MASSIVE DATA

Introduction to

Working with Massive Geodata Objects

with

TNTmips® and TNTedit

Before Getting Started

The key to working with large data is to get rid of any unnecessary largeness. For raster objects, unnecessary largeness is removed by compression. Whether or not lossy compression is acceptable will depend on the intended use for the data. There are a variety of techniques used to remove unnecessary largeness from vector objects, which include database modifications, filtering, and thinning. With vector objects, this process is generally referred to as “cleanup.”

Prerequisite Skills This booklet assumes you have completed the exercises in *Displaying Geospatial Data*, *Navigating*, and *Editing Vector Geodata* tutorial booklets. The exercises in those booklets introduce essential skills and basic techniques, which are not covered again here. These prerequisites will help to provide a context and the terminology to help you better understand the concepts in this booklet.

Sample Data This booklet does not use exercises with specific, downloadable sample data to develop the topics presented. You may have acquired some of the data sets that resulted from the techniques described (for example, World Data Sets, Lincoln Property Viewer) if you are a professional TNT product user. Some of the sample data is too large to distribute except by removable hard drive.

More Documentation This booklet is intended only as an introduction to managing large data sets. Details of the processes discussed can be found in the Online Reference Manual and a variety of tutorial booklets.

TNTmips and TNTlite® TNTmips comes in two versions: the professional version and the free TNTlite version. This booklet refers to both versions as “TNTmips.” If you did not purchase the professional version (which requires a software license key), TNTmips operates in TNTlite mode, which limits the size of your project materials and does not allow preparation of atlases or export of data. None of the data discussed in this booklet will fit in TNTlite.

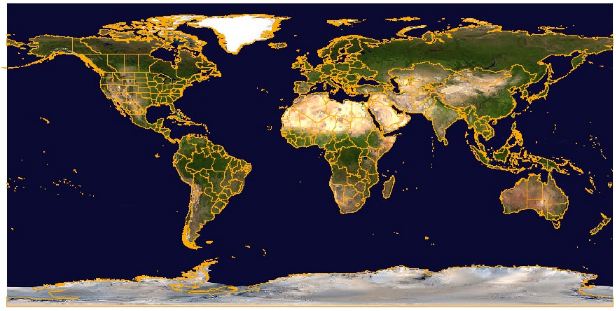
Merri P. Skrdla, Ph.D., 10 November 2004

The Blue Marble earth illustration on the cover was assembled at the NASA Goddard Space Flight Center from MODIS satellite data.

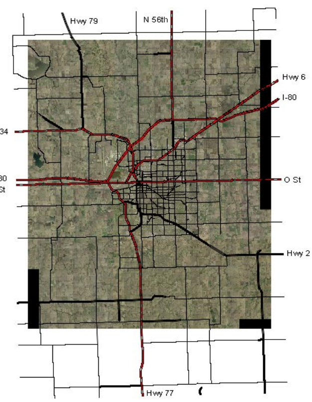
You can print or read this booklet in color from MicroImages’ web site. The web site is also your source of the newest tutorial booklets on other topics. You can download an installation guide, sample data, and the latest version of TNTlite.

http://www.microimages.com

Working with Massive Geodata



What is massive geodata? It is not defined by the geographic extents of an object but rather by its storage requirements. The type of object must also be taken into account. For raster data, we will define massive as 2 GB or greater. The TNT products currently have a 4 TB storage limit on raster objects whether uncompressed or compressed with any of the supported compression methods, which include JPEG2000, Run Length Encoded, Standard Lossless, and Huffman. This limit applies to a single raster object stored in a single file and does not include its pyramids.



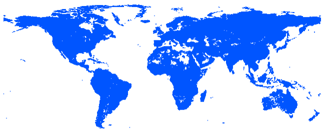
Massive raster geodata does not translate into slow display performance in the TNT products. The use of pyramids and quick JPEG2000 decompression displays the 500 MB JPEG2000 lossless compressed raster (above right with a vector overlay) in less than one second.

We arbitrarily define massive vector objects as over 100 MB. The TNT products support vector objects requiring up to 4 TB for storage, with up to 2 GB of storage per element type. 2 GB of storage is more than 134,000,000 nodes, points, or line vertices. Each 2D node, point, and line vertex requires 16 bytes. So a massive vector object, which is far smaller than the maximum size, would be made of more than 6,553,600 points or a single line with that many vertices or some combination of points, lines, and polygons that requires that amount for storage. The 4 TB limit does not include databases, style objects, and other subobjects, but these and all the other "behind the scenes" lists and other vector requirements must fit in the 16 TB Project File size limit.

Vocabulary: TB or **terabyte**. A unit of measurement for 2^{40} (approximately) 1,000,000,000,000 bytes, 1,000,000,000 kilobytes, 1,000,000 megabytes, or 1,000 gigabytes.

Pages 5–10 discuss assembling raster data and dealing with null values. Pages 11–18 describe merging and cleaning up vector data including attributes and topology. Pages 19–22 discuss the importance of map scale control and DataTips. Additional vector filtering processes are discussed on p. 23.

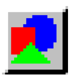
Further Introduction to Massive Geodata

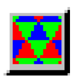


This display of world hydrology at full view took 26 seconds, but no individual elements can be discerned.

Massive vector data should be optimized to improve display times, but display at full view may be slow and meaningless. At full view, individual elements are rarely distinguishable. The use of map scale control to manage massive vector displays is discussed later.

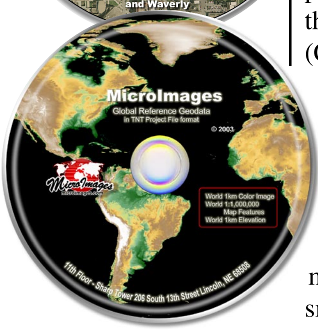
If you have not noticed, the TNT products' limits on object size are not really limitations for most people. The maximum Project File size is 16 TB. Do you have storage for 16 TB? Do you even have storage for the 4 TB raster size limit?





As long as we're talking about size limits in regard to massive data, we will complete the list of primary object types in TNTmips even though MicroImages has acquired no data that could be considered massive in these formats. CAD blocks can be up to 4 TB and the object must fit within the 16 TB limit. A block can contain up to 2,147,483,648 elements.

TIN objects can also be up to 4 TB. Each TIN node requires 24 bytes of storage, edges 16 bytes, and triangles 40 bytes. Databases can be up to 16 TB with each table up to 4 TB. There is no size limit on ODBC linked databases, although file based databases usually have a 2 or 4 GB limit imposed by the software that created them or by the ODBC driver.



Generally, you are not working with an object in isolation; massive data frequently comes in sets. A practical definition for a massive data set is anything that doesn't fit on your chosen distribution medium (CD or DVD). In the case of vector data, massive data may fit on CD or DVD but is unnecessarily slow to display and larger than it needs to be, generally because of unnecessary nodes and database records. The Global Reference Geodata DVD distributed with the V6.9 release of the TNT products is a massive data set that consists of a massive raster, some massive vectors, some very small vectors, and others in between.

