Stratified One-stage Cluster Sampling using GIS for Surveys

Mena Paghadar
Manager
NDDB, Anand
menap@nddb.coop

Dr. Jignesh Shah
Sr. Manager
NDDB, Anand
jgshah@nddb.coop

Rural Sample Surveys

- Rural sample surveys are important in dairy sector, which provide essential inputs for various business/ operational planning
- Conventionally, multi-stage stratified random sampling methodology is used for the conduct of such surveys
- However, this methodology does not take into account the nature and shape of the geography and therefore, proper spread of the sample cannot be ensured

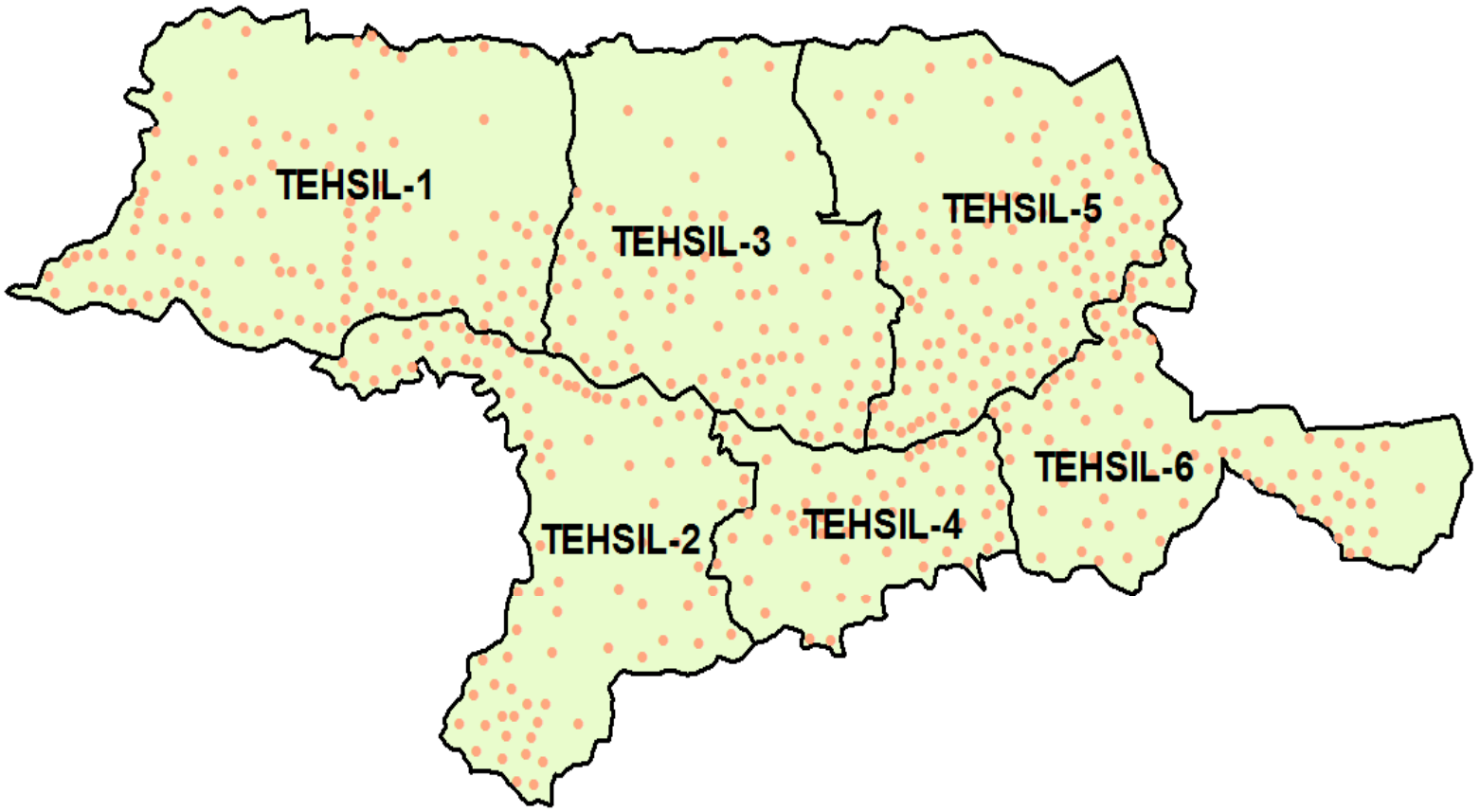
Sampling methodology followed using GIS

- In first stage, the tehsil is divided into quadrant on the basis of area sampling
- In second stage, the villages in a quadrant are divided into two categories based on the village size (i.e., households *in a village)*
  - Villages having village size above average
  - Villages having village size below average
- 2 villages are selected at random from each of the two categories formed as above and all the households in sample villages will be surveyed

