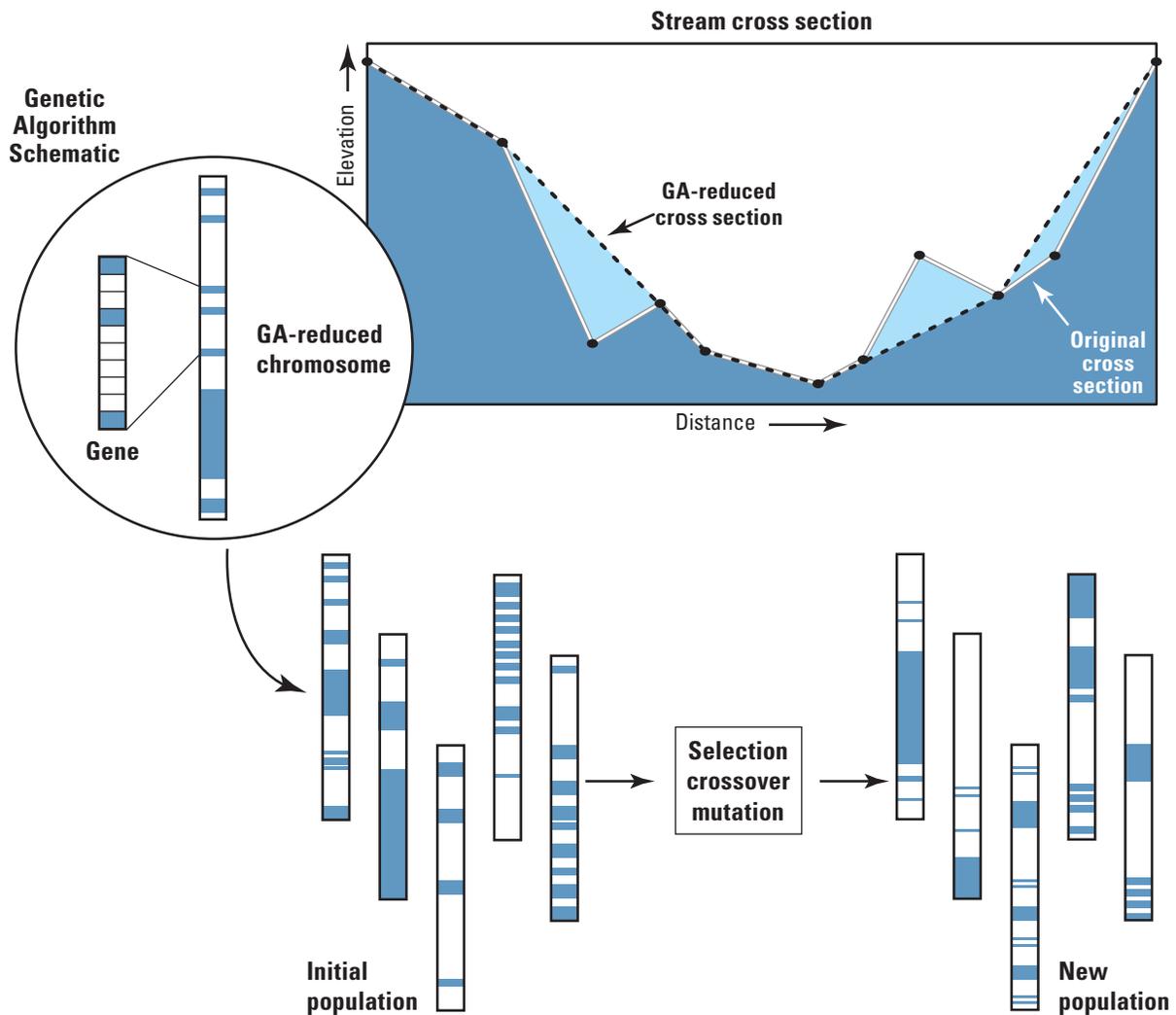


Reducing Cross-Sectional Data Using a Genetic Algorithm Method and Effects on Cross-Section Geometry and Steady-Flow Profiles



Scientific Investigations Report 2015–5034

Reducing Cross-Sectional Data Using a Genetic Algorithm Method and Effects on Cross-Section Geometry and Steady-Flow Profiles

By Charles Berenbrock

Scientific Investigations Report 2015–5034

U.S. Department of the Interior
U.S. Geological Survey

U.S. Department of the Interior
SALLY JEWELL, Secretary

U.S. Geological Survey
Suzette M. Kimball, Acting Director

U.S. Geological Survey, Reston, Virginia: 2015

For more information on the USGS—the Federal source for science about the Earth, its natural and living resources, natural hazards, and the environment—visit <http://www.usgs.gov> or call 1–888–ASK–USGS.

For an overview of USGS information products, including maps, imagery, and publications, visit <http://www.usgs.gov/pubprod/>.

Any use of trade, firm, or product names is for descriptive purposes only and does not imply endorsement by the U.S. Government.

Although this information product, for the most part, is in the public domain, it also may contain copyrighted materials as noted in the text. Permission to reproduce copyrighted items must be secured from the copyright owner.

Suggested citation:

Berenbrock, Charles, 2015, Reducing cross-sectional data using a genetic algorithm method and effects on cross-section geometry and steady-flow profiles: U.S. Geological Survey Scientific Investigations Report 2015–5034, 16 p., <http://dx.doi.org/10.3133/sir20155034>.

Contents

Abstract.....	1
Introduction.....	1
Reduction Methods	2
Standard Reduction.....	3
Genetic Algorithm Reduction.....	3
Comparison of Reduction Results	5
Visual Analysis of Cross-Section Reductions	6
Hydraulic Modeling Analysis of Reduction Methods	11
Summary.....	14
References Cited.....	15

Figures

1. Graph showing fitness calculation for a hypothetical cross section	4
2. Graphs showing effects of data-point reduction on cross-sectional shape.....	7
3. Graphs showing effects of data-point reduction on cross-sectional area	8
4. Graphs showing reduction-error (RE) curves for 10 cross sections resulting from the genetic algorithm (GA)-reduction method, Kootenai River, Idaho.....	10
5. Graphs showing the effects of reduced cross sections on simulated water-surface elevation at five river discharges	12
6. Graphs showing comparisons between the original and 10 point, 20 point, and 30 point reduced cross sections produced by two reduction methods for cross-section 154.575.....	13

Tables

Table 1. Comparison of best fitness between several reduction methods for 10 cross sections on the Kootenai River, Idaho.....	5
Table 2. Reduction error (RE) values for 10 cross sections resulting from the standard and genetic algorithm (GA)-reduction methods, Kootenai River, Idaho.....	9

Conversion Factors

[International System of Units to Inch/Pound]

Multiply	By	To obtain
	Length	
meter (m)	3.281	foot (ft)
	Area	
square meter (m ²)	10.76	square foot (ft ²)
	Flow rate	
cubic meter per second (m ³ /s)	35.31	cubic foot per second (ft ³ /s)

Supplemental Information

Elevation, as used in this report, refers to distance above the vertical datum.

River miles are used to measure horizontal distance along a river from a downstream confluence and to designate cross-section locations, which provided consistency with river-mile marker locations as designated by the U.S. Army Corps of Engineers, as shown on topographic maps, and as cited in earlier studies.

A cross section is a series of distance and elevation data point pairs that describe the channel shape perpendicular to the mean flow direction.

Distance is the length across the channel from the left-most data point while facing downstream.

Datum

Vertical coordinate information is referenced to the North American Vertical Datum of 1988 (NAVD88).

Abbreviations

ABC	area between the cross sections
DEMs	digital elevation models
FEMA	Federal Emergency Management Agency
GA	Genetic Algorithm
GPS	global positioning system
LiDAR	light detection and ranging
m/m	meters per meter
npr	number of reduced data points
1-D	one dimensional
RTK	real-time kinematic
RE	reduction error

Reducing Cross Sectional Data Using a Genetic Algorithm Method and Effects on Cross-Section Geometry and Steady-Flow Profiles

By Charles Berenbrock

Abstract

Reduction of cross-sectional data using a genetic algorithm method, and the effects of data reduction on channel geometry and steady-flow profiles, were analyzed. Two reduction methods—standard and genetic algorithms—were used to reduce cross-sectional data from the Kootenai River in northern Idaho. Cross sections that are representative of meander, straight, braided, and canyon reaches were used to evaluate the reduction methods. Visual and hydraulic analyses were used to assess the methods. The genetic algorithm-reduced cross sections approximated the shape of the original cross sections better than the standard-reduced cross sections. A greater number of cross-sectional data points were needed for reduced cross sections in the straight reach, and even more in the braided reach, because a greater amount of data points are needed to adequately define cross sections that have greater topographic variability. For the genetic algorithm-reduction method, about 40 data points were needed to adequately define the shape of a reduced cross section in the braided reach compared to 10 to 20 data points in the meander and canyon reaches. The standard-reduction method needed about 70 data points for the braided reach and more than 30 points for the meander and canyon reaches. The genetic algorithm can effectively reduce data while staying within the threshold set by the maximum number of points to be included in the reduced dataset.

The effects of reduced cross-sectional data points on steady-flow profiles were also determined. Thirty-five cross sections of the original steady-flow model of the Kootenai River were used. These two methods were tested for all cross sections with each cross section resolution reduced to 10, 20 and 30 data points, that is, six tests were completed for each of the thirty-five cross sections. Generally, differences from the original water-surface elevation were smaller as the number of data points in reduced cross sections increased, but this was not always the case, especially in the braided reach. Differences were smaller for reduced cross sections developed by the genetic algorithm method than the standard algorithm method.

