Ladder versus star: Comparing two approaches for generalizing hydrologic flowline data across multiple scales

Kevin Ross
kevin.ross@psu.edu

Paper for *Seminar in Cartography: Multiscale Hydrography*
GEOG 512, taught by Dr. Cynthia A. Brewer, cbrewer@psu.edu
Department of Geography, Pennsylvania State University
December 16, 2009

1 INTRODUCTION & BACKGROUND

Generalization of spatial data, such as the National Hydrography Dataset (NHD), is an important process to increase the usability and utility of the data and any maps that include the data. Generalizing the data helps to improve them for use at various scales and levels of detail, from high resolution and complex detail to low resolution and simple detail. There are various techniques of generalization, which can be used for different types of data and different objectives (Regnauld & McMaster, 2007). There are also different methods for applying these techniques in order to improve a dataset for a range of scales. The broad question explored in this paper is whether it is better to apply generalization techniques in an incremental fashion, where the dataset at each scale is a generalization of the previous, larger scale dataset (ladder approach), or whether it is better to derive the dataset at each scale by applying the generalization techniques to the original dataset (star approach) (Stoter, 2005).

In both the ladder and the star approaches to generalization, the process begins with the largest scale at which the data are available. For example, the largest scale at which the NHD is widely available is 1:24,000. For various reasons, an analyst may need to use these data at a range of different scales, such as 1:50,000, 1:100,000, 1:250,000, etc. The original 1:24,000 scale data would not necessarily be suitable for these different tasks, which is where the need for generalization comes in. Generalizing the original data for use at smaller scales can be beneficial for several reasons, including decreased storage requirements, increased application performance, improved cartographic representations, and enhanced perceptibility of the information for map readers (Bertin, 1983; Cecconi & Galanda, 2002; Sarjakoski, 2007).

In the ladder approach for generalizing the original dataset, each of the smaller scales is derived from the next largest scale. For example, the 1:50,000 scale dataset would be a generalization of the original 1:24,000 scale dataset. The 1:100,000 scale dataset would then be a generalization of the 1:50,000 scale dataset. This incremental process would continue, stepping through each scale level like the rungs of a ladder.

The star approach is a different way of deriving datasets for these same smaller scales from the original, large scale dataset. As opposed to the incremental technique of the ladder approach, in the star approach each different scale is generalized directly from the original dataset. The 1:50,000 scale dataset is a generalization of the 1:24,000 dataset. The 1:100,000 scale dataset is also a generalization of the 1:24,000 dataset, and so on. Figure 1 shows an illustration of the ladder and star approaches to generalization.

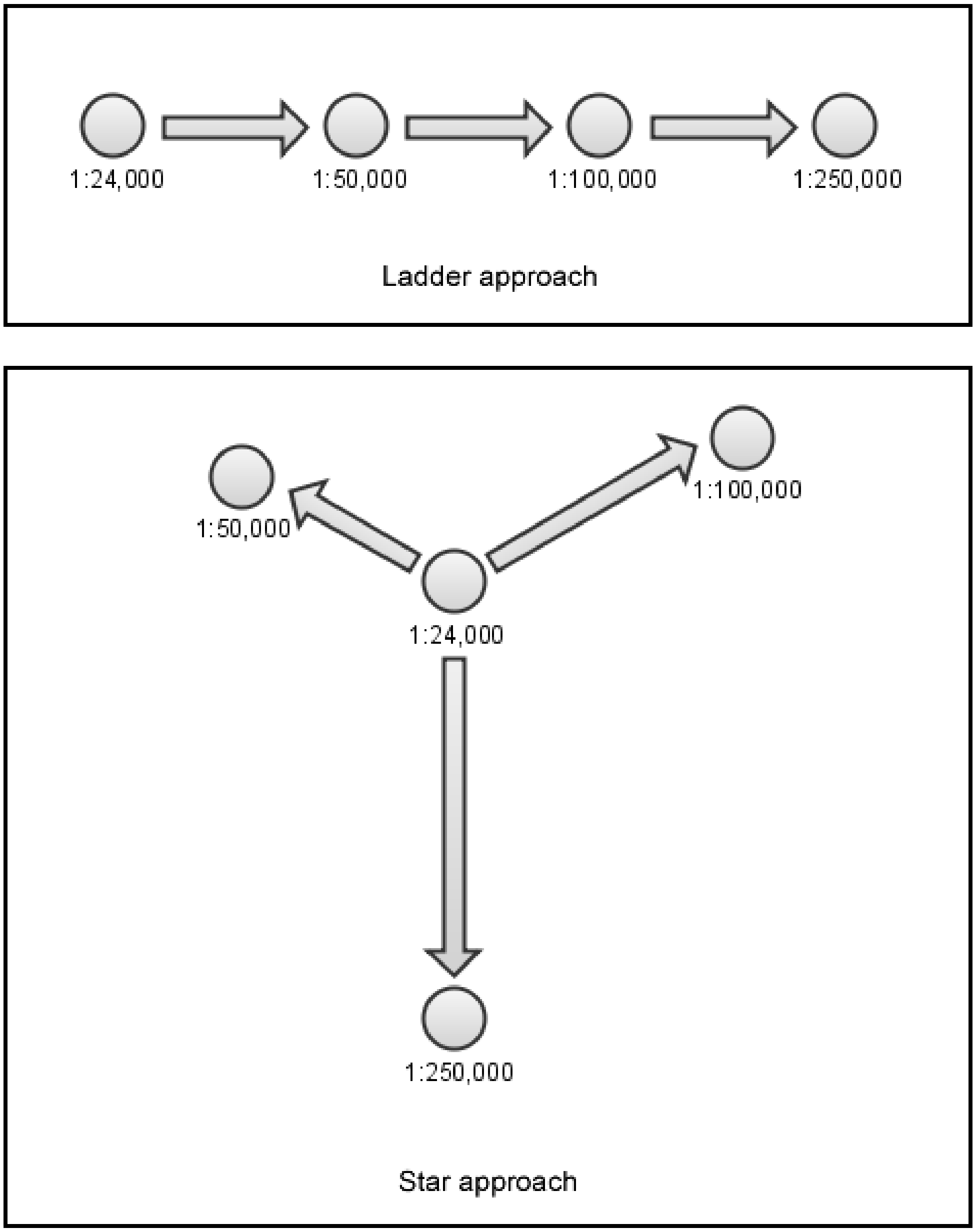

Figure 1. Illustrations showing the ladder and star approaches for generalization