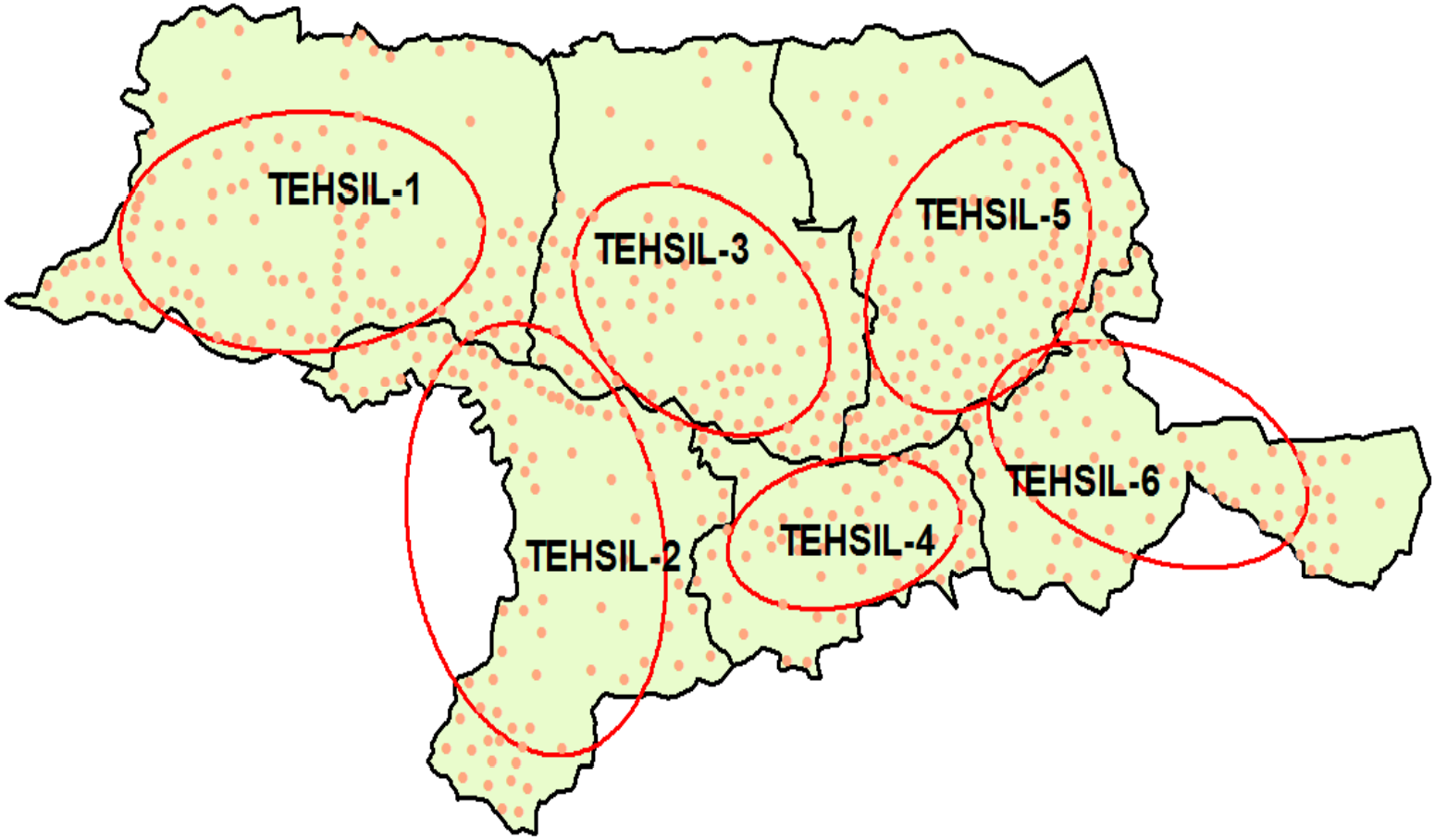
Therefore, 16 villages from each tehsil are selected for the survey