Introduction

Cross-sectional data are used for many purposes, such as the investigations of flood plain delineation, flow patterns, shear stress, sediment mobility and transport, channel evolution, and aquatic habitat conditions. Accuracy of cross-section data is important because it could affect channel-geometry determinations and water-surface profile calculations. For example, the consequence of errors in the water-surface elevation has a major effect on computations of velocity, shear stress, and sediment transport. Water-surface profiles in many studies are computed by using one-dimensional (1-D) step-backwater models such as HEC-RAS (Brunner, 2010), which uses the standard step method for steady flow. The standard step method uses the energy, continuity, and flow resistance (for example, Manning's) equations between cross sections to compute the water-surface elevation and streamflow velocity (Chow, 1959).

All models have a limit to the number of points allowed in a cross section. For example, the HEC-RAS version 4 model has a 500-point limit. This limit might seem large enough, but when cross sections are computer generated from Light Detection and Ranging (LiDAR) or Digital Elevation Models (DEMs), or data are collected with equipment such as an echo sounder, the number of cross-sectional data points tends to be quite substantial. In an earlier study (Barton and others, 2004), for example, approximately 400 cross sections were surveyed on the Kootenai River in northern Idaho in order to understand the hydraulic characteristics of the river and to promote hydraulic conditions that improve spawning conditions for the endangered Kootenai River white sturgeon. The number of data points for each cross section ranged from about 500 to more than 2,500 points (Moran and Berenbrock, 2003; Barton and others, 2004). Only a few cross sections had more than 2,000 data points. More than half of the surveyed cross sections were included in a HEC-RAS model of the Kootenai River (Berenbrock, 2005, 2006a), and most cross sections were reduced to less than 150 data points.

2 Reducing Cross Sectional Data Using a Genetic Algorithm Method and Effects on Cross-Section Geometry and Steady-Flow Profiles

Large datasets must be reduced to meet the limitations of the programs being used. Reduced datasets improve run-time performance and facilitate data transmission and storage. Selecting the appropriate data points to keep from among the hundreds or thousands of data points can be both challenging and tedious. However, reducing the number of cross-sectional data points can result in significant changes to the reduced cross section, which could affect computed water-surface elevations, streamflow velocity, shear stress, and sediment transport. Considerable care must be taken when reducing data so that computed errors and uncertainties remain small or within acceptable limits. It is important to understand the effects that reduced cross sections can have on National Flood Insurance Program, flood-inundation, habitat, and sediment-transport studies. The Federal Emergency Management Agency (FEMA) has indicated that there is no point minimum—the number of data points for defining a cross section—as long as the actual shape of the cross section is well defined (Federal Emergency Management Agency, 2007).

Much research has been done on error and uncertainty analysis in surface-water hydraulics. Research has been carried out on determining the optimal spacing between cross sections (Samuels, 1989; Castellarin and others, 2009); developing cross sections from topographic maps, LiDAR, and DEM data (Burnham and Davis, 1990; Pasternack and others, 2004; Cook and Merwade, 2009); and interpolating cross sections between known cross sections (Traver and Miller, 1993). Travis and Lokey (1999) developed a method to reduce cross-section data to 100 data points, the maximum limit of the HEC-2 model. Berenbrock (2006b) developed a genetic algorithm (GA) computer program that reduces the number of data points in a cross section to any size. He compared the GA-reduced cross sections to cross sections developed by standard reduction methods—selecting every 10th, 20th, or n th point and omitting the rest—for the same number of data points in a cross section. Reduced cross sections developed from standard and GA methods were compared to the original cross-sectional data, and results showed that the GA method produced smaller differences from the original cross-sectional data than those obtained by using standard procedures. Unfortunately, no research to date has been done to determine the optimal number of points that are needed in a cross section or the effects on cross-section geometry and steady-flow profiles. The optimal number of points depends on the degree of topographic variability and the scale of topography that is of interest. It also depends on laws, regulations, or the requirements of the funding party—for example, FEMA (Federal Emergency Management Agency, 2007).

The purpose of this report is to describe an application of a GA to the reduction of cross-sectional data points, demonstrate that the GA is a viable approach, and to evaluate the effects of reduced cross sections on channel geometry and steady-flow profiles. First, the study compared the accuracy of two reduction approaches, standard and genetic algorithm

methods. Data from 10 cross sections covering 4 different channel types on a river were reduced by standard and genetic algorithm methods. These reduction methods were employed because the raw (original) data are preserved, not averaged, interpolated, or extrapolated. Second, the study identified the sources and spatial distribution of error in different channel types and determined the requisite sample size for different scales of resolution and application. Third, the study examined the effects of data reduction on steady-flow profiles. From these components, the amount of reduction can be tailored to the goals of an application.

Reduction Methods

There are many data-reduction methods available. For this study, only methods that preserve the original data were considered. The advantages of preserving the original data are that the original features, including vertical banks and discontinuities, are maintained. The original cross section—whether it contains 2,000 data points or 20 data points—is more accurate than anything generated in part from those data.

Stream-channel cross-section data from the Kootenai River (Barton and others, 2004) were used in the data-point reduction methods. Data for the streambed part of these cross sections were collected by connecting continuous Real-Time Kinematic (RTK) Global Positioning System (GPS) equipment to an echo sounder, and bank data were collected by connecting a RTK GPS to a laser rangefinder equipped with an angle encoder (Moran and Berenbrock, 2003). Berenbrock (2006b) specifically used 10 cross sections from the Kootenai River to substantiate a genetic algorithm (GA) for data-point reduction. Data points from these 10 cross sections were also used in this study because the original data were still available. The cross sections are 107.658, 151.438, 152.019, 154.972, 163.027, 185.394, 199.727, 212.227, 216.622, and 219.881, which are defined by a station number in river miles¹ that corresponds to its location on the river. The number of data points in these sections ranged from 497 in cross-section 152.019 to 2,521 in cross-section 154.972. Original data from the other cross sections on the Kootenai River were not available. However, there were still enough data points in the reduced cross sections by Barton and others (2004)—usually more than 100 data points per cross section, with some sections containing several hundred data points—for further reduction. For this study, cross-section data were reduced to as few as 10 data points per cross section.

¹River mile locations are based on Columbia Basin Inter-Agency Committee (1965) river-mile index for the Kootenai/Kootenay River. River mile 0 is at the confluence of the Kootenay River and Columbia River near Castlegar, British Columbia, Canada, and river mile 152 is upstream on the Kootenai River near Bonners Ferry, Idaho.

Standard Reduction

The standard-reduction method to reduce data points is to keep every 10th, 20th, 30th, or n^{th} point and discard (omit) the rest. This procedure was used in this study because of the standard-reduction method's simplicity, ease, and quickness. The value of the n^{th} data point for each cross section was different because the total number of data points in the cross sections was different. For example, cross-section 152.019 has 497 data points. If the number of points was reduced to 20, then every 26th point would be selected with 2 points remaining ($1+(26 \times 19)=495$; then $497-495=2$). The first data point in the cross section is always kept, so a value of 1 is added to the number of intervals (19 for this example). Fewer points exist along the banks than on the streambed for this dataset because the bank data were collected manually with a laser rangefinder. The remaining two points in this example were inserted into two different intervals that spanned the streambed. Thus, the size of those intervals was increased to 27. This procedure ensures that the last point in the cross section is kept and, for this example, the 497th point (last point) was kept ($1+(26 \times 17)+(27 \times 2)=497$).

Genetic Algorithm Reduction

Reducing the number of data points in a cross section is a non-linear combinatorial problem and, therefore, is well suited for heuristic algorithms such as genetic algorithm (GA). GAs mimic the natural selection and survival of the fittest and are well suited for solving combinatorial optimization problems in which there is a large set of candidate solutions (Fisher, 2013). Koza (1992, p. 18) provides the following definition of a GA:

The genetic algorithm is a highly parallel mathematical algorithm that transforms a set (population) of individual mathematical objects (typically fixed-length character strings patterned after chromosome strings), each with an associated fitness value, into a new population (i.e., the next generation) using operations patterned after the Darwinian principle of reproduction and survival of the fittest and after naturally occurring genetic operations (notably sexual recombination).

In a GA, a population is represented by a number of individuals called genes (strings of chromosomes). Individuals are produced by 'mating' (crossover of chromosomes) two individuals together and 'mutating' a chromosome. The fittest individuals in the new population are selected to breed and mutate again, passing their genetic information to their children to create a newer population, and the least fit individuals are discarded. The newer population is then used in the next iteration of the algorithm. This process is repeated until a number of iterations has been reached or the maximum number of consecutive iterations without any improvement to the best fit individual is exceeded. Note that each individual is a solution to the problem. In essence, the GA represents an "intelligent"

exploitation of the search space in a random fashion to solve a problem. A more complete discussion on genetic algorithms is given in Goldberg (1989), Grefenstette (1990), Davis (1991), and Mitchell (2002).

The GA developed by Berenbrock (2006b) was used in this study to reduce cross-section point data because it is easy to use, fast, and preserves the original data. The fitness function in Berenbrock's (2006b) GA, however, is biased because the x -value (distance) for each data point does not contribute to the fitness function—only the y -values (elevation) do. The function does not account for the varying distances between the data points (irregularly spaced data) and, thus, data points need to be regularly spaced along a cross section. However, most cross sections are composed of irregularly spaced points. To account for irregularly spaced points, the fitness function was modified to calculate the area between the original and reduced cross sections—noted as the area between the cross sections (ABC). Thus, ABC accounts for the contribution from both x and y values. The value of ABC is always positive, regardless of how the cross sections cross one another. To solve this mathematically, the absolute value of ABC is employed and is denoted as $|ABC|$. The modified two-conditional fitness function is as follows:

$$f(i) = \begin{cases} \sum_{j=1}^n |ABC_j| & \text{if } inc_i \leq plimit \\ \left[\left(\sum_{j=1}^n |ABC_j| \right) + 1 \right] \times 10^{(inc_i - plimit)} & \text{otherwise} \end{cases} \quad (1)$$

where

$f(i)$	is the value of fitness for individual i ;
n	is the number of data points in the original dataset;
ABC_j	is the area between the cross sections, original and reduced, for trait j ;
inc_i	is the number of included data points in individual i ; and
$plimit$	or point limit is the maximum number of points to be included in the reduced dataset.

