than that in the Yala yields. This could be due to several reasons. Since less rainfall is received in the Yala season there is a greater dependancy on the irrigation schemes than the adequacy of pre-cultivation rainfall may be a factor to be considered. There has been much more variation in paddy production at Anuradhapura than at Ratnapura. The results seem to indicate that this could be due to the increasing extent at which the cultural practices are affected by technology, particularly at Anuradhapura.

A summary of the general overall indications of the results are:

1. That the yields have been positively influenced by changes in technology, which affect the cultural practices, such as increases in transplanted extents, extents under pureline seed, extents under irrigation and fertiliser issues. Where there has been a marked extension of these changes in cultural practices then the corresponding effect on yield is more pronounced than the effect of weather.
2. Rainfall during certain growth periods affects yields especially in the case of Anuradhapura Maha. The detrimental effect of rainfall in the ripening period was highlighted in 3 of the 4 cases studied. (Anuradhapura Yala season was the only case where this was not shown).
3. Differences in the temperature range during the flowering period and the temperature during the ripening period were also shown to be important variables in Yala cultivation.
4. Where extents under pureline seed has been chosen as being important then so has fertiliser issues. The interaction of these two factors cannot be studied here and would be an important aspect to follow up.
5. The effect of the climate on the paddy production depends on the cultural practices which are used. When most of the extents cultivated are irrigated, as in the Anuradhapura Yala case, then rainfall is not a limiting factor and variations in temperature are important.

6. Conclusion

6.1 Areas of further work indicated by the results

Developing a statistical model to describe how variations in paddy yield are affected by certain factors is an ultimate aim. However, the intention here was to perform a preliminary investigation before conducting an analysis which would be concerned with model development. The main objectives of this preliminary investigation were:

- (a) to assist in developing an appropriate methodology for such an analysis
- (b) to provide information on whether certain areas need to be investigated in more detail and if so which
- (c) to indicate the important/unimportant factors
- (d) to indicate any problems which could arise.

Consequently the results presented above will be discussed in the context of meeting these objectives and the implications these results have on modelling variations in paddy yield.

7. Discussion

(a) Methodology

The statistical techniques used in this preliminary investigation were based on regression analysis, which is what the majority of crop-climatic studies use.² Certainly regression analysis is appropriate for such studies, however, forward selection for model development and estimation of effects may not be the best method.¹⁰ Kendall recommends looking at all of the possible combinations of explanatory variables.

The results of this investigation have indicated that the effects of the technological factors are more pronounced than those of climatic factors. To gain a better understanding of how climate affects yields it may be constructive to remove the effect due to technology from the yields and study the variation of these residuals with respect to climate. However this may not be possible in the presence of weather and technology interactions and consequently their presence needs to be checked first.

The interpretation of coefficients when unknown interactions are present can be misleading. Due to the form of data used it was not possible in this investigation to investigate interactions but the results have indicated that these are present. Subsequent modelling should investigate these possibilities.

Another aspect of methodology, to be investigated further, is the aggregation of the meteorological data. In this investigation the average of 5 stations, in the Anuradhapura district, and 3, in the Ratnapura district, was used to give an indication of the weather in each district. If the weather experienced varies greatly within a district, then a straightforward average may not give a very representative figure, depending on how the meteorological stations are situated in relation to the areas cultivated with paddy.

(b) Areas where a more detailed study may be needed

One important result that emerged was that when water was not a limiting factor i.e. Anuradhapura Yala, then temperature was shown to be important. This suggests that the results will differ depending on the different levels of irrigation used and separate models may need to be developed for each. This is not possible with the data set used here and a breakdown of the district data is needed where yield figures can be matched with the different levels of irrigation.

A further breakdown of the district figures to enable the presence of interactions to be studied, is another area where a more detailed study could be conducted. Information on the yields of the different varieties and the corresponding figures of fertiliser used could also be studied with weather. Such data would also enable an assessment of which varieties are best adapted to which environments. The Department of Census and Statistics crop cutting survey data gives a more detailed breakdown of paddy yield from which it would be possible to investigate the presence of some interactions. This data also allows other factors to be investigated such as the effect of fungicide and weedicide.

