Classification in Geography

Order out of Chaos

Spatial Order / Regularity → The Spatial Pattern of Elements over the Earth Surface:

This can be defined, identified and analysed with a scientific understanding of geographical knowledge.

In space – time frame, it can be measured, monitored, mapped and modelled.

It is this that forms the philosophical foundation of the discipline of 'Geography'.

Naturally, it is the Geographer who discovers this Spatial Order.
Spatial Order → Order-forming Processes → Order-forming Factors for scientific geographical explanation.

Areal Differentiation

Classification

= fundamental technique for ordering and organizing huge volume of real world data.
= grouping data into classes makes it more easy to comprehend and manipulate.

therefore, it is one of the basic tools of a geographer who deals with robust set of spatial data for identifying 'spaces' based on
resemblance (i.e., certain affinities),
homologous character (i.e., a common origin),
common line of descent (i.e., similar evolutionary history).

the procedure is such that —
no 'space' is left outside the classification (i.e., the property of exhaustiveness)
no 'space' is assigned simultaneously to more than one class (i.e., the property of mutual exclusiveness)

Classification

Attribute Classification (i.e., grouping of attributes only)
Spatial Classification (i.e., grouping of spaces on earth, commonly known as 'regionalisation')

regionalisation

to

identify
Backward Areas/Deficit Areas/Negative Areas
through
'Spatial Mapping'
for
'inputs' of 'development'
to eliminate 'regional disparity'

Classification/ Regionalisation

Natural or 'general' Classification —

based on 'apparent'

similarity

common origin

common evolution

Artificial or 'statistical' Classification —

Univariate (1 variable case)

class interval = arbitrary, or statistical

Bivariate (2 variable case)

four classes (groupings using either mean or median)

Multivariate (more than 2 variables case)

using PC Scores, Factor Scores, Similarity Coefficients, Discriminant Functions

Data Acquisition

Physical Database

field data, lab data (using RS / GIS)

Socio-economic Database

GDM using attribute Data

Mapping

Thematic Data Layers (physical)
Thematic Data Layers (social)
Thematic Data Layers (economic)

Data Integration : using RS/GIS adopting appropriate 'project design' and 'management' with proper 'process models'.

Univariate Classification (1 variable situation)

arbitrary classes

based on range and number of class

statistical

mean/standard deviation (mean ± n.σ)
standard scores (0 ± n.1z)
etc.

spatial index

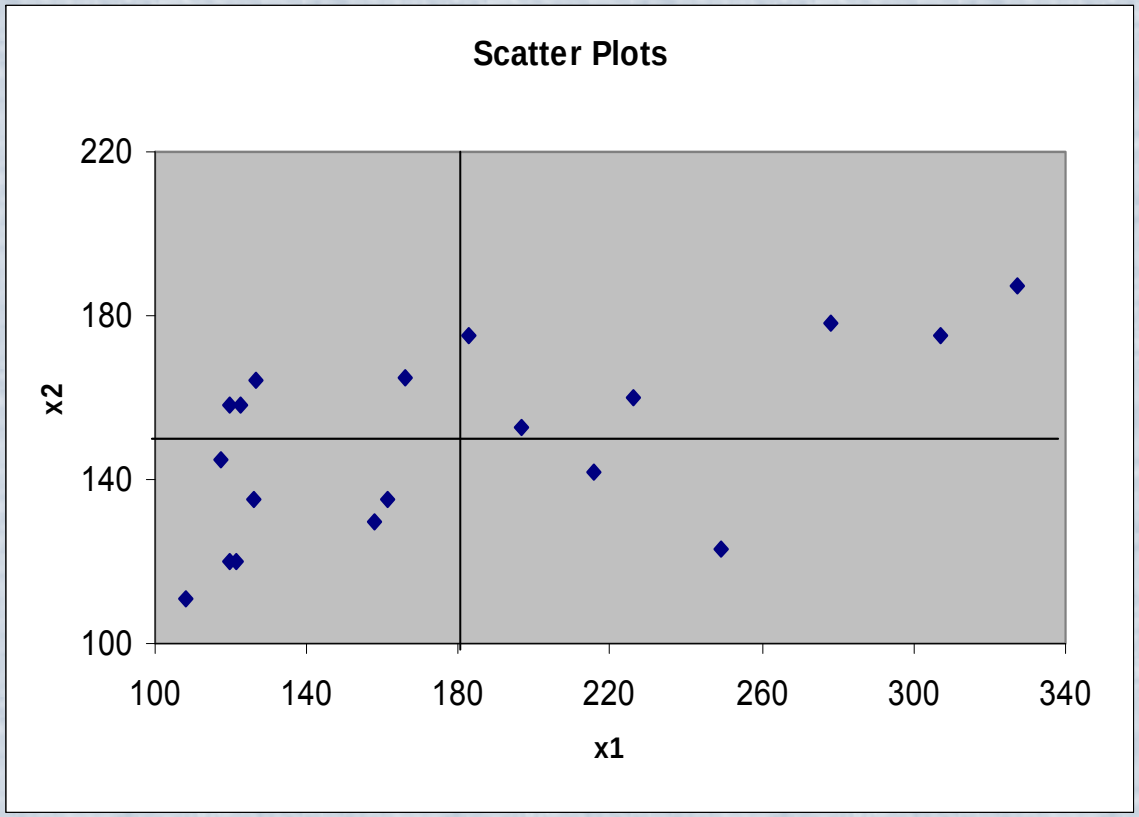
linkage (groupings based on CM or CGA)
location quotient (class interval = 1)
inequality (Gini coefficient with class interval = 0.20/ 0.25 /0.30)