Each individual in the GA represents a reduced cross section, and the fitness of an individual (thus, a fitness of traits or data-point combinations) is represented as a value from the fitness function. The GA minimizes the fitness function, *Minimize* $f(i)$, to identify the best-fit or optimal individual from all possible data-point combinations.

A sample calculation of fitness for a hypothetical cross section is shown in figure 1. Reduced cross sections are composed of included and excluded points. Included points are data points that are kept from the original cross section, and excluded points are data points that are discarded from the original cross section. For the reduced cross section shown in figure 1, the included data points are at points 1, 2, 5, 6, 9, and 11, and excluded data points are at points 3, 4, 7, 8, and 10. The area between the cross-sections (ABC), original

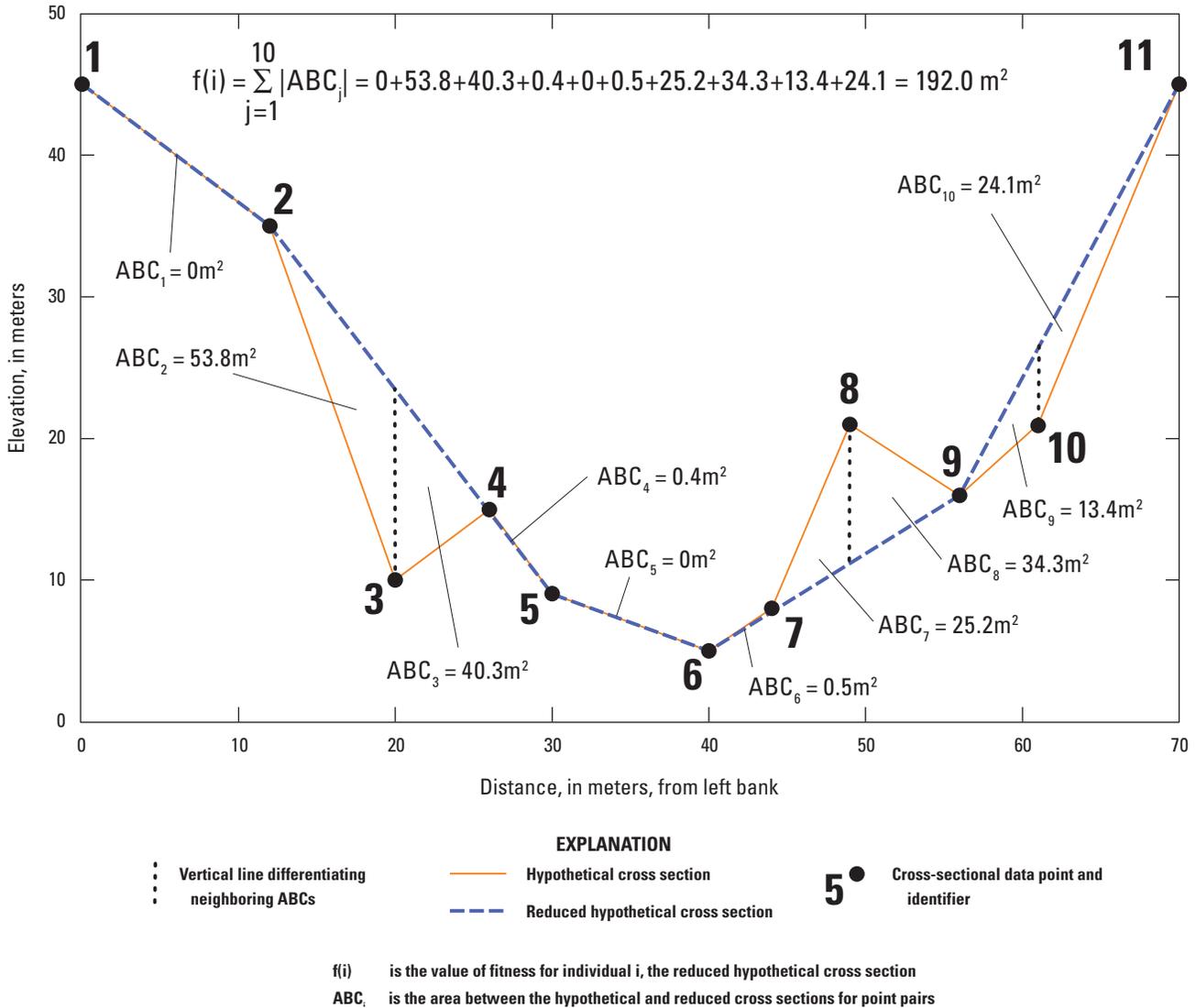


Figure 1. Fitness calculation for a hypothetical cross section. Fitness is calculated according to equation 1 as shown at the top of the figure. m^2 , square meters.

and reduced, is calculated for each closest pair of points. For the first pair of points 1–2, ABC is calculated to be 0 square meters (m^2) because the data points for both cross sections are the same. For the second pair (2–3), ABC is calculated to be 53.8 m^2 . For point pairs 3–4 and 4–5, ABC is calculated to be 40.3 m^2 and 0.4 m^2 , respectively. For point pairs 5–6, ABC is calculated to be 0 m^2 because the data points for both cross sections are the same. For point pairs 6–7, 7–8, 8–9, 9–10, and 10–11, ABC is calculated to be 0.5 m^2 , 25.2 m^2 , 34.3 m^2 , 13.4 m^2 , and 24.1 m^2 , respectively. The total ABC is 192.0 m^2 , which is the fitness value for this reduced hypothetical cross section. Fitness serves to aggregate the errors of an individual into a single measure. It is a good measure of accuracy, but only between other individuals in the population, as it is scale dependent.

To validate the modifications made to the GA, the 10 cross sections that were used to validate the original GA (Berenbrock, 2006b) were used. The GA was run with the same parameter values as in the original GA. The best fitness results are shown in table 1. The reduction method used by Barton and others (2004), however, was based on selecting points every 1 to 2 meters (m), plus additional user-specified points to capture any important missing topography as determined from visual inspection (Berenbrock, 2006b, p. 388); this results in a variable number of reduced data points per cross section (table 1, column 3), rather than a fixed number of points, as described by the standard-reduction method. To compare these methods, the fitness for the best fit reduced cross sections from Barton and others (2004) and original GA (Berenbrock, 2006b) were recalculated by using equation 1 and presented in table 1.

Table 1. Comparison of best fitness between several reduction methods for 10 cross sections on the Kootenai River, Idaho.

[GA, genetic algorithm]

Cross section	¹ Number of data points in original dataset	¹ Number of reduced data points (npr)	Percentage reduction of data points	² Best fitness, in square meters			GA method percentage lower than	
				³ Barton and others (2004) method	³ Original GA (Berenbrock, 2006b) method	GA method	Barton and others (2004) method	Original GA (Berenbrock, 2006b) method
107.658	886	126	85.8	4.26	3.87	2.52	40.8	34.9
151.438	987	181	81.7	7.02	14.08	3.53	49.7	74.9
152.019	497	117	76.5	6.37	8.27	3.47	45.5	58.0
154.972	2,521	444	82.4	8.72	10.27	5.69	34.7	44.6
163.027	1,723	123	92.9	1.34	3.57	1.05	21.6	70.6
185.394	548	94	82.8	1.10	0.92	0.56	49.1	39.1
199.727	762	81	89.4	0.96	1.37	0.51	46.9	62.8
212.227	1,534	126	91.8	1.12	1.15	0.60	46.4	47.8
216.622	696	86	87.6	1.96	2.99	1.16	40.8	61.2
219.881	2,024	130	93.6	1.63	5.63	1.36	16.6	75.8
Average =			86.4				39.2	57.0

¹Barton and others (2004).

²Calculated from equation 1, which is based on the area between the original and reduced cross sections (ABC).

³To compare these methods, the fitness for each best fit reduced cross section from Barton and others (2004) and original GA (Berenbrock, 2006b) were recalculated using equation 1.

The best fitness values for the GA runs were consistently less (better performance, more accurate) than those from Barton and others (2004) and original GA (Berenbrock, 2006b) methods. On average, the GA fitness was 39.2 percent and 57.0 percent less than Barton and others (2004) and original GA methods, respectively. These results indicate that the GA cross sections are significantly more (better) defined or representative of the original cross section than cross sections by the other two reduction methods. Note that the original GA fitness values were usually greater (lower performance) than Barton and others (2004) fitness values because the original GA fitness function was optimized for regularly spaced data, not irregularly spaced data as constitute most cross sections.

The GA-reduction method preserves the detailed character of the original cross section better than Barton and others (2004) and original GA (Berenbrock 2006b) methods. The GA cross sections were generally more defined where the original cross section had more topographic variability. The GA cross section matched the original cross section quite well, especially in the relatively smooth areas. The point density in the relatively smooth areas was far more reduced in the GA section than in Barton and others (2004) and original GA sections. Conversely, the point density in the relatively rough areas (topographic variability) was increased more in the GA section than in Barton and others (2004) and original GA sections. For the most part, the detailed character of the original cross section was better preserved by using the GA

method than using the other two reduction methods. Therefore, the modification that was made to the GA method is the more appropriate genetic algorithm-reduction method and was used throughout this study.