Given these two different approaches for generalizing geographic data for use at various scales, a decision has to be made about which one to use (or whether to use a mixture of the two). Stoter (2005) discusses this decision in her review of how various National Mapping Agencies (NMAs) handle generalization of geographic data. Different NMAs have chosen different approaches (ladder, star, or mixed), but no justification for how this decision was reached has been given in this literature. There seems to be a general lack of research about the actual differences between the ladder and star approaches to generalization. Without knowing how these the results of these two approaches differ, it is impossible to make an informed decision about which one to use.

As the decision of whether to use the ladder or star approach is already being made in large generalization projects, it is important to understand more about them. It is possible that one approach is appropriate for some tasks while the other approach is more appropriate for different tasks. For example, one approach might be faster and overall more efficient and the other might produce results that are more visually appealing or accurate. It is also possible that one approach is simply superior to the other and should be used in most or all instances. A third possibility is that there is little or no noticeable difference between the two approaches. Evaluating ladder and star generalization techniques will provide insight about their differences, if any, and help with deciding which to use in future generalization projects.

The rest of this paper will discuss a study that seeks to compare the results of the ladder approach to those of the star approach for generalizing hydrographic flowline data. This study is intended as a first step toward understanding the differences between these two approaches to generalization. It is limited in scope to reviewing two possible generalization methods: Douglas-Peucker and bend simplify, for line simplification. This is obviously only a small subset of all of the possible generalization operators. The goal is to provide some preliminary results on the subject, to learn about some differences between the approaches when performing line simplification, and to provide some guidance for the future research that will be necessary to develop a full understanding of the ladder vs star concept. Section 2 of this paper describes the methods used, section 3 provides an overview of the results and section 4 wraps up with the key lessons and potential directions for future research.

2 METHODS

In order to evaluate and compare the ladder and star approaches, the first step is to choose and obtain the master dataset(s). The original source data will be the same for each approach so that direct comparisons can be made about the process and the outcomes. It is not feasible to perform an evaluation of all possible generalization techniques for this research project, so the focus of this paper is narrowed to generalization of polyline hydrographic data. Various line simplification techniques will be performed on the data using both the ladder and star approaches, while keeping all other variables consistent.

The data include a sample of high resolution flowline data from the NHD, which are intended for use at 1:24,000 scale (NHD Data Availability, 2009). To get a reasonably representative sample of NHD data, five different subregions in various areas of the United States were chosen. Specifically, the data used for this study are flowline data from subregions 0207 in West Virginia, 0311 in Florida and Georgia, 1029 in Kansas and Missouri, 1112 in Texas, and 1405 in Colorado. The data were selected from these regions of the country to get a good sample from various climates and terrain, which have significant effects on the characteristics of the hydrography.

The first generalization technique evaluated is a basic line simplification. The simplification method used here is the Douglas-Peucker algorithm and it is a commonly used generalization technique (Douglas & Peucker, 1973; McMaster & Shea, 1992). The Douglas-Peucker line simplification algorithm was applied to the flowline data for each of the subregions, starting with a generalization from 1:24,000 to 1:50,000.

Another generalization technique for simplifying line features is the bend simplify algorithm. Developed by Wang and Lee (1999, cited in ESRI, 2000), the bend simplify algorithm detects and removes extraneous bends from lines. The goal with this algorithm is to reduce the original line data while maintaining the main shape of the features. ESRI (2000) claims that bend simplify preserves the main shape of line features better than the Douglas-Peucker algorithm. Dal Santo and colleagues (Dal Santo, Wosny, & de Oliveria, 2008) evaluated the bend simplify and Douglas-Peucker algorithms for line simplification and found that bend simplify does result in topology more similar to the original data, but Douglas-Peucker reduces the data volume to a greater degree. Bend simplify was applied to the original datasets in the same way as the Douglas-Peucker algorithm, starting with generalizing the 1:24,000 dataset to 1:50,000.

When performing line simplification a tolerance threshold value is required, which is used by the algorithms in determining the maximum allowable distance between points and lines. To decide what values to use for the tolerance threshold, there are two factors to consider: (1) the size of the smallest detectable feature and (2) the spatial resolution. The size of the smallest detectable feature is equal to the denominator of the scale level divided by 1000, while the spatial resolution is half of that (Tobler, 1987). For instance, at a scale of 1:50,000, the smallest detectable feature size is 50 meters and the spatial resolution is 25 meters. I used the range of the spatial resolution to the smallest detectable feature size to determine appropriate threshold values. The values that produced good results without being overly simplified are 25 meters, 60 meters, and 180 meters for scales of 1:50,000, 1:100,000 and 1:250,000 respectively. The key is that these parameters were kept consistent through all of the procedures so that the only variable was the approach used (ladder versus star).

After performing all of the generalization processes on all of the datasets, the next steps were to evaluate and compare the generalized datasets from each approach. The first important factor to consider is the visual representation of the generalized data. Frequently the goal of generalization is to improve the appearance of the data when viewed on a map. If the generalized data look different when derived through the ladder approach compared to data generalized through the star approach, that is very important to keep in mind when deciding which approach to use. To make this comparison I viewed each of the derived datasets in ArcGIS; layering the data derived through the ladder approach on top of the data from the star approach and vice versa.

Another comparison to be made is on the efficiency of each approach. In a large generalization project, or in a project where generalizations are done in a time critical manner, it is important to consider how long the overall generalization process will take. It is possible that one approach might be faster than the other. While performing each generalization process, logs were recorded for each step, including the amount of time taken to process the data. With these data it is easy to determine which approach is faster.