Bivariate Classification (2 variable situation)

Identification of Groups from—

Scatter Plots of $x_1 - x_2$ with lines of means of x_1 and x_2

Scatter Plots of $x_1 - x_2$ with lines of medians of x_1 and x_2



Group – 1 : x_1 high, x_2 low
Group – 2 : x_1 high, x_2 high
Group – 3 : x_1 low, x_2 low
Group – 4 : x_1 low, x_2 high

No. of Groups = 4
Colour Patch Mapping

Combinatorial Method: Map Algebra (Nominal Data)

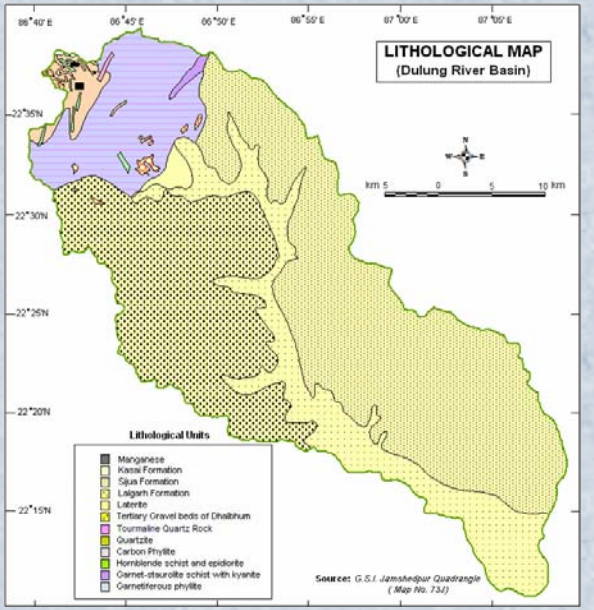
Let there be two sets defining the attributes of two variables, e.g., lithology (L) and geomorphology (G) as :

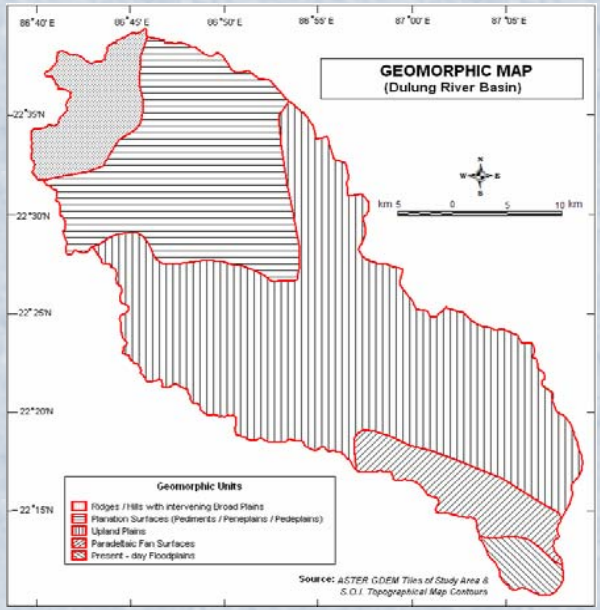
$L_i = \{L_1, L_2, L_3, \ldots L_n\}$, and

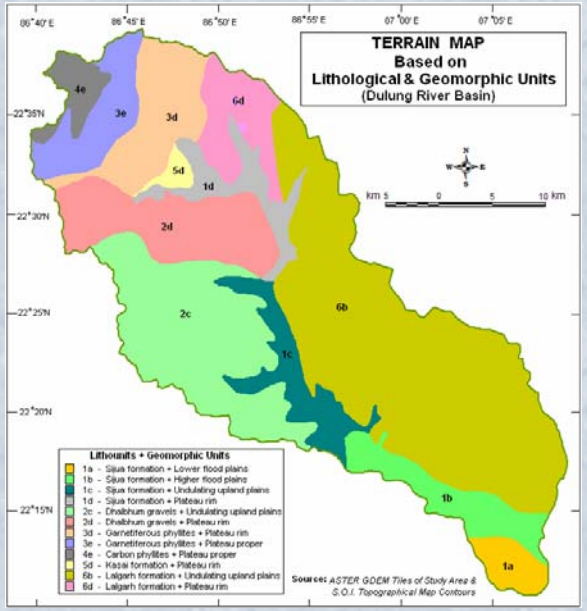
$G_j = \{G_1, G_2, G_3, \ldots G_k\}$.

