Pruning of Hydrographic Networks: A Comparison of Two Approaches

Larry Stanislawski, Center of Excellence for Geospatial Information Science (CEGIS), United States Geological Survey, Rolla Missouri

Sandro Savino, Department of Information Engineering, University of Padova
Outline

• Objectives for comparison of network pruning approaches
• Description of Two Pruning Approaches
  o Length and density pruning (LaDP)
  o Stratified pruning (SP)
• Brief overview of coefficient of line correspondence (CLC)
• Maine subbasin test results
• Results of tests on four Iowa subbasins
• Summary Statements
Objectives

• Compare stratified pruning results with the length and density pruning results to identify:
  o Strengths and weaknesses of each approach
  o Possible enhancements for each approach

• Hydrographic network pruning removes less prominent features until a desired number, density, or length is achieved. Must maintain
  o Network connectivity (topology)
  o Attribution rules (eg. full reach codes)
  o Natural density variations, but remove or diminish density variations from data compilation differences.
Length and Density Pruning (LaDP)

• **Enrichment**
  1. Detect sources of the network and, following the flow direction downhill, compute length to furthermost source, number of branches uphill, number of edges uphill, Strahler order
  2. Build river courses using attributes and enriched data
  3. Compute density: build a buffer around each river course and calculate the ratio total area of the buffer overlapped with other buffers / buffer area

• **Pruning**
  1. Remove dangling river courses shorter than minimum length
  2. Analyze all the river courses with density larger than a threshold: if deemed not relevant (on the basis of the semantic data and the enriched data) remove one, update the density of the neighbors and repeat
Length and Density Pruning (LaDP)

River course reconstruction
Deletion of short river courses
Calculation of density (buffer overlap)
Deletion of river courses with high density
Stratified Pruning (SP)

- **Enrichment**
  1. Catchment area estimate for each feature
  2. Prominence: Upstream drainage area assigned to each feature in network
  3. Partition features into line-density classes (stratification)

- **Pruning**
  1. Determine target density for each density partition (Radical Law, or match existing data)
  2. Separately prune each partition until target density achieved (prune reaches by upstream drainage area)
  3. Return features needed for connectivity between partitions.
  4. Remove dangling tributaries shorter than minimum length
  5. Adjust target densities, and re-prune
Stratified Pruning (SP)

1. Catchments

2. Upstream Drainage Area

3. Line density partitions

Enrichment

Pruning

1. Target Density

Radical Law

\[ \sqrt{\frac{RF_{source}}{RF_{target}}} \]

or

Existing Data

2. Prune

3. Remove Short Dangles

(Adjust as needed)

Average Density
(Kilometers per Square Kilometer)

- 0.621
- 1.325
- 2.421
- 4.245

14th Workshop of the ICA Commission on Generalization and Multiple Representation, Paris, France, 30 June to 1 July, 2011
Comparison to Benchmark: coefficient of line correspondence (CLC)

Commission Errors

• Buffer around benchmark lines

• Buffer is two times US National Map Accuracy Standards at scale of generalized dataset and scale of benchmark dataset.

• Buffer for pruned 24K lines compared to 100K benchmark: 126 meters

Note: This example is not Maine data.
Comparison to Benchmark: coefficient of line correspondence (CLC)

Omission Errors

Buffer around target (pruned) lines

Omission error where > 50% confluence-to-confluence feature outside of buffer

Note: This example is not Maine data.
Coefficient of Line Correspondence (CLC)

\[ \text{CLC} = \frac{M}{(O+C+M)} \]

- \( M \) = sum of length of matching features from benchmark (1:100,000-scale) dataset
- \( O \) = sum of length of omission error features from the benchmark (1:100,000-scale) dataset
- \( C \) = sum of length of commission error features from the pruned (high-resolution NHD) line dataset *

Proportion Commission Errors = \( \frac{C}{(O+C+M)} \)

Proportion Omission Errors = \( \frac{O}{(O+C+M)} \)

* Commission lengths are divided by the benchmark-to-LoD length expansion factor to compensate for higher granularity in LoD representations.
Maine Subbasin (Compilation Issues)

1:24000-scale (24K) NHD

Average Density (Kilometers per Square Kilometer)
- 0.621
- 1.406
- 2.421
- 4.245

14th Workshop of the ICA Commission on Generalization and Multiple Representation, Paris, France, 30 June to 1 July, 2011
Maine Results: Pruning to 100K
(Compilation Issues)

Stratified Pruning (SP)

Length and Density Pruning (LaDP)

100K NHD Benchmark

14th Workshop of the ICA Commission on Generalization and Multiple Representation, Paris, France, 30 June to 1 July, 2011
### Maine Results: CLC
(Compilation Issues)

