Stream order selection for model generalization of the topographic map of Indonesia

Fahrul Hidayat¹, Nugroho Purwono¹, Danang Budi Susetyo¹, Mochamad Irwan Hariyono¹, Tia Rizka Nuzula Rachma¹, Rizka Windiastuti¹

¹Geospatial Information Agency
¹Jalan Raya Jakarta-Bogor Km 46, Cibinong, Bogor, Indonesia
email: fahrul.hidayat@big.go.id

Abstract. The importance of stream networks is related to other features in the topographic map e.g. as weighted-parameter for contour derivation. The generalization of these features needs complex parameters specifically geometrical and conceptual aspects. Geometric parameters consist of stream length and vertices, while the conceptual part handles stream networks connectivity as logical consequences. Stream networks selection is a type of important step on map features analysis and in map databases. This paper proposes a new approach for stream networks generalization of Topographic Map of Indonesia (as known as RBI) for 1:5,000 to 1:25,000 of scale by using geometrical and conceptual parameters. Three stages used in this research were: data pre-processing (include resolving the topological errors), generating stream order (1:5,000 of scale as an input), comparing stream order algorithms (Strahler, Scheidegger, Shreve, and Drwal), and performing feature similarity-based analysis (comparison of stream ordering results and 1:25,000 of scale). The research result ed four different stream orders and eight different similarity-analysis values since each algorithm was tested in two scenarios (in 1st scenario, order > 1 were selected while in 2nd scenario, order > 2 were selected). Eventually, after comparing those results, the Scheidegger method obtained the highest similarity value in on both 1st and 2nd scenarios. Further, generalization by using stream order selection delivered the representation of river in constructing map elements of RBI.

1. Introduction
A stream network is essential component of hydrographic features on topographic map. A stream which has multiple tributaries also shows distinct drainage patterns. On the cartographic context (in this case refers to Topographic Map of Indonesia (RBI), hydrologic features represent area of the landscape that drains to a portion of the stream network. More specifically, a hydrologic feature defines the areal extent of surface water drainage to an outlet point on a stream network or to multiple outlet points. A stream network may represent all or only part of the total drainage area to an outlet point so that multiple stream network may be required to define the entire drainage area at a given outlet [1]. Stream networks as hydrologic features are determined based on topographic/ relief, and other relevant landscape characteristics regardless administrative, political, or jurisdictional boundaries. Stream network provides a basic hydrological and morphological partition of a watershed, thus it is a very important component as a landscape feature for stream analysis using Geographic Information System (GIS). Due to indigenous topography and geology, a stream shows a certain drainage pattern based on the form and texture of its network of channels and tributaries [2, 3]. Furthermore, output from automated generation of contours could be improved by refining alignment of the hydrography and topography data [4].

The identification and mapping of stream networks is very important to applications in cartography and any other fields [5, 6]. For instance, geomorphological studies usually involve the activity of quantitative watershed characteristics like drainage density, stream network, and stream order, similar to several
In the current GIS systems, stream networks are kept as line segments with their geographical and topological factors. In the stream networks, different patterns are often discovered and associated with alternative geographical factors. There are many kinds of drainage patterns related to stream networks. At present, a lot of analysis has been done concerning the drainage patterns [1, 2, 9]. Particularly, geometric indicators for the characterization of drainage patterns are conferred and explained on this stream network analysis. At present, several researchers have begun to concentrate to the feature of stream networks throughout the generalization method [5, 6, 9].

Difficulties in identification of stream channels due to cartographic generalizations and decision rules lead to inaccuracies in published drainage networks. For instance, [5] found that many spatial data providers in his study created their own set of decision rules and criteria for the inclusion and length of stream channels. Within that research, [5] suggests that limitations within the recognition of small channels build measures of drainage density obsessed with map scale. An important prerequisite for integrating data infrastructure is generalizing and managing data representation at various scales [4]. Data modeling has important implications for cartographic design and analysis, since RBI users expect that the data will match build upon geographical and topological aspect. The scientific use of downloaded data carries additional requirements to support reliable measurement of data at several resolutions, and to ensure that features integrate horizontally and vertically [9, 10].

Related to RBI, the generalization of stream networks from scale 1:25,000 to 1:50,000 is established by using stream length and density selection. There are 5 steps to perform generalization of rivers. The first step is omitting short unnamed rivers. The minimum length of the river channels to be retained for scale 1:50,000 is 500 meters (10 mm on the map), and for single line rivers is 250 meters (5 mm on the map). Those having length less than the thresholds will be omitted and the rivers having a specific toponym will be retained even though they do not meet the minimum length. Meanwhile, there is no standard method for the process of generalizing hydrographic features in RBI. Therefore, it is necessary to evaluate the implementation of the model generalization of 1:5,000 to 1:25,000 scale.

