Analysis to Establish Relationship among Subsystems of a Regional System: A Case of Bihar, India

Rashmi Kumari [1] and V. Devadas [2]

[1] *Ph.D. Scholar, Department of Architecture and Planning, IIT Roorkee*
[2] *Department of Architecture and Planning, IIT Roorkee*

ABSTRACT

In this study, the relationship among the physical, social, economic, environmental, ecological, infrastructure and institutional subsystems of a regional system has been validated using correlation and regression analysis. The study utilised the time series data from the study area of Bihar, India, collected from the secondary sources for the quantitative research on tangible and intangible parameters of regional planning. The correlation coefficients have been calculated and the graphs have been generated for the statistical analysis of the selected set of variables. The paper identifies causal linkages among the subsystems of the regional system. The result of the analysis shows that the variables, considered as the indicators of different subsystems of a regional system, are highly correlated with each other, hence, advocates the integrated planning approach for the sustainable development of a region. The results of present analysis pave the way for further modelling and simulation to get optimum plan policy for the region's development.

Keywords: Correlation, Regression, Region, System, Subsystem, and planning.

1. INTRODUCTION

A planning region or an area comprises of many complex structures and behavioural phenomena which more or less change with time due to some external forces acting upon or the internal forces acting within the regional system. While planning for the development of a region we consider some well known phenomena and structures within the regional system, which includes land use; soil type; topography; people and their social condition; demography; economy in terms of gross domestic product (GDP), resources, income, finance, exports etc.; forest; flora and fauna; pollution; roads; communication; energy; governance; institutions etc. These phenomena and structures can be grouped under seven categories, subsystems of a regional system, according to the intensity of their functional and behavioural association with each other. These seven subsystems, physical subsystem; social subsystem; economic subsystem' ecological subsystem; environmental subsystem; infrastructure subsystem; and institutional subsystem, function as a whole. The purpose of this study is to test the hypothesis of relationship among regional planning variables using correlation and regression analysis. The study area of Bihar, a state in eastern India, has been

selected for the present research. It is the 12th largest state of India in terms of geographical area at 93,596 km^2 and 3rd largest by population. The urban population is 11.3 %, and literacy rate is 63.82% in Bihar. It is bounded by Uttar Pradesh to its west, Nepal to the north, Northern part of West Bengal to the east and by Jharkhand to the south.

2. CORRELATION AND REGRESSION ANALYSIS

The application of correlation and regression analysis in the regional planning is evident since long time in the form population forecast; economic growth prediction; and social sciences. The correlation analysis measures the degree of linear association between two variables and regression analysis involves the testing of a hypothesized relationship between a dependent variable and one or more independent variables, and estimation of the parameter values which have been used to develop an estimated regression equation to predict the unknown values of the dependent variables using corresponding known independent variables. In the present paper, the application of correlation and regression analysis has been done in order to affirm the hypothesis.

The correlation between two variables measures the degree of linear dependence between them. A positive correlation between two variables x and y implies the direct linear relationship and a negative correlation implies the inverse linear relationship. By definition a “correlation coefficient” close to zero indicates weak or no linear relationship between two variables; correlation coefficients close to 1 or -1 denote a strong positive or negative relationship [1, 2].

The regression analysis is a statistical process for estimating the relationships among variables. The earliest form of regression was the method of least squares. The term "regression" was coined by Francis Galton in the nineteenth century to describe a biological phenomenon [3]. His work was later extended by Udny Yule and Karl Pearson to a more general statistical context [4].

The simplest linear model involves only one independent variable and states that the true mean of the dependent variable changes at a constant rate as the value of the independent variable increases or decreases [5]. Thus, the functional relationship between the true mean of *Yi*, denoted by $E(Yi)$, and *Xi* is the equation of a straight line:

$$E(Yi) = \beta_0 + \beta_1 Xi. \quad (1)$$

β_0 is the intercept, the value of $E(Yi)$ when $X = 0$, and β_1 is the slope of the line, the rate of change in $E(Yi)$ per unit change in X.