Hence, the terrain classes are defined by the elements derived from the union of L_i and G_j as — $T_{ij} = \{L_i.G_j\}$

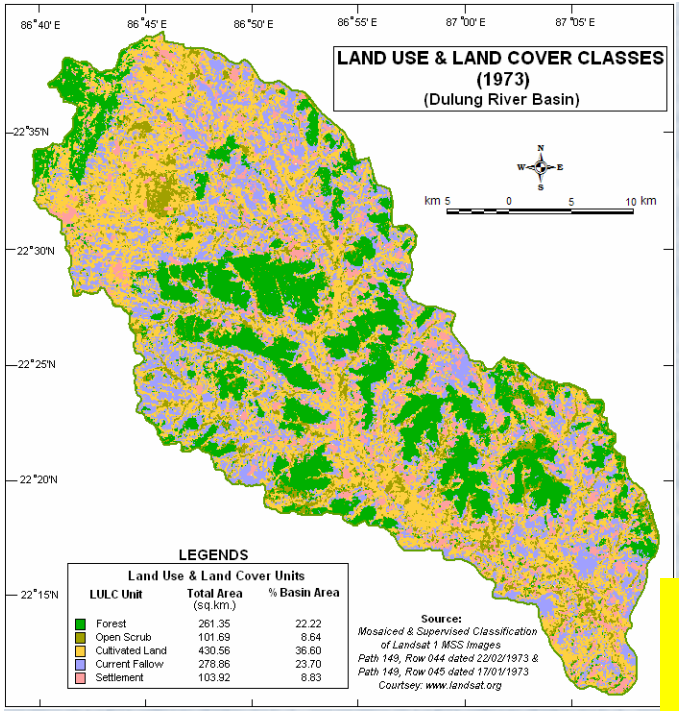
where i = 1, 2, 3, ... n and j = 1, 2, 3, ... k