Process Flow

Size, shape and distribution of villages in tehsil

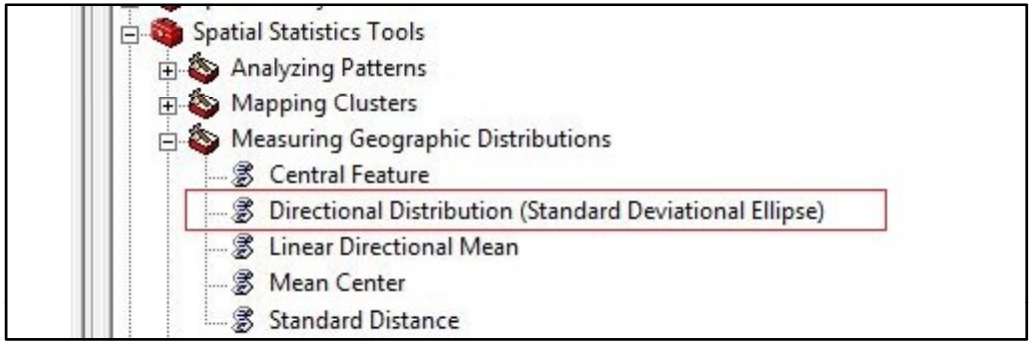


Step 1 : Preparation of standard deviational ellipse



Methodology for developing "Standard Deviational Ellipse"

Use of ESRI's Spatial Statistics tool



Input parameters for directional distribution

Parameter	Explanation	Data type	Actual Input
<Input_Feature_Class >	A feature class containing a distribution of features for which the standard deviational ellipse will be calculated.	Feature Layer	Village locations (as points)
<Output_Ellipse_Feature_Class>	A polygon feature class that will contain the output ellipse feature.	Feature Class	
<1 Standard Deviation | 2 Standard Deviations | 3 Standard Deviations >	The size of output ellipses in standard deviations. The default ellipse size is 1; valid choices are 1, 2, or 3 standard deviations.	String	1 Standard Deviation
{Weight_Field}	The numeric field used to weight locations according to their relative importance.	Field	Size of Households in a village
{Case_Field}	Field used to group features for separate directional distribution calculations. The case field can be of numeric, date, or string type.	Field	Tehsil of villages

The output parameters of resultant directional ellipse

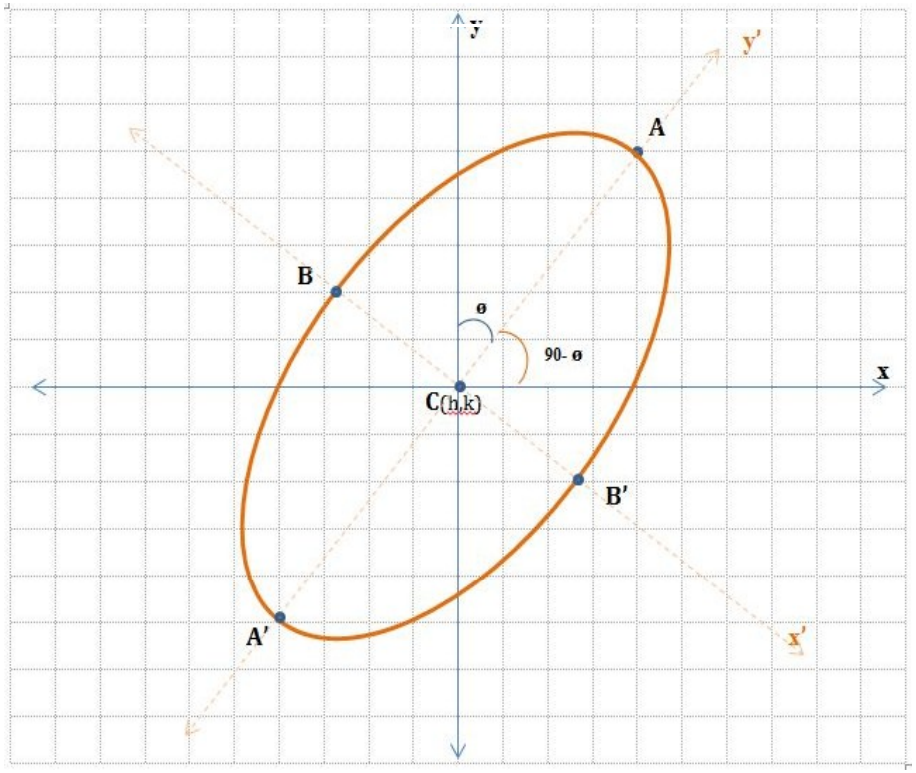
	FID	Shape	TEHSIL	CenterX	CenterY	XStdDist	YStdDist	Rotation
	5	Polygon	TEHSIL-1	531454.269	2348707.71	10201.0969	19171.8682	88.428766
	8	Polygon	TEHSIL-2	556100.721	2322444.92	18325.9663	13213.0564	165.132701
	14	Polygon	TEHSIL-3	573512.998	2342026.71	14322.8983	9580.03338	117.056591
	2	Polygon	TEHSIL-4	588483.702	2322944.51	6208.6886	12347.8852	80.623216
	13	Polygon	TEHSIL-5	602589.851	2345539.87	10096.0509	13786.5084	40.893259
	10	Polygon	TEHSIL-6	620438.134	2329704.23	17277.3455	8780.53128	106.295433