Feature generalization, or technically defined as stream ordering, can be obtained by using position of the stream channel heads as starting points for topologically connected stream networks [1, 3, 5]. This research tested automated data generalization of geodatabase from 1:5,000 to 1:25,000 map scale, to meet the best hydrography product by using stream order approach [11]. This research investigated several options of stream ordering (i.e. Strahler, Shreve, Scheidegger, and Drwa) to build, implement, and evaluate automated methods for cartographic generalization of RBI datasets without negating the main characteristic of hydrographic meaning.

2. Data and Method

2.1. Data

The main datasets used were RBI of scale 1:5,000 and 1:25,000, then consecutively abbreviated as RBI5K and RBI25K. This study tested the data of Tabuan Island, Tanggamus, Lampung Province, Sumatera Island as a case study (Figure 1). Map sheet numbers of these datasets are: 1) RBI25K 1010-323; and 2) RBI5K 1010-3233C, 1010-3233D, 1010-3235A, 1010-3235B, 1010-3235C, 1010-3235D, 1010-3236A, 1010-3236B, 1010-3236C, 1010-3236D, 1010-3237B, 1010-3237D, 1010-3238A, 1010-3238B, 1010-3238C, 1010-3239A. These datasets were produced by Geospatial Information Agency of Indonesia. RBI25K was produced in 2014 while RBI5K was in 2016.
2.2. Method

The data processing methods for the two datasets were carried out separately. Three stages for main processing RBI5K dataset were: data pre-processing, stream orders generation, comparing stream order algorithms (Strahler, Scheidegger, Shreve, and Drwal). The processing for RBI25K dataset used stream network selection in GIS software. Research stages is shown in Figure 2.

2.2.1. Pre-processing

The stream network selection process was carried out for RBI5K dataset including topology resolving errors and generating outlet points. Stream networks selection of RBI5K data was performed for the attributes of river, single line river, and river flow. Next step was topology checking and segmentation on selected data. Topology is a provision related to the relationship between spatial objects in the form of points, lines and areas of a geographical element. Topology is needed to manage the geometry of shared spatial objects (shared geometry) and to maintain data integrity. The example of topological errors and the solution for resolving are consecutive shown in Figure 3 and Figure 4. Hereafter, the next stage was determining the outlet points by using “end vertices” to “point” conversion. Outlet points were generated by performing intersection between end points and coastline (Figure 5).

2.2.2. Stream order generation

Stream order is a typical strategy to assign a hierarchy to the sections of a river network used for selection [12]. Stream ordering is a technique of allocating a numeric order to joints in a stream network. This order is a way to identify and classify types of streams depending on their numbers of tributaries. Some stream characteristics can be estimated only by knowing their order. There are several methods for generating stream orders, such as Strahler [13, 14], Scheidegger [15], Shreve [14, 16], and Drwal [17]. The river ordering method carried out in this study utilized tools in the GRASS GIS software namely v.stream.order.

Figure 1. The preview of study area
Figure 2. Research method

Figure 3. The topological error needed to be fixed due to: (left) the dangles and (right) incomplete geometries
2.2.3. Similarity – based analysis method
Similarity-based analysis was performed for two different tests. The first test was to compare stream network from RBI25K scale (as a reference dataset) and RBI5K scale. Hereafter, the second test was needed to find the similarity level of each generalized data (Strahler, Scheidegger, Shreve, and Drwal) compared to the reference map (existing RBI25K scale). The first test was an iterative process started from 12.5 meters of buffer width, considering that it was the possible displacement or accuracy at 1:25,000 map scale. Then, the buffer was increased with interval 12.5 meters and stopped at 150 meters. Further, the second test only used a single buffer width that was 12.5 meters in order to ascertain the similarity level for all ordering method on the 1:25,000 of scale accuracy or possible displacement. The formula used for similarity – based analysis in this study was the improvement from [18] which was proposed by [19]:

\[ S_O = \frac{f_A(C \cap R)}{f_A(C \cap R) + \alpha f_A(C - R) + \beta f_A(R - C)} \]  (1)
where $f_A(C \cap R)$ was the component of the intersection area of C and R, $f_A(C - R)$ represented features of the area of R to erase the evaluated object C, and $f_A(R - C)$ represented the features of the area of the evaluated object C to erase reference object R. The range of $S_O$ was 0 to 1. If the extracted and reference objects overlapped completely, then $S_O = 1$. If there was no overlap between the two objects, then $S_O = 0$ [19].