The complexity of the resulting generalized data is also an important factor. If a lot of additional analyses are going to be performed on the generalized data, the efficiency of those analyses often depend on how complex the data are. If the goal is to increase efficiency, it is better to have less complex data. For this reason, I compared the number of vertices in the generalized data from the ladder approach to those from the star approach. The more vertices there are in the data, the more complex the data are. I also calculated and compared the total length of the flowlines in all of the generalized data, to determine if one approach produced longer or more numerous line segments than the other approach.

The final comparison made between the ladder and star approaches was on the file size of the resulting data. For projects involving immense amounts of data, in situations where storage space is limited, or if efficient use of resources is of importance, the file size of the resulting data matters. The size of each generalized dataset was recorded and comparisons were made to see which approach produced smaller files.

3 RESULTS

The results from this study include the comparisons that were made between the resulting data from the ladder and the star approaches. The comparisons of the visual representation of the generalized data are presented in section 3.1. Section 3.2 includes the results detailing the amount of time taken for each generalization process. The data complexity comparisons are discussed in section 3.3. Finally, section 3.4 presents the results comparing the file sizes of the generalized datasets.

3.1 Visual Representation

When generalizing for the purpose of improved cartographic representation, it is crucial to know whether the data will look different based solely on whether the ladder or the star approach was used. After performing each generalization on each of the datasets, the resulting data from the ladder approach were compared directly to the resulting data from the star approach.

When using the Douglas-Peucker algorithm for line simplification, there was no difference in the visual representation of the generalized data whether the ladder approach or the star approach was used. This was true for all five of the datasets, and for all of the scale levels. It is a given that there will never be a difference between the two approaches in the first step of generalization (1:24,000 to 1:50,000). However, it was unknown if there would be a difference in the following steps, so it is interesting to see that there is not. An example of the output from the ladder approach (Figure 2) and from the star approach (Figure 3) show how there is no difference.

To further illustrate these results and to some extent the accuracy of the parameters used for the simplification algorithms, the resulting data were viewed along with some reference data. The medium resolution (1:100,000 scale) data were retrieved directly from NHD in order to compare to the 1:100,000 scale data derived through my chosen methods. Figure 4 shows the results of the Douglas-Peucker line simplification along with the NHD medium resolution data. Elimination of line segments was not performed in this study, so disregarding all of the additional line segments, the flowlines that exist in both datasets can be compared. The data derived through my generalization methods are more simplified than the NHD medium resolution data, which is why I used tolerance threshold values on the lower end of the range (closer to the spatial resolution than the smallest detectable feature size).

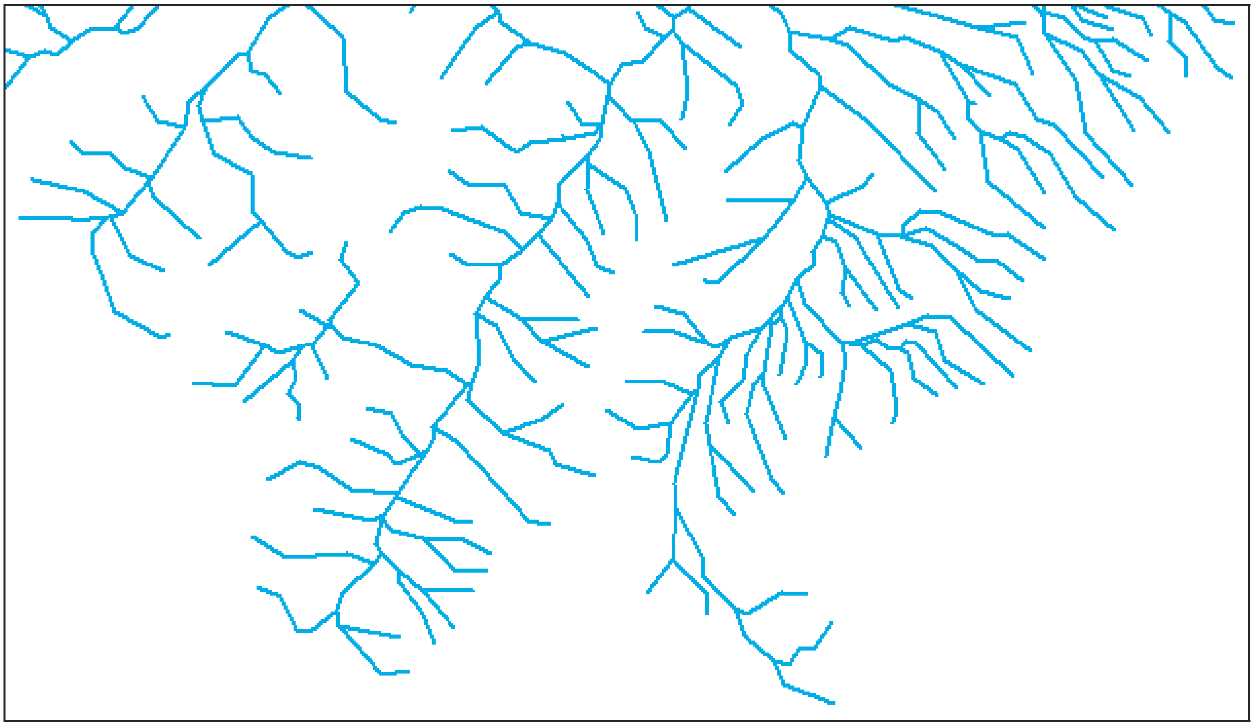

Figure 2. West Virginia flowline data generalized to 1:100,000 using the *ladder* approach. Shown at 1:100,000.

