Vector Approaches to Generalizing Faults and Polygons in 1:24,000 Geologic Maps:
Santa Rosa, California, Case Study

Craig A. McCabe
Department of Geography, The Pennsylvania State University, University Park, PA 16802
E-mail: cam509@psu.edu

1. Introduction

The city of Santa Rosa is situated in the picturesque heart of California wine country, approximately 100 km north of San Francisco. The U.S. Geological Survey recently completed field mapping efforts conducted between 2003-2006 for the Santa Rosa 7.5 minute quadrangle, and the resulting 1:24,000 map (McLaughlin et al., 2007, in review) will provide important information to the public and private sectors on the geologic processes and structures at and below the Earth's surface.

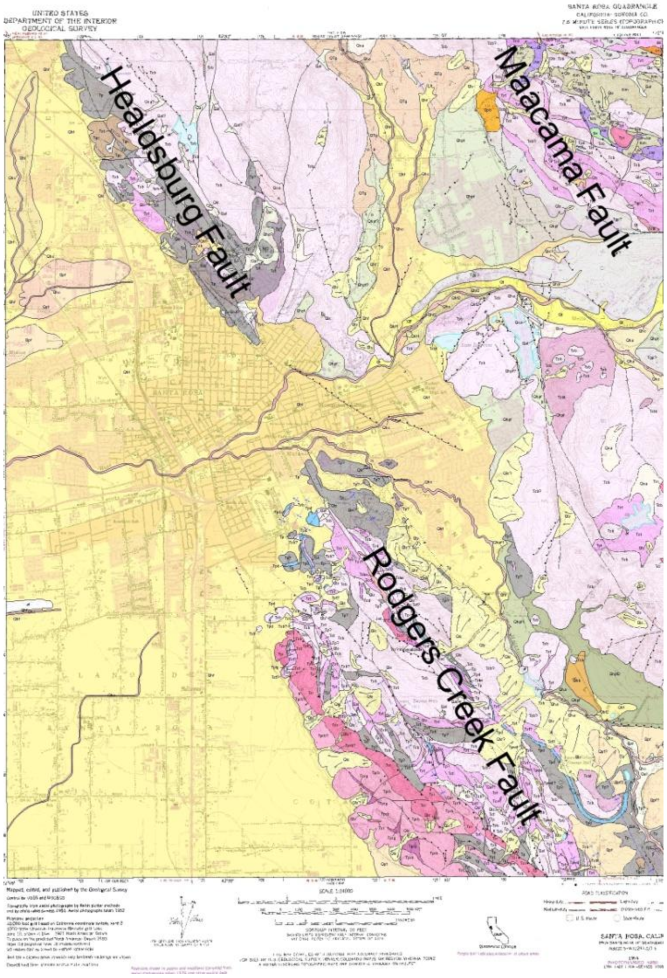


Figure 1. Major faults in the Santa Rosa 1:24,000 quadrangle.

Geologic mapping and geophysical modeling studies conducted by the USGS have shown that Santa Rosa's geologic framework is unique, and may account for the significant damage experienced after the M7.9 1906 San Francisco earthquake, despite being some 40 km to the east of the San Andreas fault rupture (McPhee et al., 2007). Santa Rosa was also shaken severely following consecutive M5.6 and M5.7 earthquakes on the nearby

Healdsburg fault in 1969, resulting in damage that was both extremely concentrated and greater than expected, considering the relatively moderate magnitude of the earthquakes. Part of the damage can be explained by the configuration of 3 major strike-slip faults within the 1:24,000 quadrangle: the Healdsburg, Rodgers Creek, and Maacama faults (Figure 1). These form a zone of horizontal differential stress that results in what is called a strike-slip or pull-apart basin. This produces a series of deep basins or "bowls" in the dense, Mesozoic basement rock that have filled with less-dense Quaternary fluvial sediments over millions of years (McLaughlin et al., 2007, in review). Ground motion simulations of the 1906 San Francisco earthquake conducted by the USGS (Aagaard et al., 2006) have shown that the density contrast of material within these basins beneath Santa Rosa effectively amplifies the intensity of shaking during earthquakes, and increases the susceptibility of Santa Rosa to severe damage during seismic events (Figure 2).

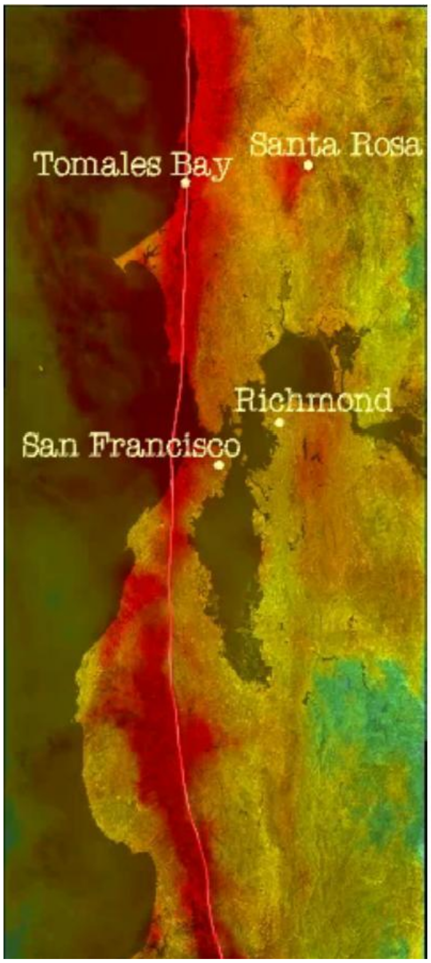


Figure 2. 1906 SF Earthquake ground motion simulation (Aagaard et al., 2006). Red areas indicate high amplitude ground motion.

Having assisted in the Santa Rosa mapping efforts, geophysical data collection (Langenheim et al., 2006) and 3D fault and geologic unit modeling, I was personally interested in working further with the data, and most importantly, had access to the ArcGIS geodatabase. On many occasions while working at the USGS, I was tasked with creating generalized, region-wide geologic maps of the North Bay Area at scales of up to 1:250,000 using original 1:24,000 digital map data. Generalizing features on geologic maps poses specific challenges: how do you create a set of rules that maintains the topological consistency and geologic significance of certain features over others? A strictly geometric approach to generalization neglects the variable levels of "importance" of certain map elements, which is particularly crucial in a geologic setting when some of those elements have the capacity to cause a natural disaster.

With this in mind, I will apply appropriate map generalization techniques learned over the course of this seminar to the Santa Rosa 1:24,000 geologic map through a broad range of scales.

2. Related Work

Map generalization has emerged as an active field of cartographic study in recent years, with a surge in GIS tools aimed at online map services and mobile device applications. Geologic maps, however, have specific topological demands, and literature related to generalization in this context is limited compared to other uses, such as in building, land use, or transportation network generalization.

One feature of geologic maps that make them unique is that the entire map space is covered by a continuous layer of polygons, like a jigsaw puzzle. Overlapping polygons or gaps between polygons are not allowed – every areal unit can be described as being composed of *something*, be it alluvium, water, volcanic breccia, etc. Likewise, no two geologic units can coexist in the same place. Bader and Weibel (1997) dealt with a variety of approaches and algorithms to detect and resolve conflicts in generalizing polygonal maps, including: the rolling ball principle, Delaunay triangulation, and polygon buffering. They concluded that to be effective, additional contextual knowledge was required to further guide the selection and application of these methods.