3. Result and Discussion

3.1. Stream order

RBI displayed the entire river on Tabuan Island. Tributaries of the small river on Tabuan Island formed several large rivers, the two largest rivers namely Way Pelus Balak and Way Pabuwiyan Balak. Overall on the Tabuan Islands, there were 4,268 and 1,412 river segments generated from RBI5K and RBI25K consecutively (Table 1). Stream ordering comparisons were made based on the methods of Strahler, Shreve, Scheidegger, and Drwal. Ordering is important because on a smaller scale map it is not necessary to display a river that has a low order, or spatially its position is located upstream.

The common method used for stream ordering is the Strahler and Streve [20]. Using Strahler method on the research data, all channels that did not have tributaries were categorized as order 1. When there was a meeting of two "order 1" channels, then the order of next channel became 2. When there were two "order 2" channels converge, a channel with order 3 was formed. When a channel met a lower order channel, the next channel remained at the same order. For example, when an "order 3" channel met an “order 1” channel, then the order of next channel remains 3.

Using Shreve methods, all channels that did not have tributaries were categorized as order 1. When there was a meeting of two channels, the magnitude was the sum of the order values that joined. For example, when an "order 1" channel met another "order 1", the following channel was "order 2". When an "order 2" channel met an "order 1" channel, an "order 3" channel was formed, and when an "order 3" channel met an "order 2" channel, the following channel was "order 5", and so on.
Figure 7. The part of stream ordering result in Tabuan Island
Scheidegger's method was similar to the Shreve method, but the Scheidegger method started with the smallest river order as order 2. The convergence of the two channels produced a channel value which was the sum of the two orders. Indirectly, on the same channels the order using the Scheidegger method was twice as much as that channel on the Shreve method.

Drwal method started with the lowest order on each stream of 1. The order of the tributaries in the Drwal method ascended in one of the following conditions: 1) the converging of two rivers with the same order; for example, "order 3" channel met another "order 3", produced an "order 4" channel; 2) the convergence of three channels, one of which had a different value than the other two; for example, the channels of "order 3", "order 2", and "order 2", produced an "order 4" channel; 3) the convergence of 4 channels where one had the highest order value, another one has a lower value, and the two remaining had the lowest value; for example, order 3, order 2, order 1 and order 1, produced an "order 4" channel; 4) the convergence of 5 channels, where one had high order, and the other four had 2 lesser value; for example, order 3, order 1, order 1, order 1, and order 1 produced channel order 4.

When a channel met another channel that had a lower order, the next channel would have the same order value; for example, order 3 channel met order 2, produced an order 3 channel, as well as order 3 channel met order 1 channel, produced order 3 channel. The comparison of the results of ordering using the four methods is shown in figure A. Although the Strahler, Shreve, and Drwal methods on the upstream of the river all started in order 1, the downstream of the river resulted a different order value. In the Scheidegger method, the value of each tributaries in one order showed twice its value in the Shreve method.

Using the Strahler method for the Tabuan Island case study, the highest order in the river downstream was order 5, the Drwal method showed the highest order namely 9th order, while using the Shreve method, the highest order reached 302, and Scheidegger reached order 604. The order distribution for all rivers on Tabuan Island is shown in Figure 8, and the order comparison table is presented in Table 1.

![Figure 8. River segments distribution based on stream--ordering method](image)
Table 1. Segment numbers comparison

| Order | Scheidegger | Shreve | Strahler | Drwal | RBI5K | RBI25K |
|-------|-------------|--------|----------|-------|-------|--------|
| Order 1 | -           | 2,216  | 2,216    | 2,216 | -     | -      |
| Order 2 | 2,216       | 473    | 1,096    | 736   | -     | -      |
| Order 3 | -           | 263    | 643      | 484   | -     | -      |
| Order 4 | 473         | 172    | 257      | 310   | -     | -      |
| Order 5 | -           | 124    | 56       | 241   | -     | -      |
| Order 6 | 263         | 101    | -        | 172   | -     | -      |
| Order 7 | -           | 87     | -        | 59    | -     | -      |
| Order 8 | 172         | 68     | -        | 41    | -     | -      |
| Order 9 | -           | 47     | -        | 9     | -     | -      |
| ... | -           | -      | -        | -     | -     | -      |
| \(\sum_{k=10}^{n} Order_r\) | 1,144   | 717    | -        | -     | -     | -      |

Total segments 4,268 4,268 4,268 4,268 4,268 1,412

3.2. Similarity-based analysis of generated stream order

The result of first test is shown in aggregated distribution functions as can be seen in Figure 9 and 10. Figure 9 illustrated the graph of correlation between referenced and tested data, which comprised the area of buffer intersection and the difference value between the buffers from two data. As seen in the graph, intersection and difference had contrast trend, in which intersection showed a consistent escalation with the increase of buffer width, while the difference of buffer areas showed opposite trend. However, to assess similarity of two objects, intersection was the most important part compared to the others.