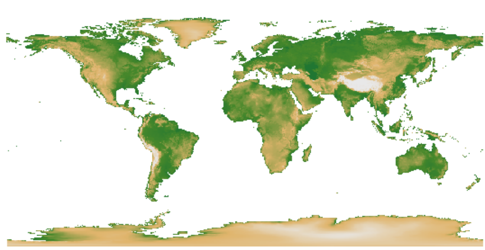
Getting the Parts

Massive data often comes in pieces. The pieces should be assembled before proceeding. You will be repeating steps unnecessarily if you begin cleanup of vector pieces before you merge them. You also want to mosaic raster objects before you compress them because you want the compression solution to be for the whole raster rather than individual compression solutions stuck together.

Before you start, you need to be familiar with your data and know its intended use. Lossy compression should not be used on rasters intended for classification or other analytic uses with results that will be altered by changing the data values. Lossy compression is suitable for base images for hardcopy or display layouts, scanned map sheets, and images used for visual (non-computational) photointerpretation. Thinning for vectors is analogous to compression for rasters and should take into account how important it is to preserve the original information. For example, if all line vertices are surveyed points, you may want to preserve them.

Most of the massive data illustrated in this booklet is available for free download. Don't let an Order Data button make you assume you have to buy the data—look further on the site. For example, you can download the 30 arcsecond elevation data illustrated on this page from http://edcdaac.usgs.gov/gtopo30/gtopo30.asp although the page has an Order Data but no Download button. If you click on the ReadMe link, you will find instructions for downloading the data under Data Distribution.

It is obvious that you would want to mosaic the 33 pieces of the 30-arcsecond elevation data (GTOPO30) shown at the left before applying a color map (shown below) because you can do it once instead of 33 times. The same rationale applies to vector cleanup as well.



USGS Offers High Resolution Imagery

STEPS

- ☑ acquire your high resolution data
- ☑ choose Process/Import/Export, select import and raster, then choose the format you acquired the data in, then import into TNTmips without compression

Little or no degradation of the image can be detected in the comparison of the original TIFF file (left) and the MrSID 10:1 compression (right).

Size adds up very quickly when working with high resolution imagery, which may often provide the backdrop for a large data set. When imagery is used as a backdrop for other data, compression is quite appropriate. We would recommend the TNT products' internal JPEG2000 compression because of the speed of its decompression. If you want the imagery to be usable in other products, you can create a linked external JP2 file with a world file for georeference, which is decompressed equally fast.

USGS provides a means to download a 30-cm (1 foot) resolution color orthophoto of your area of interest (seamless.usgs.gov). This download is limited to 100 MB, which is slightly more than one square mile. If you are interested in larger areas, you can download it in pieces or order it on CD. This orthophoto can be imported into the TNT products or used directly if downloaded in TIFF format. The imagery can also be purchased as 10:1 compressed MrSID files and likely in other formats.

The original downloaded TIFF file (left) is compared at 2X with the purchased MrSID 10:1 compressed file. Both are directly displayed (not imported) in their native format.

Lincoln Area Mosaic

MicroImages acquired the USGS high resolution orthoimages for 1360 square kilometers (525 square miles) that includes Lincoln, Nebraska and surrounding communities. The data was delivered in MrSID lossy compressed format (10:1) on two DVDs. Since the goal was to produce an atlas that fit on a single DVD, clearly greater compression was required with JPEG2000 as the format of choice. The orthoimages were decompressed on import.

The next step is not to recompress the tiles with a greater degree of loss, but to mosaic them. Compressing the images before mosaicking would result in definite seams. The issue is that JPEG2000 compression cannot produce identical losses when applied separately to pieces of a larger image. Consider the case of applying 10:1 lossy compression to two adjacent tiles making up a larger image where one has a very large, uniform lake and the other does not. In the results, the amount of detail lost in the image containing the lake is not as significant as for the image containing only detailed ground features. Thus, when you apply fixed ratio lossy or best quality lossy, the detail lost will vary slightly from tile to tile resulting in visible seams.

Internal JPEG2000 compression is available in both the Raster Extract and Mosaic processes. So if you want to retain an uncompressed version of your mosaic, you can later use Raster Extract to copy the entire image to a lossy JPEG2000 compressed raster object. You could alternatively export the raster to a JP2 file using the same lossy compression.

STEPS

☑ choose Process/Raster/Mosaic, select all your rasters and mosaic them with or without compression

- 641 images
- each 5,000 lines x 5,000 columns
- each 7.1 to 7.3 MB with MrSID 10:1 compression
- each approximately 71.5 MB when uncompressed
- mosaic 145,000 lines x 115,000 columns
- mosaic 3.98 GB with internal 12:1 JPEG2000 compression
- cell size 30 cm (1 foot)

Null Value Masks

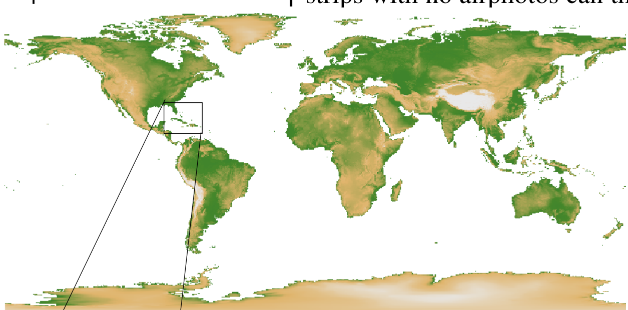
STEPS

- ☑ examine the histograms for the rasters you want to mosaic to determine their null value (if any) and its distance from valid data values
- ☑ change the null value if necessary (next exercise)
- ☑ choose Process / Raster / Mosaic and select the rasters with null areas you want to mosaic

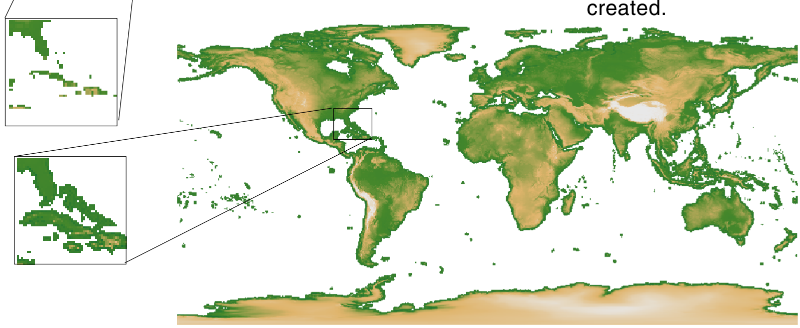
You do not need to set a null value or choose a null mask, null values/masks are handled automatically by the process. If your input rasters have null values, they need not all be the same to be included in the null mask generated for the output.

A null value mask is a raster subobject that tells whether a cell is valid (has a value) or null (has no value). Null value masks are important for JPEG2000 compressed images because JPEG2000 compression does not handle null values. In order to have a transparent background for a JPEG2000 compressed raster, you must have a null mask, which is created automatically by the Mosaic process. Many formats in which imagery is acquired, including TIFF, GeoTIFF, and MrSID, do not support null values.