In the context of database structure, Bonham-Carter and Broome (1998) dealt with geologic maps and the computational aspects of particular data structures. While only a small section of their paper is concerned with map generalization, it hits on a significant point: the type of generalization that is most pertinent to geology is the generalization of attributes, not the simplification of geometry or graphical objects. They refer to this progressive generalization of attributes as *universalization*. The important task is the creation of a clear data model to guide the process through multiple scales. Their proposed method of polygon generalization of rock units also required a somewhat rigorous use of mathematics and knowledge of relative lithologic percentages by unit, which are not available in the Santa Rosa dataset.

A closer synthesis of computational and attribute-driven generalization is given in Downs and Mackaness' *An Integrated Approach to the Generalization of Geologic Maps* (2002). Drawing on the British Geological Survey's need to produce automated 1:250,000 maps from 1:50,000 source data, the authors incorporated a 5-phase methodology, including: structure recognition, process recognition, process modeling, process execution, and data display. Their rule-based, object-oriented approach is powerful in that it looks at faults and polygons together and the relationships between lithologic units to inform generalization decisions. Faults, in particular, were considered closely and ranked by importance based on a combination of their length and the number of rock units existing on either side. This was done in the absence of fault throw (slip rate) or named faults to influence ranking. Unfortunately, their approach utilized specialized software tools built specifically to achieve their research goals.

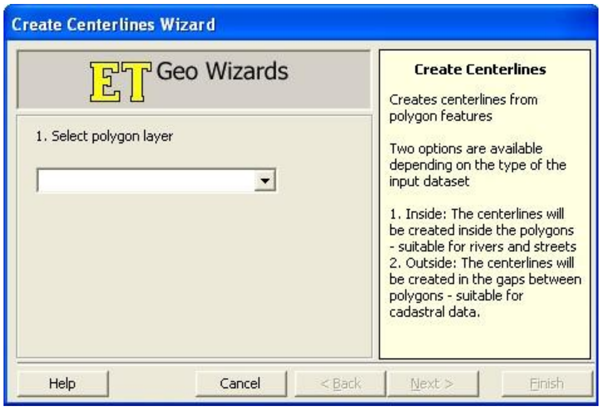
Steiniger and Weibel (2005) elaborate on the generalization of thematic maps by incorporating a 3-step process: structure analysis, generalization, and visualization. The paper builds a conceptual framework that highlights the human element in each step – from the definition of the data hierarchy in the structure analysis, to the selection of generalization tools, and finally to the visual result of the end product. The human expert is a critical component in the whole process to ensure a "good communication of information" for the given map purpose.

I draw on many of these ideas in my generalization approach for the Santa Rosa 1:24,000 geologic map. There is a clear need to devise a robust generalization framework for geologic features that combines both attribute and geometric generalization while preserving the most valuable details of the original source map.

3. Generalization Framework

ESRI's ArcGIS Desktop GIS software is the primary digital mapping tool used by the U.S. Geological Survey for 1:24,000 map creation. In generalizing the Santa Rosa 7.5 minute quadrangle, I will be using tools that are already included in the ArcGIS Toolbox, with the exception of a free plug-in called ET Geowizards (version 9.7) that provides additional functions not available in the standard ArcGIS environment. The use of only ArcGIS tools is both due to my limited programming ability, and for ease of implementation by other ArcGIS users, should anyone want to recreate the steps I have outlined on geologic maps in other locations.

I have chosen to use a combination of both attribute- and geometry-based generalization on the two main features of the geodatabase: fault/contact lines and geologic unit polygons. In the case of both the lines and polygons, this meant coding additional parameters, or *universalizing*, so that as features are represented from large to small scales, progressively more generalized attributes are used to guide the selection of particular map features. This attribute selection will then be supplemented by ArcGIS line simplification routines (point remove, bend simplify), a specialized fault buffering and centerline approach employing the ET Geowizards plug-in (Figure 3), and minimum polygon area cutoffs for the geologic units. Pre-existing line and polygon details were limited to *LTYPE* and *PTYPE* fields in the corresponding attribute tables, whose purpose was for line symbolization and geologic unit colors. In the original map creation process, *LTYPE* values were digitized and coded from the paper map first, and these values determined the line type used to symbolize the relative certainty of fault line and polygon boundary locations. In the case of fault lines, these include: *certain, approximately located, concealed, inferred*, etc. (Figure 4a), with similar descriptions for lines of contact between geologic units. The



1. Select polygon layer

Create Centerlines

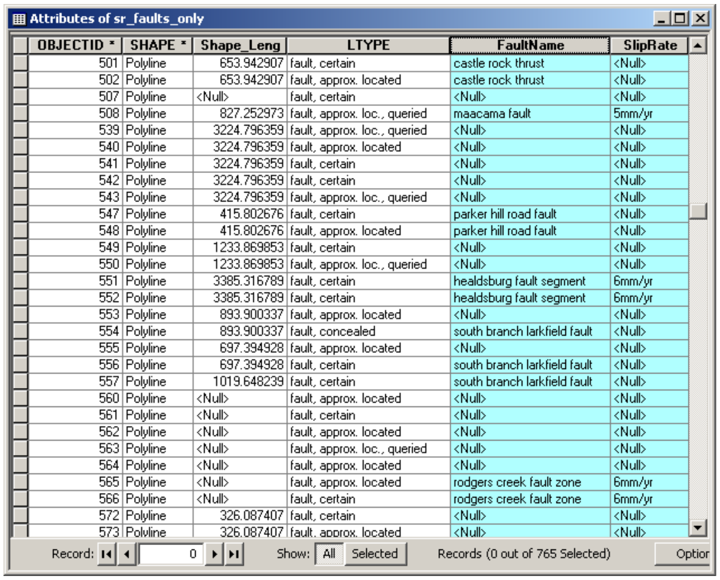
Creates centerlines from polygon features

Two options are available depending on the type of the input dataset

1. Inside: The centerlines will be created inside the polygons - suitable for rivers and streets
2. Outside: The centerlines will be created in the gaps between polygons - suitable for cadastral data.