3. METHODOLOGY

The year wise data, from 2001 to 2010, about the selected variables have been collected from the study area of Bihar (refer with: Table 1). The data collected from secondary sources have been

analysed using statistical tools like Ms. Excel and SPSS software. Further, the missing data from each variable has been predicted using the estimated equation formed by the regression analysis of those variables. The parameters of the estimated equations have been calculated and the relation between the dependent variable and independent variable has been established in the form of linear equations (refer with: Eq. 1). After generating all the unknown data, the correlation coefficient has been calculated among all the variables to affirm the positive or negative linear relationship among them.

Table 1: List of variables selected under each subsystem for the analysis

Region (System)	Physical Subsystem	1.Total geographical area (In Ha.) 2. Land put to non agricultural uses (% of total area) 3. Net sown area (% of total area)
	Social Subsystem	1. Population 2. Literacy Rate (%) 3. Literacy Rate (Males) (%) 4. Literacy Rate (Females) (%) 5. sex ratio 6. % Urban Population 7. Birth rate 8. Death rate 9. Infant Mortality Rate (IMR)
	Economic Subsystem	1. GSDP in Primary sector (Rs. Crore) at current prices 2. GSDP in Secondary sector (Rs. Crore) at current prices 3. GSDP in Tertiary sector (Rs. Crore) at current prices 4. GSDP (Rs. Crore) at current prices 5. Per Capita GSDP (Rs.) 6. NSDP (Rs. Crore) 7. Credit Deposit ratio (CD ratio)
	Environmental Subsystem	1. Concentration of pollutant SO_2 (in µg/m3) 2. Concentration of pollutant NO_2 (in µg/m3) 3. Concentration of pollutant PM_{10} (in µg/m3)
	Ecological Subsystem	1. Total Forest Area (In Sq. km)
	Infrastructure Subsystem	1. Total area irrigated (In thousand hectares) 2. Number of Health Centres (PHC, Sub Centre, APHC)

		3. Road Length (In Km) 4. Number of Electrified Villages 5. Number of commercial bank branch offices 6. Available Electricity (MW)
	Institutional Subsystem	1. Number of Judges of Chief courts 2. Number of District Judges

4. RESULT AND DISCUSSION

The observed correlation coefficients has been categorised in terms of strength of linear relationship or degree of interdependence as follows:

Range of Correlation Coefficient 'r'	Degree/Strength of linear relationship
$0.7 < \|r\| \leq 1$	High/Strong
$0.4 < \|r\| \leq 0.7$	Moderate
$0.2 < \|r\| \leq 0.4$	Low/Weak
$\|r\| \leq 0.2$	Very low or No relation

All the variables shows very strong correlation with time except the variables, *total irrigated area* (0.589); and *concentration of* PM_{10} (0.402). These two variables show moderate interdependence with time variable (refer with: Table 1). The calculated correlation coefficient of *Population; Literacy rate; Literacy Rate (Males); Literacy Rate (Females); sex ratio*; and *Infant mortality rate* with each of the variables is more than 0.7, showing higher degree of interdependence except two variables, *total irrigated area* which shows moderate correlation; and *concentration of* PM_{10} which shows weak correlation (refer with: Figure 2). The correlation coefficient between *birth rate* and each of remaining variables is more than 0.7, showing higher degree of interdependence except two variables, *concentration of* PM_{10}; and *total irrigated area* which shows moderate correlation (refer with: Figure 1). The correlation coefficient between *death rate* and each of remaining variables is more than 0.7, showing higher degree of interdependence except four variables showing moderate interdependence which are *CD ratio; net sown area*; *and concentration of pollutant* NO_2 *and* PM_{10} (refer with: Figure 2). The correlation coefficient between *percent urban population* and each of remaining variables is more than 0.7, showing higher degree of interdependence except four variables. The variables showing moderate interdependence with *Percent urban population* are *net sown area; total irrigated area; number of health centres;* and *available electricity*. The variable, *concentration of pollutant* PM_{10}, shows very less correlation with the *percent urban population* (refer with: Figure 2). The variables, *GSDP in Primary sector; GSDP in Secondary sector; GSDP*