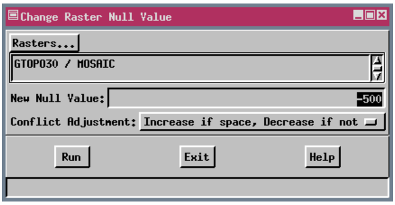
The mosaic in the previous exercise did not have a null value and was created before the Mosaic process automatically generated null masks, and generating a null mask was not selected as an option. A null value cannot be assigned after the fact in that case because there are real zeros within the image area and zeros were used to fill the rectangle where there was no image input. A null mask is now automatically generated by re-mosaicking the pieces, and the strips with no airphotos can then be transparent.



The effects of a null mask are particularly clear when using smaller pyramids. The raster object in this illustration is 21600 lines by 43200 columns. The extended coastal effect is due to sampling when pyramids are created.



Changing the Null Value



JPEG2000 compression fits a wavelet to the pixels (typically 9x9 areas) and uses the values of the pixels in its computations. Large differences in values will result in intermediate values not being "close" to any of the input values. Often the null value of a raster is at an extreme of the data range, which means it is not a good null value if the raster is going to be JPEG2000 compressed unless the data values are also near the same extreme. Pick a value that is outside the data range, either higher or lower, and assign it as the new null value in the Change Raster Null Value process before mosaicking.

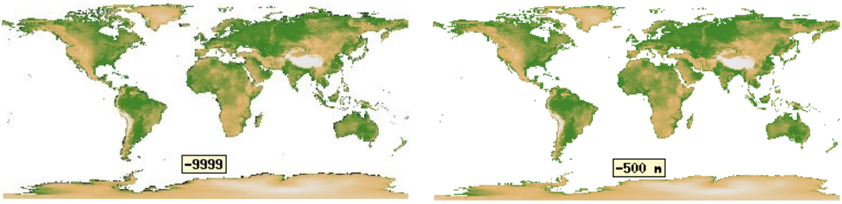
You can change the null value of rasters before or after they are mosaicked. If conflict adjustment is necessary or if you are creating JPEG2000 output, you should change the null value before mosaicking. If the raster does not have a null value set but has areas of a single value that should be null, simply set the null value using Project File Maintenance. All input rasters for mosaicking need not have the same null value. The Change Raster Null Value process makes changes in place (no new raster).

You need no longer be concerned about whether the null values for input rasters for mosaic are the same or whether a valid cell value will conflict with the automatically assigned null of the mosaic output—the null mask, not the cell value, determines whether a cell is valid or null when a raster has a null mask subobject.

STEPS

- ☑ before mosaicking, choose Process/Raster/Utilities/Change Null Value and select all rasters for a single mosaic
- ☑ enter the value that occurs in the data that you want to use for the null value, such as [0 0 0], in the New Null Value field
- ☑ choose a Conflict Adjustment method, and click [Run]
- ☑ proceed with the mosaic

Note: Conflict adjustment changes any cell values that are the same as the designated new null value to a different, very close value (for example, [0 0 0] may be changed to [0 0 1]).



The GTOPO30 data had a null value of -9999 when imported. The lowest data value in the raster is -405 m. The null value was changed to -500.

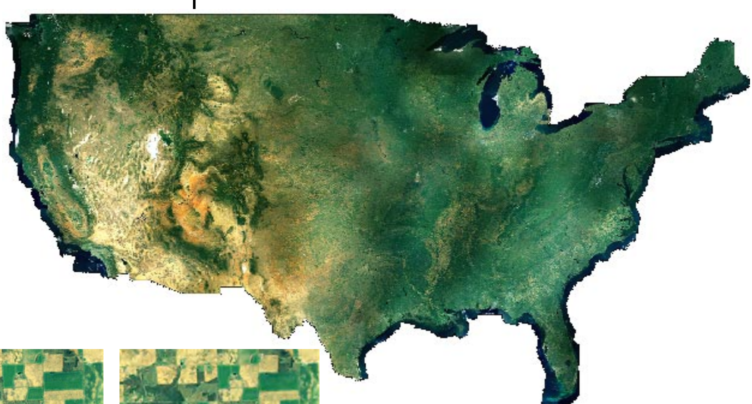
Really Massive Rasters

- 264 2° x 2° (Lat/Lon) input tiles
- each 14,400 lines x 14,400 columns
- each 593 MB uncompressed
- mosaic 187,546 lines x 525,519 columns trimmed at the 49th parallel to 180,393 lines x 513,619 columns
- mosaic 275 GB without compression, 85 GB with JPEG2000 lossless compression after trimming, 1.67 GB with 20:1 JPEG2000 lossy compression (external JP2 file)
- cell size 15 m (lin) x 12 m (col)

Really massive rasters are hundreds of gigabytes. An example is the 275 GB Millennium Mosaic acquired from Digital Globe in 2° x 2° pieces. The Millennium Mosaic is a pan-sharpened, contrast matched, seamless mosaic of natural color Landsat7 scenes covering the 48 conterminous states in the USA. Even such massive data can be gotten to a distributable size with reasonable quality as shown by the full resolution comparison of the 20:1 lossy JPEG2000 compressed raster with the lossless version shown below. The 1.67 GB lossy compressed raster leaves quite a bit of room for other data on a DVD.

Lossy compression of such large rasters requires some memory management. Windows will kill any process whose virtual memory usage exceeds 2 GB. The memory management controls are part of the export to JP2, but are not yet included in the Raster Extract process.

The Millennium Mosaic at full view.



lossless

20:1 lossy

20:1 lossy JPEG2000 compression takes the Millennium Mosaic from 85 GB to 1.67 GB, which is a much more manageable size. There is some loss of quality, but the result is certainly acceptable for the convenience of being able to distribute the data on DVD. These full resolution screen captures (left) are from eastern Michigan.

Standard Attributes and Topology

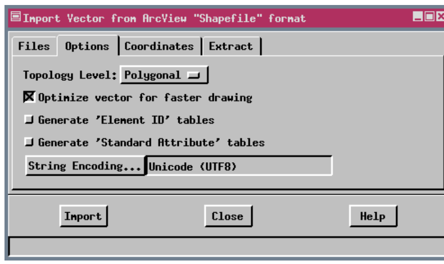
You may want standard attributes for your massive vector objects. It is useful to be able to readily tell the length of a stretch of road or the area of a lake. However, standard attributes take a fair amount of time to calculate for massive vector objects and, once calculated, have to be recalculated at every step in the cleanup process. You could actually lose hours of time by turning this option on during import if there will be several steps in the cleanup process. Calculate standard attributes at your last editing step or in the separate Standard Attributes process (Process/Vector/Attributes/Standard Attributes). Also, unless you work with element IDs, do not generate these tables.