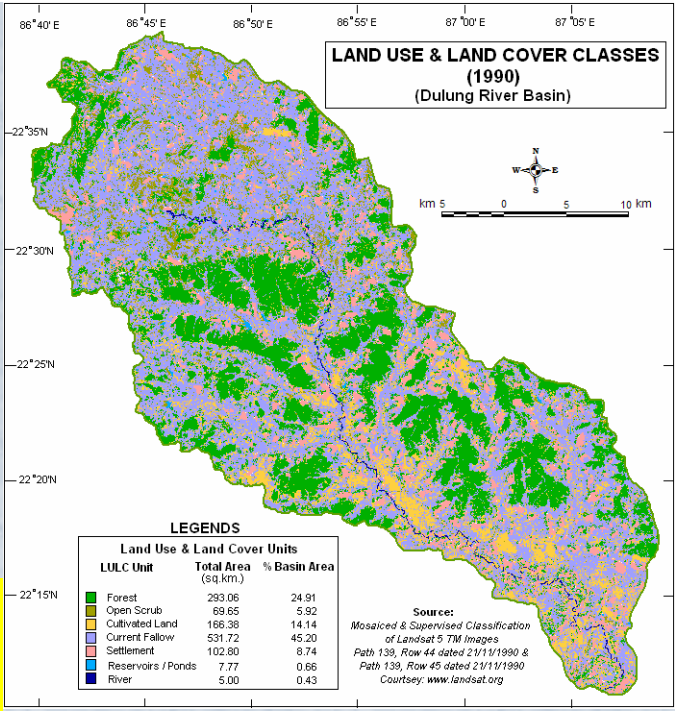


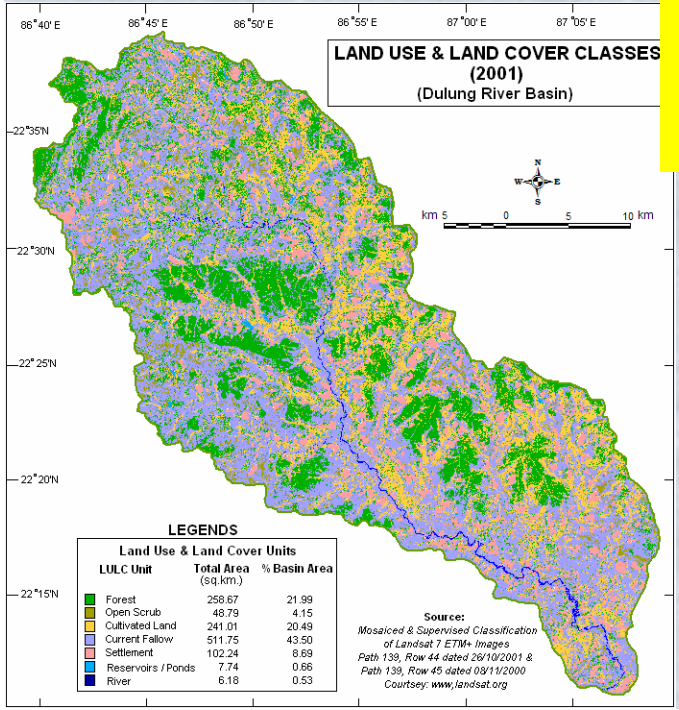


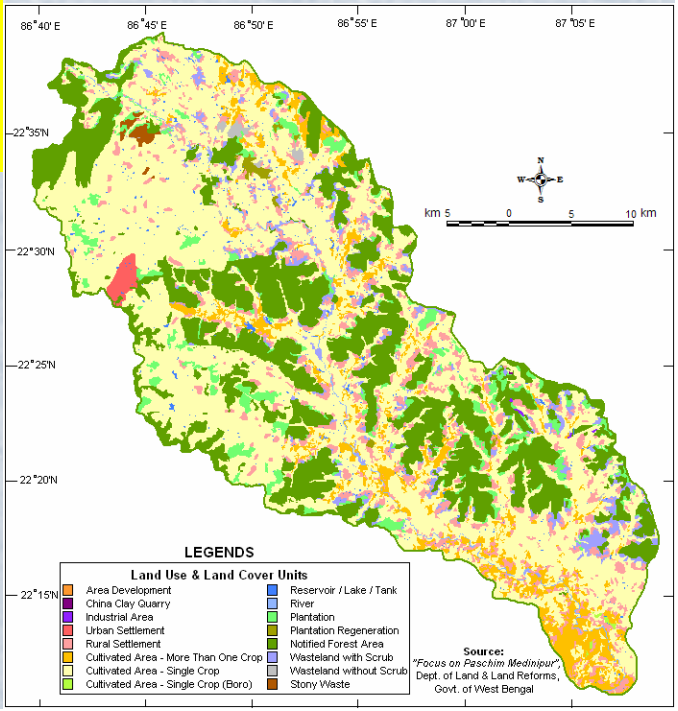


Classification (Raster Data) using Statistical Classifier









Multivariate Classification (Multi Variable Situation)

Virtually geographical events / objects are inherently multivariate, and hence suited to multivariate techniques. These allow the researcher to consider changes in several properties simultaneously in order to explore the properties of dependence, interdependence and **classification**. Softwares are now easily available: SPSS, Statistica, etc

PC Scores / Factor Scores:

to find the directions of maximum variance in the data, to use these **to ordinate data** in 1, 2, 3 or 4 dimensions and to interpret them as factors influencing the data.

Discriminant Functions:

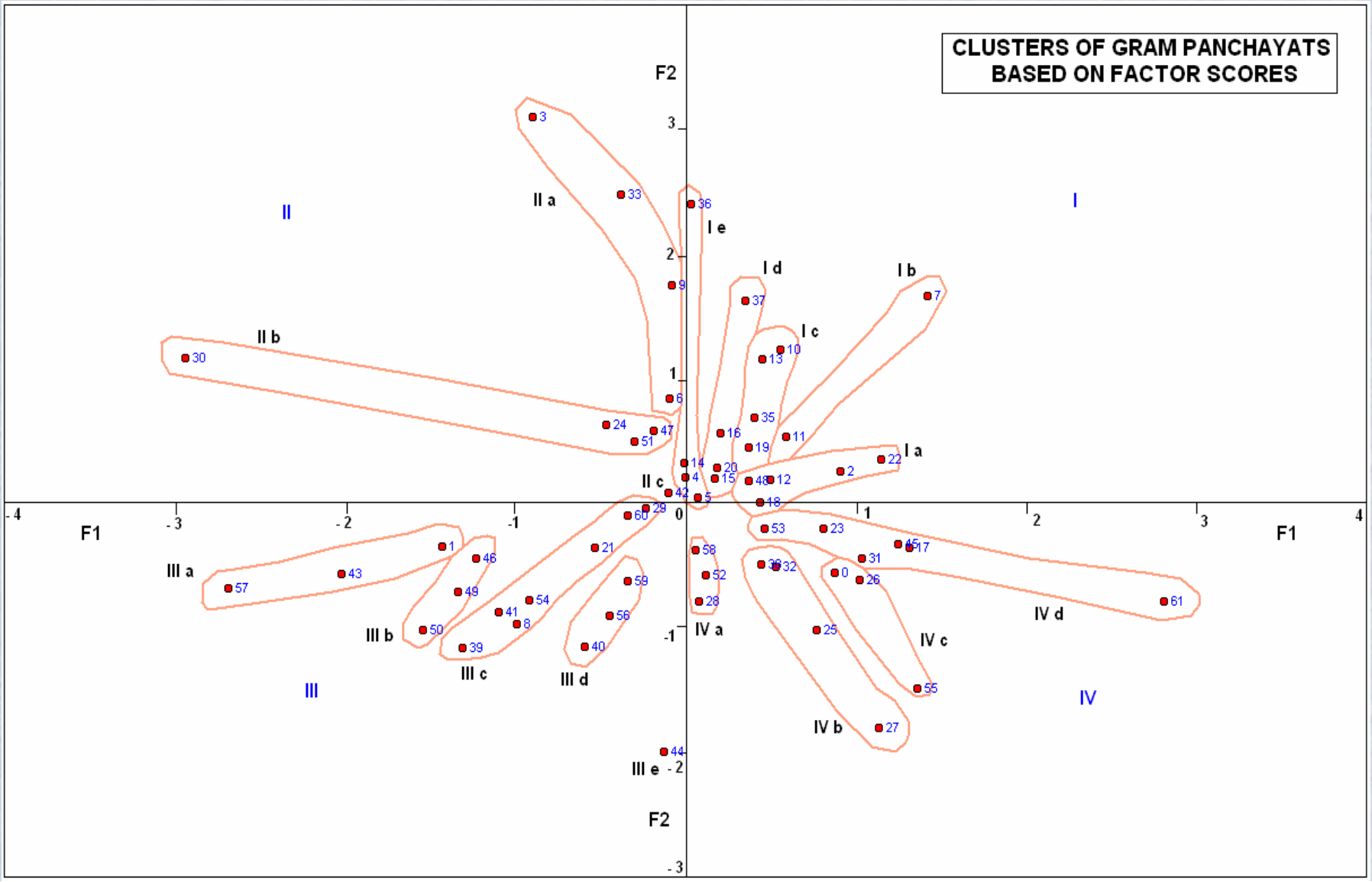
to find the equation of a line that best separates two or more user-defined **(a priori) sub-groups** within the dataset and to allocate new data to one or other of the a priori groups on this basis.