in Tertiary sector; GSDP; Per Capita GSDP; and NSDP, show very strong correlation (more than 0.7) with each of the variables except the four variable, *CD ratio*; and *concentration of pollutants* SO_2, NO_2 *and* PM_{10} with which they show moderate correlation (refer with: Figure 3). The variable, *CD ratio,* is highly or moderately correlated to each of the variables excluding a few which are, *concentration of pollutant* PM_{10} (very weak correlation); *total irrigated area;* and *number of health centres* (refer with: Figure 3). The variable, *land put to non agricultural uses,* is highly correlated to each of the variables except four moderately correlated variables which are *CD ratio; concentration of pollutant* NO_2 *and* PM_{10}; and *total irrigated area* (refer with: Figure 4). The variable, *net sown area,* is highly or moderately correlated to each of the variables except two variables, *concentration of pollutant* NO_2; and *total irrigated area* which show weak and very weak correlation respectively (refer with: Figure 4). The variable, *concentration of pollutant* SO_2, shows very strong correlation with each of the variables except *CD ratio; concentration of pollutants* NO_2 *and* PM_{10}; and *total irrigated area* with which it shows moderate correlation (refer with: Figure 5). The variable, *concentration of pollutant* NO_2, shows high or moderate interdependence with each of the variables except the variables, *net sown area; concentration of pollutant* PM_{10}; and *number of health centres* which shows weak correlation (refer with: Figure 5). The variable, *concentration of pollutant* PM_{10}, shows very weak correlation with the variable *CD ration* and shows high correlation with none of the variables.

The correlation has been observed moderate or low with the remaining variables (refer with: Figure 5). The linear dependence between the variable *total area irrigated* and the variables *death rate; and number of district judges* are high. The variable shows moderate or low correlation with each of the remaining variables except *net sown area* with which the correlation is very low (refer with: Figure 8). The variable, *number of health centres,* shows high or moderate correlation with each of the variable except the variables *CD ratio*; and *concentration of pollutant* NO_2 with which the correlation is low (refer with: Figure 8). The variable, *road length,* shows very high linear dependence with each of the variables except the three moderately related variables which are *concentration of pollutant PM10; total irrigated area*; and *number of health centres* (refer with: Figure 8). The variable, *number of electrified villages*, is highly correlated to each variable except the three moderately related variables which are *concentration of pollutant* NO_2 *and* PM_{10}; and *total area irrigated* (refer with: Figure 8). The variable, *number of commercial bank branch offices*, is highly correlated to each variable except the two moderately related variables which are *concentration of pollutant* PM_{10}; and *total area irrigated* (refer with: Figure 8). The variables, *available electricity; number of Judges of Chief courts*; and *number of district judges*, show high or moderate correlated to each of the variables and do not show low correlation with any of the variable (refer with: Figure 7, Figure 8).

It can be clearly stated from the above results of regression and correlation analysis that there exists the direct (positive) or inverse (negative) linear relationship among the variables. Further, the graphs generated also validate the same. Hence, the hypothesis of interdependence of subsystem of a regional system is validated with this quantitative analysis (refer with: Figure 1).