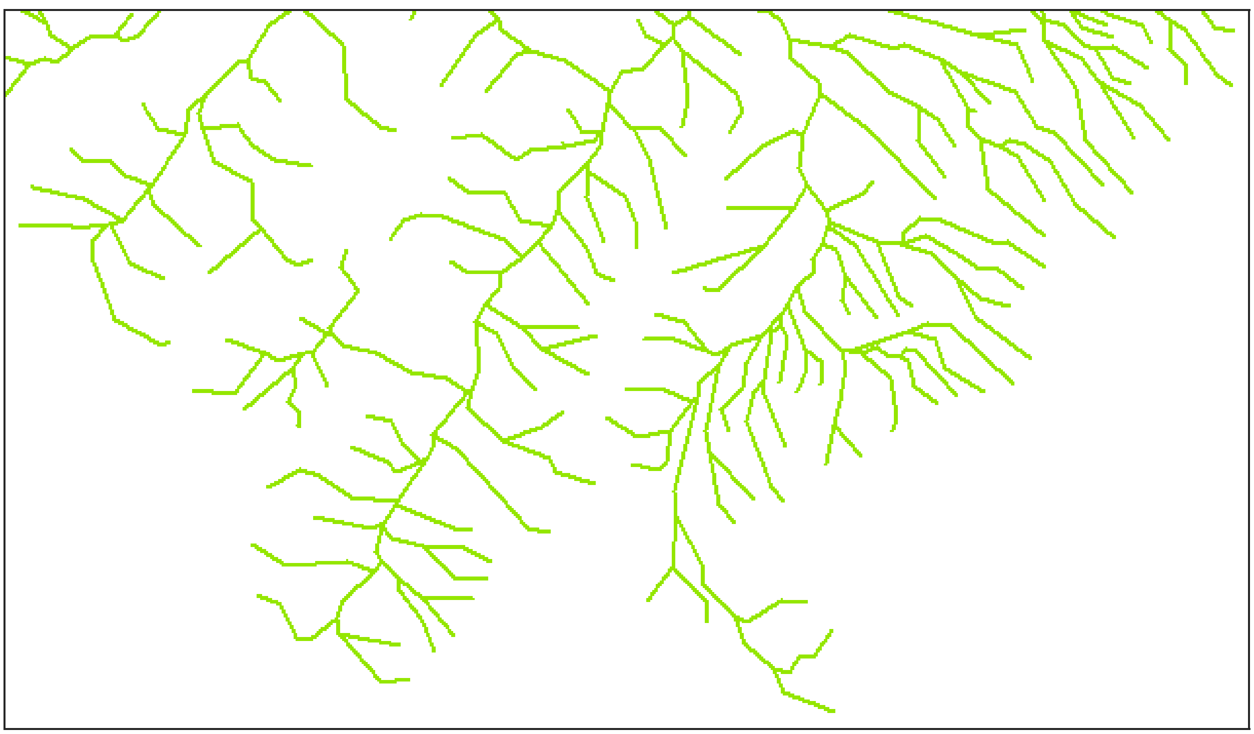

Figure 3. West Virginia flowline data generalized to 1:100,000 using the *star* approach. Shown at 1:100,000.

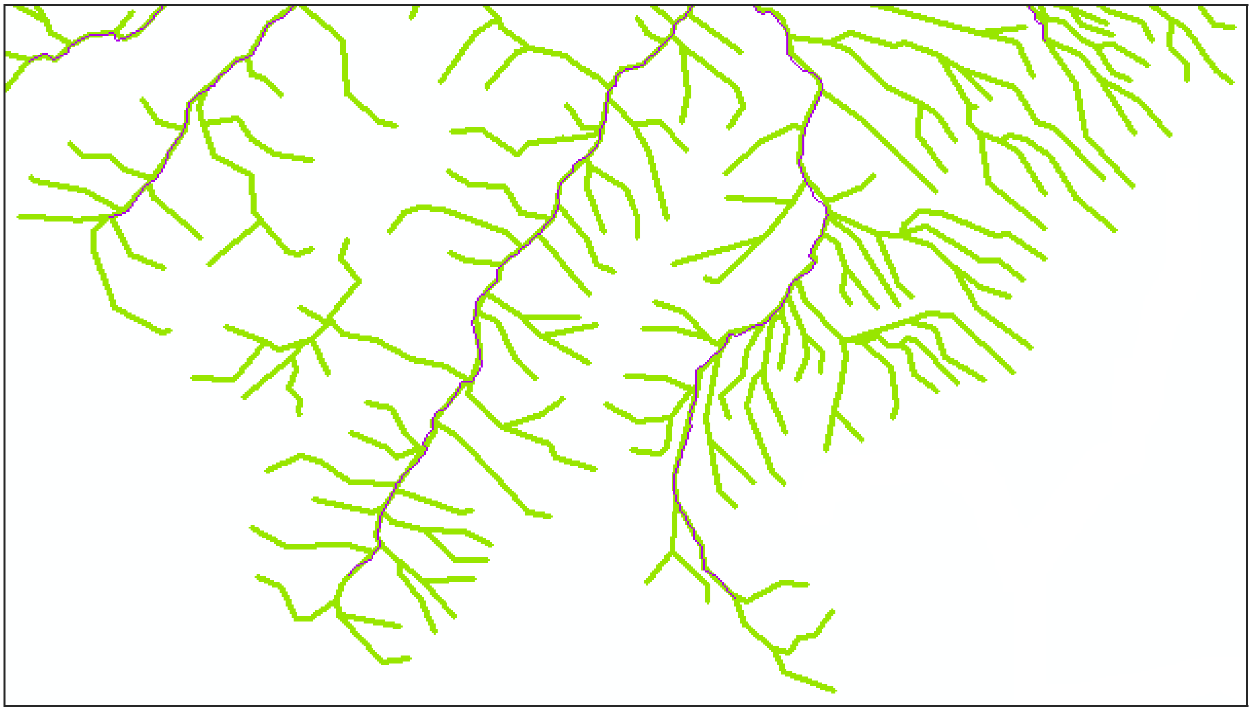

Figure 4. West Virginia flowline data generalized to 1:100,000 (green), along with original NHD medium resolution data (purple). Shown at 1:100,000.