Topology is another consideration for import. TNTmips supports three levels of topology: polygonal, planar, and network. For more information about topology, see the *Vector Analysis Operations* booklet. Some sources of vector data, such as shapefiles, do not support lines and polygons within a single object. Vectors that were stored as lines only should be imported with planar topology. Planar topology saves about 15% in object size because polygon information is not maintained. If you want to import vectors with a mix of topology types, import all with polygonal topology then change the topology type in the Spatial Data Editor for those that need it.

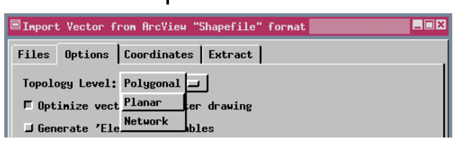
STEPS

☑ choose Process/Import-Export, set the Object Type to Vector and the Operation to Import

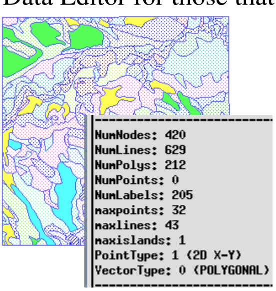
☑ be sure that Optimize vector for faster drawing is on and Generate element ID tables and Generate Standard Attribute tables are toggled off

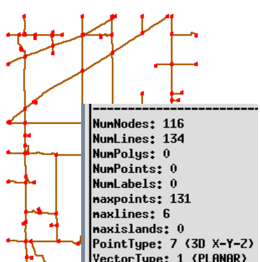


☑ set the appropriate Topology Level and Import

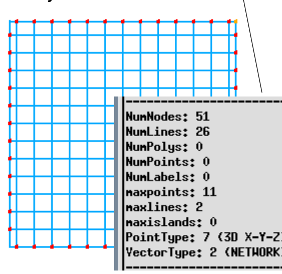


Note: you can skip this step if you are working with shapefiles.





Project File Maintenance Object Information



Merge Vectors First

STEPS

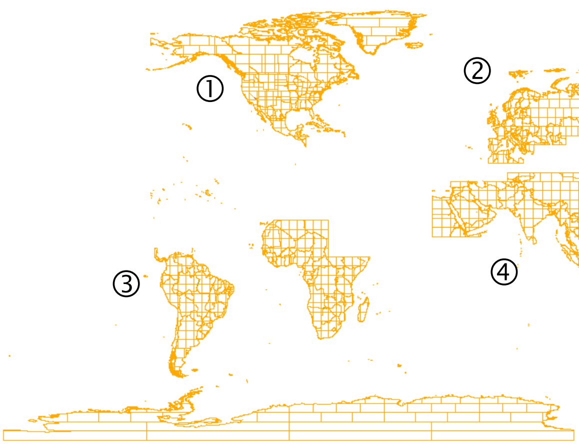
☑ after import or before import if using shape files, choose Process/ Vector/Merge

☑ be sure that the Optimize vector when saving toggle is on

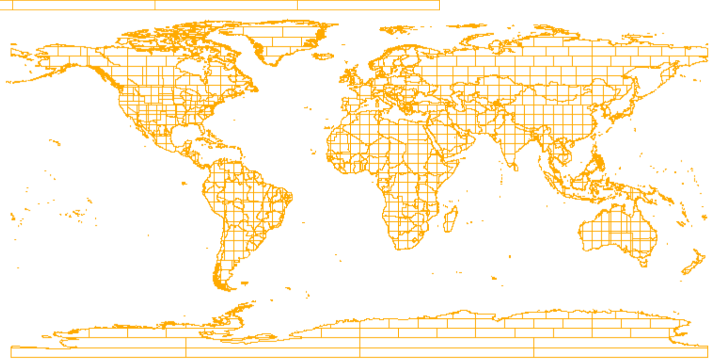
Note: if you begin cleanup before merging you will have to repeat the process for each piece.

Even vector data is often delivered in pieces. The Digital Chart of the World (DCW), and VMap0 themes are distributed by the National Geospatial-Intelligence Agency (NGA) in four pieces. The VMap1 themes are distributed in even more pieces. Your first task after import is to merge the pieces for the area you want. If your vectors were acquired in shapefile format, you do not even need to import them before merging them into a single vector.

This merged vector at just over 30 MB is not a massive vector itself but is part of a massive data set that includes a vector that is nearly 500 MB with approximately 1,500,000 lines and 425,000 polygons after cleanup.



DCW and VMap0 are provided in four pieces: North America, Europe and Northern Asia, South America and Africa, and Southern Asia and Australia.



The merged vector has almost 35,000 lines and over 30,000 polygons.

Dissolving Polygons

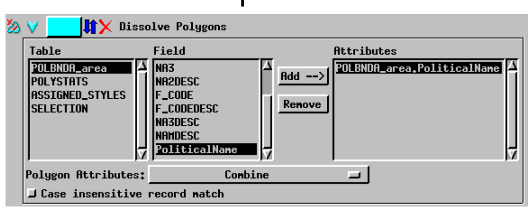
A few of the DCW and VMap0 themes contained grid lines, which are distracting and unnecessary in a GIS system. Grid lines were removed using the Dissolve Polygons filter after the quadrants were merged into a single vector object (previous exercise). In the resulting vector (shown below), each polygon represents a country, province (Australia, Canada), state (Brazil, USA), island, or continent (Antarctica). Use Spatial Data Display to determine the names of the table(s) and field(s) you want to define the polygons that result from the filtering process.

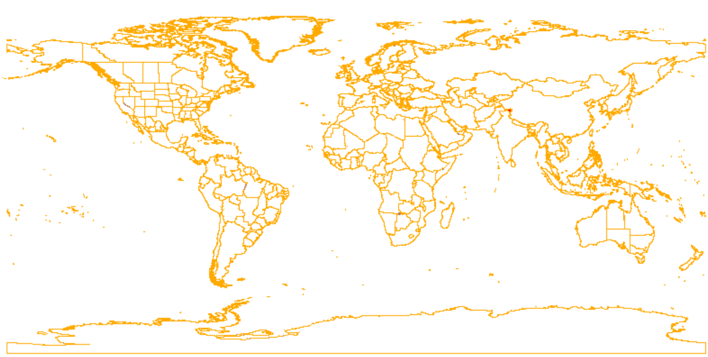
Once you have dissolved polygons, you are likely to have excess nodes, or nodes that divide two lines with the same attributes, where the grid lines used to cross the political borders. You can set the Remove Excess Nodes filter to run after the Polygon Dissolve filter completes. The process runs the filters in the order they are shown from top to bottom.

Although a test option is available in the vector filters process, you probably do not want to use it with massive data because the job will take twice as long. Use Spatial Data Display to determine if you got the expected results.