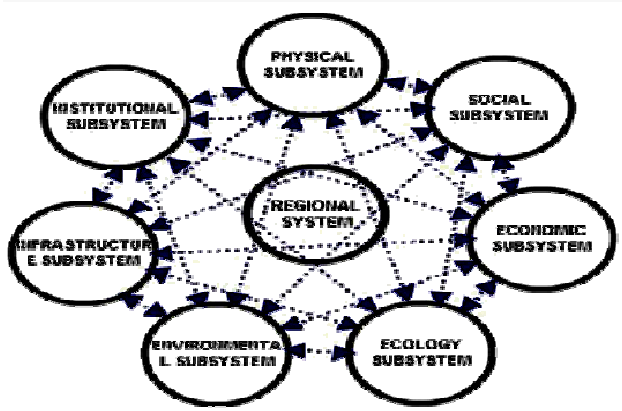
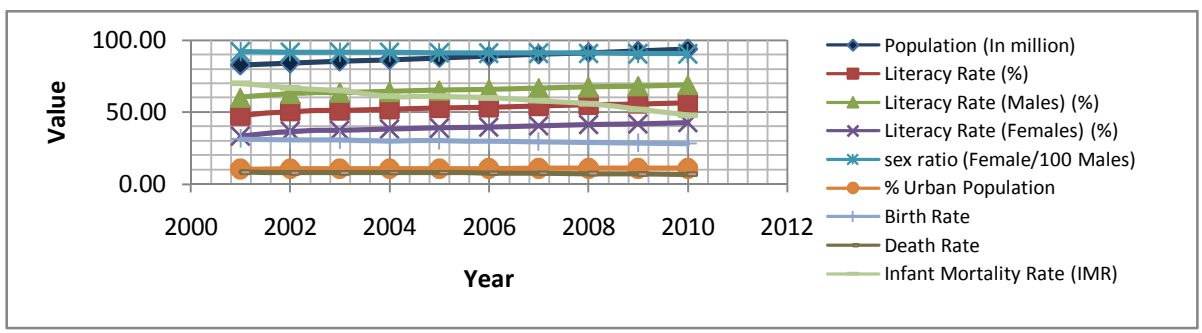
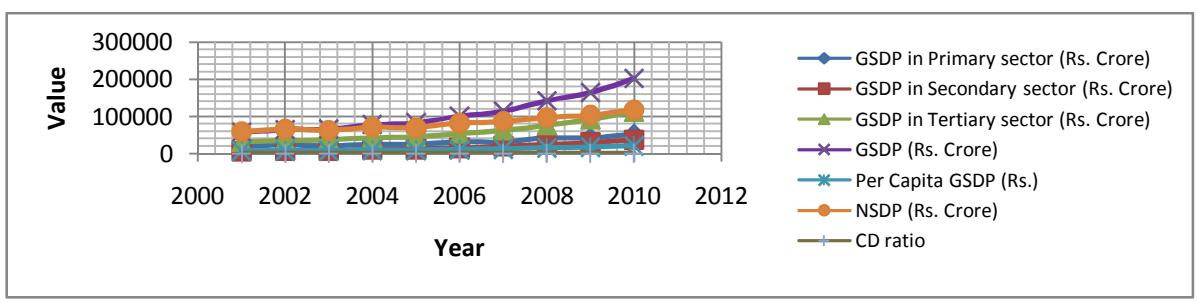


Figure 1. Interrelation between subsystems of a regional system



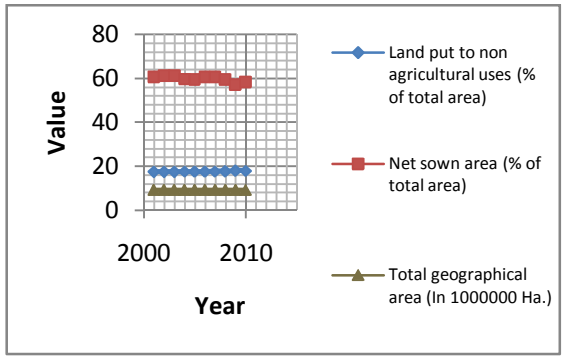
Sources: Census of India; Directorate of Economics and Statistics, Bihar, Patna.

Figure 2. Graph showing trend of change in social condition in Bihar (2001-2010)



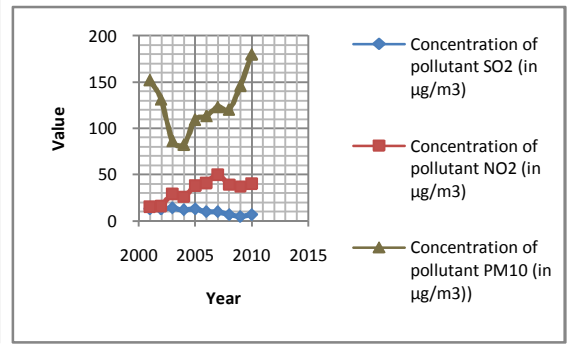
Sources: Directorate of Economics and Statistics, GOB; State level bankers committee.
Note:-GSDP and NSDP are at current prices

Figure 3. Graph showing trend of change in economic condition in Bihar (2001-2010)



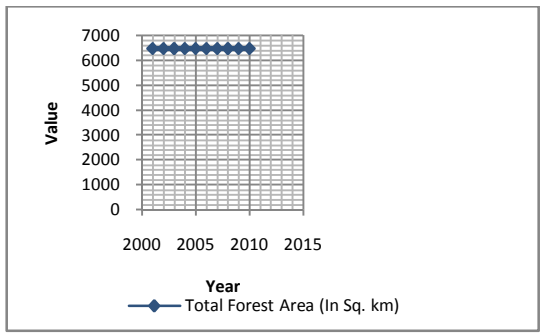
Source : Department of Agriculture, GOB.