The story is a little different when using the bend simplify algorithm. In all cases, there actually are slight differences in the visual representation of the data depending on whether the ladder approach or the star approach is used. Using the ladder approach produces results that have slightly more detailed representation, where there are more curves than there are in the results produced by the star approach. These differences are very subtle, but they seem to become more exaggerated as the data are generalized out to smaller and smaller scales.

Figure and Figure show how the resulting data look at their intended scale levels of 1:100,000 and 1:250,000 respectively. The differences are noticeable at their intended scales, but again they are quite subtle. To get an idea of what the actual difference is, these same data are shown at a scale of 1:24,000 in Figure 5 and Figure 6. This exaggerated view makes it clear that the data derived through the ladder approach are more detailed than the data derived through the star approach.

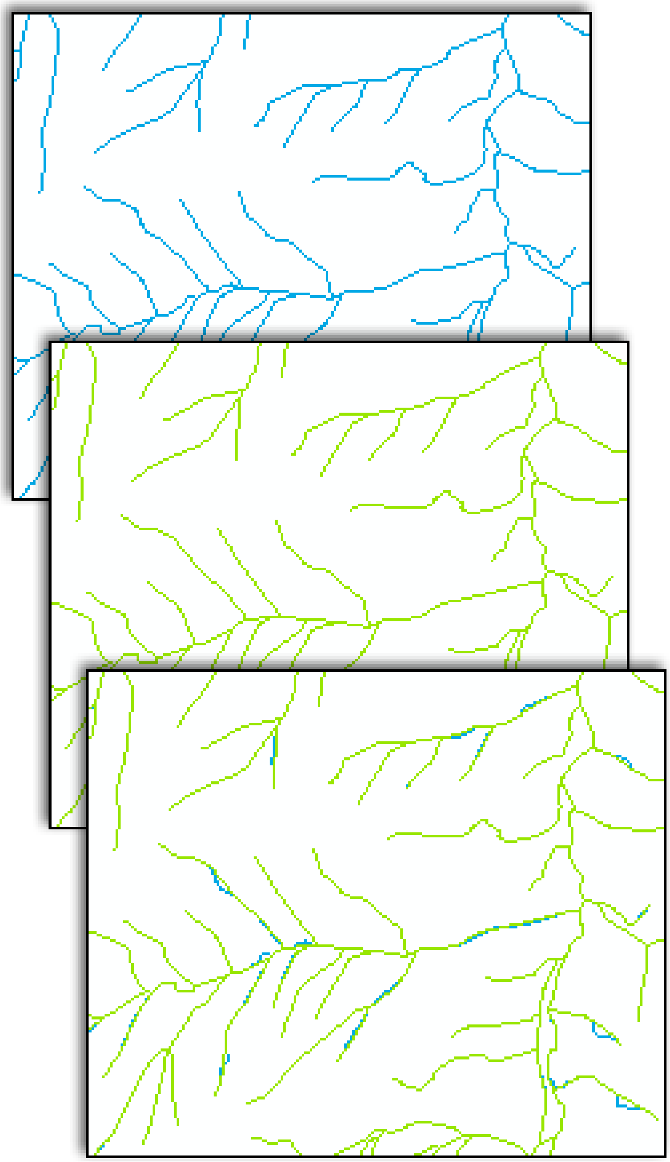

Figure 5. Colorado flowline data generalized to 1:100,000 scale using the ladder approach (blue) and the star approach (green). Shown at 1:100,000.

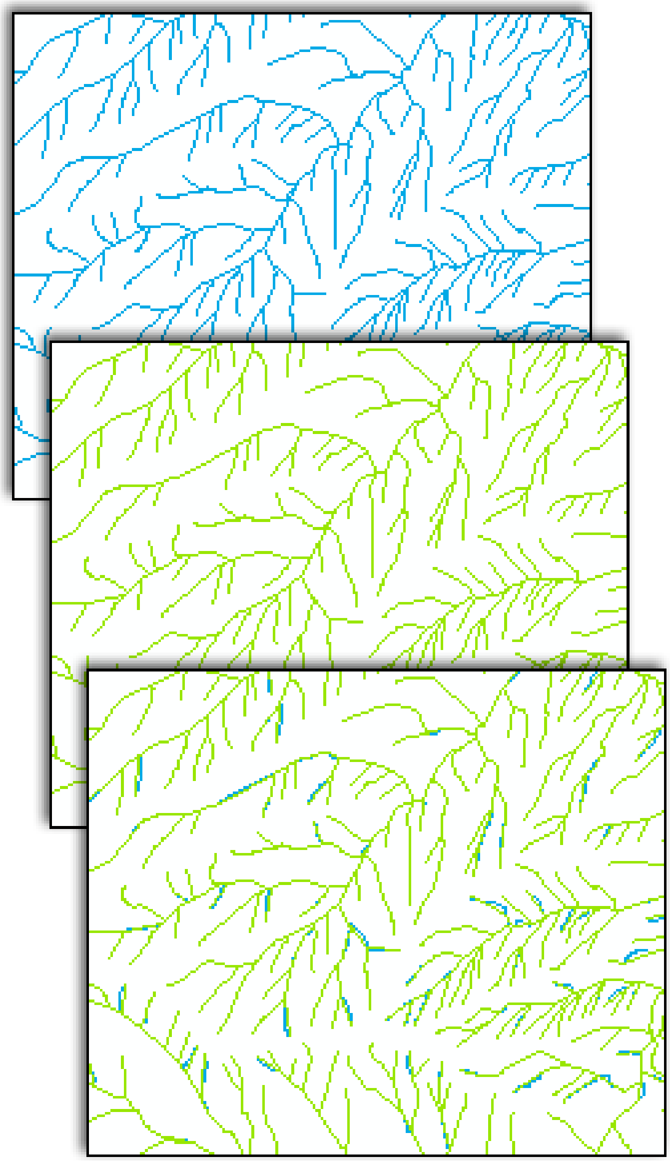

Figure 6. Colorado flowline data generalized to 1:250,000 scale using the ladder approach (blue) and the star approach (green). Shown at 1:250,000.