Figure 3. ET Geowizards ArcGIS plug-in

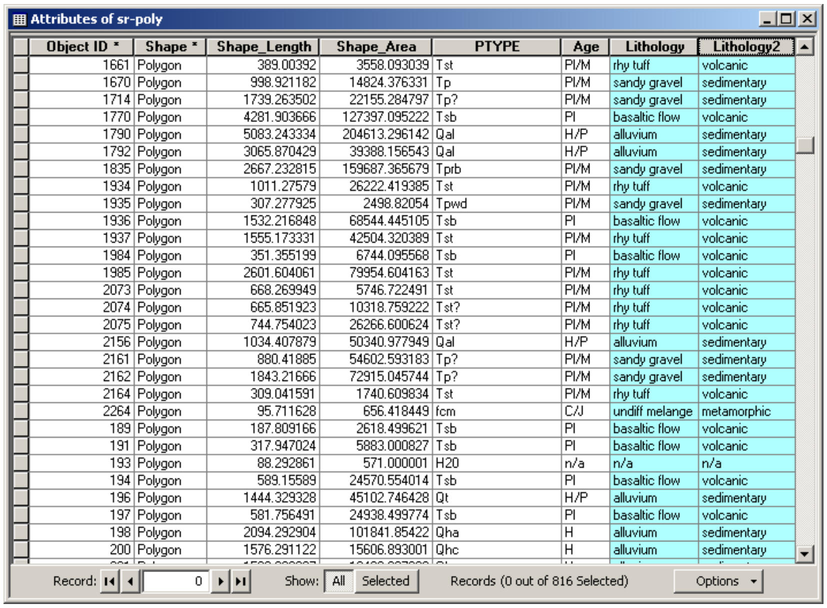
resulting line types showed up on the printed 1:24,000 map as solid, dashed, dashed with question marks, dotted, and so on. From this feature class of attributed lines, polygons were then built, using the “Create Polygons from Lines” topology tool in ArcGIS. The resulting polygon feature class was then coded according to the geologic unit that it represented, and these unit descriptions were added to the newly-created *PTYPE* field. Standard geologic unit abbreviation practices were used for the resulting polygons and result in units names that represent their age and a brief 1-3 letter description (e.g. Qt=Quaternary terrace deposit, Qal=Quaternary alluvium, Qls=Quaternary landslide deposit, Tsb=Tertiary Sonoma basalt, etc.) (Figure 4b).


Attributes of sr_faults_only

OBJECTID *	SHAPE *	Shape_Leng	LTYPE	FaultName	SlipRate
501	Polyline	653.942907	fault, certain	castle rock thrust	<Null>
502	Polyline	653.942907	fault, approx. located	castle rock thrust	<Null>
507	Polyline	<Null>	fault, certain	<Null>	<Null>
508	Polyline	827.252973	fault, approx. loc., queried	maacama fault	5mm/yr
539	Polyline	3224.796359	fault, approx. loc., queried	<Null>	<Null>
540	Polyline	3224.796359	fault, approx. located	<Null>	<Null>
541	Polyline	3224.796359	fault, certain	<Null>	<Null>
542	Polyline	3224.796359	fault, certain	<Null>	<Null>
543	Polyline	3224.796359	fault, approx. loc., queried	<Null>	<Null>
547	Polyline	415.802676	fault, certain	parker hill road fault	<Null>
548	Polyline	415.802676	fault, approx. located	parker hill road fault	<Null>
549	Polyline	1233.869853	fault, certain	<Null>	<Null>
550	Polyline	1233.869853	fault, approx. loc., queried	<Null>	<Null>
551	Polyline	3385.316789	fault, certain	healdsburg fault segment	6mm/yr
552	Polyline	3385.316789	fault, certain	healdsburg fault segment	6mm/yr
553	Polyline	893.900337	fault, approx. located	<Null>	<Null>
554	Polyline	893.900337	fault, concealed	south branch larkfield fault	<Null>
555	Polyline	697.394928	fault, approx. located	<Null>	<Null>
556	Polyline	697.394928	fault, certain	south branch larkfield fault	<Null>
557	Polyline	1019.648239	fault, certain	south branch larkfield fault	<Null>
560	Polyline	<Null>	fault, approx. located	<Null>	<Null>
561	Polyline	<Null>	fault, certain	<Null>	<Null>
562	Polyline	<Null>	fault, approx. located	<Null>	<Null>
563	Polyline	<Null>	fault, approx. loc., queried	<Null>	<Null>
564	Polyline	<Null>	fault, approx. located	<Null>	<Null>
565	Polyline	<Null>	fault, approx. located	rodgers creek fault zone	6mm/yr
566	Polyline	<Null>	fault, certain	rodgers creek fault zone	6mm/yr
572	Polyline	326.087407	fault, certain	<Null>	<Null>
573	Polyline	326.087407	fault, approx. located	<Null>	<Null>

Records (0 out of 765 Selected)

Figure 4a. Fault line attribute table.


Attributes of sr-poly

Object ID *	Shape *	Shape_Length	Shape_Area	PTYPE	Age	Lithology	Lithology2
1661	Polygon	389.00392	3558.093039	Tst	Pl/M	rhy tuff	volcanic
1670	Polygon	998.921182	14824.376331	Tp	Pl/M	sandy gravel	sedimentary
1714	Polygon	1739.263502	22155.284797	Tp?	Pl/M	sandy gravel	sedimentary
1770	Polygon	4281.903666	127397.095222	Tsb	Pl	basaltic flow	volcanic
1790	Polygon	5083.243334	204613.296142	Qal	H/P	alluvium	sedimentary
1792	Polygon	3065.870429	39388.156543	Qal	H/P	alluvium	sedimentary
1835	Polygon	2667.232815	159687.365679	Tprb	Pl/M	sandy gravel	sedimentary
1934	Polygon	1011.27579	26222.419385	Tst	Pl/M	rhy tuff	volcanic
1935	Polygon	307.277925	2498.82054	Tpwd	Pl/M	sandy gravel	sedimentary
1936	Polygon	1532.216848	68544.445105	Tsb	Pl	basaltic flow	volcanic
1937	Polygon	1555.173331	42504.320389	Tst	Pl/M	rhy tuff	volcanic
1984	Polygon	351.355199	6744.095568	Tsb	Pl	basaltic flow	volcanic
1985	Polygon	2601.604061	79954.604163	Tst	Pl/M	rhy tuff	volcanic
2073	Polygon	668.269949	5746.722491	Tst	Pl/M	rhy tuff	volcanic
2074	Polygon	665.851923	10318.759222	Tst?	Pl/M	rhy tuff	volcanic
2075	Polygon	744.754023	26266.600624	Tst?	Pl/M	rhy tuff	volcanic
2156	Polygon	1034.407879	50340.977949	Qal	H/P	alluvium	sedimentary
2161	Polygon	880.41885	54602.593183	Tp?	Pl/M	sandy gravel	sedimentary
2162	Polygon	1843.21666	72915.045744	Tp?	Pl/M	sandy gravel	sedimentary
2164	Polygon	309.041591	1740.609834	Tst	Pl/M	rhy tuff	volcanic
2264	Polygon	95.711628	656.418449	fcm	C/J	undiff melange	metamorphic
189	Polygon	187.809166	2618.499621	Tsb	Pl	basaltic flow	volcanic
191	Polygon	317.947024	5883.000827	Tsb	Pl	basaltic flow	volcanic
193	Polygon	88.292861	571.000001	H20	n/a	n/a	n/a
194	Polygon	589.15589	24570.554014	Tsb	Pl	basaltic flow	volcanic
196	Polygon	1444.329328	45102.746428	Qt	H/P	alluvium	sedimentary
197	Polygon	581.756491	24938.499774	Tsb	Pl	basaltic flow	volcanic
198	Polygon	2094.292904	101841.85422	Qha	H	alluvium	sedimentary
200	Polygon	1576.291122	15606.893001	Qhc	H	alluvium	sedimentary

Records (0 out of 816 Selected)

Figure 4b. Geologic unit polygon attribute table.

