Trans. Inst. Indian Geographers
Vol. 6, No. 2, July 1984

ASSOCIATION BETWEEN AGRICULTURAL LANDUSE AND PHYSICO–SOCIO–ECONOMIC PHENOMENA : A MULTIVARIATE APPROACH

V.S. Datye and S.C. Gupte, Pune

ABSTRACT : An attempt is made here to express quantitatively the influence of some selected physico-socio-economic phenomena on agricultural landuse. With the help of ICL 1904 S Computer unit (Package ICL 1900 Statistical Analysis XDS 3/26) the effect of fourteen variables on the landuse types of Poona District with respect to hundred thirty five observations (physical regions) was explained using the techniques simple correlation, multiple regression and Principal Component analysis. The results of the correlation analysis and multiple regression bring out the importance of factors of physical environment viz., distance from crest, slope$<3^{o}$ and slope$>20^{o}$. The influence of socio-economic factors is not that significant except on the cash crops. The Principal Component analysis helps to define seven 'dimensions' of agricultural landuse. The loadings and eigen values on the components for each of the fourteen variables and mapping of the component scores for one hundred thirty five physical regions help understand the characteristics of spatial variations.

INTRODUCTION

The location of agriculture and the possible explanation of landuse types has attracted many scholars working in the field of agricultural geography. With the advent of sophisticated quantitative techniques, Geographers are better equipped to deal with the locational problems. In agricultural geography quantitative techniques have been used for quite a long time. These techniques provide the means by which one can study a large number of variables and understand their significance in the explanation of occurrence of phenomena. Multiple regression and correlation models have been used by McCarty et al. (1956), Hartman and Hook (1956), Garrison (1956 b), Thomas (1960 a), Yeates (1965), King (1961), Salisbury and Hart (1965), Lakshmanan (1964), Russwurm (1964) and many others. The principal component analysis is not very popular with the geographers. Moser and Scott (1961), Ahmad (1965) and King (1966) have applied this technique to study various interrelationships. In this paper an attempt is made to explain how and to what extent do the factors of physico-socio-economic environment influence the agricultural landuse in Poona district.

The district is located in the central part of Western Maharashtra. It is triangular in shape with a part of the crest of the 'Sahyadri' mountains as its base. The district receives maximum rainfall (over 3000 mm), in the 'ghats' section (western parts) and the rainfall decreases progressively eastwards to less than 500 mm. The transitional zone receives rainfall between 1000 mm and 600 mm. The western most part of the district is characterised by rugged relief and steep slopes. This is the source region of the major rivers in the district. Eastwards in the transitional zone and semi arid part of the district the valleys open out with low devides and gentle slopes. The cropping pattern of the district is dominated by jowar and bajra followed by oilseeds, pulses, cash crops, rice, wheat and inferior millets.

Initially the district was divided into 135 physical units based on variations in rainfall, relief and soils. Data for major landuse types and other variables were collected at village level for the year 1974-75 and later grouped into 135 regions.

It is observed that rice is grown in the western parts of the district with rainfall above 1000 mm. A variety of crops is grown in the transitional zone and jowar dominates the cropping pattern of semi arid eastern parts with rainfall below 500 mm. Areas not suitable for rice cultivation are given to grass in the western hilly parts. Cash crops including vegetables, fruits, cotton, sugarcane are generally not attempted without irrigation.

METHODOLOGY

In order to investigate the associations between landuse types on the one hand and physico-cultural elements on the other, following fourteen variables – X_1 to X_6 dependent and X_7 to X_{14} independent – were carefully selected from the set of available variables.

X_1 – Net sown area (NSA) (Percentage of total area)

X_2 – Grass (Percentage of total area)

X_3 – Rice (Percentage of NSA)

X_4 – Jowar (Percentage of NSA)

X_5 – Bajra (Percentage of NSA)

X_6 – Cash crops (Percentage of NSA)

X_7 – Irrigated area (Percentage of NSA)

X_8 – Accessibility

X_9 – Owner Cultivators (Percentage of total agricultural workers)