Finally, the data derived through each approach for use at 1:100,000 scale were compared to the original NHD medium resolution data. An exaggerated view of this comparison is shown in Figure 7. Being more detailed, the data derived from the ladder approach are a closer match to the NHD medium resolution than the data derived from the star approach.

There is also a noticeable difference between the amount of variation in the 1:100,000 scale data and in the 1:250,000 scale data. The variation is more pronounced in the 1:250,000 scale data, which is one step further in the overall generalization process. This may suggest that the more smaller-scale datasets that are derived, the larger the differences will be based on which approach is used.

Figure 5. Colorado flowline data generalized for use at 1:100,000 scale with the ladder approach (blue) and the star approach (green) layered on top. Shown at an exaggerated 1:24,000 scale.

Figure 6. Colorado flowline data generalized for use at 1:250,000 scale with the ladder approach (blue) and the star approach (green) layered on top. Shown at an exaggerated 1:24,000 scale.

Figure 7. Colorado flowline data generalized for use at 1:100,000 scale with the ladder approach (blue) and the star approach (green), along with the original NHD medium resolution data (purple). Shown at an exaggerated 1:24,000 scale.

3.2 Processing Time

The second comparison I made between the end results of the ladder approach and the end results of the star approach was based on the amount of time the overall process took to complete. Overall, the ladder approach is faster than the star approach for performing line generalization using both the Douglas-Peucker algorithm and the bend simplify algorithm. This is true for all five of the sample flowline datasets.

In running the Douglas-Peucker simplification algorithm, the ladder approach was faster than the star approach in every instance. The relevant numbers are all shown in Table 1. The first step of the process (generalizing the source 1:24,000 scale data to 1:50,000 scale) is ignored since it is exactly the same for each approach. The time taken for the following steps are shown for each dataset, along with the difference in seconds between the ladder approach and star approach, and finally the difference in time as a percentage. The difference is then also calculated for the overall process for each dataset, since that is when it matters which approach was used; these percentages are shown in bold for each of the five datasets. The final result is that on average, the ladder approach is 19.43 percent faster than the star approach when performing line simplification using the Douglas-Peucker algorithm. The results from the bend simplify algorithm were very similar and are shown in Table 2. On average, the ladder approach is 18.86 percent faster than the star approach when performing line simplification using the bend simplify algorithm.

Table 1. The amount of time, in seconds, taken to perform the Douglas-Peucker line simplification algorithm for each step of the process.

Time (in seconds) to complete the Douglas-Peucker line simplification				
Layer	Ladder	Star	Difference	**Percent Difference**
CO-100k	15	18	3	16.67
CO-250k	14	17	3	17.65
CO	29	35	6	**17.14**
FL-100k	49	63	14	22.22
FL-250k	48	59	11	18.64
FL	97	122	25	**20.49**
MO-100k	20	24	4	16.67
MO-250k	19	24	5	20.83
MO	39	48	9	**18.75**
TX-100k	32	44	12	27.27
TX-250k	31	40	9	22.50
TX	63	84	21	**25.00**
WV-100k	32	39	7	17.95
WV-250k	32	37	5	13.51
WV	64	76	12	**15.79**
Average difference:				**19.43%**

Table 2. The amount of time, in seconds, taken to perform the bend simplify line simplification algorithm for each step of the process.

Time (in seconds) to complete the bend simplify line simplification				
Layer	Ladder	Star	Difference	**Percent Difference**
CO-100k	19	23	4	17.39
CO-250k	16	24	8	33.33
CO	35	47	12	**25.53**
FL-100k	54	64	10	15.63
FL-250k	52	66	14	21.21
FL	106	130	24	**18.46**
MO-100k	25	27	2	7.41
MO-250k	23	26	3	11.54
MO	48	53	5	**9.43**
TX-100k	39	50	11	22.00
TX-250k	35	51	16	31.37
TX	74	101	27	**26.73**
WV-100k	43	47	4	8.51
WV-250k	36	45	9	20.00
WV	79	92	13	**14.13**
Average difference:				**18.86%**

3.3 Data Complexity

The next evaluation is of the complexity of the resulting data after performing the line simplification processes using the ladder and star approaches. To measure the complexity, the number of vertices in the resulting flowline data were counted and the total length of the flowlines were calculated. Unlike the processing time, the results of the complexity measurement differed between the two simplification algorithms.

With the Douglas-Peucker algorithm, there was hardly any difference in the number of vertices between the resulting data of the ladder approach and that of the star approach. This is not surprising, since there was no difference whatsoever in the visual representation of these resulting data. When the cartographic display is the same we can expect that the number of vertices will also be the same. The actual results are shown in **Table 3** and we can see that there were actually some very slight differences in the vertex counts of some of the datasets. However, the biggest difference is 27 vertices, which out of a total of over 7000 vertices is insignificant. Overall, the data derived from the star approach had, on average, 0.06 percent fewer vertices than the data derived from the ladder approach.

The bend simplify algorithm again produced some different results than the Douglas-Peucker algorithm. In this case there actually was some variation in the complexity of the resulting data, based on the approach used. The star approach consistently produced less complex data than the ladder approach. As shown in Table 4, the data derived from the star approach had 11.59 percent fewer vertices than the data derived from the ladder approach on average. These results are in line with the visual representation results, where the star approach produced less detailed data than the ladder approach.