![Figure 9](image-url)
Buffer width 12.5 meters only gave a similarity result about 27 percent, and always increased as long as the buffer width was enlarged. From buffer width of 12.5 meters until 62.5 meters, the line graph was steep, with the similarity gap between two intervals reached at least 10 percent. Hereafter, the similarity result was more leaning, where buffer width in that range indicated average displacement of evaluating data. According to the graph, between the ranges of 100 to 150 meters the line could be considered as stable position, with the similarity result between 89-94 percent. To conclude, based on reference data used, the average similarity level of existing stream dataset at RBI5K scale was stable on the range of 100 meters until 150 meters buffer width. The wider buffer would generate a higher similarity level due to the increase of the overlapped buffers polygons for both referenced and tested datasets.

Two scenarios were used to test the similarity levels on the second test which is visualized in Figure 11 and 12. Scenario 1 was comparing “order > 1” on ordered streams (tested dataset) to RBI25K (referenced dataset) while scenario 2 is comparing “order > 2” on ordered streams (tested dataset) to RBI25K (referenced dataset). All components of similarity – analysis varied for both scenario 1 and 2. Surprisingly, the intersection areas and the features of the area of the evaluated object “ordered streams” to erase reference object “RBI25K” had mirrored pattern (see Figure 11). In the other hand, features of the area of “RBI25K” to erase the evaluated object “ordered streams” had different trend compared to these two components. Based on the calculation, Scheidegger ordering method generated the highest intersection areas among all. In this case, Scheidegger method started from “order 2” so it produced same as input (RBI5K) for the selection parameter “order > 1”. In contrast, Strahler on scenario 2 generated the lowest intersection areas.

The second test was done using two scenarios to test the similarity level of selected stream order as tested dataset toward RBI25K as reference dataset. In the first scenario selected stream order was obtained through to eliminate order 1 of the stream network system, while the second scenario was obtained by eliminating the order 2. The result of both was tested toward RBI25K dataset to acquire the similarity level. Data testing in each scenario followed the stream order techniques as explained in the research method, and the results of each scenario varied particularly in the components of similarity. The results of this second test is depicted in Figure 11 and 12. The correlation between tested and referenced data consisting of the area of buffer intersection and the difference value is illustrated in

Figure 10. Similarity result for the first test (RBI5K fitted to RBI25K)
According to this graph (Figure 11), it was surprising that trend of intersection shaped a reversal (mirrored) pattern with trend of the difference of buffer areas. Related to the result of the test, it could be analyzed that the process of data comparison was very influenced by the number of segments that resulted from stream ordering. Furthermore, different techniques were used in stream ordering to determine the number of segments. For instance, the different numbers of segment resulted from Strahler, Scheidegger, Shreve, and Drwal (see Table 1).

The results of intersection and the difference between buffer area in each stream order continued with the similarity assessment. This assessment was done through iteration process by increasing the buffer interval for every 12.5 meters until it stopped at 150 meters, with the result that is illustrated in Figure 12. According to the graph (Figure 12), the value of similarity produced to first scenario (order>1) tended to be higher than second scenario (order>2). The results of the first scenario were 23.09% and 25.65% while the second were 13.88% until 23.09%. Thus, it could be interpreted that the elimination of several segments of stream order affected the result value of similarity. However, in this study an unexpected result was found where the result using Scheidegger technique had the same value as when it was compared to other methods, in particular Shreve (see the 1st and 5th histograms in the Figure 12). By comparing these several stream order assessment, the similarity level of Scheidegger’s stream order would result the same value as with Shreve’s in the previous order of stream ordering.
4. Conclusion
Similarity-based analysis was performed for two different tests. The first test was to compare stream network from RBI25K scale (as a reference dataset) and RBI5K scale. In the first test, we concluded that the wider buffer would generate a higher similarity level due to the increase of the overlapped buffers polygons for both referenced and tested datasets. Hereafter, the second test was needed to find the similarity level of each generalized data (Strahler, Scheidegger, Shreve, and Drwal) when compared to the reference map (existing RBI25K scale). The second test resulted four different stream orders and eight different similarity-analysis values since each algorithm was tested in two scenarios (in 1st scenario, order > 1 were selected while in 2nd scenario, order > 2 were selected. Eventually, after comparing those results, the Scheidegger method obtained the highest similarity value in on both 1st and 2nd scenarios. Nevertheless, Scheidegger method started from “order 2” so it produced same ordered streams as input (RBI5K) for the selection parameter “order > 1”. Even the 2nd scenario proved that Scheidegger had the highest similarity level but it had the same level as Strahler, Shreve, and Drwal on the 1st scenario. So, it can be concluded that Strahler, Shreve, and Drwal ordering methods/algorithms generated the same similarity level for “order > 1” as Scheidegger ordering method for “order > 2” selection. Further, generalization by using stream order selection delivered the representation of river selection in constructing map features of Indonesian Topographic Map.