X_{10} – Density of population

X_{11} – Slope less than 3^{o}

X_{12} – Slope greater than 20^{o}

X_{13} – Distance from crest

X_{14} – Distance from major stream

Among the independent variables, distance from crest indirectly represents the variations in the rainfall distribution. It is assumed that amount of rainfall decreases as the distance from crest increases. This is backed by the fact that all the isophyets are almost parallel to the crest line. Thus, this variable in fact represents the rainfall pattern. Secondly increasing distance from crest also means change of relief from hilly, rugged areas to broad open valleys, low divides and low plateaus. It, therefore, reflects the variations in climate as well as in relief. Another variable distance from the major stream (of order 6th and above), is defined in order to bring out the variations in the landuse with changes in the slope characteristics and soils across the river valleys. Two more variables of slope types represent also 'relief' factor. The amount of slope in degrees was obtained from the available one inch topographic maps. The variables irrigated area and accessibility in terms of route length per unit area, are included to estimate the relation between these two economic elements and the landuse types. Lastly the variables population density and proportion of owner cultivators may throw light upon the impact of social factors upon landuse.

The relationships are further studied and analysed applying quantitative techniques like simple correlation, multiple regression and principal component analysis. It would have been impossible to handle a large volume of data consisting of 14 variables and 135 observations without the help of ICL 1904 S Computer unit of the Regional Computer Centre, Pune. Package ICL 1900 Statistical Analysis XDS 3/26 was used for the above analysis.

CORRELATION ANALYSIS

The correlation analysis brings out clearly the importance of physical factors like slope and distance from crest (rainfall) and economic factors like irrigation and accessibility as major factors influencing the landuse patterns (Table I). The increasing proportion of cultivators in the hilly tracts with low NSA and the decreasing proportion of owner cultivators in the irrigated areas with cash crops is clearly brought out. However, the distance from major streams which differentiates between valley locations and divide locations does not show any significant relationship with the landuse variables except that for bajara.

MULTIPLE REGRESSION

In all six multiple regression equations were derived in order to estimate the influence of physico-socio-economic variables (X_7 to X_{14}) on the location of agricultural landuse (X_1 to X_6) in 135 regions of the district defined earlier. The partial correlation coefficients derived indicate the intercorrelation of one independent variable with the dependent variable, with all other independent variables held constant. Thus, these coefficients indicate the relative importance of each independent variable in explaining the total variance. The significance of the individual multiple regression coefficients was tested at 0.01 and 0.05 levels of significance. Lastly the level of 'explained variation' indicated by R^2 was calculated. The results of the analysis are summarised as follows:

1. Dependent variable X_1 NSA

   Independent variables X_7 to X_{14}.

Among the variables included in the regression set distance from crest appears to be very important, followed by slope $< 3^{o}$ and population density. The coefficient is negative for the variable slope $> 20^{o}$. Slope above 20^{o} and distance from stream are not significant though they are included in the regression set. About seventy eight per cent of the variation is explained by the five above mentioned variables. Though the correlation analysis reveals significant relationship of NSA with all variables (except distance from stream) only five variables were included in the regression set and only three were observed to be significant in explaining the occurrence of NSA. However, the economic factors having significant positive correlations with the NSA as revealed by correlation analysis do not occur in the regression set.

TABLE I : CORRELATION MATRIX

	X_1	X_2	X_3	X_4	X_5	X_6	X_7	X_8	X_9	X_{10}	X_{11}	X_{12}	X_{13}	X_{14}
NSA	1.0000	–0.7997	–0.7564	0.6963	0.4914	0.4914	0.4926	0.3008	–0.5115	0.3159	0.8257	–0.7040	0.8289	0.0434
Grass		1.0000	0.6742	–0.5779	–0.4328	–0.4784	–0.4452	–0.3445	0.4677	–0.2155	–0.6701	0.5662	–0.6865	–0.1209
Rice			1.0000	–0.6031	–0.5665	–0.5577	–0.4204	–0.2640	0.4987	–0.2846	–0.7242	0.6959	–0.6383	–0.1249
Jowar				1.0000	–0.0595	0.1780	0.3063	0.2289	–0.4520	0.1528	0.7336	–0.6240	0.7767	0.3114
Bajara					1.0000	0.4494	0.1631	0.1988	–0.1407	0.2257	0.3120	–0.4253	0.1416	0.1646
Cash Crops						1.0000	0.7315	0.2412	–0.5161	0.6649	0.4499	–0.4709	0.2886	–0.0458
Irrigated Area							1.0000	0.1149	–0.5450	0.5544	0.4817	–0.3724	0.4456	–0.0561
Accessibility								1.0000	–0.2379	0.3214	0.3338	–0.3275	0.2250	0.0781
Owner Cultivators									1.0000	–0.3788	–0.5657	0.4207	–0.5224	0.0101
Population Density										1.0000	0.3120	–0.3647	–0.1386	–0.2108
Slope $<3^{\circ}$											1.0000	–0.7925	0.8127	–0.0491
Slope $>20^{\circ}$												1.0000	–0.6240	0.0345
Distance Crest													1.0000	0.0487
Distance Stream														1.0000

