Spatial Calculation of Locus Allele Frequencies

Using ArcView 3.2

This instruction set applies to calculating allele frequency from point data of DNA analysis results within ArcView 3.2. To calculate the frequency of an allele across space a moving window is applied in a neighborhood analysis to sum allele presence, and map calculator is used to take the log of the ratio of the sum divided by twice the number of individuals according to the following equation:

$$f = \text{Log} \left[(\Sigma A_1 + \Sigma A_2)/2N_i \right]$$

where f = allele frequency
A_1 = presence of one allele from one parent (1)
A_2 = presence of two alleles from both parents (2)
N_i = number of individuals

Using the Spatial Analyst extension the resulting output is a raster grid of cells in which each cell or pixel contains the frequency value. If the required output is to be in tabular form containing cell centroid coordinates and frequency values then the Grid Pig and Add XY extensions can be used. The Grid Machine extension fills in the entire extent.

The spatial calculation steps are outlined as follows:
- Add the required extensions and data to an ArcView project
- Specify the project properties
- Query the data to be included in the analysis (i.e. omit missing data)
- Perform a neighborhood analysis to sum allele presence
- Calculate the log ratio using map calculator
- Create point shapefile of cell centroid and add coordinates

The point data is assumed to be in tabular format such as a dBase (.dbf) file. The attribute field of interest referred to is the "Alleles" field which contains values of 0=no alleles found, 1= allele from one parent, and 2=allele from both parents, as well as empty records that contain no data.

Add the required extensions and data to an ArcView project

1. Open ArcView with a new view but do not add any data yet
   Choose FILE → EXTENSIONS…
2. Place a check beside the following extensions:
   - Add XY
   - Grid Machine
   - Grid PIG Tools
   - Spatial Analyst
3. Add any background themes (e.g. roads or natural features)
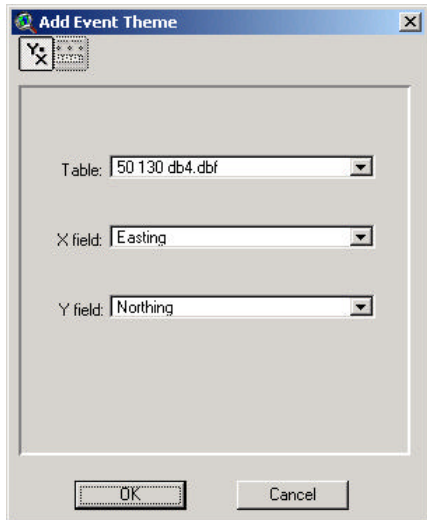
Add the point locations:

4. In the project window, click the TABLES icon and then click ADD
5. Navigate to the drive/directory containing the .dbf file
6. Select the file and click OK (e.g. **50 130 db4.dbf**)
7. You may opt to close any tables that appear

Now add the event theme:

8. Choose VIEW → ADD EVENT THEME
9. Select the table, X field, and Y field (e.g. **50 130 db4.dbf**, Easting, and Northing)
10. Click OK
11. Add any other tables/event themes as desired
12. In the view window, turn the theme on

Specify the project properties

To work in the proper map/distance units:

1. Choose VIEW → PROPERTIES
2. Select **Meters** for both Map and Distance Units (if data is in UTM coordinates)
3. Optionally, provide a new name for the view
4. Click OK

To specify the default extent and cell size:
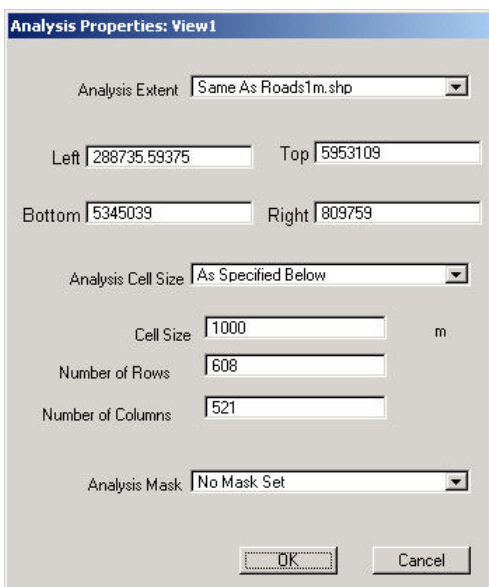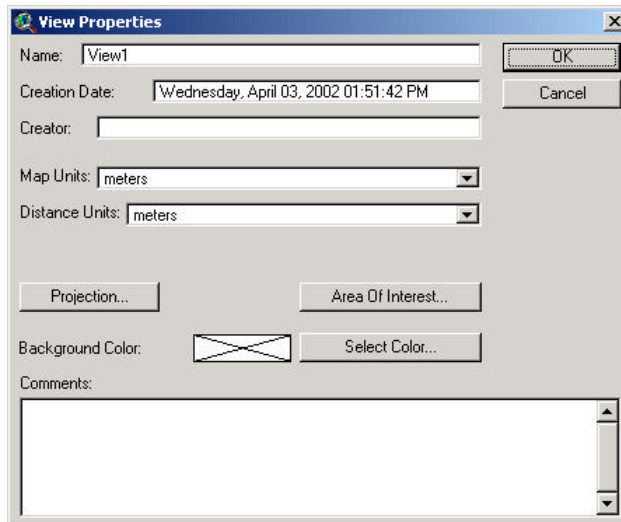
5. Choose ANALYSIS → PROPERTIES
6. Select the desired Analysis Extent (e.g. the same as a roads layer or other data extent)
7. Select "As Specified Below" for the Analysis Cell Size
8. Enter **1000** m as the Cell Size (or other desired size)
9. Optionally, set an analysis mask
10. Click OK