Similarity Coefficients (CA):

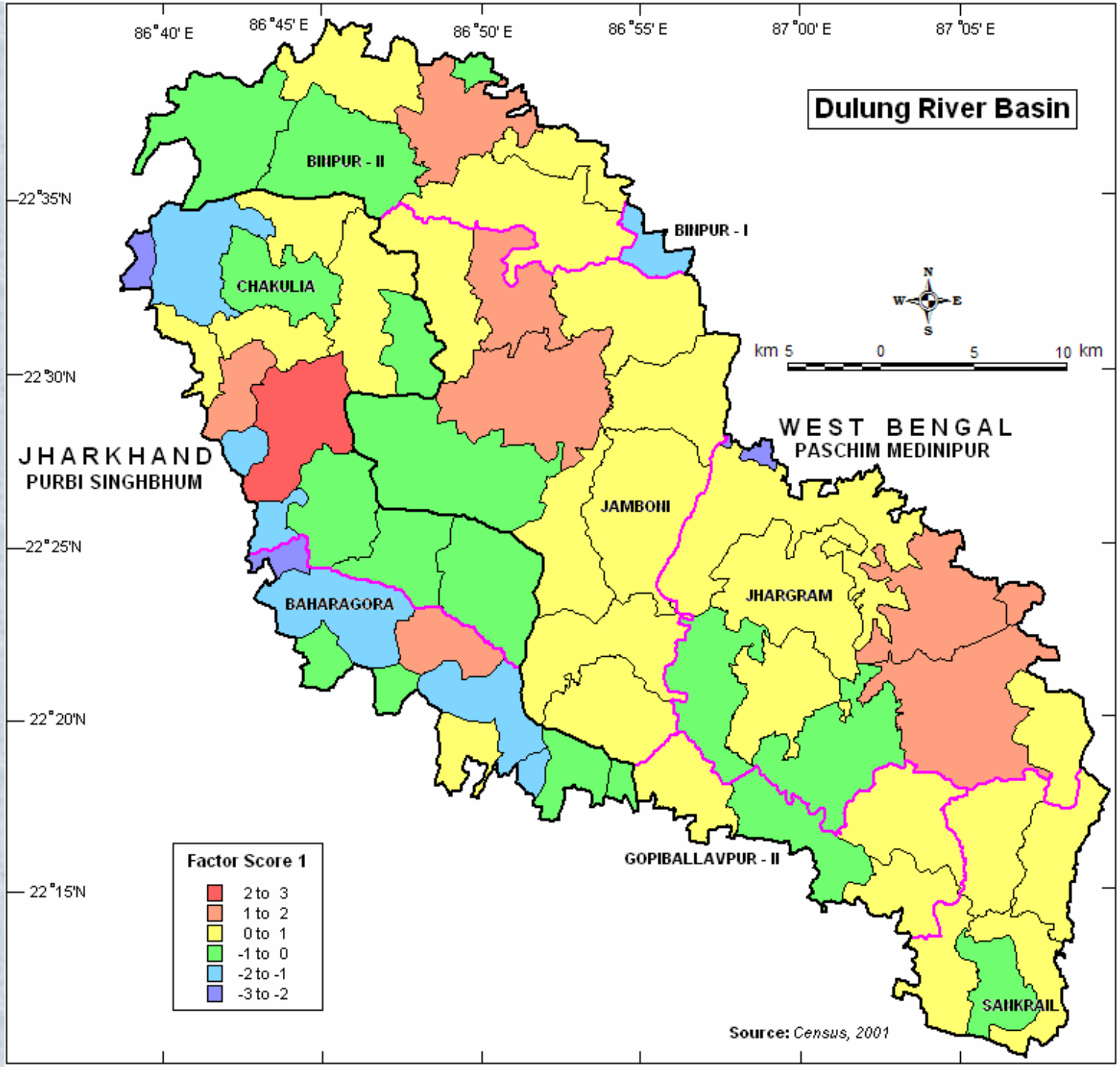
to find the magnitude of similarity between pairs of objects or observations and to use this to produce an **empirical classification**.

Scatter Plots of Factor Score – 1 and 2

Linear Clusters can be identified, which are regarded as Groups in the Classification Scheme.