STEPS

- ☑ choose Process/Vector/Filter, select your vector, then click on the Add Filter button 
- ☑ choose Dissolve Polygons, click on Show Details, select the table and field to be used to separate polygons, and click on [Add] 
- ☑ select additional table and field combinations if needed and set the appropriate option for Polygon Attributes
- ☑ click on Add Filter again and choose Remove Excess Nodes 
- ☑ click on Run, then click on [View Log File] when the process is done and note the extent of change





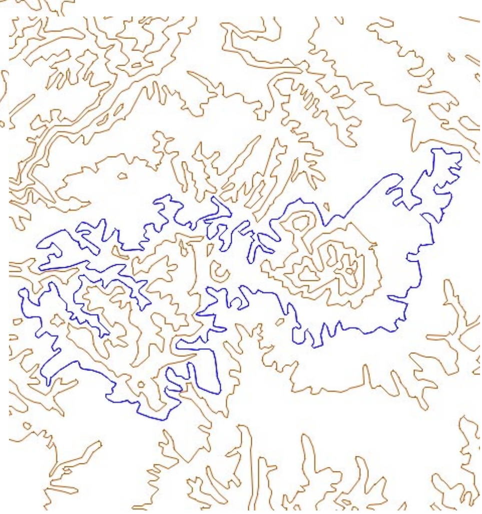
Removing Excess Nodes

STEPS

- ☑ open Spatial Data Display and choose the Select tool 
- ☑ click on some lines and see if the full extent of the line you logically expect is highlighted
- ☑ determine if the lines are uniquely attributed
- ☑ if not, run the Remove Excess Nodes filter as described in the previous exercise

Vector objects can have excess nodes without removing lines. In the VMap0 data downloaded, all themes with line attributes had logically continuous lines, such as contour lines, broken into a number of segments. ESRI's E00 and Coverage formats have a 500 vertex limit per line. The Import process will remove excess nodes if the lines are not uniquely attributed. Unfortunately, ESRI's use of element IDs makes such lines uniquely attributed. How to deal with this problem is discussed in the next exercise, which concerns removing duplicate records.

The single contour line selected (blue) in the World Map Features Global Data Set that was supplied with V6.9 of the TNT products (below, right) was 17 separate lines in VMap0 (left), one of which is shown selected. The number of lines went from 1,099,572 to 655,769, approximately a 40% reduction when excess nodes were removed.



The Remove Excess Nodes filter is also available in the Spatial Data Editor, but opening a massive vector in the Editor is time consuming. If you are not going to be doing additional editing on the vector object, you will save time by running the Vector Filters process.

Removing Duplicate Records

When logical lines, such as contour lines or a stretch of road between intersections, are broken into a number of pieces with unique attributes, you should remove the fields that make them unique from the table unless they are something meaningful like "under construction." After removal of these fields, run Remove Duplicate Records. If you determined in the previous exercise that your broken lines are uniquely attributed, you will need to make duplicate records so you can remove them. You could theoretically remove all the fields you don't particularly care about, but you only need to remove the fields that make these lines artificially unique. Any field that contains "ID" in its name is a likely candidate for removal.

In this exercise, we are looking at contour lines, so the only really important field is the Z value (ZV2 field). However, it is only the three fields with ID in their names that make the contour line segments unnecessarily unique. Removal of these fields followed by deletion of duplicate records reduces the number of records in the table from 1,099,572 to 419.

STEPS

☑ choose Display/Spatial Data, open a new 2D group, and add the vector that has split lines that are uniquely attributed

☑ choose the select tool, open relevant table(s), and see which fields differ between lines that should be continuous

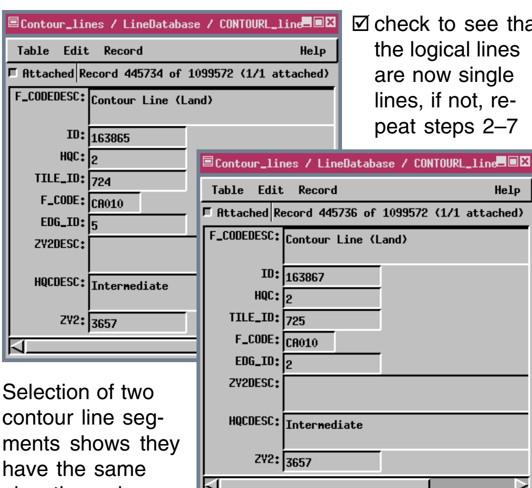
☑ choose Table/Edit Definition for each open table that needs fields removed (one at a time)

☑ highlight a field that needs to be removed, and click on the delete icon

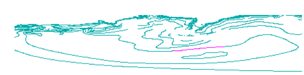
☑ repeat until all problem fields are deleted

☑ choose Edit Relations from the Make Table/Form icon, right click on the table, and select Delete Duplicate Records

☑ after exiting display, choose Process/Vector Filter, select the Remove Excess Nodes filter, and Run



☑ check to see that the logical lines are now single lines, if not, repeat steps 2–7



ID 163865

ID 163867

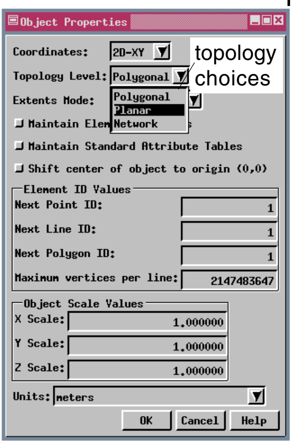
These two lines are part of the same elevation contour, but unique ID numbers make them separate lines.

Selection of two contour line segments shows they have the same elevation value (ZV2), as well as some other field values, but differ in the ID, TILE_ID, and EDG_ID fields. After removing these fields and excess nodes, the contour line was continuous.

Changing to Planar Topology

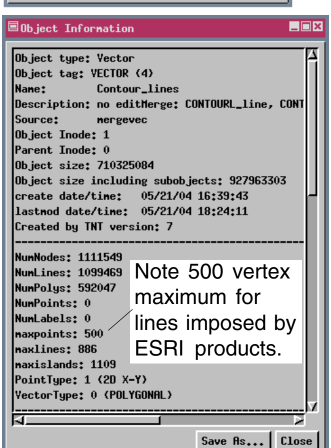
STEPS

- ☑ choose Edit/Spatial Data and select the object for topology change
- ☑ select Layer/Properties in the Spatial Data Editor window and change Topology Level to Planar
- ☑ do any other editing tasks necessary
- ☑ generate standard attributes if no further manipulation is required

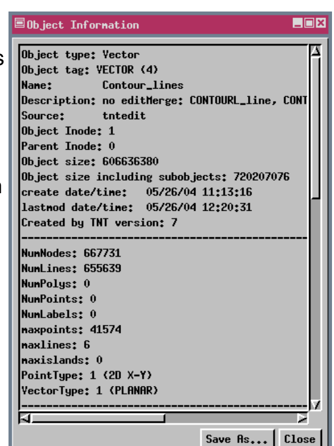


If you are importing a single vector you know contains only line data, you should assign it planar topology on import. If you are not sure or you are importing multiple vectors with mixed topologies, import them all with polygonal topology, then change those with only line attributes to have planar topology. There may be a clue to the appropriate topology in the file name, such as if all names end in a *p*, *l*, or *a* for point, line, and area (polygon), respectively. There is no size benefit for making an object that contains only points planar, so there is no need to do so. For objects that contain attributes for lines only that have a structure that includes topological polygons, switching from polygonal to planar topology saves about 15% in size.