For this project, I have opted to represent the original 1:24,000 geologic map at three successive levels of generalization: 1:50,000, 1:100,000, and 1:250,000. 1:50,000 maps continue to be used by many agencies, such as the Army (Topographic Line Maps), and the British Ordnance Survey. It represents a roughly 2x coarser level of detail from the original 1:24,000 map. Reducing the detail by another factor of 2 results in the next level, 1:100,000, which is currently used in USGS 30 x 60 minute map products.

(Figure 5**).** The final level of generalization, at 1:250,000, represents the extreme end of the spectrum, and is used in representing very large areas. Maps at this level of detail would typically be used in conjunction with many other quadrangles, as the 7.5 minute map at 1:250,000 is roughly the same size as a credit card. Map features at this scale are only significant in the context of their surroundings.

Figure 5. Santa Rosa 1:24,000 quadrangle within the greater Napa 30 x 60 minute map area, outlined in black.

In order to use attributes to guide the generalization process through the scales of 1:50k, 1:100k, and 1:250k, additional information had to be created and added to the attribute tables for both lines and polygons. In the case of fault lines, this involved successively high degrees of generalization, progressively omitting features of less importance. Two new attributes were created from *LTYPE* to determine this line progression:

1) *FaultName:* Faults that are named. Faults are given names based both on their "big picture" significance (how they fit into the broader, regional fault framework), or because they are of local importance to the field geologist. Faults that have names are presumed to be more significant than those that do not.
2) *SlipRate:* Whether the fault has a known lateral slip-rate. Slip-rates of major strike-slip faults are calculated based on the lateral offset of geologic units. Recent slip-rates were available for 3 major fault networks that run through the Santa Rosa quadrangle: Healdsburg (6mm/yr), Rodgers Creek (6mm/yr), and Maacama (5mm/yr) (McLaughlin, personal communication, 2007).

Polygon features also gain additional attributes, but instead of using the attributes to progressively omit features through the scales, polygon attributes serve to amalgamate features based on increasingly generalized lithology descriptions. Again, two new levels of attributes were created from the original *PTYPE* descriptions:

1) *Lithology:* A first level of lithology generalization. For example, Qha, Qhb, Qhc, Qal, Qpf in *PTYPE* all fall under the broader umbrella term of "alluvium" and are grouped as such. This strikes a balance between extreme lithologic detail at 1:24,000 and that of *Lithology2:*
2) *Lithology2:* A more aggressive level of lithology generalization, resulting in just 3 categories: Sedimentary, Volcanic and Metamorphic. With the exception of water bodies, all geologic units on the 1:24,000 map can be characterized as falling into one of these three categories.

Table 1 below summarizes the overall framework of attribute generalization for faults and polygons, and associates newly-created attributes with specific scales for representation (also, see the highlighted attribute table columns in Figure 4a & 4b). Geometric generalization will be used on existing *LTYPE* and *PTYPE* values for the 1:50,000 scale, while the new generalized attributes, in addition to traditional geometric generalization, will be used to create the 1:100,000 and 1:250,000 scale maps:

Scale	**Fault Attribute**	**Polygon Attribute**
1:24,000	LTYPE	PTYPE
1:50,000	LTYPE	PTYPE
1:100,000	FaultName	Lithology
1:250,000	SlipRate	Lithology2

Table 1. Attribute generalization framework.

My generalization approach falls under the "star" category, as the operators for each scale function on the original 1:24,000 line and polygon data. The alternate methodology is referred to as a “ladder”, where successive levels of generalization operate on the results of the previous level. The use of a star approach is beneficial in that obtaining a 1:250,000 map does not require that you generate all the intermediate scale levels to get your end product. In the case of attribute-based generalization, the star approach happens to be the only possible method – beyond the 1:50,000 scale, the underlying attributes of each feature change, so the data between scales loses the consistency required for the ladder method.

The final underlying principle of my generalization approach is that no manual manipulation of individual features is involved. All operators function on the entire map area. The only functions which could be considered manual are the initial coding of the attribute generalization levels, and feature selection for minimum area elimination.

4. Implementation

For the sake of simplicity and to better see the effects of my generalization choices, I decided to start the simplification with a pre-processing step that effectively removed many of the additional features and information found on a typical 1:24,000 geologic map. In actual practice, I expect many of these features would be retained through the 1:50,000 and 1:100,000 scales, but for the purposes of this study, I opted to immediately remove the following and focus primarily on the faults and polygons:

> Contact lines, geologic unit labels, leaders, cross-section lines, cross-section annotation, landslide directional arrows, structural axes, strike/dip symbols, up-thrown/down-thrown structural block annotation, rock sample localities, tephrochronology localities, Digital Raster Graphic (DRG) background

4.1 1:50,000

Faults

I decided to use the existing 1:24,000 fault lines with only a symbology change, removing line types (dashed, queried, solid, etc.) and making all fault lines solid. The move from 1:24,000 to 1:50,000 did not prove significant enough to warrant geometric simplification of the fault lines, so they maintain the complexity of the original map. This decision is mainly due to the linear nature of fault features – they are not prone to drastic direction changes and for the most part are composed of straight line segments (Figure 6).

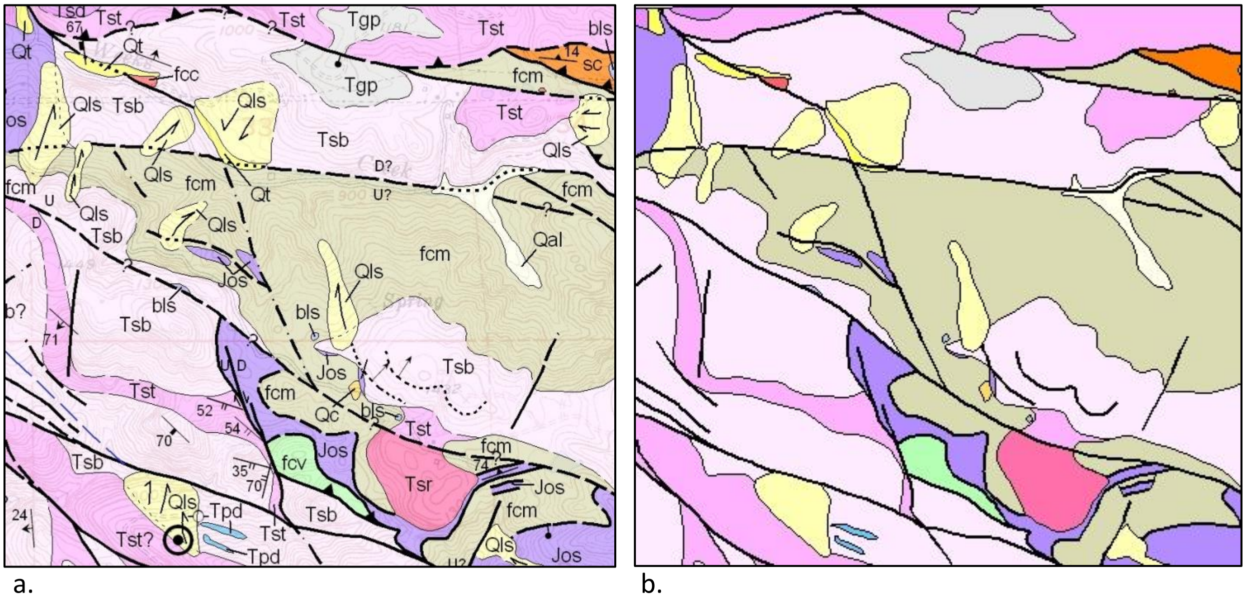


Figure 6. Fault symbology is simplified to solid lines and additional map information is dropped in pre-processing step. Original 1:24,000 map is shown at 1:50,000 on the left (a), resulting generalized map is on the right (b).