a) CenterX & CenterY : **X & Y coordinates of center of ellipse; C(h,k)**

b) XStdDist & YStdDist : Length of semi-major semi-/minor axis (i.e., length between the center and vertices of the ellipse); **CA**= **C⟷A** and **CB**= **C⟷B**

c) Rotation : Rotation angle of ellipse; **Ø**

Illustrative directional ellipse



Step 2 : Ascertaining Vertices & Co-vertices

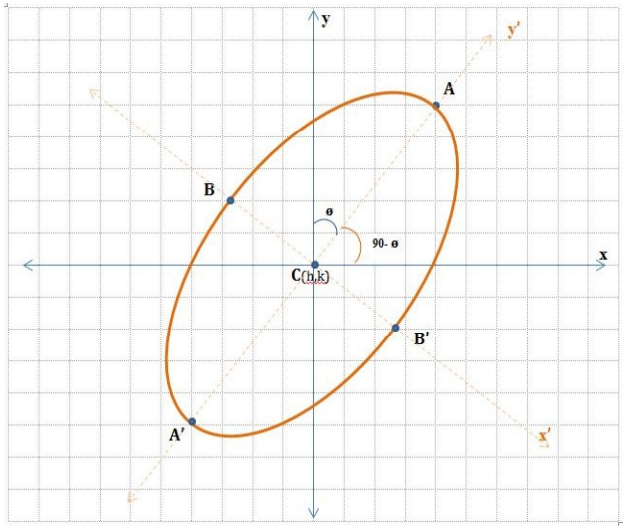
The co-ordinates for vertices (A & A') and co-vertices (B & B') are calculated as –

If ø < 90 and Y > X i.e., Y is the major axis

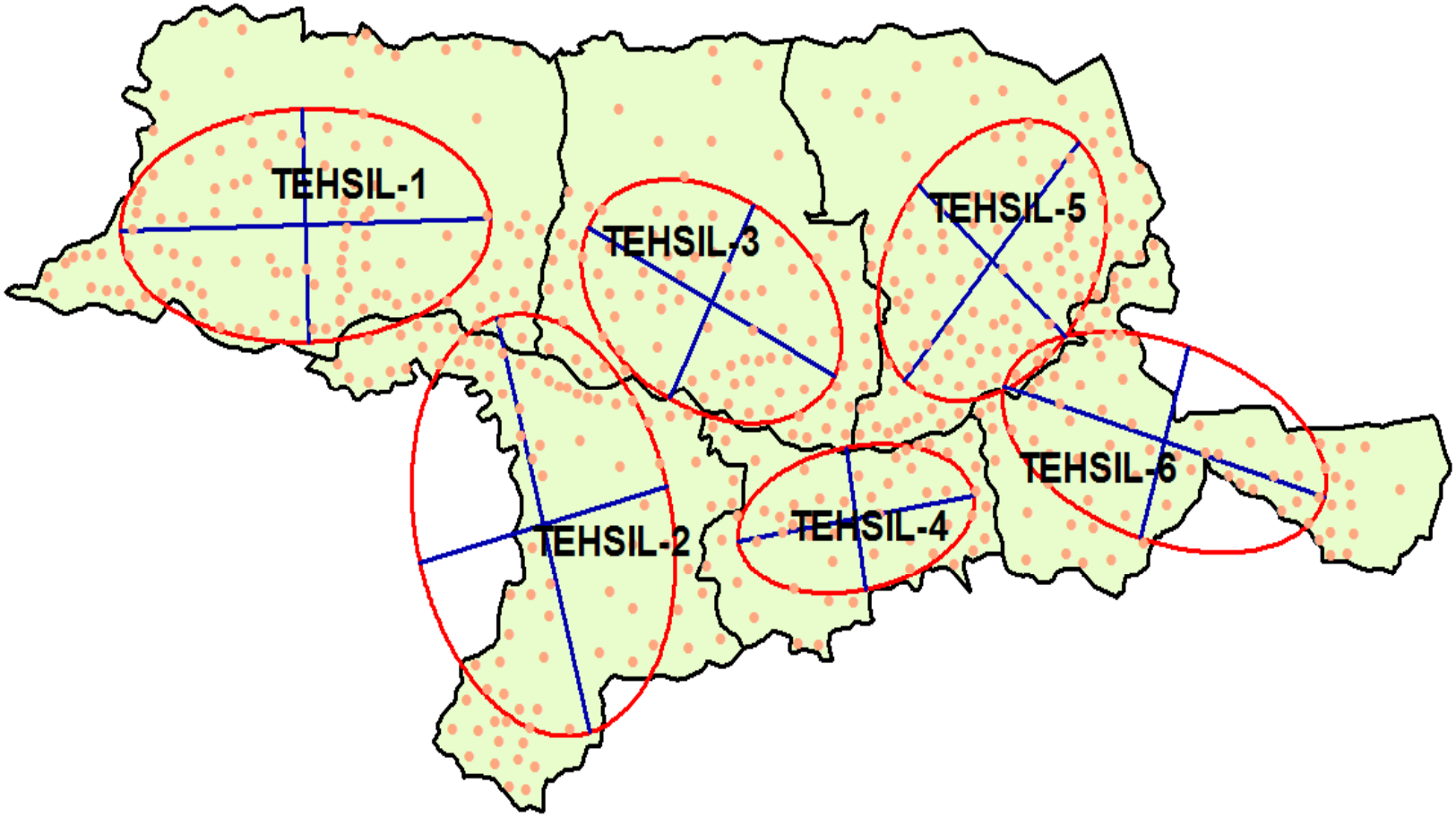
Ax = h + CB Cos(90-ø)	and	Ay = k + CB Sin(90-ø)
A'x = h - CB Cos(90-ø)	and	Ay = k - CB Sin(90-ø)
Bx = h + CA Cos(180-ø)	and	By = k + CA Sin(180-ø)
B'x = h - CA Cos(180-ø)	and	B'y = k - CA Sin(180-ø)