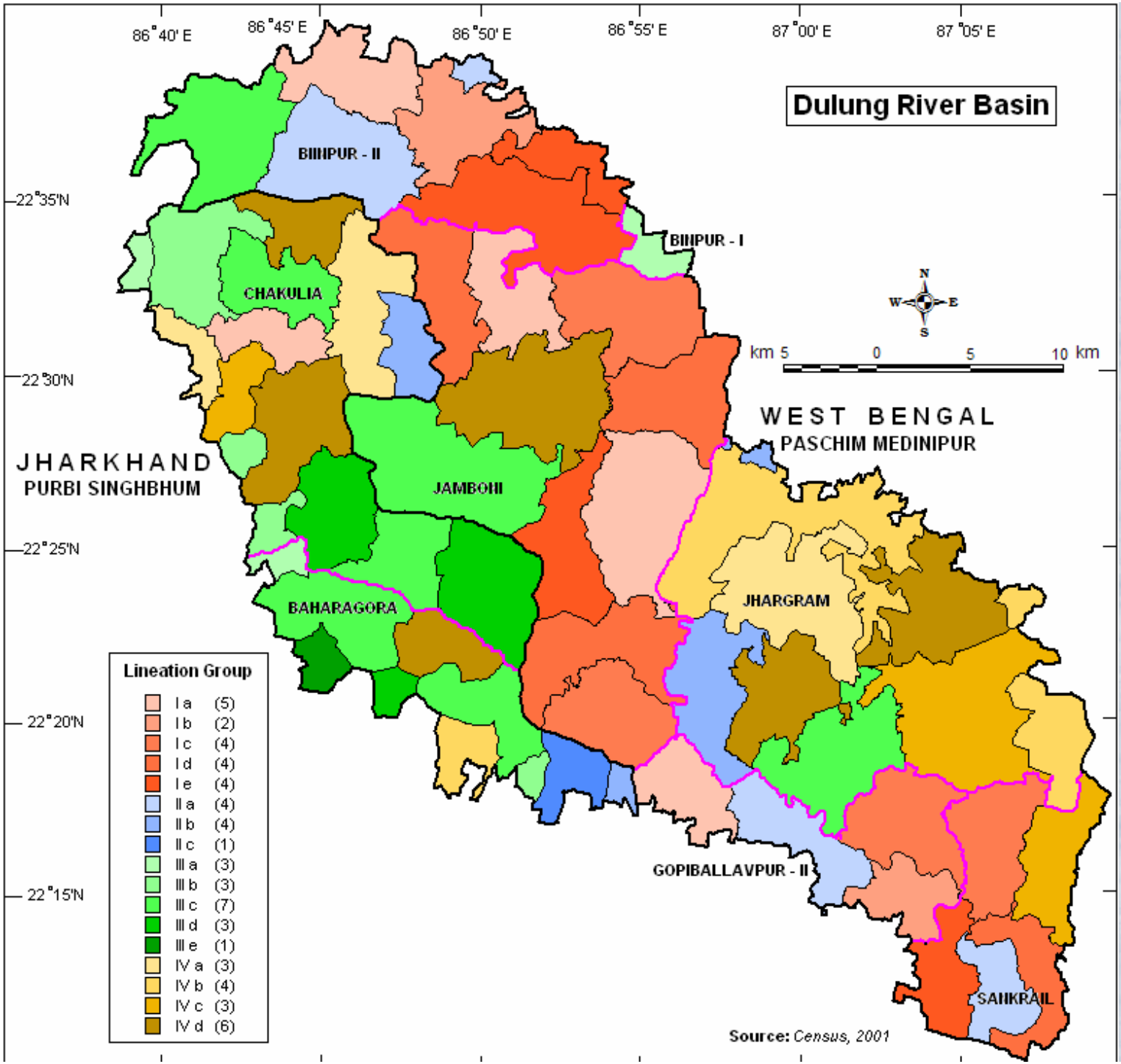


Multivariate Classification

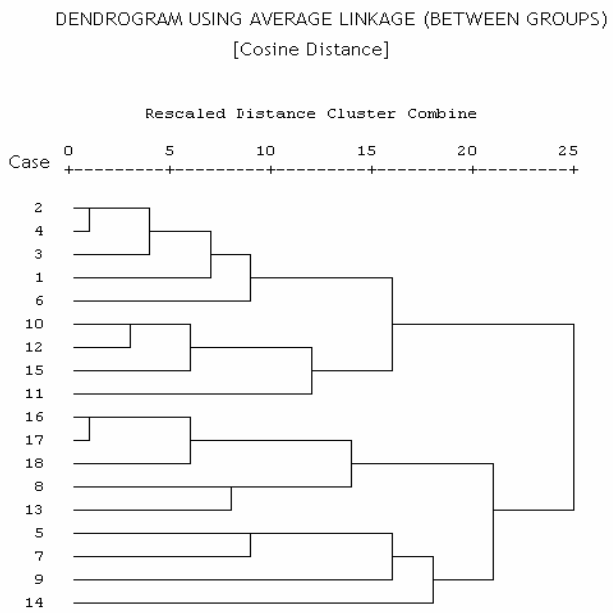


Factor Score – 1 may also form the basis of Numerical Classification of the GPs in terms of the selected 21 variables.