Polygons

In removing the contact line types in my line feature class in the pre-processing step, I was left without borders around the polygon units. For all succeeding scales, I applied a solid black outline to each polygon type in the symbology menu to better visually distinguish the unit boundaries. Further generalization of unit polygons at the 1:50,000 scale involved three main operations: feature selection, elimination, and line simplification. The initial selection was dictated by the areal size of individual geologic units in square meters. This value was determined through trial and error, checking the resulting map for readability when printed to scale. I tested areas of 2,000, 4,000, 5,000, 7,500, and 10,000 m^2 before settling on a minimum area of 7,500 m^2 for the 1:50,000 map. With all geologic polygons <7,500 m^2 selected, I used the Eliminate feature in the Data Management toolset of ArcGIS to remove them. The selected polygons were then merged with the neighboring polygon whose shared border was the longest (Figure 7).

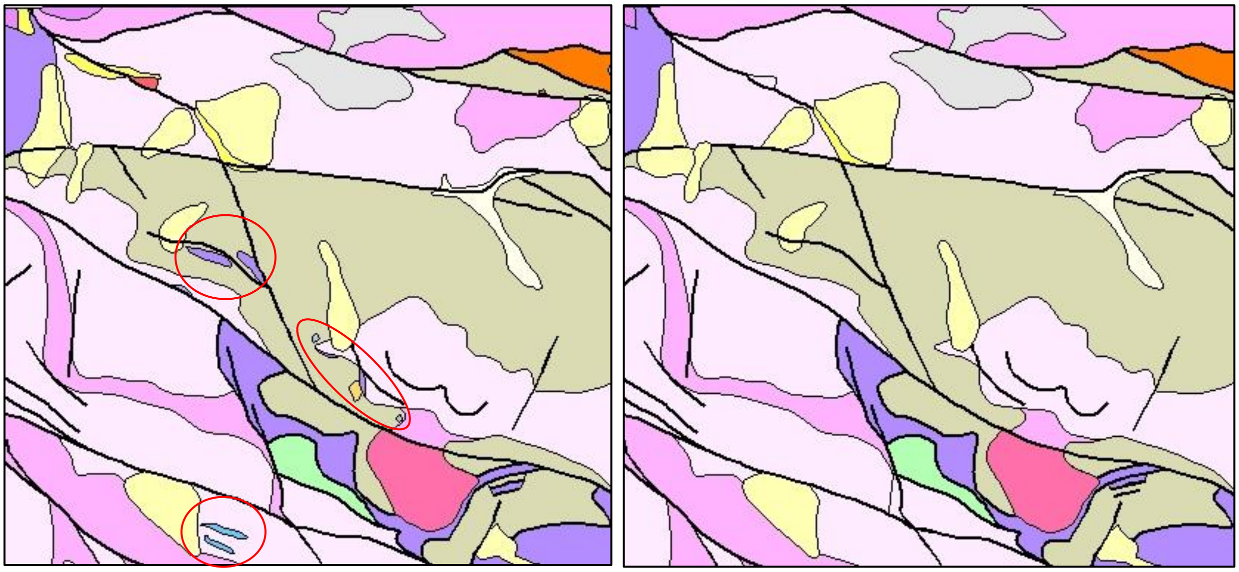


Figure 7. Polygon areas less than 7500 m^2 are incorporated into neighboring units, as in red circled regions.

Polygon line simplification was the last step, and I employed the Simplify Polygon tool in the ArcGIS Data Management toolbox. I opted for Bend Simplify over the alternative, Point Remove, to reduce the angularity of the resulting polygons. Additionally, I turned on RESOLVE_ERRORS in the topology drop-box to ensure that the simplification process would not introduce overlapping polygon areas or voids in the output. The simplification tolerance I chose was, again, determined through trial and error. The final tolerance value I settled on was 50m. Figure 8 shows the effects of the bend simplification on an alluvial feature:

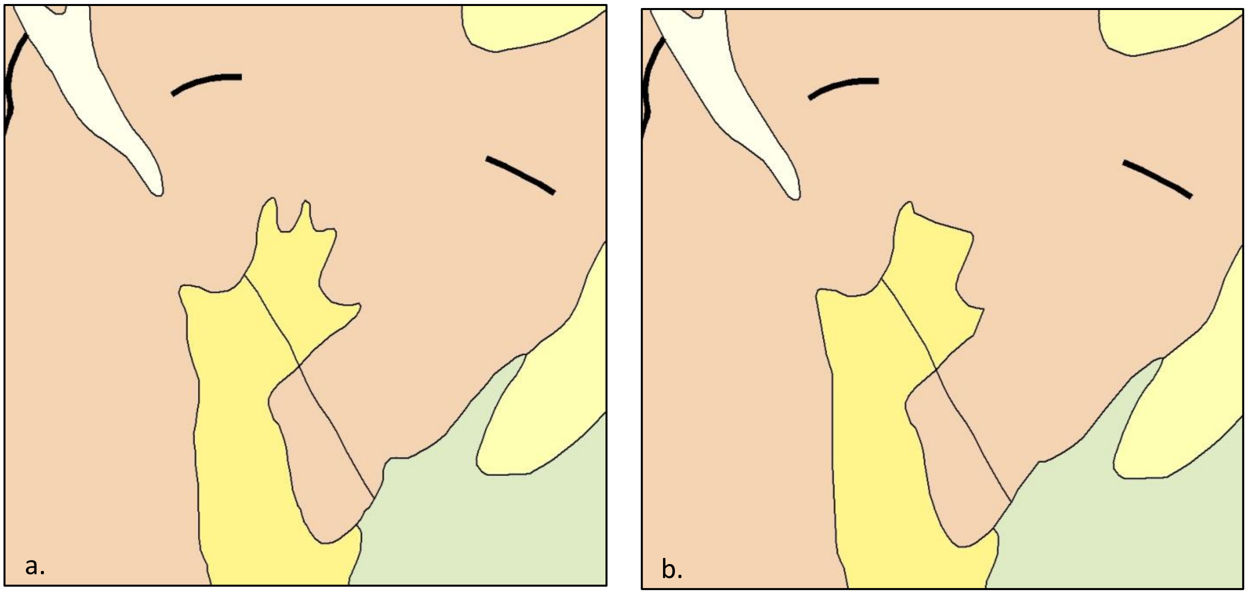


Figure 8. Local effect of bend simplify on a Quaternary unit in 1:24,000 map (a). “Fingers” of the hand are removed (b).

4.2 1:100,000

Faults

Fault line generalization at 1:100,000 employed the first level of attribute generalization, moving from *LTYPE* in previous scales to those faults with names only, shown in Figure 4a and Table 1. Figure 9 shows the Northeast portion of the Santa Rosa quadrangle fault network with the surface geology removed. The black lines represent the faults that were coded with the *FaultName* attribute, while the green lines were not named, and therefore removed at this scale level.