All values of $r > 0.17$ are significant at 0.05 level

$r > 0.22$ are significant at 0.01 level

2. Dependent variables X_2 Grass

Independent variables X_7 to X_{14}

The occurrence of grass is strongly influenced by distance from crest and slope $<3^{o}$. Regression coefficients for both are negative. The factors distance from stream and slope $> 20^{o}$ are less important in the regression set. These four variables account for about fifty three per cent of the total variation. Like NSA, grass also has significant correlations ('r' values) with all the remaining variables except distance from stream but only four were considered important enough to be included in the regression set. Though grass has a strong positive correlation with slope$>20^{o}$, in the regression analysis the effect of slope $>20^{o}$ upon grass is not that significant. The economic factors which have significant negative correlations with grass do not occur in the regression set.

3. Dependent variable X_3 Rice

Independent variable X_7 to X_{14}

The occurrence of rice is explained by six independent variables. Of these 'b' coefficients for three i.e. slope $>20^{o}$, distance from stream and slope$<3^{o}$ are important. Regression coefficients for the rest are not significant and thus, the contribution of the variables owner cultivators, distance from crest and irrigated area to the total variation is not important. The total variation explained by six variables is about sixty per cent. The correlation analysis reveals that distance from crest has a strong negative correlation and distance from stream does not have a significant relation with rice. However, in the regression analysis distance from stream, explaining the location of rice confined to valleys appears to be more important than the distance from crest.

4. Dependent variable X_4 Jowar

Independent variable X_7 to X_{14}

Four variables in the regression set explain the occurrence of jowar. The only variable which contributes significantly to the total variation explained is distance from crest. The regression coefficients for the remaining variables i.e. slope $< 3^{o}$ irrigated area, and slope $> 20^{o}$ are not significant indicating their lesser importance. About sixty five per cent of the total variation is explained by these four variables. Slope$<3^{o}$, which has a very strong positive correlation with jowar does not appear that significant in this regression analysis. Another interesting relationship is between irrigation and jowar which is positive in correlation analysis but because of the combined effect of other variables it has become negative in the multiple regression analysis.

5. Dependent variable X_5 Bajara

Independent variable X_7 to X_{14}

The variables slope $>20^{o}$, distance from stream and distance from crest are more or less equally important in explaining the occurrence of bajara as indicated by the respective partial correlation coefficient values. Other variables like slope $<3^{o}$ and population density are less important. The total variation explained is rather low at twenty six per cent. Bajara does not have very strong correlations with other variables as indicated in the correlation analysis, though some important relationships are singled out in the regression analysis. Bajara has a positive correlation with the distance from crest as revealed by correlation analysis but the regression coefficient slopes negatively.

6. Dependent variable X_6 Cash crops

Independent variables X_7 to X_{14}

The regression coefficients for five variables in this case are significant. Only for the variable distance from stream 'b' is not significant. The contribution of the variable irrigated area is most important in explaining the occurrence of cash crops followed by population density, slope$>20^o$, distance from crest and owner cultivators. The total variation explained is about sixty eight per cent. The correlation coefficients for distance from crest and accessibility are nearly equal but in the regression set accessibility has not been included. Besides, the correlation coefficient is positive for distance from crest and negative for distance from stream but the regression coefficients are negative and positive respectively.

The application of multiple regression model to explain the influence of various factors upon agricultural landuse reveals that among the set of selected independent variables, by and large, the variables related to physical environment like slope $< 3^o$, slope$>20^o$, distance from crest and distance from major stream appear to be more important than the factors chosen to represent the socio-economic environment like irrigated area, accessibility, proportion of owner cultivators in the total agricultural workers and population density. Only in the case of cash crops the most important variable contributing to the total explained variation is irrigated area. In this particular case other variables like population density and proportion of owner cultivators are also important.