Range of Factor Score – 1	No. of Gram Panchayats	Gram Panchayat ID	Remarks
>2	1	61	Highly Developed
1 to 2	8	7, 55, 17, 45, 22, 27, 31, 26	Fairly Developed
0 to 1	26	2, 34, 23, 25, 11, 10, 32, 12, 13, 53, 38, 18, 35, 48, 19, 37, 16, 20, 15, 52, 28, 5, 58, 36, 4, 14	Developed
-1 to 0	17	6, 42, 44, 47, 29, 51, 59, 60, 33, 56, 24, 21, 40, 3, 54, 8	Backward
-2 to -1	6	41, 46, 39, 49, 1, 50	Fairly Backward
< -2	3	43, 57, 30	Very Backward



Clusters / Classes derived from Scatter Plots



Dendrogram Data

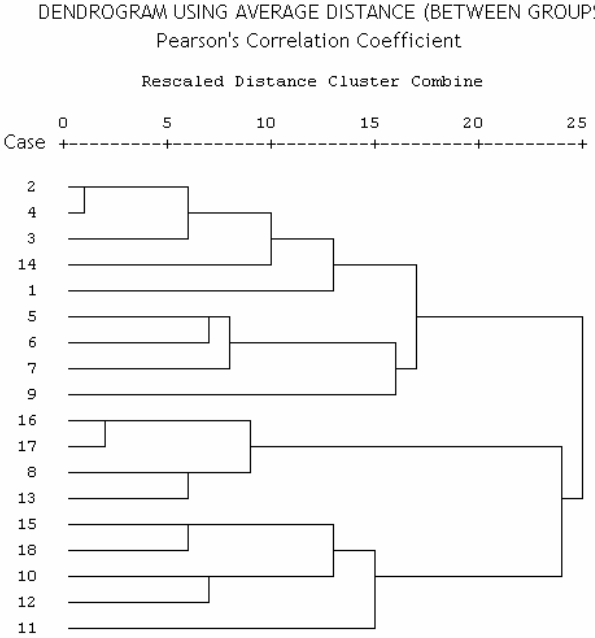
from Cluster Analysis

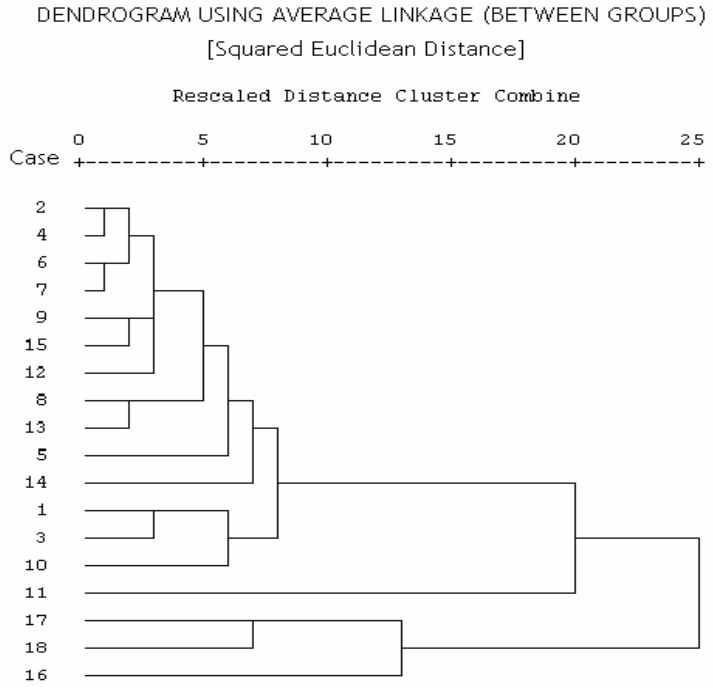
may be used to prepare

Thematic Maps

showing

Spatial Classes / Regions





Classification (Spatial Data)

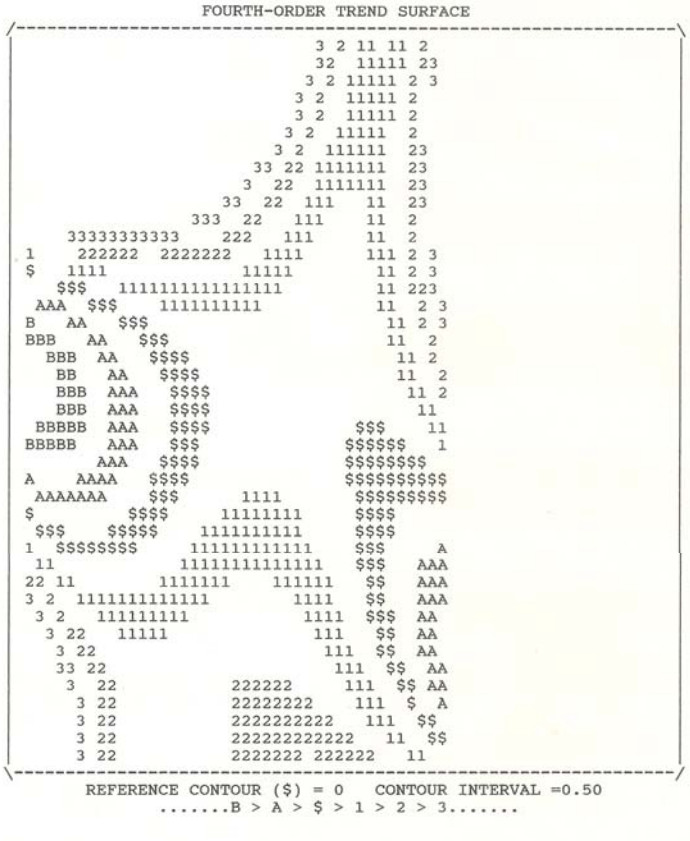
Trend Surface Analysis: z = f (x, y)