To specify the default output directory:

11. Choose FILE → SET WORKING DIRECTORY
12. Type in the default output directory for results
13. Click OK
14. Save the ArcView project

Query the data to be included in the analysis (i.e. omit missing data)

To perform the calculations on selected data, first query attributes of the data. You may wish to omit records containing "No Data" so select only those records that have data (e.g. Alleles field values of 0=no alleles found, 1= allele from one parent, and 2=allele from both parents).
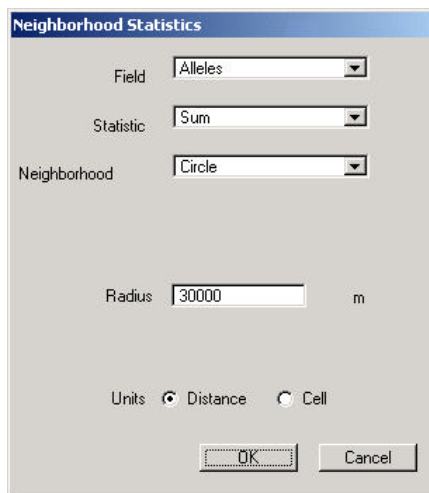
1. In the table of contents of the view window, highlight the desired point theme
2. Choose THEME → QUERY…
3. Enter the expression "([Alleles] >= 0) to omit empty records
4. Click NEW SET
5. Close the Query calculator

If you want to make sure that the appropriate records are selected then open the theme table by choosing THEME → TABLE…
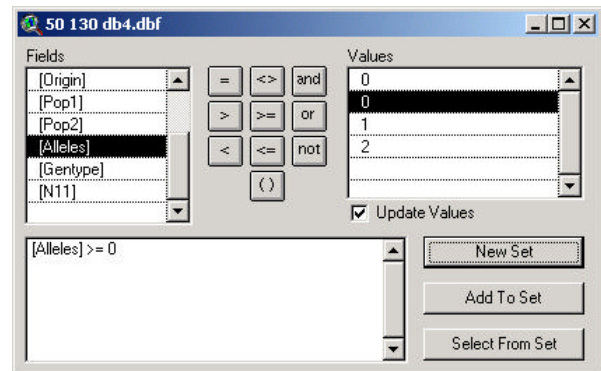
Perform a neighborhood analysis to sum allele presence

1. Choose ANALYSIS → NEIGHBORHOOD STATISTICS…
2. Select "**Alleles**" as the Field
3. Select "**Sum**" as the Statistic
4. Use a **Circle** for the Neighborhood
5. Type in **30000** m as the Radius

You may wish to use a rectangular or other neighborhood, so enter alternate types and size values as needed. Also available is the option to specify units by distance as set in the View/Properties or by number of cells, which takes into account the cell size you specified in Analysis/Properties.
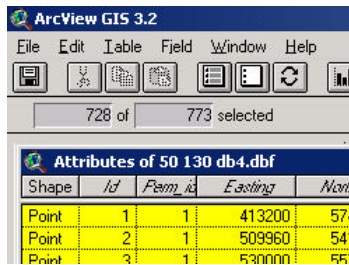
6. Click OK

The resulting raster grid is a map of pixels that contain a value for the sum of all selected allele presence values.

7. Check beside the default name "NbrSum of …" to turn the theme on
8. Double click on the legend to modify how the cells are symbolized
9. Select "Graduated Color: for the Legend Type
10. Select "Value" for the Classification Field
11. Click on the CLASSIFY button and change the parameters (e.g. 10 classes using the Equal Interval method)
12. Select a Color Ramp of choice
13. Click Apply then close the Legend Editor window

Calculate the log ratio using map calculator

The numerator of the ratio was calculated by the neighborhood statistic. Now you will use map calculator to calculate the remainder of the allele frequency equation. First you will need the denominator of the ratio.

1. Open the table for the point theme of interest
2. In the toolbar below the main menu you will find two boxes containing information on the number of records selected (first box) out of the total number of records (second box).
3. Read the value of the first box – this is the number of individuals used in the calculation
4. Multiply the value in the first box by 2 to obtain $2N_i$

Spatial Analyst is picky on the syntax of expressions and how they are to be evaluated. If improper form is used then the output may not be what is expected or desired. For this calculation, a raster grid of floating point values is needed because a logarithmic function is used. Therefore, to force the output as a floating point grid you must enter decimal values. This is achieved by multiplying by the inverse of $2N_i$ to the desired decimal places instead of dividing by the whole integer.