Because the change from polygonal to planar topology must be done in the Spatial Data Editor, you may as well do filtering and any database changes needed for your merged objects in the Editor as well. Thus, the most efficient use of your time when cleaning up large data sets, in particular vector objects, requires some forethought. In fact, the previous three exercises, if all were necessary, can be done in the Spatial Data Editor.



The object information at the left is for the polygonal vector as imported with standard attributes. Note the reduction in elements and size for the planar vector with database editing and excess nodes removed. Standard attributes were recalculated at the end of the editing session.



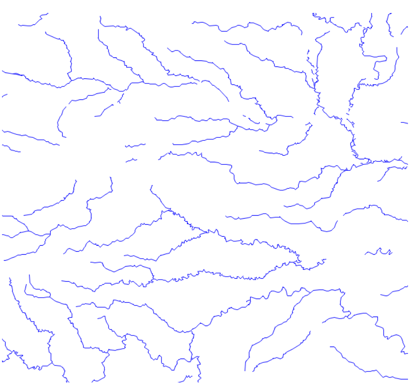
Merge Line and Polygon Objects

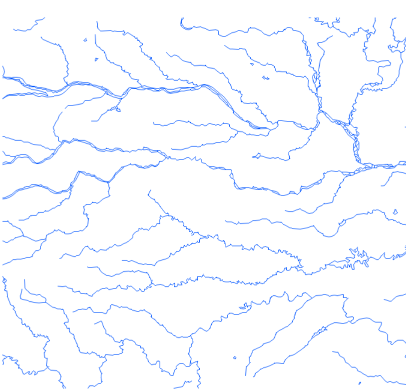
A number of difficulties are presented when objects are artificially separated into line- and polygon-only objects. In the hydrology example for the World Data derived from VMap0, rivers do not flow into lakes because there are no lakes in the object containing rivers and there are no rivers in the object containing lakes. If displayed together, the view is reasonable but when displayed separately, they appear disjointed with gaps (rivers) or contain isolated lakes and wide rivers (water bodies).

Although each of these vectors qualified as a massive vector object, it is not a problem to merge them in TNTmips. The result is a vector with polygonal topology and both line and polygon attributes.

STEPS

- ☑ choose Process/Vector/ Merge
- ☑ select the lines-only and polygons-only vector objects that should logically be a single object
- ☑ check that all three toggles are on in the Vector Merge window
- ☑ click on Run





VMap0 divides its hydrography into water courses (above left) and inland water areas (above). When viewed separately, it looks like there is some problem with the data because it is not connected. Luckily, these two vector objects fit perfectly when merged leaving no gaps between rivers and the lakes they empty into or run from or between the wide and narrow representations of the same river (left).

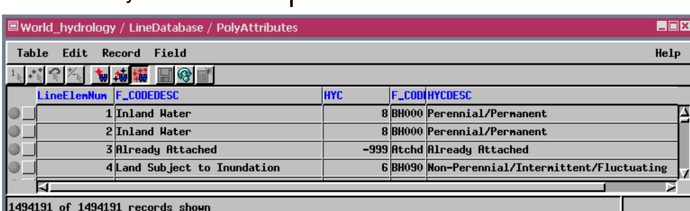
Transfer Attributes from Polygons to Lines

STEPS

- ☑ choose Display/Spatial Data and select your merged vector object from the previous exercise
- ☑ add a new table (e.g., Attached) to the line database with implied one-to-one attachment and a single computed field (e.g., Num) with the expression SetNum(TableName[*]) where TableName is the name of the table that contains the line attributes of interest
- ☑ start your virtual field expressions to be used for attribute transfer with *if (Attached.Num==0)*, then choose the desired attributes from the polygon database
- ☑ do cleanup and union as necessary

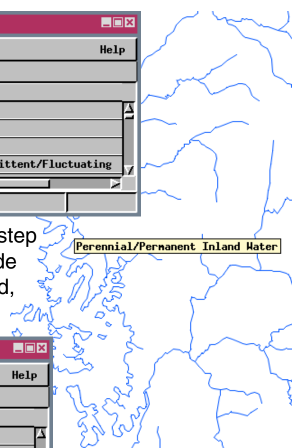
In the previous exercise, a massive vector object was created (approximately 650 MB with 1.5 million lines) with some lines that have attributes and some lines that do not. The lines that have attributes are from the planar rivers object and the lines without attributes are from the polygonal inland water areas object. If both lines and polygons are selected for drawing, drawing takes longer since you are drawing all of the polygon lines twice. The time factor is not noticeable when drawing smaller vector objects but can be significant with massive vector objects. However, if an element type is not selected for drawing, DataTips for that element type are not available.

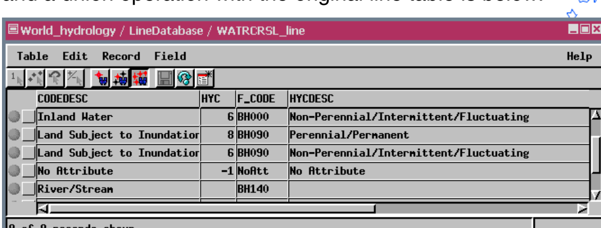
The vector can be manipulated by introducing a progression of computed fields to transfer polygon attributes to the lines that make up the polygons, followed by database cleanup and joining tables. You first need to determine if an element already has a record attached because you will take no further action on that element.



LineElemNum	F_CODEDESC	HYC	F_CODE	HYCDESC
1	Inland Water	8	BH000	Perennial/Permanent
2	Inland Water	8	BH000	Perennial/Permanent
3	Already Attached	-999	Atchd	Already Attached
4	Land Subject to Inundation	6	BH090	Non-Perennial/Intermittent/Fluctuating

The table with virtual fields for polygon attribute transfer (step 3) is shown above. The table after virtual fields were made permanent, duplicate and selected other records removed, and a union operation with the original line table is below.





CODEDESC	HYC	F_CODE	HYCDESC
Inland Water	6	BH000	Non-Perennial/Intermittent/Fluctuating
Land Subject to Inundatior	8	BH090	Perennial/Permanent
Land Subject to Inundatior	6	BH090	Non-Perennial/Intermittent/Fluctuating
No Attribute	-1	NoAtt	No Attribute
River/Stream		BH140	

Resulting DataTip for polygon with only lines displayed is shown above.

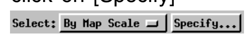
Map Scale Visibility Control

Map scale control specifies the map scale at which an element, layer, or group is visible. Map scale control for elements is set using Select: By Map Scale in the Vector Layer Controls. Choosing this option sets up a scale selection table that can subsequently be edited like any other database table. Map scale control by element is generally used for fairly dense vectors that contain a few elements that are suitable for display at full view and many elements that are not. Elements are mostly only suitable for display at full view if they can be distinguished from one another. Other elements should be turned off until they are distinguishable. Map scale control for layers and groups is set as part of a layout—it has no effect on display of the object by itself.