(Polynomial Surface of order 2 – 6 can be fitted)

Residual Maps of ($z - z_c$) are used to identify the 'regions'

Backward Regions :: negative residuals

Developed Regions :: positive residuals

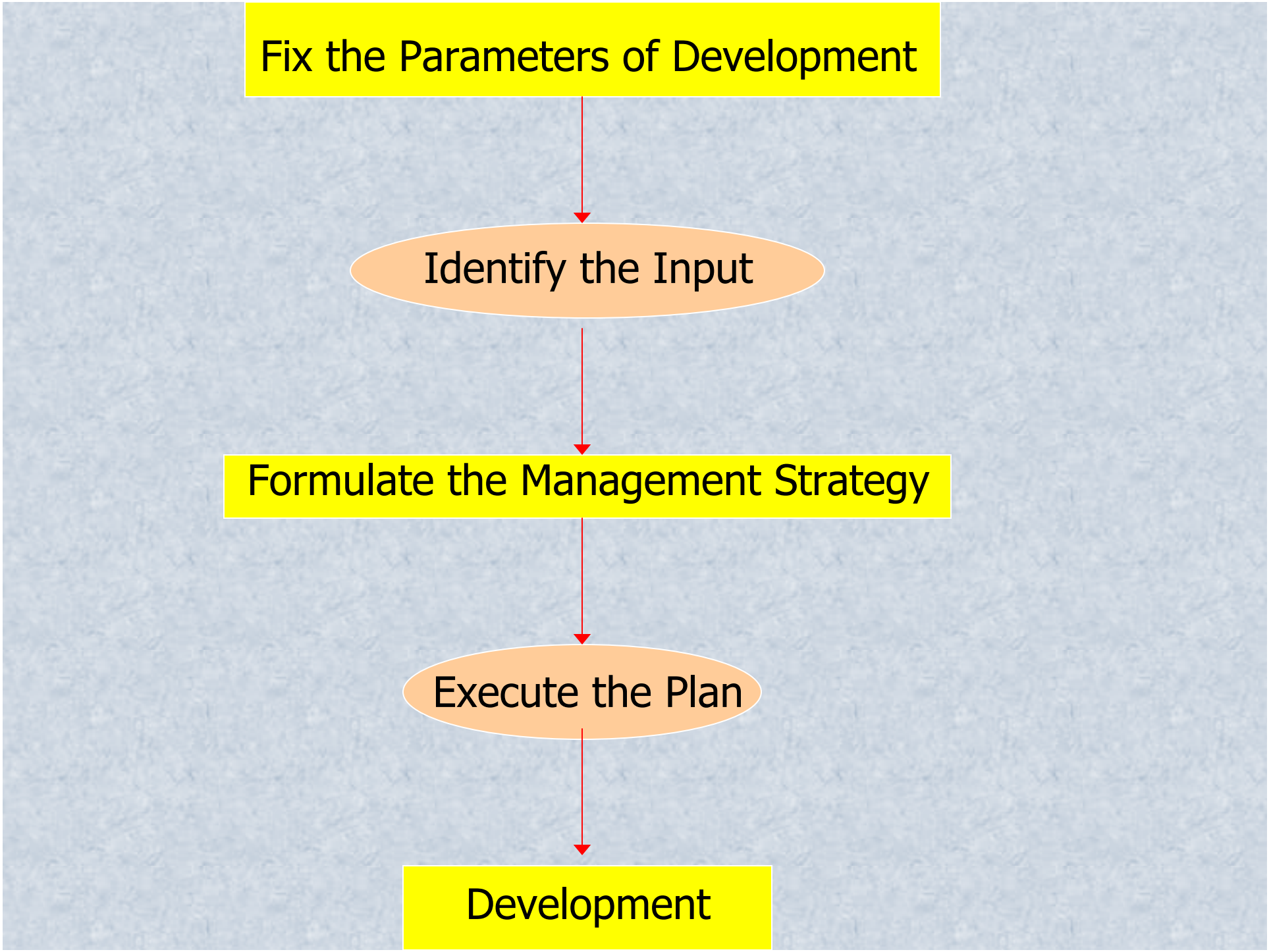


Trend surface of 4th Order : Equation

$$Z_i = -48.524 + 50.936\,u + 38.736\,v - 6.715\,u^2 - 52.567\,uv - 3.168\,v^2 - 6.031\,u^3 + 17.739\,u^2 v + 8.987\,u v^2 - 1.587\,v^3 + 1.377\,u^4 - 0.852\,u^3 v - 3.104\,u^2 v^2 + 0.346\,u v^3 + 0.103\,v^4$$

Table - II : General ANOVA for TSA : Fourth Degree Surface

Sources of Variation	Sum of Squares	df	Mean Squares	F - Test	Goodness of Fit	r
Polynomial Regression	20.60	14	1.47			
Deviation from Polynomial	0.40	7	0.06	24.500	0.98	0.99
Total Variation	21.00	21				


Fix the Parameters of Development
Identify the Input
Formulate the Management Strategy
Execute the Plan
Development

Thank You

Prof Ashis Sarkar
Presidency University, Kolkata
profdrashis@gmail.com