If ø > 90 and X > Y i.e., X is the major axis

Ax = h + CA Cos(90-ø)	and	Ay = k + CA Sin(90-ø)
A'x = h - CA Cos(90-ø)	and	Ay = k - CA Sin(90-ø)
Bx = h + CB Cos(180-ø)	and	By = k + CB Sin(180-ø)
B'x = h - CB Cos(180-ø)	and	B'y = k - CB Sin(180-ø)

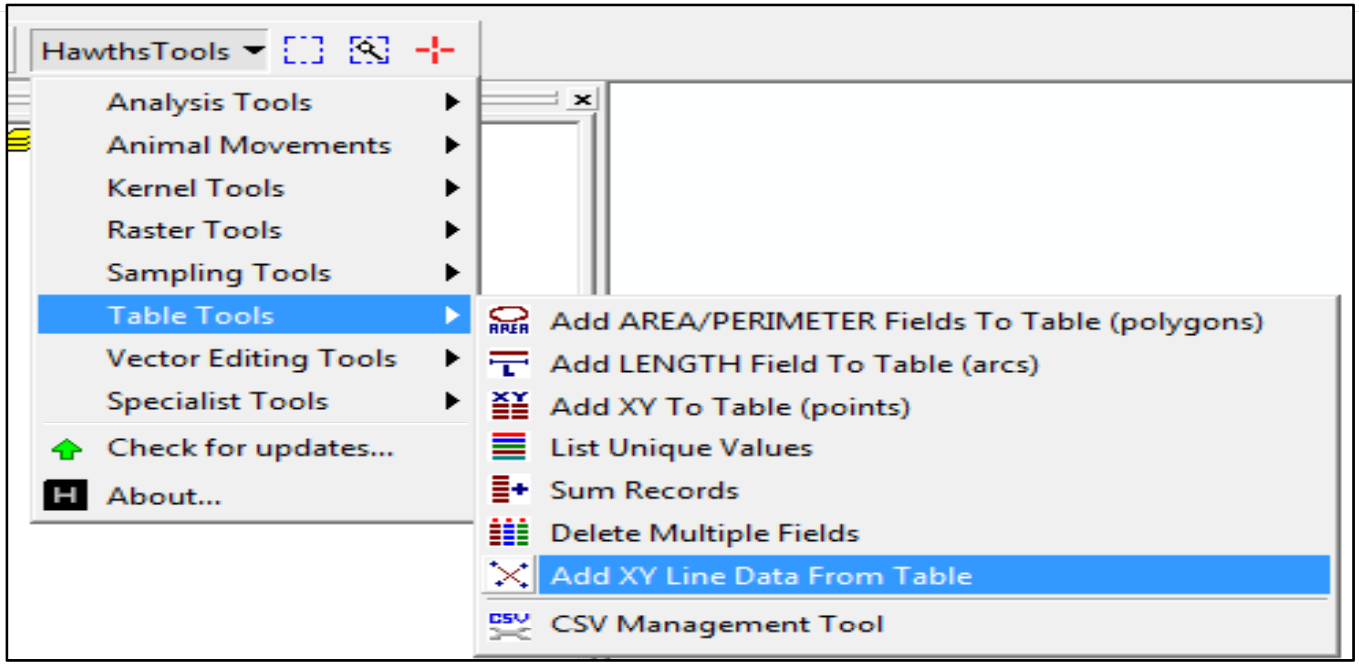


Joining vertices & co-vertices of ellipse

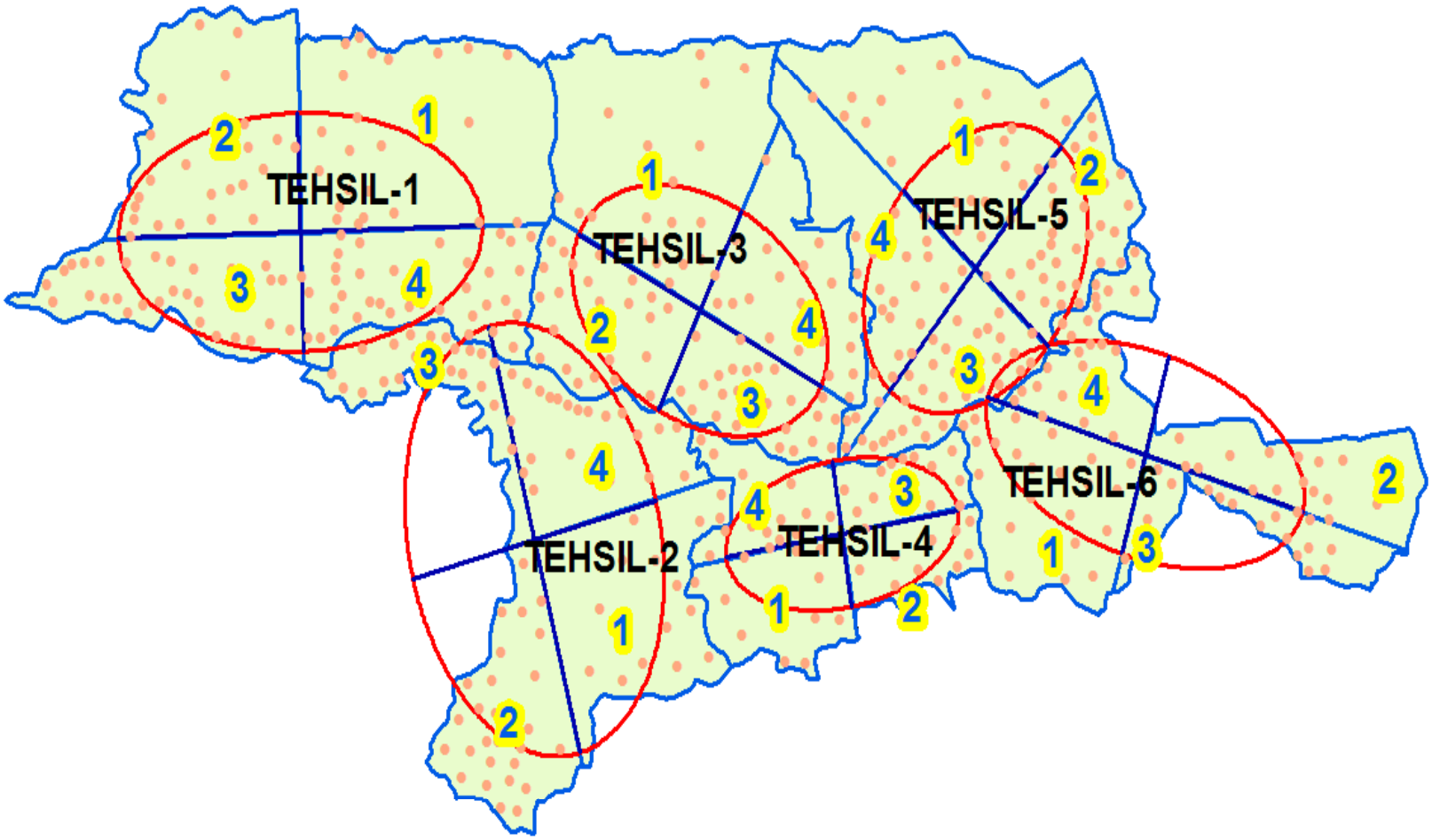


Use of Hawth's tool

- The major and minor axes are drawn by joining the vertices (A & A') and co-vertices (B & B') using the functionality 'Add XY Line Data From Table' of **Hawth's Tool** provided under 'Table Tools'