5. Calculate $1/2N_i$ e.g. $2N_i = 2$ x $728 = 1456$, therefore $1/2N_i = 0.0006868$
6. Choose ANALYSIS → MAP CALCULATOR…
7. Enter the expression: **([NbrSum of 50 130 db4.dbf] * 0.0006868).Log10**
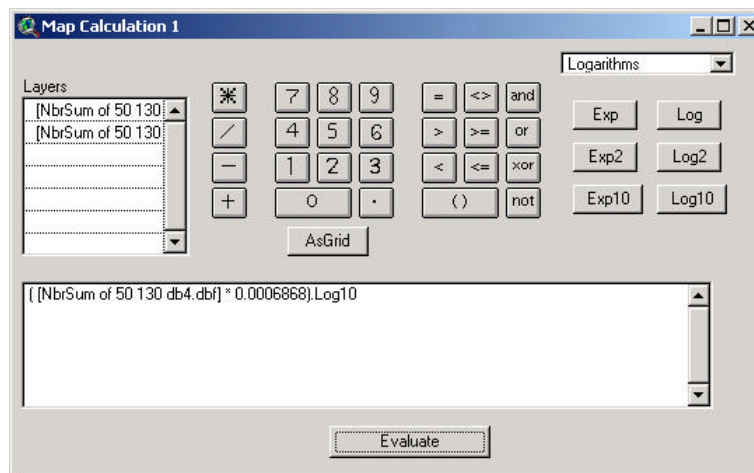
8. Click EVALUATE
9. Turn on the Map Calculation 1 theme to examine the output and symbolize appropriately

Make this new grid theme permanent by saving it with a descriptive name.

10. Choose THEME → SAVE DATA SET…
11. Navigate to the directory and type in a new name (e.g. **freq50130**)
12. Click OK
13. Choose THEME → PROEPRTIES and change the name to reflect the change in the table of contents

## *Create point shapefile of cell centroid and add coordinates*

The Grid Machine extension is used to replace values of "NoData" across the study area.

1. Choose GRID MACHINE → NODATA CHANGER
2. Select the grid you want to change the values for (e.g. **freq30150**)
3. Click OK

Input and output grids must be stored in separate folders, so you may need to go into Windows Explorer and create a New Folder. In this example, it is called "newgrids."

4. Navigate to a new directory
5. Click OK

The Grid PIG extension has a handy tool that enables you to batch several grid files at once into new point shapefiles that designate the cell centroids.
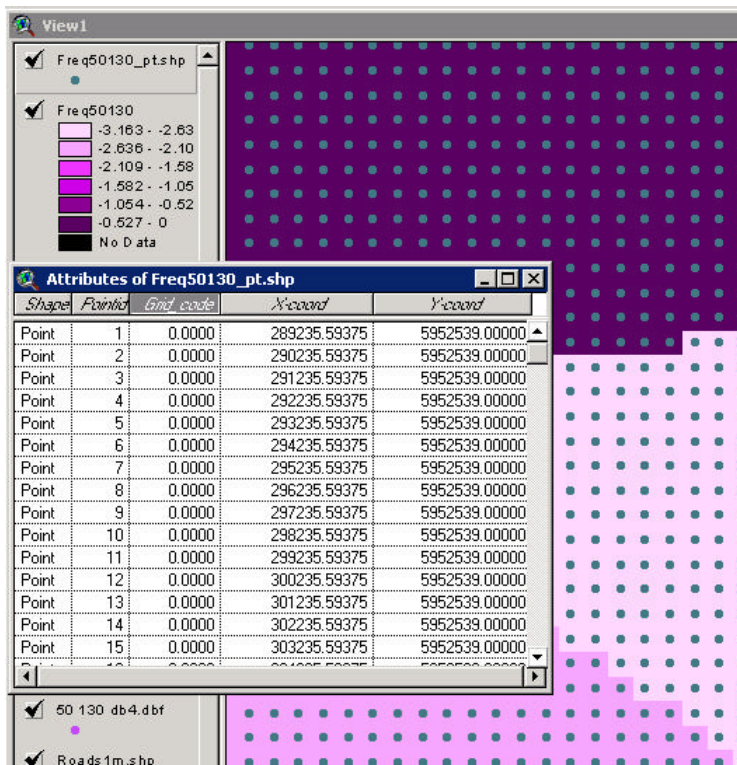
6. Choose GRID PIG TOOLS → BATCH RASTER TO GRID POINT
7. Select the grid file(s) and click OK
8. Click ADD THEME
9. Add the new point shapefile to the view

This new theme represents the cell centroids and associated grid value. The Add XY extension is used to append the coordinates to the table.

10. In the table of contents, highlight the point theme that you want to add the XY coordinates to
11. Choose THEME → ADD XY
12. Open the point theme's attribute table
13. Optionally, export as a dBase or delimited text file for use in a statistical software package

---

TIP: Create a single table of all frequencies for centroid locations

- Add the GET GRID VALUE extension
- Use the first point shapefile as the point theme (make sure all grids have same the extent)
- Name each new field same as the grid that the values are extracted from (i.e. locus allele)
- Optionally, export as a dBase or delimited text file for use in a statistical software package

---