5. Acknowledgments
Authors gratefully acknowledge Geospatial Information Agency of Indonesia due to financial support and gives all materials for the research i.e. datasets and technical specification of medium to small scale map generalization.
6. References

[1] Zhang, L. and E. Guilbert, Automatic drainage pattern recognition in river networks. International Journal of Geographical Information Science, 2013. 27(12): p. 2319-2342.

[2] Howard, A.D., Drainage analysis in geologic interpretation: a summation. AAPG bulletin, 1967. 51(11): p. 2246-2259.

[3] Zhang, L. and E. Guilbert, A study of variables characterizing drainage patterns in river networks. 2012.

[4] Arundel, S.T., P.T. Thiem, and E.W. Constance, Automated extraction of hydrographically corrected contours for the conterminous United States: the US Geological Survey US Topo product. Cartography and Geographic Information Science, 2018. 45(1): p. 31-55.

[5] Heine, R.A., C.L. Lant, and R.R. Sengupta, Development and comparison of approaches for automated mapping of stream channel networks. Annals of the Association of American Geographers, 2004. 94(3): p. 477-490.

[6] Stanislawski, L.V., B.P. Buttenfield, and A. Doumbouya, A rapid approach for automated comparison of independently derived stream networks. Cartography and Geographic Information Science, 2015. 42(5): p. 435-448.

[7] Supangat, A.B., Karakteristik hidrologi berdasarkan parameter morfometri DAS di kawasan Taman Nasional Meru Betiri. Jurnal penelitian hutan dan konservasi alam, 2012. 9(3): p. 275-283.

[8] Sukiyah M., Morfometri Daerah Aliran Sungai Pada Bentangalam Vulkanik Kwarer Terdeformasi. Bulletin of Scientific Contribution: Geology, 2007. 5 (3): p. 1-8.

[9] Brewer, C.A., B.P. Buttenfield, and E.L. Usery. Evaluating generalizations of hydrography in differing terrains for The National Map of the United States. in Proceedings, 24th International Cartographic Congress. 2009.

[10] Bobzien, M., et al., Multi-representation databases with explicitly modeled horizontal, vertical, and update relations. Cartography and Geographic Information Science, 2008. 35(1): p. 3-16.

[11] Sandro, S., et al., Model Generalization of the Hydrography Network in the CARGEN Project, in Advances in Cartography and GIScience. Volume 1. 2011, Springer. p. 439-457.

[12] Horton, R.E., Erosional development of streams and their drainage basins; hydrophysical approach to quantitative morphology. Geological society of America bulletin, 1945. 56(3): p. 275-370.

[13] Strahler, A.N., Quantitative analysis of watershed geomorphology. Eos, Transactions American Geophysical Union, 1957. 38(6): p. 913-920.

[14] Tarboton, D.G., R.L. Bras, and I. Rodriguez-Iturbe, On the extraction of channel networks from digital elevation data. Hydrological processes, 1991. 5(1): p. 81-100.

[15] Scheidegger, A.E., Statistical description of river networks. Water Resources Research, 1966. 2(4): p. 785-790.

[16] Shreve, R.L., Statistical law of stream numbers. The Journal of Geology, 1966. 74(1): p. 17-37.

[17] Jan, D., Wyksztalcenie i organizacja sieci hydrograficznej jako podstawa oceny struktury odpływu na terenach młodoglacjalnych. 1982: Uniwersytet Gdański in Gdańsk.

[18] Tversky, A., Features of similarity. Psychological review, 1977. 84(4): p. 327.

[19] Cai, L., et al., Accuracy assessment measures for object extraction from remote sensing images. Remote Sensing, 2018. 10(2): p. 303.

[20] Gleyzer, A., et al., A Fast Recursive Gis Algorithm For Computing Strahler Stream Order In Braided And Nonbraided Networks 1. Jawra Journal of the American Water Resources Association, 2004. 40(4): p. 937-946.