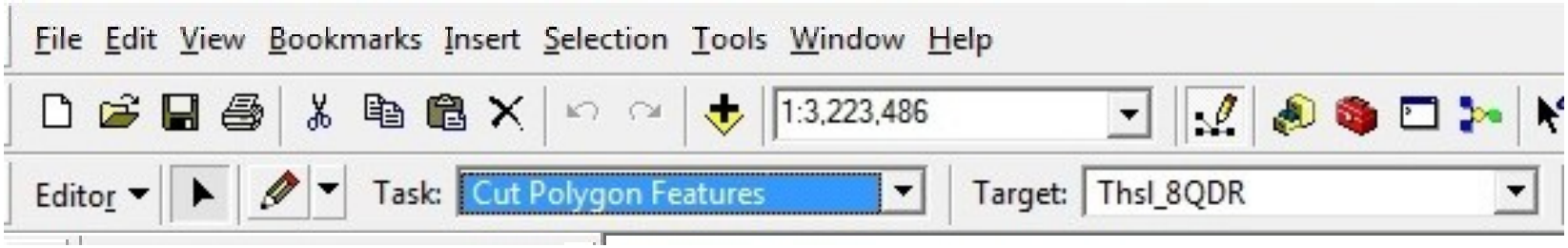


Step 3 : Cutting the tehsil polygon by taking the major/minor axes of ellipse as reference