<table>
<thead>
<tr>
<th>Pruning method</th>
<th>Weighted CLC</th>
<th>Weighted omissions</th>
<th>Weighted commissions</th>
</tr>
</thead>
<tbody>
<tr>
<td>Stratified</td>
<td>0.779</td>
<td>0.080</td>
<td>0.141</td>
</tr>
<tr>
<td>Length and Density</td>
<td>0.775</td>
<td>0.130</td>
<td>0.095</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Pruning method</th>
<th>Bootstrapped 90 Percent Confidence Interval</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Lower bound</td>
</tr>
<tr>
<td>Stratified</td>
<td>0.763</td>
</tr>
<tr>
<td>Length and density</td>
<td>0.757</td>
</tr>
</tbody>
</table>
Maine Results: CLC
Comparison to 100K NHD Benchmark

100K Stratified Prune (SP)

100K Length and Density Prune (LaDP)

Weighted CLC
- 0.0000 - 0.0020
- 0.0020 - 0.0036
- 0.0036 - 0.0052
- 0.0052 - 0.0066

0 20 40 Kilometers

14th Workshop of the ICA Commission on Generalization and Multiple Representation, Paris, France, 30 June to 1 July, 2011
Maine Results: CLC

Compare SP to LaDP

Stratified Pruning (SP)

Red omitted from LaDP

Weighted CLC
- 0.0000 - 0.0010
- 0.0011 - 0.0025
- 0.0026 - 0.0040
- 0.0041 - 0.0055
- 0.0056 - 0.0066

Length and Density Pruning (LaDP)

Pink omitted from SP
Four Iowa Subbasin Results: Pruning to 100K (Natural Density Variations)

Average Density (Kilometers per square kilometer)

- 0.326
- 1.385

24K NHD

0 10 20 40 Kilometers

14th Workshop of the ICA Commission on Generalization and Multiple Representation, Paris, France, 30 June to 1 July, 2011
Four Iowa Subbasin Results: Pruning to 100K
(Natural Density Variations)

14th Workshop of the ICA Commission on Generalization and Multiple Representation, Paris, France, 30 June to 1 July, 2011
### Four Iowa Subbasin Results
CLC comparing to 100K Benchmark
(Natural Density Variations)

<table>
<thead>
<tr>
<th>Pruning method</th>
<th>Weighted CLC</th>
<th>Weighted omissions</th>
<th>Weighted commissions</th>
</tr>
</thead>
<tbody>
<tr>
<td>Stratified</td>
<td>0.914</td>
<td>0.048</td>
<td>0.038</td>
</tr>
<tr>
<td>Length and Density</td>
<td>0.841</td>
<td>0.140</td>
<td>0.019</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Pruning method</th>
<th>Bootstrapped 90 Percent Confidence Interval</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Lower bound</td>
</tr>
<tr>
<td>Stratified</td>
<td>0.907</td>
</tr>
<tr>
<td>Length and Density</td>
<td>0.830</td>
</tr>
</tbody>
</table>
Four Iowa Subbasin Results: CLC
Comparison to 100K NHD Benchmark

100K Stratified Prune (SP)
Weighted CLC
- 0.0000 - 0.0015
- 0.0015 - 0.0040
- 0.0040 - 0.0052
- 0.0052 - 0.0063

100K Length and Density Prune (LaDP)
Weighted CLC
- 0.0000 - 0.0015
- 0.0015 - 0.0040
- 0.0040 - 0.0052
- 0.0052 - 0.0063

0 30 60 Kilometers
Four Iowa Subbasin Results: CLC

Compare SP to LaDP

SP

LaDP

Red omitted from LaDP

Pink omitted from SP

WEIGHTED CLC

0.0000 - 0.0016
0.0017 - 0.0036
0.0037 - 0.0050
0.0051 - 0.0056
0.0057 - 0.0063

14th Workshop of the ICA Commission on Generalization and Multiple Representation, Paris, France, 30 June to 1 July, 2011
Four Iowa Subbasin Results
Disconnects in the LaDP

14th Workshop of the ICA Commission on Generalization and Multiple Representation, Paris, France, 30 June to 1 July, 2011
Summary Statements

• The SP and LaDP approaches work equally well at removing data inconsistencies, but the SP approach better retains density variations that depict natural terrain variations.

• Enrichment processing is faster for the SP approach than the LaDP approach, and tools associated with the SP enable quick testing of various pruning alternatives.

• Automated selection of target densities remains a topic for future development for the SP approach.

• “One size fits all” concept implemented through the LaDP approach is wrong. Additional development may test alternatives that apply different parameters to areas with different physical conditions. (Connectivity issue must be resolved).

• Additional tests should compare smaller scale (500K or smaller) generalization results from these two approaches.
Thank you! Question?