Calculate a Distance Matrix of Locations along River Network

These instructions enable you to measure the length of line segments between points, which is much more useful than simple straight-line distances when you need to work in a network environment such as along water courses, or the more traditional applications on roads and utility features. Using ESRI’s ArcGIS 9.x and Microsoft’s Access/Excel software perform the following steps to obtain pair-wise distance matrix between all locations:

- Project hydrology skeleton and/or sites to same coordinate system e.g. LCC or UTM
- Perform a spatial join (points to hydrology) to determine maximum distance between points and lines (use this value for snapping tolerance in OD cost matrix solution)
- Create/build network dataset from hydrology using all defaults (accept length as the cost field)
- Create OD cost matrix (see ArcGIS Desktop Help Topic: “Creating an OD cost matrix”)  
- Export the OD Cost Matrix ‘lines’ attribute table
- In MS Access/Excel, create a cross-tabulation query/pivot table to organize the data into a true matrix

You need ArcGIS 9.x and the Network Analyst extension (available from ESRI with paid license).

The hydrology framework data used in this example is 1:1 million scale and available for free from: http://geogratis.gc.ca/downftp/en/Vector/National/Frameworks/hydrology_1M/analytical. NOTE: Examine any hydrology dataset that you use to make sure that rivers are continuous and do not ‘break’ at the ends of where water body polygons may exist (useless for network analysis and may require extensive editing).

ORIGINAL DATA
sites.dbf a dBase table of sample point locations recorded by GPS (this example is projected in NAD83 UTM zone 12)
canadskel_l.shp a polyline shapefile of continuously connected rivers

CREATED DATA
studysites.shp a point shapefile from the dBase table added as XY Data
studysitesoffset.shp a point shapefile resulting from a spatial join that calculates distance of points to nearest river
studysiteslines.shp a polyline shapefile of the network solution connecting all origin and destination locations with polyline features
sites_distances.dbf an exported dBase table of distances among all points
Prepare the study data:

1. Start using ArcMap with a new empty map

2. Choose TOOLS >>> ADD XY DATA

3. Click on the BROWSE button (empty folder icon) and navigate to your working directory; e.g. C:\Workspace\AVRiverDistance

4. Select the sites.dbf table, and click ADD

5. Specify the appropriate X and Y fields

6. Click on the EDIT button and SELECT the predefined coordinate system that the data were originally collected in; e.g. PROJECTED >>> UTM >>> NAD 1983 >>> NAD_1983_UTM_Zone_12N

7. Click OK

8. In the table of contents, right click on the sites Event layer and choose DATA >>> EXPORT DATA

9. Specify and output name and location for the shapefile and click OK; e.g. studysites.shp

10. Click YES to add the new layer to the map document

11. Right click on studysites and choose OPEN ATTRIBUTE TABLE

12. Choose OPTIONS >>> ADD FIELD

13. In the ADD FIELD dialog box, specify:
   - Name: ID
   - Type: Short Integer
   - Precision: default

14. Click OK

15. Right click on the new ID heading and choose CALCULATE VALUES

16. In the Field Calculator expression box, enter \([FID] + 1\) and click OK

This new unique ID field is important for joining back attributes later!!! The network analysis processes the data in the order the features are spatially encountered, which is according to the [FID], but renames each location by incrementing it by one. Use this to link back to your original data to keep track of which site is which.

17. Remove the sites Event layer – right click and choose REMOVE

18. Click the ADD DATA button and select canadskel_l.shp to add it to the map document
This data layer is originally in a geographic coordinate system, but is reprojected on-the-fly to the coordinate system defined by your sites point data. To effectively compute distance, all data layers will need to be in a projected coordinate system (distance in decimal degrees is not useful). Therefore, if your study sites were originally in Longitude/Latitude, then you will need to project to UTM or Lambert Conformal Conic, or other projection, before proceeding with these instructions. Contact the GIS Technologist if you need help in this.

19. Right click on **studysites** and choose JOINS AND RELATES >>> JOIN...
20. Choose to ‘Join data from another layer based on spatial location’
21. At step 1, select **canadskel_l**
22. At step 2, select ‘Each point will be given all the attributes of the line that is closest to it, and a distance field showing how close that line is (in the units of the target layer).’
23. At step 3, specify an output name and location; e.g. **studysitesoffset.shp**
24. Click OK

This output shapefile is spatially identical to the input, but has the useful attribute **DISTANCE**, that indicates the maximum distance of any point to the river layer. Use this value to snap the points to the lines when analyzing the network (it virtually ‘moves’ the points so that they lie exactly on the line locations – because GPS coordinate collection methods are never exact). Also, note that the distance units are calculated in the coordinate system of the target point layer – again, it is very important that you are working in a projected coordinate system such as UTM or Lambert Conformal Conic.

25. Right click on the **studysitesoffset** layer and choose OPEN ATTRIBUTE TABLE
26. Scroll all the way to the end of the fields to view the DISTANCE field.

27. Right click on the DISTANCE heading and choose SORT DESCENDING – make note of the highest value – this is the value you will need to use in the network analysis; e.g. 1432.76, or rounded up to an even 2000 meters (because the map units of the target points are in meters!)