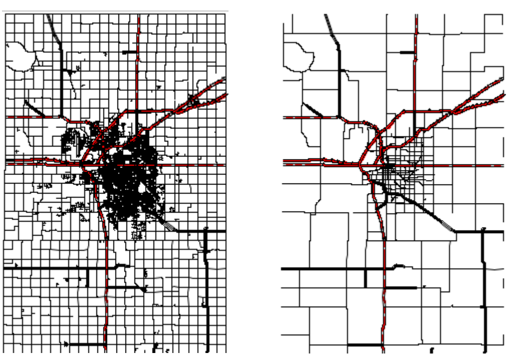
Map scale control can be used to eliminate clutter by showing only more important elements at smaller map scales or to replace coarser imagery with higher resolution imagery as you zoom in. You can control when an element will appear as you zoom in and whether it remains visible for all larger map scales or disappears when zoomed in beyond a specified scale.

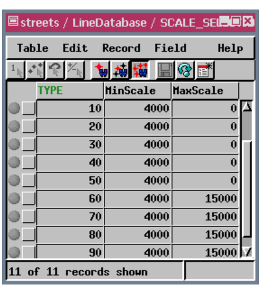
STEPS

- ☑ determine which table.field is suitable to supply map scale control (some hierarchy, such as major/minor roads or perennial/intermittent water features, is needed)
- ☑ click on the Vector  icon to open the Vector Layer Controls, choose the appropriate element tabbed panel, and set the Select option menu to By Map Scale, then click on [Specify]



- ☑ click on Attribute and choose the table.field you want to use
- ☑ enter minimum and maximum values as desired
- ☑ any subsequent editing of these minimum and maximum values should be done directly in the scale selection table





TYPE	MinScale	MaxScale
10	4000	0
20	4000	0
30	4000	0
40	4000	0
50	4000	0
60	4000	15000
70	4000	15000
80	4000	15000
90	4000	15000

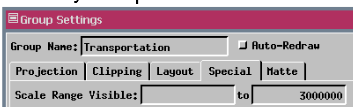
The roads in Lancaster County, NE are shown at the left with no map scale control and in the center with map scale control used to display only major roads until zoomed in beyond 1:15,000. This scale control by element was set up as part of a layout with high resolution imagery in the background, so all roads are turned off when zoomed in beyond 1:4,000. A value of zero in the MaxScale column means that the element type is displayed at all smaller map scales (larger scale values) than the MinScale value. A zero in the MinScale column means the element type does not go away when you zoom in further.

Map Scale Control in Layouts

STEPS

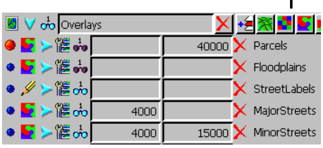
☑ set up your complex layout with massive data

☑ click on the Group Settings icon for any group whose visibility you want to control by map scale, and enter appropriate map scales



☑ turn on Options/Show Scale Ranges in the Layout Controls window

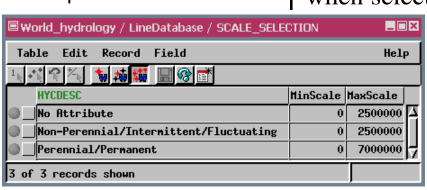
☑ expand the group with the layers you want to have map scale control, and enter the values



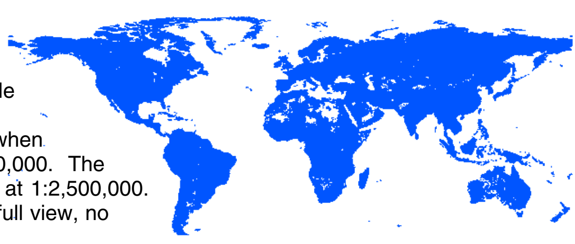
object appears when zoom in past right map scale, disappears when zoom in past left map scale

Vector layers that have no apparent hierarchy, such as soil polygons, can not make good use of map scale control by element. Such objects should have map scale control set for the layer as part of a group or layout. Another approach for complex layouts with multiple groups is to assign map scale control to individual groups.

One sample, complex layout of a massive global data set contains 57 layers organized into 11 groups. All but two of these layers are either hidden or off by map scale at full view. The hydrology layer illustrated below is one of these 57 layers. Most of the map scale control in this layout is at the group level, which is set on the Special panel of the Group Settings window. In order to set map scale control for individual layers, you need to turn on the Show Scale Ranges toggle on the Options menu in the Group or Layout Controls window; you also have to show details for the group. With these two types of map scale control, all elements selected for display appear at the same map scale. These map scale settings are saved with the group or layout and do not affect display of any of the objects when selected individually. When map scale control is set by element, it does affect display of the individual object as long as selection is set to By Map Scale, which may mean that no elements draw at full view.



World hydrology shown right without map scale control (approximately 1.5 million lines). With map scale control, the perennial water features are set to appear when you zoom in beyond 1:7,000,000. The intermittent features appear at 1:2,500,000. When displayed by itself at full view, no elements are drawn.



Show DataTips, Not Layers

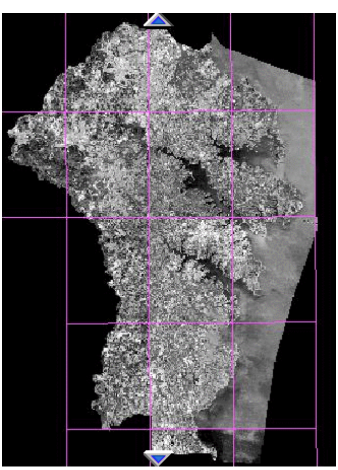
Many massive data sets have so many layers that displaying them all at one time would produce nothing but confusion. TNTmips provides two features in addition to map scale control to help you manage such chaos. You can hide layers and/or groups so that the viewer has to turn them on individually and should, therefore, know to turn some off when the view becomes too cluttered. You can also make groups or layers mutually exclusive so that when one is turned on, the others are turned off. You can have up to eight mutually exclusive group sets and any number of layers within the same group can be mutually exclusive.

A group or layer does not need to be visible to provide information to the viewer—DataTips can provide information for all layers whether visible or not. DataTip viewing options include none, active or top layer only, all visible layers, and all layers. The all layers option is recommended for massive data sets with hidden or mutually exclusive groups/layers so the viewer will know what information is available even if not currently displayed. All layers is now the default for DataTip display in TNTatlas.

STEPS

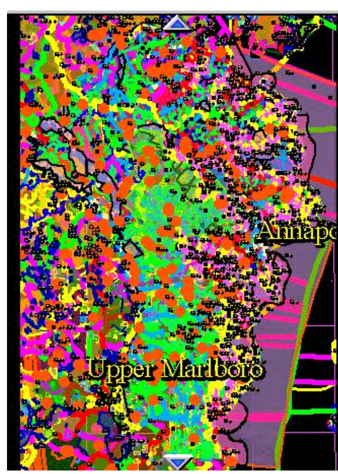
- ☑ use the Show/Hide icons to turn off some layers
- ☑ note the *Mutually exclusive group set* option menu and the *Layers are mutually exclusive* toggle on the Special tabbed panel in the Group Settings window

Only the 1-ft. resolution imagery is displayed when the atlas immediately below is viewed at full resolution (1X), but all of the information is available by DataTip.