In five of the six equations, the 'b' coefficient for distance from crest is significant and the partial correlation coefficients also indicate its importance as an explanatory variable. The variables slope$<3^o$ and slope $> 20^o$ have significant 'b' coefficients in three cases each whereas the distance from stream appears to be a less important explanatory variable with significant 'b' coefficients is only two equations. Among the socio-economic variables 'accessibility' was not found to be important enough to be included in the regression set at ninety five per cent level. Thus, it does not appear in any of the regression equations.

PRINCIPLE COMPONENT ANALYSIS

The application of principal component analysis to the data matrix of fourteen variables and one hundred thirty five observations is aimed at investigating the basic dimension of agricultural landuse.

The maximum number of possible components is equal to fourteen which is equal to the number of variables. As the first seven components explain about ninety-one per cent of the total variation, the remaining seven components which explain only nine per cent of the total variation are not extracted and ninety per cent was determined as cutoff point (Table II).

The loadings of the fourteen variables on the seven components, which indicate the correlation of each variable with each component are shown in table III.

The variables are rank-ordered with respect to their loadings on the principal component. The loadings on this component are not so strong (above 0.30 or below -0.30) for six variables. Among these, three positive loadings represent NSA, slope$<3^o$ and distance from crest which are balanced by the negative loadings on rice, grass and slope $> 20^o$. Thus, the first component appears to explain a strong physical factor NSA dimension.

The second component contains some variables like jowar and distance from crest which have positive factor loadings (above 0.30). The negative factor loadings for the variables population density, and cash crops are strong (about 0.50) and rela-

TABLE III : COMPONENT LOADINGS

Variables	I	II	III	IV	V	VI	VII	Adequacy of the Solution
NSA	0.3489	0.1364	0.0406	0.0213	0.1129	0.0832	0.2224	0.2116
Slope $<3^{o}$	0.3432	0.1482	–0.1035	–0.0892	–0.1002	–0.0734	–0.0920	0.1823
Distance from Crest	0.3125	0.3049	–0.1690	0.0877	–0.0169	0.0717	0.1583	0.2574
Jowar	0.2771	0.3619	–0.2758	–0.0450	–0.0894	–0.2800	–0.0021	0.3723
Cash Crops	0.2590	–0.4716	0.0645	0.1692	–0.0227	–0.0315	0.0634	0.3279
Irrigated Area	0.2478	–0.3417	–0.2102	0.3750	–0.1612	–0.0087	0.3371	0.5028
Population Density	0.1854	–0.5278	–0.1400	–0.1947	–0.1063	–0.3987	0.0477	0.5430
Bajara	0.1749	–0.1802	0.6333	–0.0240	0.3712	0.1454	–0.1093	0.6357
Accessibility	0.1551	–0.0621	0.1600	–0.7396	–0.5084	0.2031	0.0791	0.9067
Distance from Stream	0.0109	0.2173	0.5273	0.3685	–0.6119	–0.3630	–0.0319	0.9685
Owner Cultivators	–0.2596	0.1804	0.2093	–0.1976	0.2882	–0.4405	0.7075	0.9398
Slope $>20^{o}$	–0.3121	–0.0644	–0.0370	0.2119	–0.2254	0.4301	0.2328	0.4379
Grass	–0.3123	–0.1194	–0.1449	–0.0451	0.0178	–0.4062	–0.4187	0.4756
Rice	–0.3251	–0.0619	–0.2094	–0.0712	–0.1614	0.0749	0.2181	0.2317

TABLE II : PERCENTAGE OF VARIANCE ACCOUNTED FOR BY EACH COMPONENT

Component Number	Per cent of total Variance	Accumulated Per cent
1	48.44	48.44
2	12.51	60.95
3	9.72	70.67
4	6.72	77.39
5	6.40	83.79
6	3.64	87.43
7	3.40	90.83

tively less strong for irrigated area. Thus, this second component describes jowar – distance from crest dimension where cash crops and irrigated area are less significant. The third component where bajara and distance from stream are strongly positively loaded (0.63 and 0.52 respectively) has weak negative loadings on variables jowar, irrigated areas rice, etc. This is a dimension which describes heavy concentration of bajara.