If the data available on planting and harvesting dates then the total growth period could also be studied in relation to yield and weather. Knowledge of planting dates provides a more accurate assessment of the growth periods and also will enable pre-season climate to be investigated as a possible factor. For example the extents cultivated is highly dependent on when the rains start and sometimes on the amount of irrigated water available. The type of paddy planted can also depend on when the rains start.

(c) Factors shown to be important

Further areas of work may be in the type of explanatory variables used. The results have indicated which variables appear to be important, however the form of these variables may be incorrect.

Variations in temperature in the Yala season have been shown to be important, however it may be worth investigating whether temperature is the correct variable. Sunshine hours of radiation may be the more relevant. In the case of rainfall then the maximum length of dry spells or wet spells may be more relevant variables. This would mean that daily meteorological data would need to be used.

(d) The results have indicated that problems can arise due to the presence of multi-collinearity. Also the interpretation of coefficients could be meaningless unless the possibility of interactions is properly investigated. This could mean it may be necessary to develop separate models for certain aspects of paddy cultivation, such as different levels of irrigation or different varieties, for interpretations of coefficients to be meaningful.

Overall the preliminary investigation has highlighted the difficulties which need to be considered before developing a realistic model of how paddy yields are affected by important factors. Investigating these difficulties constitutes areas of further work.

Acknowledgements

I should like to thank the Natural Resources, Energy and Science Authority for their support of the Crop-Climatic project at the Statistical Unit, University of Colombo. I should like to express my appreciation to colleagues involved on this project, namely Dr. S. Abeyasekera and Mr. S. T. Nandasara, Miss G. N. de Silva for her invaluable help and Mr. K. Jayasinghe of the Department of Census and Statistics. I should, also like to thank the Division of Agriculture at the Department of Census & Statistics and Meteorological Department in Colombo for releasing their data, and Mrs. Pushpa Gunasekera and Miss Sunanda Marasinghe for their typing.

References

1. ABEYRATNE, E. (1973). *Extension of new high yielding varieties. IDS-ARTI Seminar on Economic and Social Consequences of NHYVs, Kandy, Sri Lanka.*
2. BAIER, W. (1977). *Crop weather models and their use in yield assessments.* WMO Technical Note No. 15. (WMO - No. 458).
3. BAIER, H. (1979). *Note on the terminology of crop weather models.* *Agric. Met.* 20: 137-145.
4. CHAMBERS, R. (1978). *Water management and paddy production in the dry zone of Sri Lanka.* ARTI: Colombo.
5. DOMROS, M. (1974). *Agro - Climate of Ceylon.* Wiesbaden Franz Steiner Verlag. Geological Research 2.
6. DRAPER, N. & SMITH, H. (1966). *Applied Regression Analysis.* Wiley.
7. GOONERATNE, W. (1976). *Recent Climatic Fluctuations and Food Problems in Sri Lanka.* International Symposium on recent climatic change and food production.
8. GRIST, D. H. (1969). *Rice Production.* CIBA, Agrochemical Division, Tech. Mon. No. 1.
9. JONES, D. R. (1981). *A statistical inquiry into crop weather dependence.* *Agric. Met* 26: 291-300.
10. KENDALL, M. (1980). *Multivariate Techniques.* Griffin.
11. KUNG, M. (1970). *Rice.* FAO Publication.
12. MINISTRY OF AGRICULTURE AND LANDS (1970). *Draft Agricultural Development Plan (1971 - 1977).* Colombo.
13. MINISTRY OF PLANNING, (1972). *The five year plan.* Colombo.
14. SENEVIRATNE, S. T. & APPADURAI, R. P. (1966). *Field crops of Ceylon.* Lake House Investments, Colombo.
15. SNEDECOR, W. G. & COCHRAN, W. G. (1967). *Statistical Methods.* Oxford & IBH Publishing Co.
16. WONNACOTT, T. & WONNACOTT, R. (1981). *Regression a second course in statistics.* Wiley.