Comparison of Reduction Results

To evaluate the effect on cross-section geometry and steady flow profiles, comparisons from the standard-reduction method and GA-reduction method were conducted. The comparisons included the 10 cross sections from the Kootenai River in table 1. The original cross sections were reduced in size to 10, 20, and 30 data points using both standard-reduction and GA-reduction methods. Harrelson and others (1994) determined that at least 20 data points are needed in a cross section to accurately describe the character of the channel. The reduction to 30 data points was selected because a greater amount of data points are needed if the cross section is quite broad or complex, such as the case with braided channels. The reduction to 10 data points was selected because it is one-half of the recommended minimum (Harrelson and others, 1994). Results from the reductions’ visual and hydraulic analyses are presented in the following sections. The practical consequences from both reduction methods are investigated in the “Hydraulic Modeling Analysis of Reduction Methods” section by the use of one-dimensional (1-D) steady-flow profiles.

Visual Analysis of Cross-Section Reductions

After reducing the cross-section data, the reduced datasets were viewed graphically and compared to the original data. Although the analysis of fitness is useful as given in [table 1](#), it is a black-box approach that reports aggregate results without providing an understanding of the spatial details and does not distinguish which parts of the cross section are causing the error. Visual analysis provides further insight regarding the spatial distribution of error for the two reduction methods.

Synder and Minshall (1996) identified three geomorphic reaches in the Kootenai River—a meander reach, a braided reach, and a canyon reach. Barton and others (2005) defined a fourth geomorphic reach—a straight reach. The meander reach is a single channel with gentle bends. The streambed consists primarily of fine sand. Water depths usually exceed 12 m, and the water-surface slope is about 2×10^{-5} meters per meter (m/m), less than one-twentieth the slope in the braided reach. Sand dunes—as high as 1.4 m and as long as about 23 m (Barton and others, 2005)—also occur throughout the meander reach. The straight reach is a transitional reach between the meander and braided reaches, and its streambed consists primarily of sand, gravel, and cobbles. The braided reach usually consists of multiple channels, and the streambed is composed primarily of gravel and cobbles. Water depths usually are less than 2 m, and water-surface slope is about 4.6×10^{-4} m/m. The canyon reach consists of a long, straight single channel with steep canyon walls and is incised into bedrock. The streambed consists primarily of cobbles and boulders. Water depths are usually about 6 m, and water-surface slope is about 3×10^{-4} m/m.

Examples for each geomorphic or channel type—meander, straight, braided, and canyon (Czuba and Barton, 2011)—are shown in [figure 2](#). Cross-section 107.658 is in the meander reach, cross-section 152.019 is in a straight reach, cross-section 154.972 is in a braided reach, and cross-section 163.027 is in a canyon reach. The effects of standard data-point reduction on cross-sectional shape are shown in [figures 2A, 2C, 2E, and 2G](#), and effects of GA reduction on cross-sectional shape are shown in [figures 2B, 2D, 2F, and 2H](#).

At first glance, the reduced cross sections composed of 10 data points showed that it was the worst shaped case for both reduction methods. The graphs in [figure 2](#) indicate that more than 30 points were needed when using the standard-reduction method, whereas 20 to 30 data points were adequate when using the GA-reduction method, except for cross-section 154.972 ([fig. 2D](#)). The GA-reduced cross sections approximated the shape of the original cross sections better than the standard-reduced cross sections for the same number of reduced data points (npr). The standard-reduction method produced greater bank errors than the GA method; also there were greater streambed errors in the braided and straight reaches. At cross-section 154.972 (braided reach), a greater number of data points were needed for both reduction methods because there was much more topographic variability in this cross

section than in the other cross sections. Additional analysis indicated (not shown in [fig. 2](#)) that about 70 data points were needed to adequately define the shape of cross-section 154.972 when using the standard-reduction method and about 40 points when using the GA-reduction method.

A visual comparison was performed on cross-sectional area plots. Area was used instead of conveyance because the performance of conveyance can be confounded by the uncertainty in coefficients in Manning's flow equation (Chow, 1959). The effects of data-point reduction on cross-sectional area for the cross sections in [figure 2](#) are shown in [figure 3](#). The cross-sectional area curves for the reduced cross sections composed of 10 points were furthest from the original area curves for both reduction methods, and were closest to the original curves for the 30-point cross sections for both methods. Cross-sectional areas for the GA-reduced cross sections approximated the shape of the original cross-sectional areas better than the standard-reduced cross sections for respective data-point reduction. But for cross-section 154.972 (braided reach), none of the reduced area curves for both methods closely agreed with the original area curves ([figs. 3E, F](#)). The original area curves show a break near an elevation of 534 m that is probably caused by a greater amount of topographic variability in the original cross section ([figs. 2E, F](#)) compared to the reduced cross sections. For the reduced cross sections, this departure at the break in slope indicates that there are not enough data points to adequately define the topographic variability in this area.

The performance of the cross-section reduction is quantified by the measured error in the reduced cross section. This error is designated as reduction error (RE) and is defined as the total area between the original and reduced cross-section curves (ABC) divided by the number of reduced data points (npr) and is expressed as follows:

$$RE = \frac{ABC}{npr} \quad (2)$$

The RE normalizes the reduction methods with respect to the number of npr and allows for direct comparison of different cross sections and reduction methods. The RE has units of area per point; for the datasets used in this study, it was square meters per reduced data point (m²/point). The value for RE depends on the number of points in the original and reduced cross sections and the topography of the cross section. The smaller the RE, the greater similarity between the original and reduced cross sections. The larger the RE, the greater the disparity between the topography of the original and reduced cross sections. Equation 2 represents the gain in cross-sectional likeness due to the reduction of the area between the cross-section curves with the increase in the number of reduced data points. The RE values are unique and depend on cross-section topography and the number of data points in the original and reduced cross sections. The accuracy of the cross-section reduction is likely to be sensitive to the quality and quantity of the original data and how it represents the different scales of topography that are present. For example,

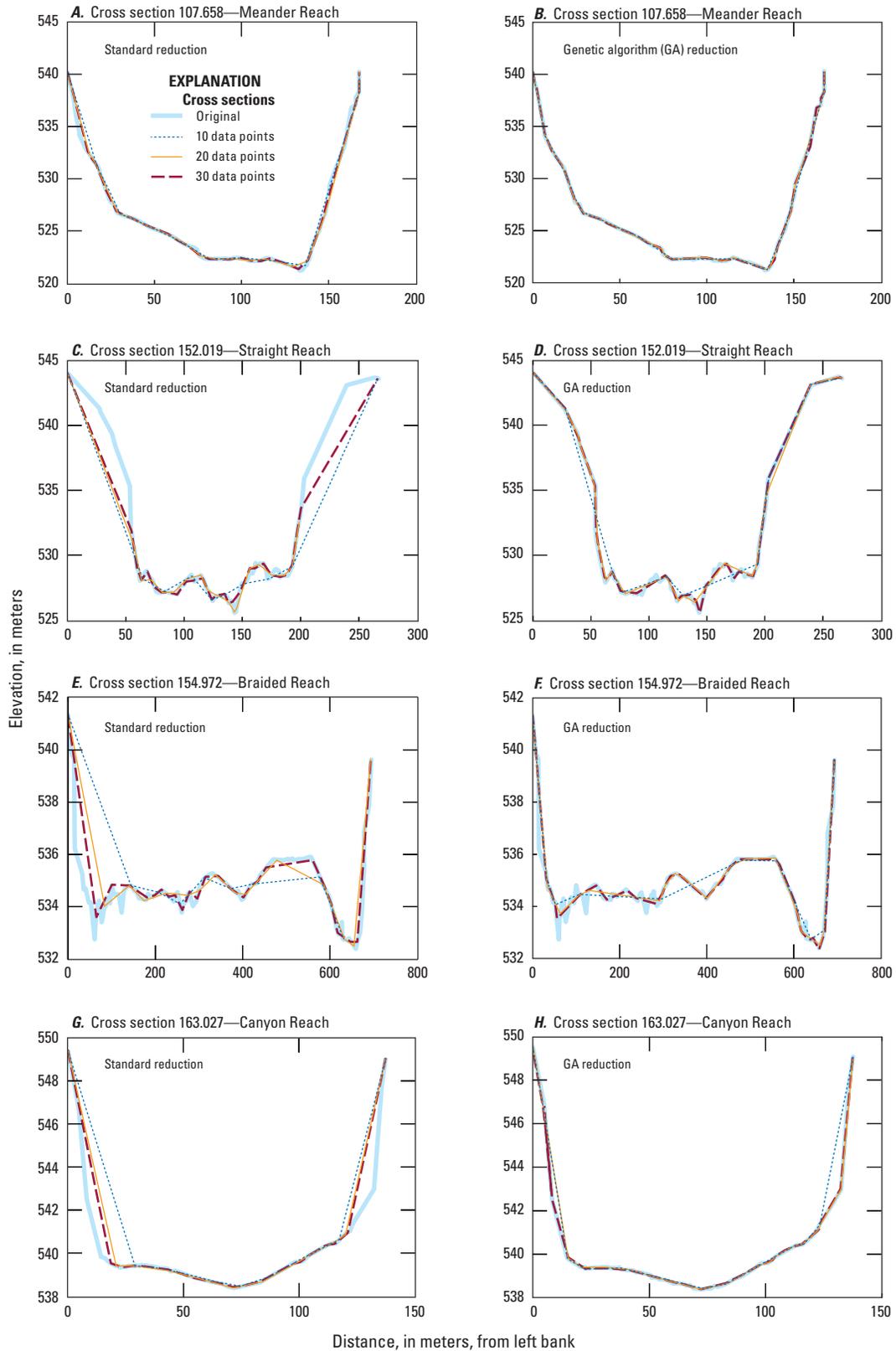


Figure 2. Effects of data-point reduction on cross-sectional shape. *A*, Cross-section 107.658, meander reach, standard reduction. *B*, Cross-section 107.658, meander reach, genetic algorithm (GA) reduction. *C*, Cross-section 152.019, straight reach, standard reduction. *D*, Cross-section 152.019, straight reach, GA reduction. *E*, Cross section 154.972, braided reach, standard reduction. *F*, Cross section 154.972, braided reach, GA reduction. *G*, Cross section 163.027, canyon reach, standard reduction. *H*, Cross section 163.027, canyon reach, GA reduction.