In the fourth component irrigated area and distance from stream have moderate positive loadings (0.37 and 0.36 respectively) whereas it has a very strong negative loading on variable accessibility (-0.73). It can be labelled as an irrigated area dimension. The fifth component has only two variables i.e. bajara and proportion of cultivators with positive factor loading (0.36 and 0.28 respectively) and two strong negative loadings on the variables distance from stream (-0.61) and accessibility (-0.50). This can be called as distance from stream and bajara dimension. Component sox has a moderate positive loading (0.43) on variable slope 20^{o} and two very weak positive loadings on variables accessibility and bajara (0.20 and 0.14 respectively). The significant negative loadings describe variables owner cultivators, grass, population density and distance from stream. It can be called as a slope $> 20^{o}$ dimension. The seventh component contains a very strong positive loading (0.70) on variable owner cultivators followed by not so strong positive loading on irrigated area (0.33), slope $> 20^{o}$ (0.23), NSA (0.22) and rice (0.21). On the negative side this is counter balanced mainly by grass (-0.41). It can be identified as owner cultivators dimension.

Based on the above observations the following seven 'dimensions' can be identified:

To know the degree to which the variation in the individual variables has been subsumed by the seven components the squares of the loadings of each variable on each component are summed up and the sum is expressed as a percentage. The variable which is most adequately accounted for is distance from stream followed by accessibility and owner cultivators. The variable that is least accounted for is slope $<3^{o}$ and

i) Component I – NSA, slope 3^{o}, distance from crest

ii) Component II Jowar, distance from crest

iii) Component III – Bajara, distance from stream

iv) Component IV – Irrigated area, distance from stream

v) Component V – Bajara, owner cultivators

vi) Component VI – Slope 20^{o}

vii) Component VII – Owner cultivators

in this case only eighteen per cent of the variation in the variable is subsumed by the seven components.

The component scores on each of the above seven components were obtained for one hundred thirty five regions. The scores were mapped for the first three components at the discrete interval of 1 S.D. on positive and negative sides.

Component one separates the eastern and transitional zones of the district from the western parts (Fig. 1). The positive loadings on this component for NSA, slope $<3^{o}$ and distance from the crest results in high scores for regions having high percentage of their total area under cultivation with increasing proportion of gently sloping land and distance from the crest. The negative loadings on this component for rice, grass, slope $>20^{o}$ and owner cultivators result in low scores for regions having higher proportion of rice and grass, high relief indicated by higher slopes and higher proportion of cultivators to the total agricultural workers.

The positive loadings on component two for jowar and distance from crest produce a different spatial pattern of scores (Fig. 2). Eastern and central parts of the district are separated from the western part but the division is not as clear as revealed by the scores on component one. The strong negative loadings for variables population density, cash crops and irrigated area are responsible for lower negative scores. Obviously these are the regions which have higher density of population and higher proportion of cultivated area under irrigation and area under cash crops.

Figure 3 reveals a slightly different spatial pattern with high scores resulting due to strong positive loadings on component three for bajara and distance from stream, showing a concentration of bajara in the central portions of the district. Here the proportion of bajara to the total cultivated area is higher. To the east and west from this central part are regions with low scores resulting from the negative loading for variables jowar and irrigated area. Bajara generally is grown on inferior soils on divides and as the distance from stream increases, proportion of area under bajara also increases.

CONCLUSION

The quantitative techniques used here to find out the association between characteristics of variation in the landuse and socio-economic as well as physical environment are of great value as tney