Figure 4. Graph showing trend of change in physical condition in Bihar (2001-2010)



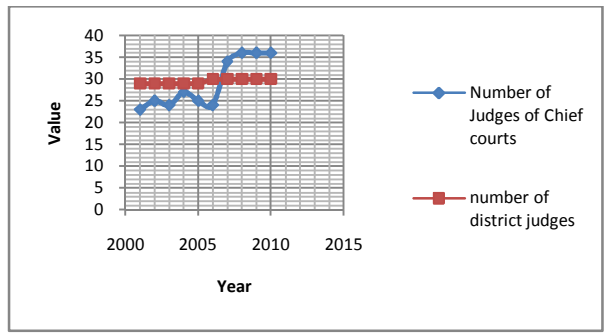
Source: Data reported by CPCB/SPCBs/PCCs/NEERI

Figure 5. Graph showing trend of change in environmental quality in Bihar (2001-2010)



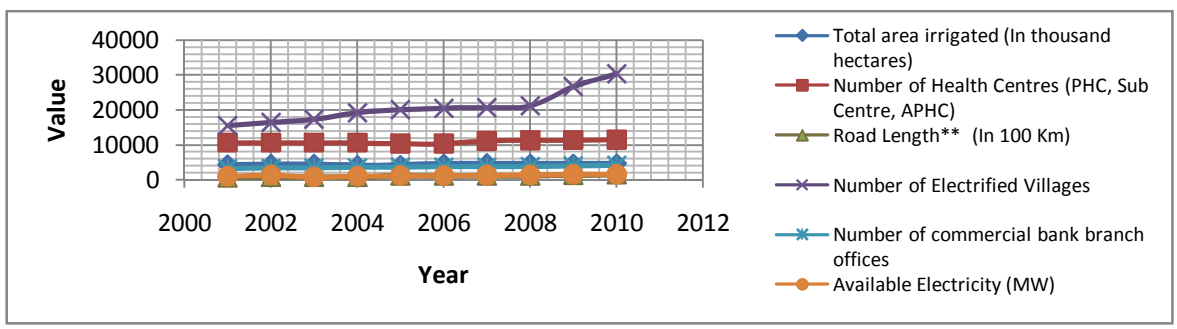
Source: Chief Conservator of Forests, Bihar.

Figure 6. Graph showing trend of ecological condition in Bihar (2001-2010)



Source: Registrar, Patna high court

Figure 7. Graph showing trend of change in institutional facility in Bihar (2001-2010)



Sources: Directorate of Economics and Statistics, Bihar, Patna; State Health Society; Basic Road Statistics published by TRW, M/o Road Transport & Highway; State Electricity Board, Bihar, Patna; and Statistical Tables Relating to Banks in India, 2010-11, RBI.

**** Excludes Roads Constructed under JRY and PMGSY**

Figure 8. Graph showing trend of change in the infrastructure availability in Bihar (2001-2010)

5. CONCLUSION

Over the years, the study area has been going through a phase of functional and structural changes. The study concludes that in a regional system, all the subsystems, physical, social, economic, ecological, environmental, infrastructure and institutional are interconnected and interdependent to each other which means, if one of the subsystems defunct or partially function or takes lead role in its function over a period of time its effects would be reflected in the whole regional system. The result of the analysis can be used for further research via system approach to get optimum plan policy for the region's development.

REFERENCES

[1] Simon, J.L. and Burstein, P. (1988) *Basic research methods in social science.* 3rd ed. New York: Random House USA Inc.

[2] Belsley, D.A., Kuh, E. and Welsch, R.E. (1980) *Regression Diagnostics: Identifying Influential Data and Sources of Collinearity*. Wiley, New York.

[3] Yule, G.U. (1897) *On the Theory of Correlation*. Journal of the Royal Statistical Society: Blackwell Publishing 60 (4): 812–54.

[4] John, O.R., Sastry, G.P. and David, A.D. (1998) *Applied Regression Analysis: A Research Tool.* 2nd ed. New York: Springer-Verlag New York, Inc.

[5] Galton, F. (1989) *Kinship and Correlation (reprinted 1989).* Statistical Science, Institute of Mathematical Statistics 4 (2): 80–86.