Table 3. The number of vertices in the flowline data after performing Douglas-Peucker line simplification

Resulting vertex count after the Douglas-Peucker line simplification				
Layer	Ladder	Star	Difference	**Percent Difference**
CO-100k	9984	9984	0	0.00
CO-250k	7311	7284	27	0.37
CO	17295	17268	27	**0.16**
FL-100k	30746	30746	0	0.00
FL-250k	27734	27724	10	0.04
FL	58480	58470	10	**0.02**
MO-100k	15120	15120	0	0.00
MO-250k	11889	11887	2	0.02
MO	27009	27007	2	**0.01**
TX-100k	25212	25206	6	0.02
TX-250k	18124	18106	18	0.10
TX	43336	43312	24	**0.06**
WV-100k	23102	23102	0	0.00
WV-250k	17575	17560	15	0.09
WV	40677	40662	15	**0.04**
Average difference:				**0.06%**

Table 4. The number of vertices in the flowline data after performing bend simplify line simplification

Resulting vertex count after the bend simplify line simplification				
Layer	Ladder	Star	Difference	**Percent Difference**
CO-100k	55277	50300	4977	9.00
CO-250k	40172	33446	6726	16.74
CO	95449	83746	11703	**12.26**
FL-100k	69317	62742	6575	9.49
FL-250k	41428	38626	2802	6.76
FL	110745	101368	9377	**8.47**
MO-100k	52677	47014	5663	10.75
MO-250k	28654	24879	3775	13.17
MO	81331	71893	9438	**11.60**
TX-100k	92532	82026	10506	11.35
TX-250k	45030	37905	7125	15.82
TX	137562	119931	17631	**12.82**
WV-100k	114794	103787	11007	9.59
WV-250k	61302	52488	8814	14.38
WV	176096	156275	19821	**11.26**
Average difference:				**11.59%**

There was even less variation in the total length of the flowlines, as shown in Table 5 and Table 6. There was no difference at all when using Douglas-Peucker line simplification. When using bend simplify, the data derived from the star approach had 0.64 percent shorter flowline lengths than the data derived from the ladder approach. Resulting flowline length is clearly not an important factor to consider when performing either of these line simplification algorithms using either the ladder or star approach.

Table 5. The total flowline length after performing Douglas-Peucker line simplification

Resulting flowline length (KM) after the Douglas-Peucker line simplification				
Layer	Ladder	Star	Difference	**Percent Difference**
CO-100k	3096.33	3096.33	0	0.00
CO-250k	3022.41	3022.42	-0.01	0.00
CO	6118.74	6118.75	-0.01	**0.00**
FL-100k	4018.08	4018.08	0	0.00
FL-250k	3903.66	3903.57	0.09	0.00
FL	7921.74	7921.65	0.09	**0.00**
MO-100k	3253.77	3253.77	0	0.00
MO-250k	3152.39	3152.35	0.04	0.00
MO	6406.16	6406.12	0.04	**0.00**
TX-100k	4982.3	4982.3	0	0.00
TX-250k	4685.06	4685.12	-0.06	0.00
TX	9667.36	9667.42	-0.06	**0.00**
WV-100k	5852.05	5852.05		0.00
WV-250k	5683.63	5683.65	-0.02	0.00
WV	11535.68	11535.7	-0.02	**0.00**
Average difference:				**0%**

Table 6. The total flowline length after performing bend simplify line simplification

Resulting flowline length (KM) after the bend simplify line simplification				
Layer	Ladder	Star	Difference	**Percent Difference**
CO-100k	3130.68	3123.61	7.07	0.23
CO-250k	3078.07	3066.37	11.7	0.38
CO	6208.75	6189.98	18.77	**0.30**
FL-100k	4072.5	4053.38	19.12	0.47
FL-250k	3950.17	3932.15	18.02	0.46
FL	8022.67	7985.53	37.14	**0.46**
MO-100k	3314.23	3299.94	14.29	0.43
MO-250k	3227.97	3209.18	18.79	0.58
MO	6542.2	6509.12	33.08	**0.51**

***Table 6** continued*

Layer	Ladder	Star	Difference	**Percent Difference**
TX-100k	5012.65	4976.09	36.56	0.73
TX-250k	4782	4742.07	39.93	0.84
TX	9794.65	9718.16	76.49	**0.78**
WV-100k	5954.1	5936.11	17.99	0.30
WV-250k	5936.11	5791.58	144.53	2.43
WV	11890.21	11727.69	162.52	**1.37**
Average difference:				**0.64%**

3.4 Data Size

The final comparison made between the results of the ladder approach and the star approach is on the file size of the generalized data. The file size is the measure of disk space consumed by the resulting data, in megabytes. Like with the vertex counts, there was some difference in the resulting file size between the Douglas-Peucker algorithm and the bend simplify algorithm.

As expected after reviewing the visual representation and data complexity results, there is very little difference in the file size results between the two approaches when performing the Douglas-Peucker algorithm. The results are shown in Table 7. The star approach results in data that are 0.37 percent smaller on average than data produced by the ladder approach.

Table 7. The file size of the data after performing Douglas-Peucker line simplification

Resulting file size (MB) after the Douglas-Peucker line simplification				
Layer	Ladder	Star	Difference	**Percent Difference**
CO-100k	1.43	1.43	0	0.00
CO-250k	1.35	1.35	0	0.00
CO	2.78	2.78	0	**0.00**
FL-100k	5.33	5.33	0	0.00
FL-250k	5.24	5.24	0	0.00
FL	10.57	10.57	0	**0.00**
MO-100k	2.23	2.23	0	0.00
MO-250k	2.14	2.13	0.01	0.47
MO	4.37	4.36	0.01	**0.23**
TX-100k	3.54	3.54	0	0.00
TX-250k	3.33	3.22	0.11	3.30
TX	6.87	6.76	0.11	**1.60**
WV-100k	3.35	3.35	0	0.00
WV-250k	3.18	3.18	0	0.00
WV	6.53	6.53	0	**0.00**
Average difference:				**0.37%**

With the bend simplify algorithm, the results are again slightly different between the two approaches. Though the differences are subtle, the file sizes of data produced by the star approach are in all cases smaller than those of the data produced by the ladder approach. As shown in Table 8, the overall average data size is 5.01 percent smaller with the star approach.