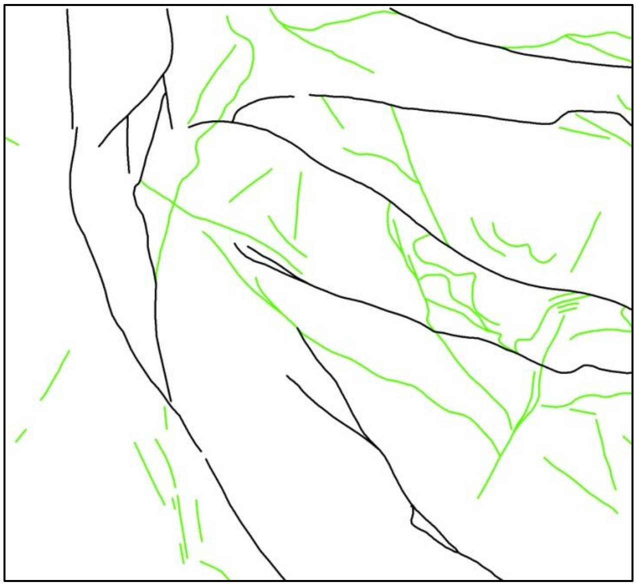


Figure 9. Named faults (black) are retained at 1:100,000, while un-named faults (green) are removed.

Additional geometric generalization was employed on the named faults using the ArcGIS Simplify Line tool. The Point Remove algorithm was selected with a tolerance value of 50 meters.

Figure 10 shows the results of the line simplification. The original lines are represented in red, while the simplified versions are shown in black.

Figure 10. Point remove simplification on named faults (red) results in finished fault lines (black) for 1:100,000 generalization.

Polygons

Polygon simplification at 1:100,000 began with the elimination of areal units that were less than 20,000 m^2 using the Eliminate tool. Again, a qualitative approach was employed in determining the cutoff, testing a number of values until the minimum polygon area was still distinguishable by its color at the printed scale. Subsequently, the first level of lithological attribute generalization, *Lithology*, was employed (Figure 4b, Table 1) reducing the number of geologic units to display. At this point, a new color scheme was introduced, and additional symbolic generalization was used, for example, grouping basaltic and andesitic/rhyolitic volcanics into separate color schemes, while maintaining the underlying differentiation in the attribute table. (Figure 11).

Polygons that shared borders and also had the same *Lithology* attribute were then combined using the ArcGIS Dissolve tool, cleaning up some of the contact line clutter. The final step was to run Simplify Polygon, using Bend Simplify, with a 100 m tolerance, and RESOLVE_ERRORS turned on for topological consistency. The results of these steps are summarized in the Northeast corner of the Santa Rosa quadrangle in Figure 12.

⊟ ☑ sr_polys_lith100k_Elim20k_Di
- ⬜ <all other values>
- Lithology
- 🟨 alluvium
- 🟧 and/dac tuff
- 🟧 and/rhy breccia
- 🟪 basalt
- 🟪 basalt breccia
- 🟪 basaltic flow
- 🟪 blueschist
- 🟪 chert
- 🟧 dacitic flow
- 🟨 fluvial gravel
- 🟨 landslide
- 🟦 n/a
- 🟧 rhy flow
- 🟧 rhy tuff
- 🟨 sandy gravel
- 🟪 serpentine
- 🟪 silica carbonate
- 🟪 undiff melange

Figure 11. Symbolic generalization of Lithology

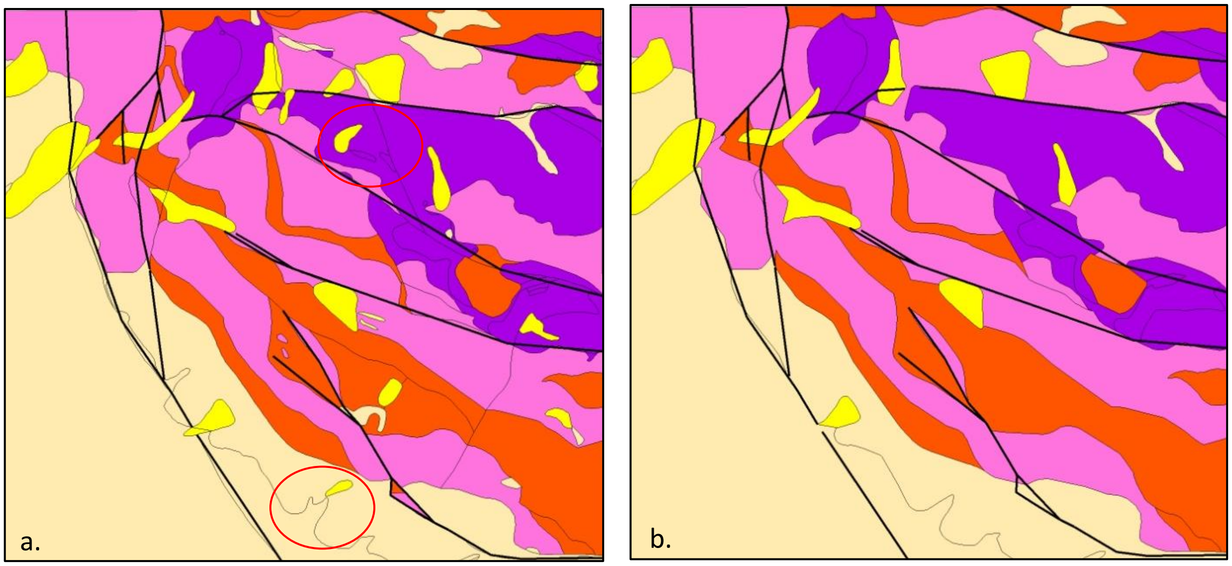


Figure 12. ***Lithology*-based polygon generalization is employed in the 1:100,000 map (a). Right image (b) is the result of eliminate, dissolve, and bend simplify operations. Red outlines show notable areas of change.**

4.3 1:250,000

Polygons

Polygon generalization for the 1:250,000 scale employs the second level of geologic attribute generalization, *Lithology2*. This results in 3 types of geologic units, the most aggressive reduction possible, short of just calling everything “rocks”. All units are lumped into Sedimentary, Volcanic, or Metamorphic, with the remaining “n/a” polygons representing water bodies. Following the attribute generalization, the ArcGIS Eliminate tool is used, this time selecting polygons with less than 40,000 m^2 area. The remaining polygons are then Dissolved based on the *Lithology2* attribute, and the polygon outlines are simplified using a 200m Bend Simplify tolerance, again using RESOLVE_ERRORS for topological consistency. The results of these steps are summarized in Figure 13:

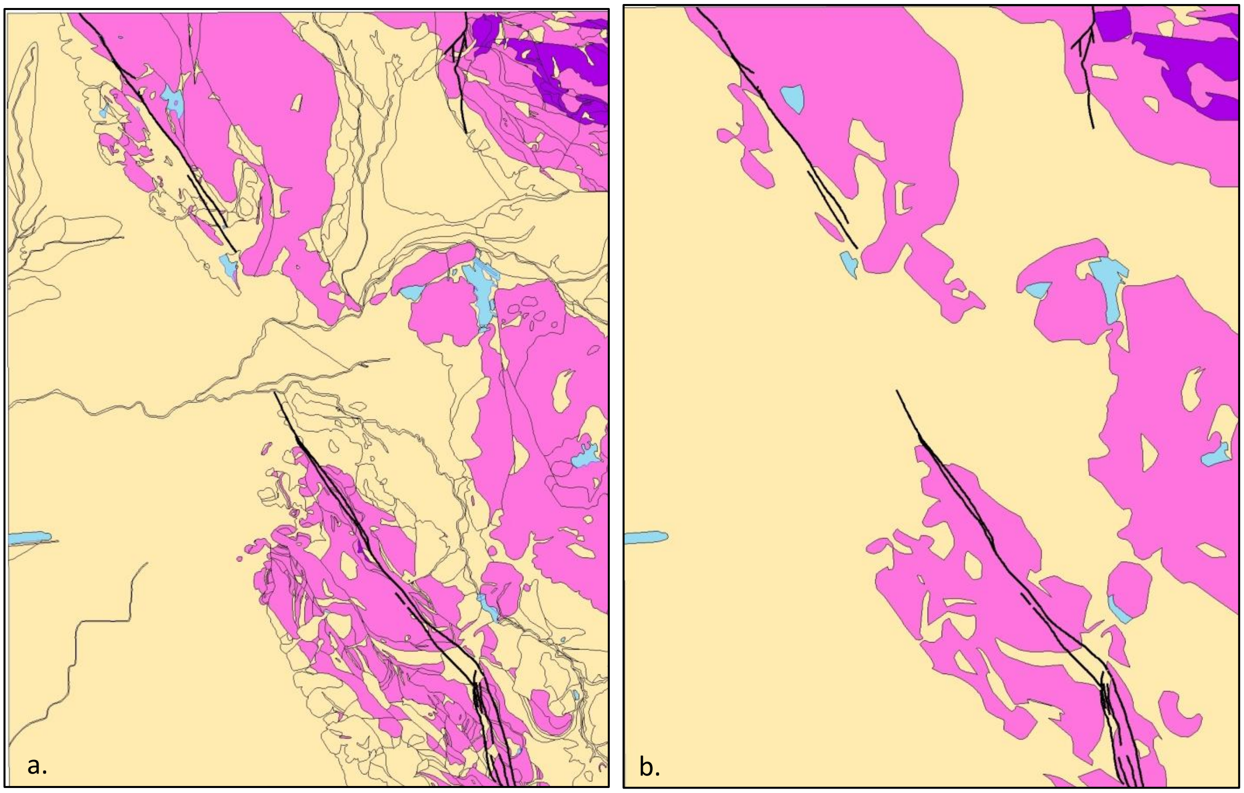


Figure 13. *Lithology2* attribute generalization in the image at the left (a), with original 1:24,000 polygon detail shown in outlines. Minimum area elimination, followed by dissolve, and bend simplify results in right image (b).

Faults

The second level of attribute generalization for faults, *SlipRate*, further reduced the number of features to display on the 1:250,000 scale map. Unfortunately, while the number of features was reduced, the complexity of the faults with slip rates remained the same as the original 1:24,000 map. Fault zones are mapped not just as straight lines, but as networks of nested line segments, complicating the generalization goal. Straightforward geometric simplification algorithms proved ineffective. The solution stemmed from correspondence with Charlie Frye at ESRI, who suggested I try start with a buffering approach on each fault network. Once I had created buffers for each fault network, the clear solution was to create a centerline for each buffer as the most accurate representation of the fault networks. Fortunately, the ET Geowizards ArcGIS plug-in (Figure 3) had the centerline tool I needed, and the process was fairly simple. Figure 14 summarizes the order of operations for fault generation based on the *SlipRate* attribute:

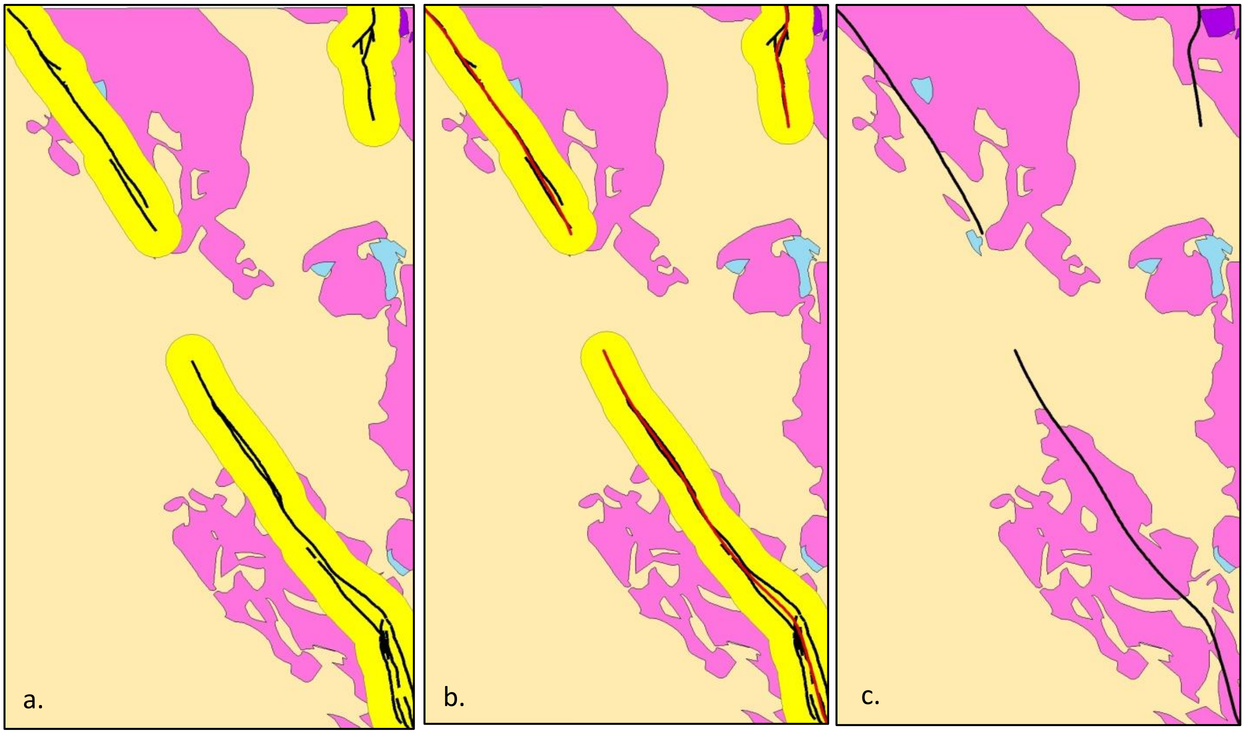


Figure 14. Fault zones with known slip rates are buffered by 500 meters (yellow regions in a, b), an internal polygon centerline is computed (red lines in center image), resulting in the final representation on the right (c).

5. Conclusions and Outlook

The process of generalizing a 1:24,000 geologic map over a broad range of scales has shed light on a number of issues that I had not considered going into the project. These insights can be separated into a two main categories: 1) the impact of my (naïve?) geologic assumptions in attribute generalization, and 2) limitations of the specific generalization tools I employed.