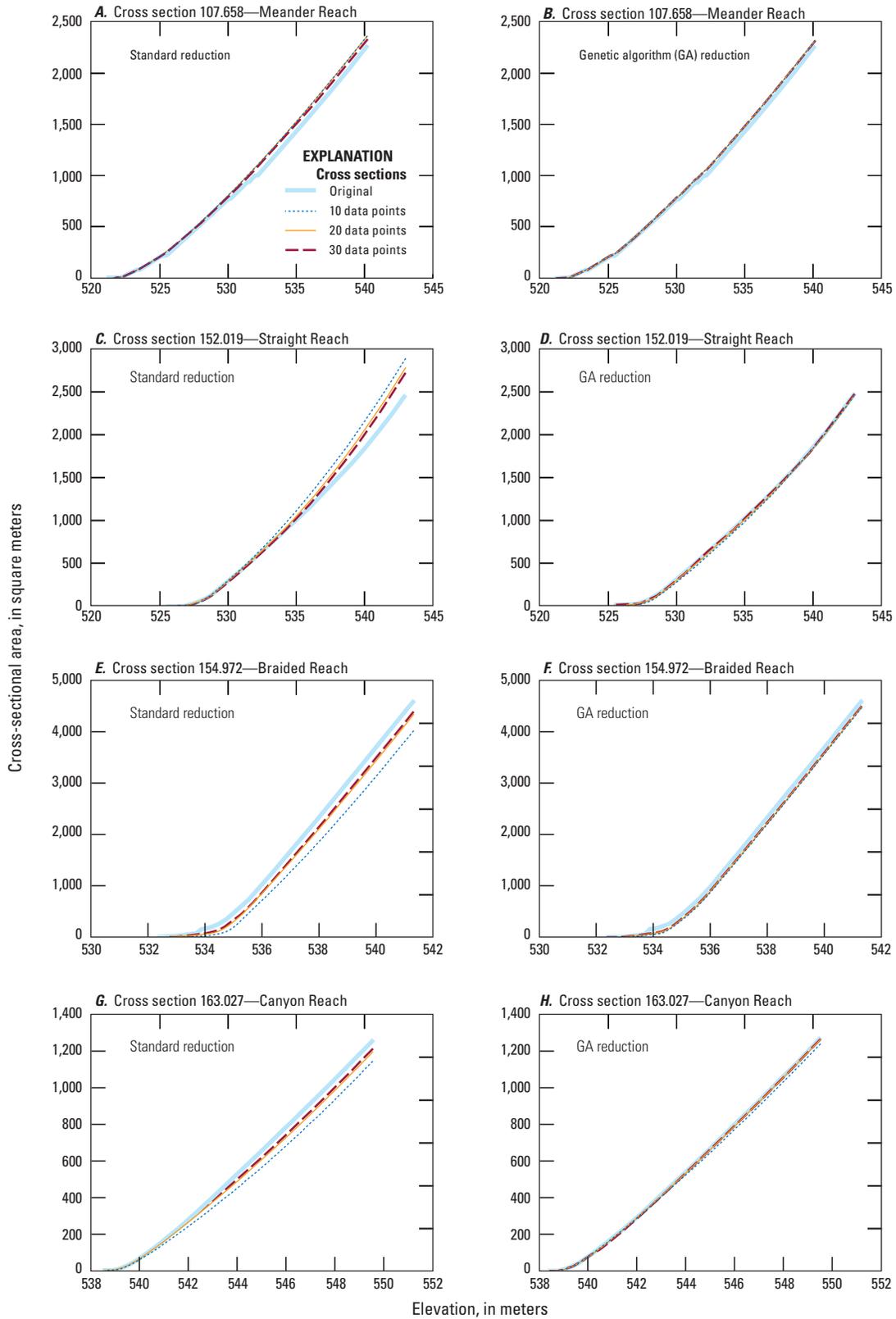


Figure 3. Effects of data-point reduction on cross-sectional area. *A*, Cross-section 107.658, meander reach, standard reduction. *B*, Cross-section 107.658, meander reach, genetic algorithm (GA) reduction. *C*, cross-section 152.019, straight reach, standard reduction. *D*, Cross-section 152.019, straight reach, GA reduction. *E*, cross section 154.972, braided reach, standard reduction. *F*, Cross section 154.972, braided reach, GA reduction. *G*, Cross section 163.027, canyon reach, standard reduction. *H*, Cross section 163.027, canyon reach, GA reduction.

the performance of the cross-section reduction could differ if a topographically complex braided channel was originally surveyed by 40 data points as opposed to 400 data points. This is not an issue for the Kootenai data, given the high density of original data points, but it could be an issue in other studies.

Values of RE for reduced cross sections that contained 10, 20, and 30 data points were calculated for the 10 cross sections on the Kootenai River by using the standard- and GA-reduction methods (table 2). For the GA-reduced cross sections, fitness or ABC was calculated by the GA program. For the standard-reduced cross sections, ABC was calculated using the first part of equation 1, similar to the calculation shown in figure 1. Then, equation 2 was used to calculate the RE for both methods. As shown in table 2, as the number of reduced data points increase, the RE decreases. Also the RE values for the GA-reduced cross sections were always less than the RE values for the standard-reduced cross sections, indicating that the GA-reduced cross sections are more representative of the original data than the standard-reduced cross sections. The RE values for the canyon-reach cross sections were usually consistent with one another for the number of data points for both methods. For the GA method, the canyon cross sections had the lowest RE values, and cross-section 154.972 (braided reach) had the highest RE values. For the standard-reduction

method, cross-section 107.658 (meander reach) had the lowest RE values because the banks in this cross section were gently sloping, thus, reducing its ABC value (fig. 2A). Also, for the standard-reduction methods, cross-section 152.019 (straight reach) had the highest RE values because the method selected only a few points on the banks, thus, causing ABC to be quite large (fig. 2C). Similarly, that is the reason for the large RE values for cross-section 151.438 (straight reach).

The GA program was used to develop RE curves (fig. 4) for the 10 cross sections listed in table 2. Each curve was developed by running the program for a selected number of reduced data points (npr) starting at 10 data points and incrementing by 10 until reaching 100 data points. The program was run 10 times at every npr to ensure that near optimal results were reached. The lowest RE value (calculated from equation 2) at each npr was retained and used to develop the RE curve for every cross section. Also, the ABC and npr values from table 1 for the modified GA were used, and, for seven cross sections, they were used to extend the RE curves beyond an npr of 100. These curves represent near optimal solutions. All RE curves in figure 4 are concave upward to the right. The RE curve for cross-section 154.972 (braided reach) is more upward toward the right than all the other RE curves because its cross section was more complex (greater topographic variability). In contrast, cross sections in the meander and canyon reaches were less complex (less topography variability), causing the RE curves to be in the lower part of the plot. The RE curves for cross sections in the straight reach were between the canyon and braided curves. The RE plot (fig. 4) shows that as topographic variability in a cross section increases, the RE curve will be more upward toward the right in the plot, thereby indicating that increasing the number of points in a cross section needs to be increased to adequately represent the original cross section.

The RE curve also allows one to visually judge where an increase in npr does not result in significant RE reduction (called the point of diminishing returns). This location on a RE curve is at the point of diminishing returns (breakpoint). For this study, the breakpoint's location was determined by a two-phase linear regression where two straight lines are fitted to the data by minimizing the residual sum of squares. Above the breakpoint (to the left on the curve), the RE value increases quite rapidly as the number of data points decrease; below the breakpoint (to the right on the curve), the increase in the number of data points does not lower the RE value as rapidly. The break points for the 10 RE curves are shown in figure 4. The location of the breakpoint for cross-section 107.658 (meander reach) is at an RE of about 0.4 m²/point and at a number of reduced data points of 22 points, about 1 m²/point and 22 points for cross-section 152.019 (straight reach), about 3 m²/point and 29 points for cross-section 154.972 (braided reach), and about 0.4 m²/point and 23 points for cross-section 163.027 (canyon reach). Breakpoints (points of diminishing returns) for cross sections in the meander and canyon reaches were less than those in the straight and braided reaches because their topography is less varied (fig. 4). By 50 data

Table 2. Reduction error (RE) values for 10 cross sections resulting from the standard and genetic algorithm (GA)-reduction methods, Kootenai River, Idaho.