28. Remove the studysitesoffset layer – right click and choose REMOVE.

**Create and build the network dataset:**

1. Open ArcCatalog and navigate to your working directory; e.g. C:\Workspace\AVRiverDistance

2. Choose TOOLS >>> EXTENSIONS and make sure there is a check beside NETWORK ANALYST.

3. In the Contents View, right click on canadskel_l.shp and choose NEW NETWORK DATASET.

4. Accept all defaults by clicking NEXT in each wizard dialog box.

5. Click YES to add a cost attribute based on the shape length.

6. Continue clicking NEXT until you can click FINISH.

7. Click YES to build the network dataset – this will take several minutes!

*Read all about creating a network dataset, and the various parameters and options, in the ArcGIS Desktop Help topic “Creating a network dataset.”*

8. Close ArcCatalog when finished.

**Solve the network for an Origin-Destination (OD) cost matrix:**

1. Back in ArcMap, choose TOOLS >>> EXTENSIONS and make sure there is a check beside NETWORK ANALYST.

2. Choose VIEW >>> TOOLBARS and make sure there is a check beside NETWORK ANALYST.

3. Click the ADD DATA button and add the canadskel_l_ND.nd network dataset.

4. Click NO to add only the edges.

5. In the NETWORK ANALYST toolbar, click on the Show/Hide Network Analyst Window to show it – dock it anywhere you wish in the map document.
6. Choose NETWORK ANALYST >>> CREATE NEW OD COST DISTANCE MATRIX

The table of contents automatically sets up the symbology for the categories of layers that will participate in the analysis. The network analysis window shows the available options to set. See the ArcGIS Desktop Help topic “Creating an OD cost matrix” for more details.

7. In the network analyst window, right click on ORIGINS and choose LOAD LOCATIONS…
   - Load from: studysites.shp
   - In the Location Position frame, specify to Use geometry with a SEARCH TOLERANCE of 2000 meters (the value you determined from the spatial join)

8. In the network analyst window, right click on DESTINATIONS and choose LOAD LOCATIONS…; e.g. studysites.shp and use the same search tolerance as above

9. In the table of contents, right click on the OD Cost Matrix layer name and choose PROPERTIES

10. Optionally, in the ANALYSIS SETTINGS tab, specify any settings — this is normally applicable to transportation and utility type networks, so ignore this step

11. Click the SOLVE button

Distances between all origins and destinations are calculated and graphically displayed by their straight-line connections. The number of lines should total to # origin locations X # destination locations, if every possible pair-wise distance was solved for.
12. In the table of contents, right click on the **Lines** layer and choose **DATA >> EXPORT DATA**

13. Export the data using the same coordinate system as the data frame, specify an output name; e.g. **studysiteslines.shp**, and click OK.

14. Right click on **studysiteslines** and choose **OPEN ATTRIBUTE TABLE**

15. Click on the **OPTIONS** button and choose **ADD FIELD**

16. In the **ADD FIELD** dialog box, specify:
   - Name: **Str_Length**
   - Type: Double
   - Precision: 20
   - Scale: 2

17. Click OK

18. Right click on the new heading **STR_LENGTH** and choose **CALCULATE VALUES**

19. Check the **ADVANCED** box

20. In the **Pre-Logic Script Code** window, enter the following:
   ```vba
   Dim Output as double
   Dim pCurve as ICurve
   Set pCurve = [shape]
   Output = pCurve.Length
   ```

21. In the bottom box, enter the word **Output**

   To learn more about this calculation, click on the **HELP** button.

22. Click OK

There are now two fields referring to distance between site locations in the attribute table. The first field, **TOTAL_LENG**, was calculated along the river network, and the second field, **STR_LENGTH**, was calculated as the straight-line connections from each origin to destination. You may wish to join the attributes from the original **studysites** layer to make identification of each location easier.
23. In the table of contents, right click on the *studysites* layer and choose JOINS AND RELATES >>> JOIN...
24. Specify that you want to ‘Join attributes from a table.’
25. At step 1, choose the *OriginID* field.
26. At step 2, choose the *studysites* table.
27. At step 3, choose the *ID* field.
28. Click OK.

*In the studysites* attribute table, you will see the appended fields from the original point data. This attribute table is also what you may export for further statistical analysis.

29. Click OPTIONS >>> EXPORT.
30. Specify an output name and export all records; e.g. *sites_distances.dbf*.
31. Click OK.
32. Save the map document and then close ArcMap when finished.
**Import the table for use in MS Excel/Access:**

The `sites_distances.dbf` table has a record for each and every pair-wise distance. To set this up in matrix form, import the table into a spreadsheet or database application. To create a cross-tabulation query in MS Access for the distance matrix, see page 3 in:


The following explains how to create the related pivot table in MS Excel:

1. Open MS Excel
2. Choose FILE >>> OPEN
3. Navigate to your working directory and select Files of type: dBase Files
4. Select the `sites_distances.dbf` table and click OPEN
5. Choose FILE >>> SAVE AS
6. Select Files of type: Microsoft Excel Workbook
7. Specify a new name; e.g. `sites_matrix.xls` and click SAVE
8. Choose DATA >>> PIVOTTABLE AND PIVOTCHART REPORT
9. In step 1, select Microsoft Excel list or database and PivotTable (i.e. the defaults for this step); click NEXT
10. In step 2, click on the range button and highlight the columns to use (i.e. the Origin, Destination, and Total_Leng columns); click NEXT
11. In step 3, select to put the PivotTable report in a New Worksheet; click FINISH
12. In the new worksheet, find the PivotTable Field List
13. Click and drag ORIGINID to the DROP ROW FIELDS HERE area
14. Click and drag DESTINATIO to the DROP COLUMN FIELDS HERE area
15. Click and drag TOTAL_LENG to the DROP DATA ITEMS HERE area
16. Double click where it says ‘Count of Total_Leng’ and choose to Summarize by SUM (Average, Min, and Max will work, too).

You now have a distance matrix showing the origin IDs and their lengths to destination IDs. You may opt to repeat the pivot table procedure using the Str_Length data in a new distance matrix.

17. Save the .xls file.