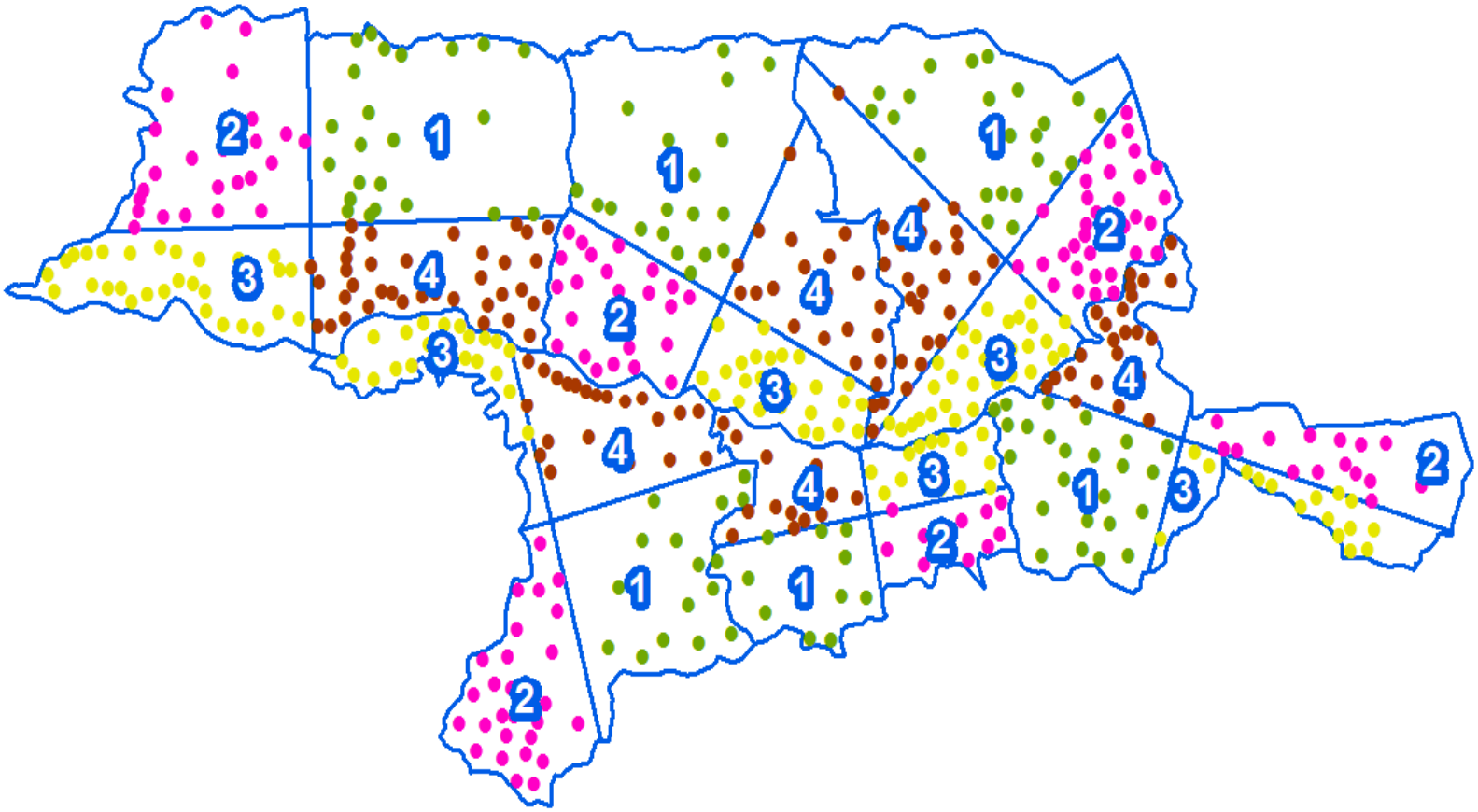


Formation of quadrant

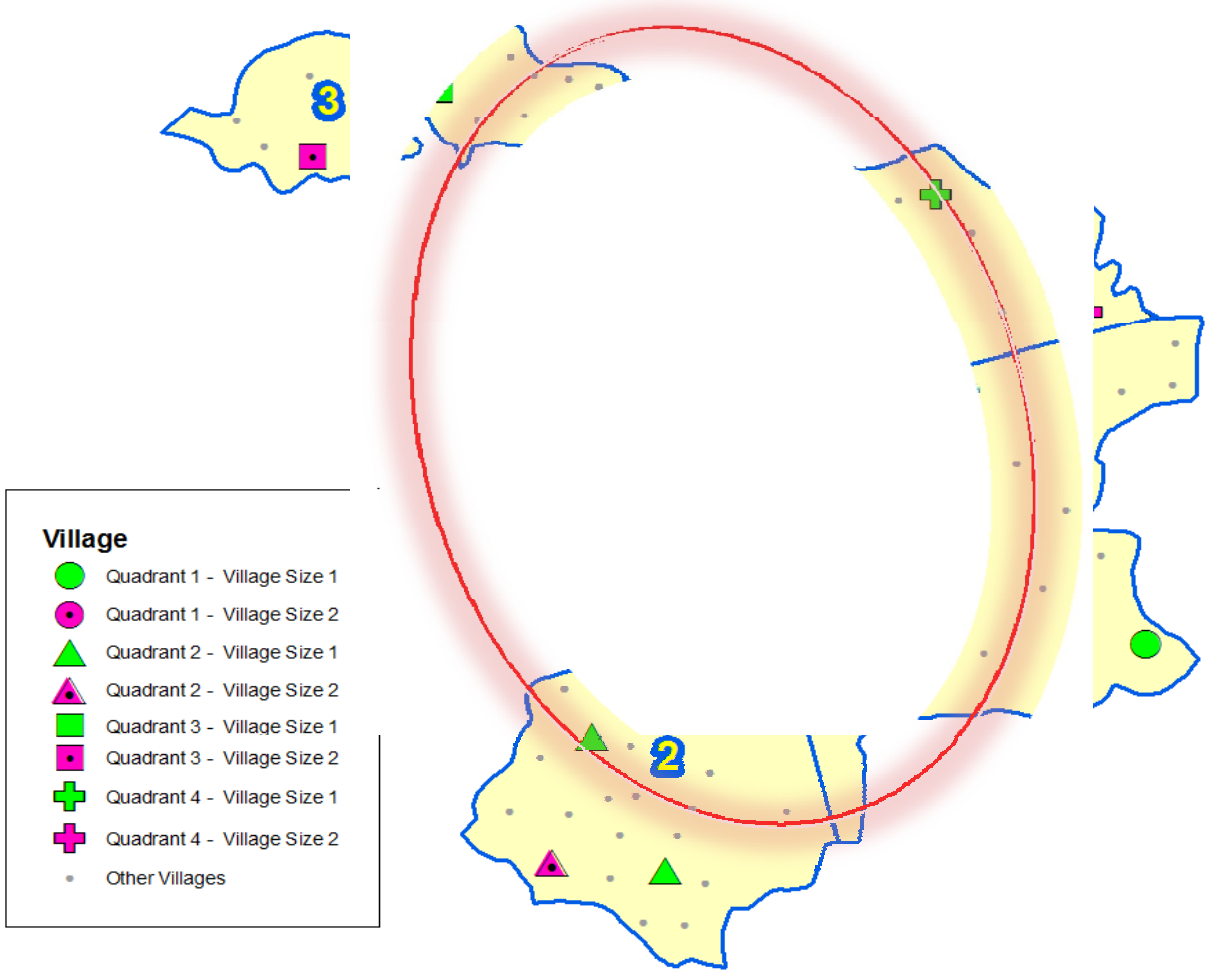
- Taking these lines drawn as reference, each polygon was cut manually by using ESRI's "Cut Polygon features" task under the 'Modify Tasks' of Editor toolbar of ArcGIS desktop.



Assigning villages to respective quadrant



Finally, the resultant outcome is …



Summing up …

- The results of the sample survey undertaken following above sampling technique provides statistically robust estimates at the tehsil level

- The percentage standard error of the estimate is in the range of ± 10-20 percent

Thank you