← a simple atlas front end showing a SPOT mosaic of a county and map quad grid lines

the same front end with all layers on →



Use Map Scale Control for On-the-Fly Labels

STEPS

☑ set up on-the-fly labels in Spatial Data Display as one of the last steps in layout preparation because other layers contribute greatly to the legibility of labels

☑ click on the Vector icon for the layer you want to label 

☑ click on the tab for the element type you want to label, set the label to by attribute or by script, and specify the attribute or enter the script

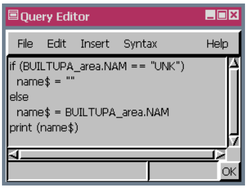
☑ specify the desired scale range, click on Label Style and specify font, size, and so on

☑ display the layout at a scale where the labels are visible and determine if the size is appropriate

☑ check that the labels are nearly legible when they become visible

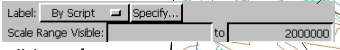
☑ make any necessary adjustments

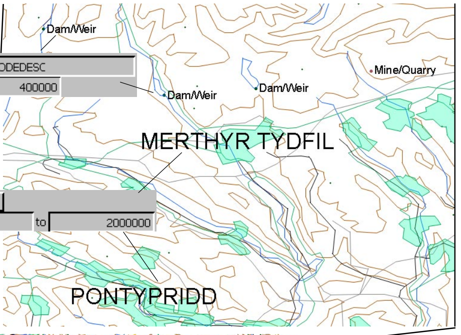
Note: Any data set, particularly large data sets, may have missing information. Rather than leaving a field blank, it often says *Unknown* or *UNK* or *no entry*. You generally do not want to see these entries as labels, so use a script for labeling and eliminate them.

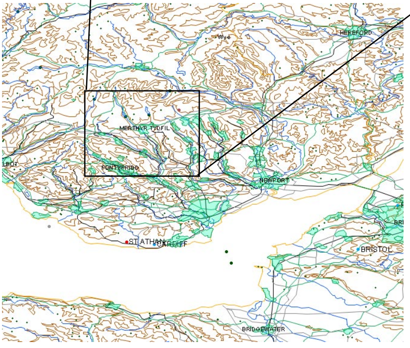


The TNT products offer two kinds of element labels: label elements, which are created in the Spatial Data Editor and add to the storage requirements of the vector, and on-the-fly labels, which are set up in Spatial Data Display and are not stored with the vector. The fact that on-the-fly labels are not stored with the vector does not mean that you have to set them up every time the object is displayed; this information is saved in the vector's display parameters for use in subsequent displays. It is also saved as part of any layout the vector objects are in so that changing this parameter when displayed alone will not affect how the labels display in the saved layout.

Map scale control is very important for labels in massive data sets. You do not want to draw labels until they would be legible or nearly legible.







The illustrations show on-the-fly labels that are visible at different map scales. There are no labels drawn at full view in this layout. The lower view shows labels that appear when you zoom in far enough for the underlying image to no longer draw. The upper view shows additional labels that appear when you zoom in again.

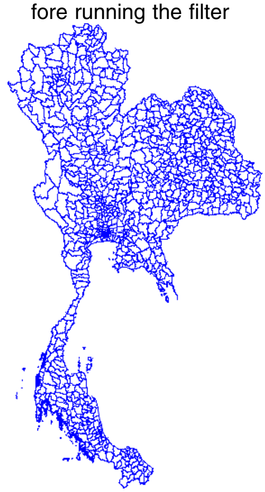
Thinning Vector Lines

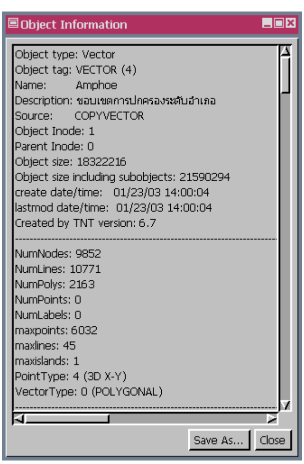
The decision on whether to thin, or simplify, vector lines depends on the origin of the data and its intended use. If all vertices are surveyed points and the vector will be used to resolve legal boundary issues, the lines should not be thinned. If the vector will be used for a printed map, thinning is appropriate. If the vector was created by raster to vector conversion of a scanned map, thinning is generally appropriate. Thinning is an empirical process that generally requires several iterations before you are satisfied with the result.

Although the vector object chosen for illustration here does not fall under our definition of massive, it illustrates the point well. This vector also had excess nodes and polygons less than 120,000 m^2 removed. This latter modification used the sliver polygons filter and was needed to divide the complete vector object into four TNTlite-sized districts. The polygons removed were tiny islands off the coast of the southern district.

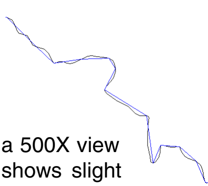
STEPS

- ☑ choose Edit/Spatial Data or Process/Vector/Filter, and select your vector
- ☑ choose the Line Simplification filter
- ☑ choose a method and a distance and a thinning ratio if necessary
- ☑ run a test and zoom up on the results; modify the values if needed before running the filter

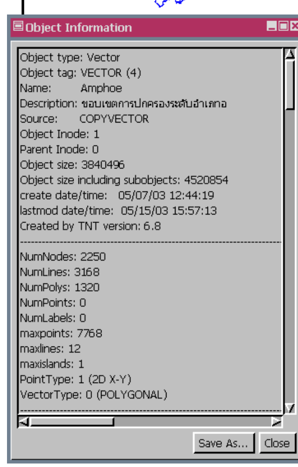




Object size reduced by 80% with nearly an 80% reduction in the number of nodes and a 70% reduction in lines while retaining 60% of polygons.



a 500X view shows slight differences in line detail



Advanced Software for Geospatial Analysis

MASSIVE DATA

MicroImages, Inc. publishes a complete line of professional software for advanced geospatial data visualization, analysis, and publishing. Contact us or visit our web site for detailed product information.

TNTmips TNTmips is a professional system for fully integrated GIS, image analysis, CAD, TIN, desktop cartography, and geospatial database management.

TNTedit TNTedit provides interactive tools to create, georeference, and edit vector, image, CAD, TIN, and relational database project materials in a wide variety of formats.

TNTview TNTview has the same powerful display features as TNTmips and is perfect for those who do not need the technical processing and preparation features of TNTmips.

TNTatlas TNTatlas lets you publish and distribute your spatial project materials on CD-ROM at low cost. TNTatlas CDs can be used on any popular computing platform.

TNTserver TNTserver lets you publish TNTatlases on the Internet or on your intranet. Navigate through geodata atlases with your web browser and the HTML TNTclient or the TNTclient Java applet.

TNTlite TNTlite is a free version of TNTmips for students and professionals with small projects. You can download TNTlite from MicroImages' web site, or you can order TNTlite on CD-ROM.

Index

MicroImages, Inc.

email: info@microimages.com
Internet: www.microimages.com