Cross section	Reduction error, in square meters per reduced point					
	Number of reduced data points in cross section					
	Standard method			GA method		
	10	20	30	10	20	30
Meander Reach						
107.658	4.9	1.7	0.8	2.0	0.6	0.2
Straight Reach						
151.438	33.9	9.0	4.3	11.6	2.1	0.9
152.019	49.8	18.1	10.4	10.9	1.7	0.6
Braided Reach						
154.972	41.1	10.4	4.2	18.2	4.7	2.6
Canyon Reach						
163.027	11.8	6.1	3.9	3.5	0.7	0.1
185.394	13.0	3.4	1.5	1.8	0.2	0.1
199.727	11.1	4.0	2.5	3.4	0.1	<0.1
212.227	9.7	2.5	1.5	4.1	0.9	0.2
216.622	7.6	2.7	1.3	3.5	0.6	0.2
219.881	17.0	4.6	2.1	2.7	1.2	0.6

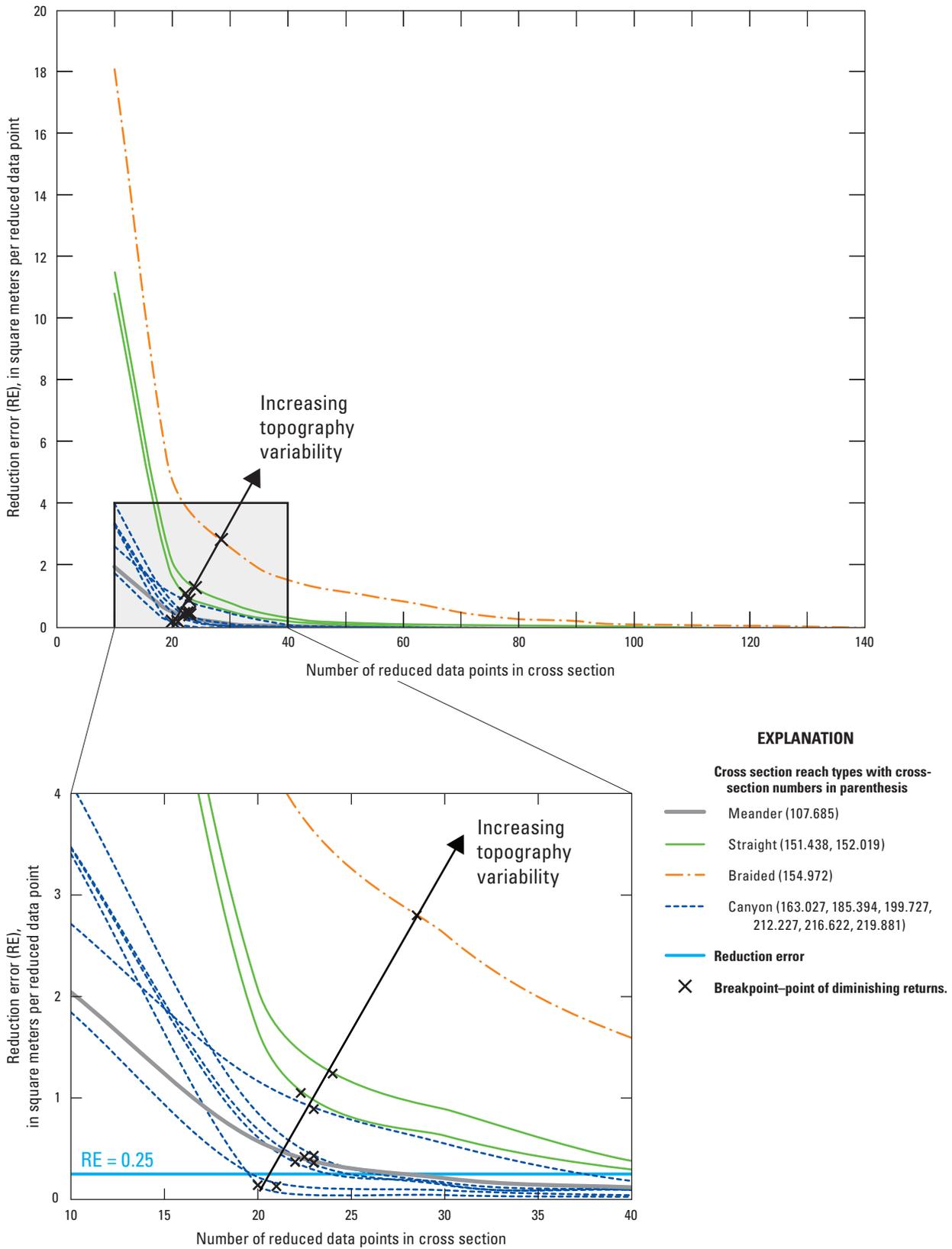


Figure 4. Reduction-error (RE) curves for 10 cross sections resulting from the genetic algorithm (GA)-reduction method, Kootenai River, Idaho.

points, all cross sections except 154.972 (braided reach) had RE values equal to or less than $0.25 \text{ m}^2/\text{point}$ (fig. 4). At this value, the RE curves for the meander, straight, and canyon cross sections are nearly asymptotic to the x-axis. Cross-section 154.972 (braided reach) did not reach an RE value of $0.25 \text{ m}^2/\text{point}$ until 93 data points. This again indicates that cross sections having varied topography such as the braided reach, require more data points to define them adequately.

Hydraulic Modeling Analysis of Reduction Methods

The effects of cross-sectional data-point reduction along a reach were examined on steady-flow water-surface profiles. The 1-D hydraulic-flow model of the Kootenai River in Idaho (Berenbrock, 2005, 2006a) was used to evaluate these effects. The HEC-RAS model, version 4.1 (Brunner, 2010), was used to compute the steady flow profiles (water-surface elevations at cross sections). Only part of the original model (164 cross sections) was used. Cross-sectional data points for only 35 cross sections were reduced—starting at cross-section 149.910 (meander reach) and stopping at cross-section 156.861 (braided reach). This reach was selected because it is the focus of a habitat-restoration project for the recovery of the endangered Kootenai River white sturgeon (*Acipenser transmontanus*) population, and misrepresentation of cross-sectional data could have major effects on computed steady-flow profiles. This reach included meander, straight, and braided reaches. The model was not extended into the canyon reach because reduction in cross-sectional data points had little effect on the computed cross sections, as shown by the low RE values (table 2, fig. 4). The transition from meander to straight occurs near river mile 151, straight to braided at river mile 153.3, and braided to canyon near river mile 161. A total of 30 simulations were run using the combinations of 5 discharges, 3 data-point reduction levels, and 2 reduction methods. The five discharges (170, 283, 850, 1,416, and 1,982 cubic meters per second, or m^3/s) represent the objective discharges, with respect to habitat restoration (Berenbrock, 2006a), and cross sections were reduced to 10, 20, and 30 data points because they span the breakpoint values for cross sections in figure 4. The two data-point reduction methods used were standard and GA.

Results for water-surface elevations are given as differences from the original (fig. 5). Generally, results from the simulations showed that the standard-reduced cross sections had greater water-surface differences from the original than did simulations from GA-reduced cross sections. Also differences were greater for the 10-point simulations than for the 20-point and 30-point simulations in respective methods and discharges. For all simulations, the greatest water-surface elevation differences were at cross sections in the braided reach, probably because the reduced cross sections in that reach were not as accurate in representing the original, as shown by the higher RE values. For the standard-reduction simulations,

effects from the reduced cross section were seen upstream of river mile 161 (not shown on fig. 5), and for the GA-reduced simulations, no effects were seen upstream of river mile 159 (fig. 5). In simulations where the RE value for every cross section was equal to or less than $0.25 \text{ m}^2/\text{point}$, differences in water-surface elevations (steady flow profile) from the original were very small.

Some cross sections had greater differences in water-surface elevation when more data points were used (fig. 5). This is contrary to the paradigm that more data are better. For example, as seen in figure 5A, model results for the standard-reduction method showed that water-surface differences for the 20-point simulation were less than for the 30-point simulation. This occurred in and around cross-section 154.575 (braided reach). The error depends on which cross-sectional data points are captured during reduction and their importance in defining the cross section. The original cross section of 154.575 has four braided channels—three shallow secondary channels and one deep main channel—when total discharge is $170 \text{ m}^3/\text{s}$ (fig. 6). The left most secondary channel was characterized by 6 and 10 data points in the 20-point and 30-point standard-reduced cross sections, respectively (fig. 6A). However, water-surface differences were less for the 20-point simulation than for the 30-point simulation (fig. 5A). This also occurred at other cross sections, at other discharges, and for both reduction methods. Figure 6A shows that both point-reduction levels poorly fit the braided bar landforms in the middle of the cross section, but both point-reduction levels fit the original quite well in the deep main channel (right). The braided bar landforms were reduced in size in the 20-point and 30-point standard reduced cross sections. For the 10-point standard reduced cross section, only two channels were seen—a shallow and wide secondary channel (left), and one deep main channel (right) in which the braided bar landforms from the original cross section are missing.

At discharges of $850 \text{ m}^3/\text{s}$ and greater, model results for the standard-reduction method showed differences in water-surface elevation were less for the 20-point simulations than for the 30-point simulations in the braided reach, specifically in and around cross-section 156.604 (fig. 5C–E). At this cross section, the 20-point and 30-point standard-reduced cross sections excluded data points in the thalweg of the secondary channel, which caused the elevation of the reduced cross section in this area to be 0.5 m higher than the original. However, the 10-point standard-reduced cross section included it, but excluded other secondary thalweg points in the cross section. Even though the 30-point standard-reduced cross sections had more points than the 20-point cross section, the 20-point standard-reduced cross section had points in locations that represent the cross-section topography more accurately.