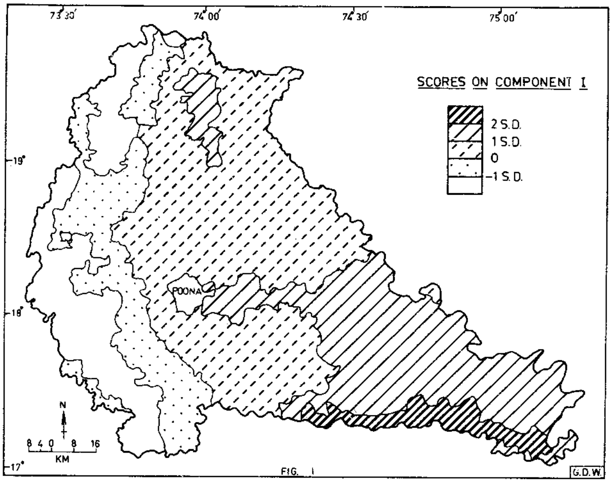

FIG. 1

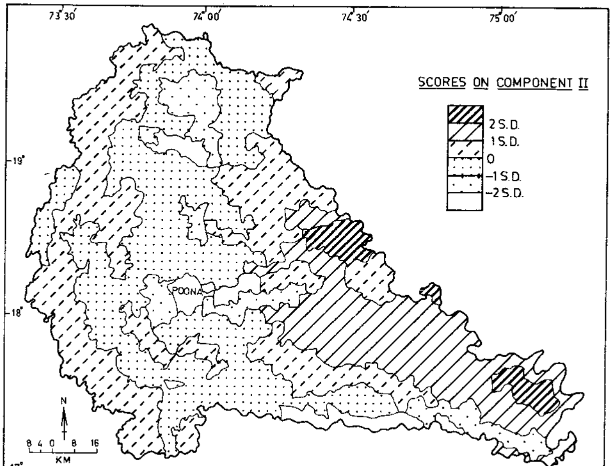

FIG. 2

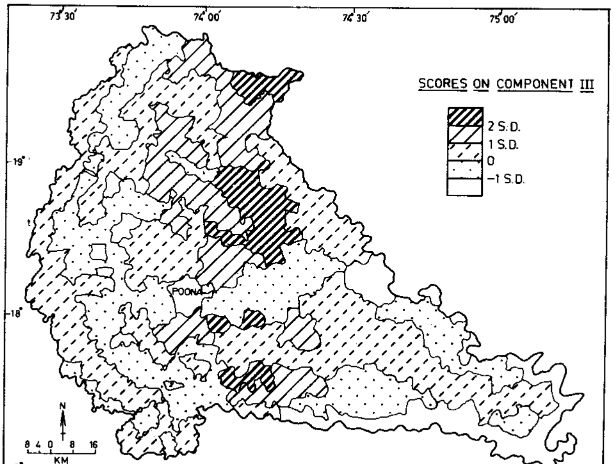

FIG. 3

throw light upon the contribution of each independent variable towards explaining the occurrence of landuse types. The results of the correlation analysis and multiple regression bring out the importance of the factors of physical environment which have a strong influence on the landuse. The principal component analysis also has brought out these relationships more clearly defining seven dimensions of landuse and has further provided a comparatively deeper appreciation of the variation in regional characteristics. The loadings and eigen values on the components for each of the fourteen variables and the component scores for 135 physical units also have brought out the characteristics of spatial variations. In fact, mapping of the component scores for the three components have very clearly brought out the validity of the regional frame developed on the basis of relief and rainfall variations.

REFERENCES

Ahmad, Q. (1965). "Indian Cities : Characteristics and Correlates", *Research Paper 102*, Department of Geography, University of Chicago.

Garrison, W.L. (1956 b). *The Benefits of Rural Roads to Rural Property*. Seattle : Washington State Council for Highway Research.

Hartman, G.W. and Hook, J.C. (1956). "Substandard Housing in the United States : A Quantitative Analysis", *Economic Geography*, **32** : 95–114.

King, L.J. (1961). "A Multivariate Analysis of the Spacing of Urban Settlements in the United States". *Annals*, Association of Ameriacan Geographers, **51** : 222–233.

King, L.J. (1966). "Cross – Sectional Analysis of Canadian Urban Dimensions : 1951 and 1961", *Canadian Geographer*, **10** : 205–224.

Lakshmanan, T.T. (1964). "An Approach to the Analysis of Intraurban Location Applied to Baltimore Region", *Economic Geography*, **40** : 348–370.

McCarty, H.H., Hook, J.C. and Knos, D.S. (1956). *The Measurement of Association in Industrial Geography*, Iowa City : Department of Geography, University of Iowa.

Moser, C.A. and Scott, W. (1961). *British Towns : A Statistical Study of their Social and Economic Differences*, London, Oliver and Boyd Ltd.

Russwurm, L.H. (1964). "The Central Business District Retail Sales Mix, 1948–1958", *Annals*, Association of American Geographers, *54* : 524–536.

Salisbury, N.E. and Hart, J.F. (1965), "Population Change in Middle Western Villages : A Statistical Approach", *Annals*, Association of American Geographers, **55** : 140–160.

Thomas, E.N. (1960). Areal Association Between Population Growth and Selected Factors in the Chicago Urbanized Area", *Economic Geography,* **36** : 158–170.

Yeates, M.H. (1965). "Some Factors Affecting the Spatial Distribution of Chicago Land Values, 1910–1960", *Economic Geography,* **41** : 57–70.

ADDRESS OF THE AUTHORS

Mr. V.S. Datye
Lecturer
Department of Geography
University of Poona
Pune - 411 007

Mr. S.C. Gupte
Professor
Department of Geography
University of Poona
Pune - 411 007