Table 8. The file size of the data after performing bend simplify line simplification

Resulting file size (MB) after the bend simplify line simplification				
Layer	Ladder	Star	Difference	**Percent Difference**
CO-100k	2.82	2.66	0.16	5.67
CO-250k	2.02	1.85	0.17	8.42
CO	4.84	4.51	0.33	**6.82**
FL-100k	6.51	6.31	0.2	3.07
FL-250k	5.77	5.69	0.08	1.39
FL	12.28	12	0.28	**2.28**
MO-100k	3.38	3.21	0.17	5.03
MO-250k	2.7	2.58	0.12	4.44
MO	6.08	5.79	0.29	**4.77**
TX-100k	5.6	5.28	0.32	5.71
TX-250k	4.22	4	0.22	5.21
TX	9.82	9.28	0.54	**5.50**
WV-100k	6.15	5.81	0.34	5.53
WV-250k	4.58	4.31	0.27	5.90
WV	10.73	10.12	0.61	**5.68**
Average difference:				**5.01%**

4 CONCLUSIONS

There are a few key lessons to come away with after this study. Ultimately the goal is to fully understand the differences between the star and ladder approaches to generalization, in order to be able to make informed decisions about which approach to use. The results of this study are a step in that direction, specifically in performing line simplification. They also serve as evidence based motivation for more research on the topic.

The first finding is that the variation between the results of the two approaches differs depending on the generalization algorithm used. It is quite interesting that there is basically no difference in the resulting data when using the Douglas-Peucker algorithm, but there is variation when using the bend simplify algorithm. First of all, this shows that unless the processing time is important, the decision between using the ladder approach or the star approach is not very critical if the Douglas-Peucker algorithm is going to be used. It also shows that this type of testing needs to be performed for each generalization algorithm individually if one wishes to truly understand the differences between the two approaches; it is not possible to extrapolate these results to other processes and the results of the bend simplify process demonstrate that it is possible to have differences in the resulting generalized data based on the approach used.

Another interesting finding is that in results of the bend simplify process, the visual differences became more pronounced between the generalized data as more smaller-scale datasets were derived. This serves as motivation to conduct further research where far more steps are included to derive more smaller-scale datasets, in order to determine if there is in fact a trend where the amount of visual variation increases as the scale level decreases.

From the perspective of the efficiency of these two approaches, it was interesting to look at the results of processing time. First of all, there is an obvious lesson to be learned about performing the Douglas-Peucker algorithm. Given that there was basically no difference in the resulting data from the two approaches, the faster processing of the ladder approach makes it an easy choice for a generalization project where multiple smaller-scale datasets will be derived using the Douglas-Peucker algorithm. The ladder approach was similarly faster than the star approach when using bend simplify, so that should be kept in mind if speed is of utmost importance.

The other factors that may be important in planning a multiple scale generalization project, from the perspective of resource efficiency, are the data complexity and file size. These two facets of the resulting data are where the star approach has the potential to outperform the ladder approach. Though there was little difference when using Douglas-Peucker, the star approach consistently resulted in less complex, smaller data than the ladder approach when using bend simplify.

The main lesson learned from this study, as pointed to in various ways throughout this paper, is that further research is needed in order to come to a full understanding of the implications of choosing either the ladder or the star approach for generalization. It has been shown that there are, in fact, differences in the results depending on which approach is used. It is therefore an important decision when planning a multiple scale generalization project. Though many of the differences in this study are subtle, it is possible that in other generalization methods the differences will be more pronounced. Until this concept has been fully researched, when planning a multiple scale generalization project it is probably best to test each approach with the data and the generalization operators that will be used in the project in order to evaluate the differences, if any, and make an informed decision about whether to use ladder or star.

REFERENCES

Bertin, J. (1983). *Semiology of Graphics: Diagrams, Networks, Maps.* Madison, WI: University of Wisconsin Press.

Cecconi, A., & Galanda, M. (2002). Adaptive zooming in web cartography. *Computer Graphics Forum, 21* (4), 787-799.

Dal Santo, M. A., Wosny, G. C., & de Oliveria, F. H. (2008). Algorithms for automated line generalization in GIS. *Proceedings of the Twenty-Eighth Annual ESRI User Conference.* San Diego, CA: ESRI.

Douglas, D. H., & Peucker, T. K. (1973). Algorithms for the reduction of the number of points required to represent a digitized line or its caricature. *Cartographica: The International Journal for Geographic Information and Geovisualization, 10* (2), 112-122.

ESRI. (2000). *Map Generalization in GIS: Practical Solutions with WorkStation ArcInfo Software.* Redlands, CA: ESRI.

McMaster, R. B., & Shea, K. S. (1992). *Generalization in digital cartography.* Washington, DC: Association of American Geographers.

NHD Data Availability. (2009, March 10). Retrieved September 20, 2009, from National Hydrography Dataset: http://nhd.usgs.gov/data.html

Regnauld, N., & McMaster, R. B. (2007). A synoptic view of generalisation operators. In W. A. Mackaness, A. Ruas, & L. T. Sarjakoski (Eds.), *Generalisation of Geographic Information: Cartographic Modelling and Applications* (pp. 37-66). Elsevier Science.

Sarjakoski, T. L. (2007). Conceptual models of generalisation and multiple representation. In W. A. Mackaness, A. Ruas, & L. T. Sarjakoski (Eds.), *Generalisation of Geographic Information: Cartographic Modelling and Applications* (pp. 11-35). Elsevier Science.

Stoter, J. E. (2005). Generalisation: The gap between research and practice. *8th ICA WORKSHOP on Generalisation and Multiple Representation.* A Coruña, Spain.

Stoter, J. E. (2005). Generalization within NMA's in the 21st century. *International Cartographic Conference (ICC).* A Coruna, Spain.

Tobler, W. (1987). Measuring spatial resolution. *Proceedings, International Workshop on Geographical Information Systems*, (pp. 42-47). Beijing.

Wang, Z., & Lee, D. (2000). Building specification based on pattern recognition and shape analysis. *Proceedings of the Ninth Spatial Data Handling Symposium*, (pp. 58-72).