Generally, the GA-reduced cross sections incorporated the shallow channel thalweg and better represented the original cross section. However, at discharges of 170 and $283 \text{ m}^3/\text{s}$, water-surface differences for the 10-point GA-reduced simulation were less than the 20-point simulation in and around cross-section 154.575 (fig. 5A–B). Although the

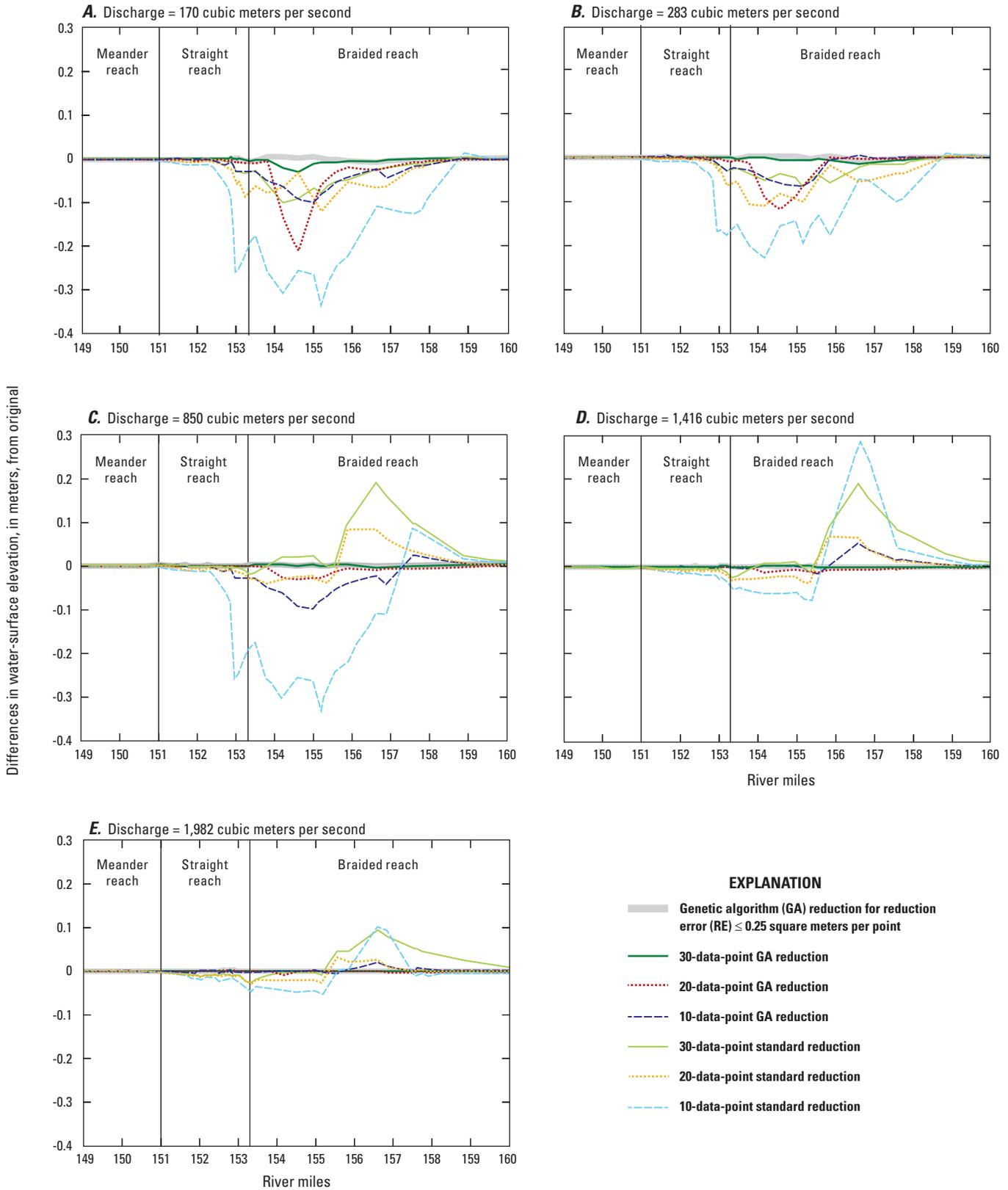


Figure 5. Effects of reduced cross sections on simulated water-surface elevation at five river discharges. *A*, Discharge of 170 cubic meters per second (m^3/s). *B*, Discharge of 283 m^3/s . *C*, Discharge of 850 m^3/s . *D*, Discharge of 1,416 m^3/s . *E*, Discharge of 1,982 m^3/s . \leq , less than or equal to.

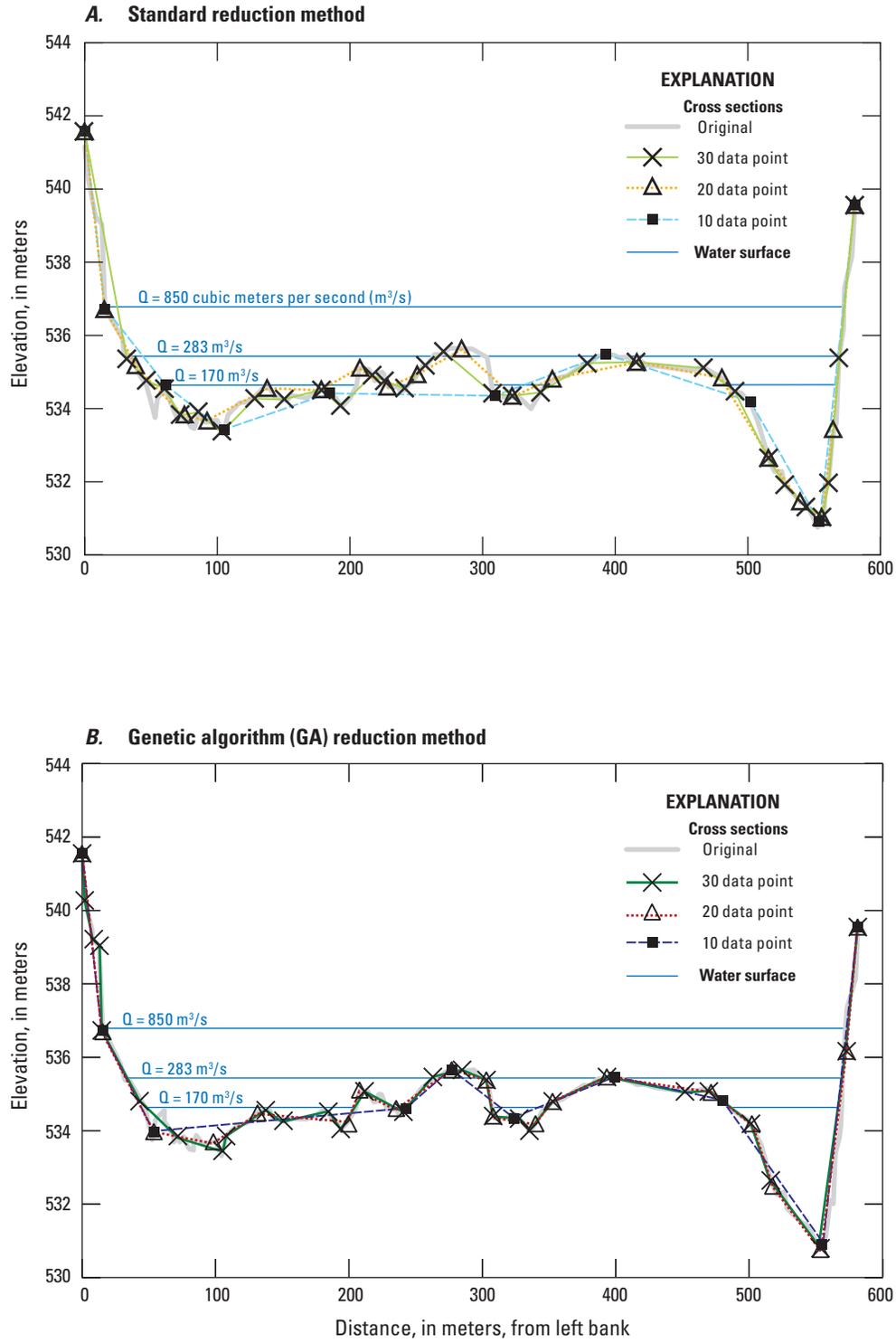


Figure 6. Comparisons between the original and 10 point, 20 point, and 30 point reduced cross sections produced by two reduction methods for cross-section 154.575. *A*, Standard reduction method. *B*, Genetic algorithm (GA) method.

20-point reduced cross section resembled the original cross section better than the 10-point reduced cross section (fig. 6B), it did worse when it came to step-backwater analysis because the water-surface elevation at cross-section 154.178, a reduced cross section just downstream of this section, contained large errors from the original. At cross-section 154.178, there were too few data points in the secondary channels to define the topographic variability adequately, especially the thalwegs in the secondary channels.

Differences from the original water-surface elevation in the steady-flow profiles were small in the GA-reduced cross sections when an RE of less than or equal to $0.25 \text{ m}^2/\text{point}$ was used (fig. 5). The number of reduced data points for this condition ranged from 20 to 40 points, but several cross sections in the braided reach contained more. Cross-section 154.972 (braided reach), for example, had the most reduced data points (93) in order to meet this condition. This value is less than 100, the limit of most hydraulic and sediment transport models such as HEC-2 and HEC-6. Note that the level of acceptable error in the water-surface elevation or flow depth depends on one's intended use (for example, floodplain maps, sediment transport, or fish habitat).

Summary

The genetic algorithm (GA) method is a viable approach for reducing data points in a cross section and produced better results than the standard method it was tested against in this study. The original GA by Berenbrock (2006b) did not account for varying distances between cross-sectional data points. To account for irregularly spaced data points, the fitness function was modified to calculate the area between the original and reduced-cross-section curves. By using 10 cross sections from the Kootenai River, best fitness values were consistently lower (demonstrating better performance) for the GA runs than for the standard-method and original GA runs. On average, the GA fitness was 39.2 percent lower than the standard method, and for several cross sections was nearly 50 percent lower. The GA-reduced cross sections approximated the shape of the original cross sections better than the standard method and, thus, the GA-reduction method should be used over the standard method.

To provide further insight regarding the spatial distribution of error for the two approaches, visual and hydraulic analyses were completed. Visual analysis (graphs) demonstrated that GA-reduced cross sections approximated the shape of the original cross section better than the standard-reduced cross sections. This was also true for the cross-sectional area. An reduction error (RE) was developed to

quantify the difference between the original and reduced cross sections. The RE values decreased as the number of reduced data points increased for both reduction methods, and as expected, RE values were lower (better) for the GA-reduced cross sections than for the standard method. The RE curves were developed for the 10 cross sections on the Kootenai River by using the GA-reduction method, and the break-points (points of diminishing returns) found. For the canyon and meander reaches, the breakpoints (about 20 data points) represent the optimal number of points needed in a cross section. However, many more cross-sectional data points were needed for cross sections in the braided and straight reaches as compared to cross sections in the canyon and meander reaches. Also, additional cross sections from other study areas are needed to draw consistent conclusions regarding the number of cross-sectional data points needed for each reach type. The GA-reduced cross sections matched the shape of the original cross section quite well when the RE was equal or less than $0.25 \text{ m}^2/\text{point}$, the point at which the RE curves become approximately asymptotic to the x-axis. Most cross sections reached this value at 20 to 40 data points, but cross-section 154.972 (braided reach) did not reach it until 93 data points. More complexly shaped cross sections need greater amounts of data points to define them adequately. Depending on the intended use of a cross section, the number of data points depends on the degree of topographic variability of the cross section and the scale of interest.

This study also investigated the practical consequences of errors due to cross-section reduction on steady-flow profiles. Thirty-five cross sections from the original steady-flow surface-water model of the Kootenai River were used. Cross-sectional data in these cross sections were reduced to 10, 20, and 30 data points for both reduction methods. Results generally indicated that differences were less for cross sections developed by the GA-reduction method than by the standard-reduction method. Also, differences from the original water-surface elevation were usually less as the number of data points in cross sections increased—except for some of the reduced cross sections in the braided and straight reaches. The exception is contrary to the paradigm that more data points are better, and is the result of the standard and GA methods not always having enough points in the secondary channels (braided) to define them adequately. The GA method did not select enough points in the secondary channels because fitness was not bettered (lower value) by doing so. To rectify this problem, the GA needs to be modified so that thalweg points in all channels, the main as well as secondary, are selected. Although the GA method is clearly a major advancement in the reduction of cross-sectional data, there are, of course, limits to its performance. In particular, cross sections having multiple channels, such as braided channels, can be problematic.

References Cited

- Barton, G.J., McDonald, R.R., Nelson, J.M., and Dinehart, R.L., 2005, Simulation of flow and sediment mobility using a multidimensional flow model for the white sturgeon critical-habitat reach, Kootenai River near Bonners Ferry, Idaho: U.S. Geological Survey Scientific Investigations Report 2005–5230, 54 p.
- Barton, G.J., Moran, E.H., and Berenbrock, Charles, 2004, Surveying cross sections of Kootenai River between Libby Dam, Montana, and Kootenay Lake, British Columbia, Canada: U.S. Geological Survey Open-File Report 2004–1045, 35 p.
- Berenbrock, Charles, 2005, Simulations of hydraulic characteristics in the white sturgeon spawning habitat of the Kootenai River near Bonners Ferry, Idaho: U.S. Geological Survey Scientific Investigations Report 2005–5110, 30 p.
- Berenbrock, Charles, 2006a, Simulations of hydraulic characteristics for an upstream extension of the white sturgeon habitat of the Kootenai River near Bonners Ferry, Idaho—A supplement to Scientific Investigations Report 2005–5110: U.S. Geological Survey Scientific Investigations Report 2006–5019, 17 p.
- Berenbrock, Charles, 2006b, A genetic algorithm to reduce stream channel cross section data: *Journal of the American Water Resources Association*, v. 42, no. 2, p. 387–394.
- Brunner, G.W., 2010, HEC-RAS, River analysis system hydraulic reference manual: U.S. Army Corps of Engineers Hydrologic Engineering Center, CPD-69, January 2010, version 4.1, 417 p.
- Burnham, M.W., and Davis, D.W., 1990, Effects of data errors on computed steady-flow profiles: *American Society of Civil Engineers Journal of Hydraulic Engineering*, v. 116, no. 7, p. 914–929, DOI: [http://dx.doi.org/10.1061/\(ASCE\)0733-9429\(1990\)116:7\(914\)](http://dx.doi.org/10.1061/(ASCE)0733-9429(1990)116:7(914)).
- Castellarin, A., Di Baldassarre, G., Bates, P.D., and Brath, A., 2009, Optimal cross-sectional spacing in Preissmann Scheme 1D hydrodynamic models: *Journal of Hydraulic Engineering*, v. 135, no. 2, p. 96–105.
- Chow, V.T., 1959, *Open-channel hydraulics*: New York, McGraw-Hill, 680 p.
- Columbia Basin Inter-Agency Committee, 1965, River mile index, Kootenai River, United States, Kootenay River, Canada, Columbia River Basin, Idaho, Montana, British Columbia: Columbia Basin Inter-Agency Committee, Hydrology Subcommittee, November 1965, 49p [114].
- Cook, Aaron, and Merwade, Venkatesh, 2009, Effect of topographic data, geometric configuration and modeling approach on flood inundation mapping: *Journal of Hydrology*, v. 377, p. 131–142.
- Czuba, C.R., and Barton, G.J., 2011, Updated one-dimensional hydraulic model of the Kootenai River, Idaho—A supplement to Scientific Investigations Report 2005–5110: U.S. Geological Survey Scientific Investigations Report 2011–5128, 36 p.
- Davis, L., 1991, Hybridization and numerical representation, *in* Davis, L., ed., *The handbook of genetic algorithms*: United Kingdom, Van Nostrand Reinhold, p. 61–71.
- Federal Emergency Management Agency, 2007, Guidelines and specifications for study contractors: Federal Emergency Management Agency, Publication 37, Washington, D.C., U.S. Government Printing Office, p. 174, accessed August 25, 2011, at <http://www.fema.gov/library/viewRecord.do?id=2238>.
- Fisher, J.C., 2013, Optimization of water-level monitoring networks in the eastern Snake River Plain aquifer using a kriging-based genetic algorithm method: U.S. Geological Survey Scientific Investigations Report 2013-5120 (DOE/ID-22224), 74 p., <http://pubs.usgs.gov/sir/2013/5120>.
- Goldberg, D.E., 1989, *Genetic algorithms in search, optimization and machine learning*: Reading, Mass., Addison Wesley, 412 p.
- Grefenstette, J.J., 1990, Genetic algorithms and their applications, *in* Kent, A., and Williams, J.G., eds., *The encyclopedia of computer science and technology*, 21 (Supp. 6): New York, Marcel Dekker, p. 139–152.
- Harrelson, C.C., Rawlins, C.L., and Potyondy, J.P., 1994, *Stream channel reference sites—An illustrated guide to field technique*: U.S. Department of Agriculture, Forest Service, Rocky Mountain Forest and Range Experiment Station, Gen. Tech. Rep. RM-245, Fort Collins, Colo., 61 p.
- Koza, J.R., 1992, *Genetic programming—Vol. 1, on the programming of computers by means of natural selection (complex adaptive systems)*: London, A Bradford Book, 819 p.
- Mitchell, M., 2002, *An introduction to genetic algorithms*: Cambridge, Mass., The MIT Press, (8th printing), p. 209.
- Moran, E.H., and Berenbrock, Charles, 2003, GPS—Time saver and functional: U.S. Geological Survey Western Water Watch, v. 1, no. 1, p. 6–7.
- Pasternack, G.B., Wang, C.L., and Merz, J.E., 2004, Application of a 2D hydrodynamic model to design of reach-scale spawning gravel replenishment on the Mokelumne River, California: *Journal of River Research and Applications*, v. 20, p. 205–225.

Samuels, P.G., 1989, Backwater length in rivers: Proceedings of Institution of Civil Engineers, pt. 2, no. 87, December, p. 571–581.

Snyder, E.B., and Minshall, G.W., 1996, Ecosystem metabolism and nutrient dynamics in the Kootenai River in relation to impoundment and flow enhancement of fisheries management: Idaho State University, Stream Ecology Center, variously paginated.

Traver, R.G., and Miller, A.C., 1993, Open channel interpolation of cross sectional properties: Journal of the American Water Resources Association, v. 29, no. 5, p. 767–776. DOI: <http://dx.doi.org/10.1111/j.1752-1688.1993.tb03236.x>.

Travis, Quentin, and Lokey, Burke, 1999, Minimizing errors due to cross-section point reduction, *in* American Society of Civil Engineers Proceedings of the 26th Annual Water Resources Planning and Management Conference (WRPMD), Tempe, Ariz., June 6–9, 1999, Wilson, E.M., ed.: Reston, Va., ASCE, 978-0-7844-0430-0 or 0-7844-0430-5, 1999, chap. 3G64, p. 1-13, DOI: [http://dx.doi.org/10.1061/40430\(1999\)142](http://dx.doi.org/10.1061/40430(1999)142).

Publishing support provided by the U.S. Geological Survey
Science Publishing Network, Sacramento Publishing Service Center

For more information concerning the research in this report, contact the

Director, California Water Science Center
U.S. Geological Survey
6000 J Street, Placer Hall
Sacramento, California 95819
<http://ca.water.usgs.gov>