Once I completed my suite of geologic maps, I compared my results to a recent 1:250,000 compilation released by the USGS (Graymer et al., 2006). The Santa Rosa portion is shown in Figure 15b below, between my 1:100,000 and 1:250,000 generalized maps:

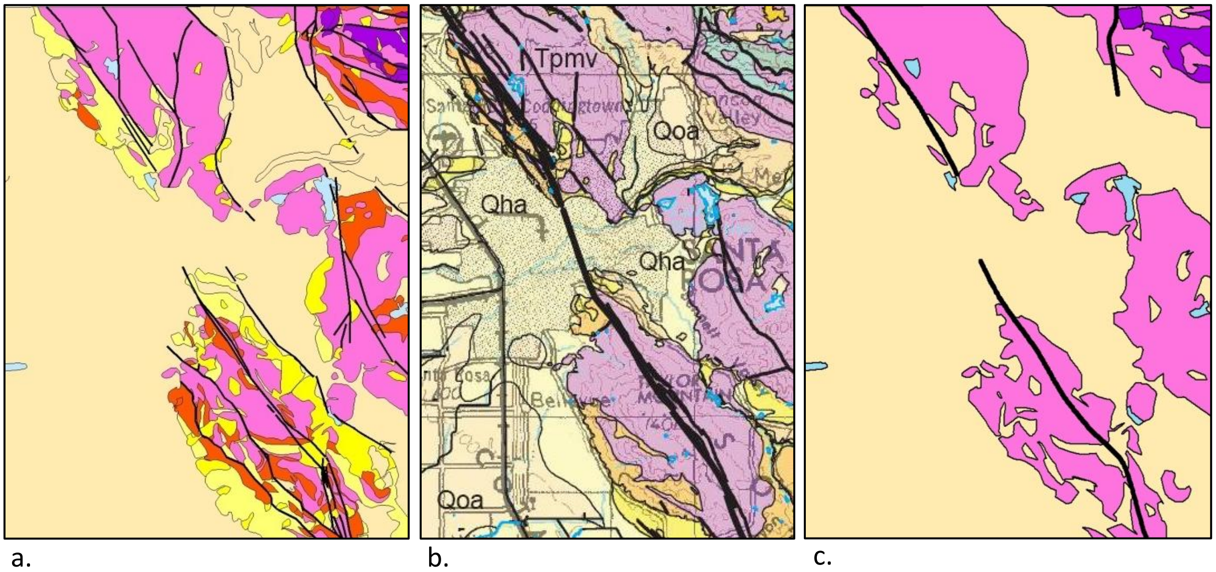


Figure 15. Published USGS 1:250,000 geologic map compilation (b) between my 1:100,000 (a) and 1:250,00 (c) versions.

The results are encouraging, though the level of fault detail on the 1:250,000 USGS version falls somewhere between my 1:100,000 and 1:250,000 ranges. In the USGS example, the Healdsburg and Rodgers Creek faults have been manually joined and the Maacama fault is missing, but the overall level of line detail is roughly equivalent to the 1:250,000 version that resulted from the *SlipRate* attribute generalization before the lines were further simplified using the buffer/centerline routine (see Figure 13).

One could even argue that by including a dense network of faults, the overlapping line segments on the USGS version has compromised readability. The geologic unit polygon detail in the USGS map also falls somewhere between my two scales. While the volcanic (pink for mine, purple in the USGS version) and metamorphic units (purple=mine, light blue=USGS) are similar, there is an additional level of sedimentary detail shown in beige and yellow that is not present on my version. This is due to a potentially over-aggressive attribute generalization for my 1:250,000 scale map. Further input by geologists is required to determine whether the USGS has under-generalized or I have over-generalized at this particular scale or use.

My use of slip rates to determine the relative "importance" of faults also has its limitations. Fault movement can be both horizontal and vertical, and earthquakes occur in both situations. Slip rate calculations can only be easily computed on strike-slip faults that have visible lateral offset, biasing my attribute generalization towards these types of faults. Given the comparison between the USGS 1:250,000 map and mine, I contend that this is still a good approach to generalization at small scales. Limitations may arise where slip rates are not available in the given area, and further input from an experienced geologist would be required to create an accurate fault generalization hierarchy through multiple scales. Another possible option would be to use the fault segment length or number of neighboring geologic units to rank fault importance, as in Downs and Mackaness (2002).

The second source of problems was due to the software limitations of the ArcGIS generalization tools I used. The first example is specific to creating polygons from lines in the original digitizing process. In Figure 16 below, the landslide polygon in the center is actually composed of two polygons separated by a buried fault contact. In each level of generalization, a minimum polygon area is selected, and the polygons that do not meet that minimum area are eliminated (merged with neighboring polygons). In this particular case, the head of the landslide should have been preserved, as it is really part of the larger polygon on the other side of the fault line. The logical solution would be to first dissolve the borders between like polygons based on their *Lithology* attribute, followed by the elimination step. Unfortunately, the ArcGIS Dissolve tool combines all polygons sharing a particular lithology into one single multi-part polygon, taking away the ability to select and eliminate individual features within that shared lithology based on a minimum area in the attribute table.

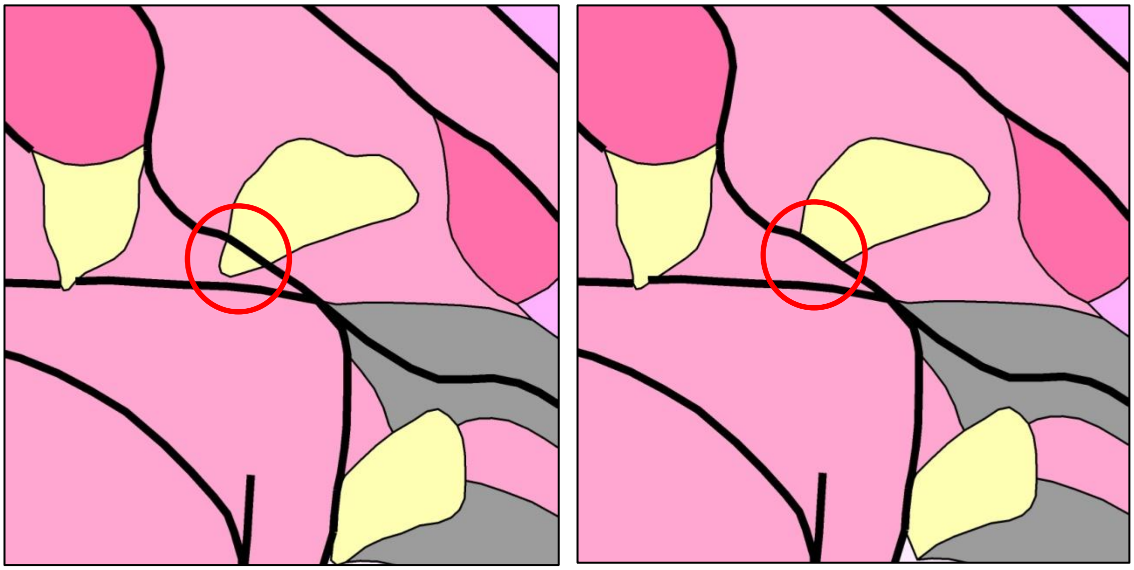


Figure 16. Area elimination problem resulting from original polygon creation from line segments.

The ArcGIS Eliminate tool also caused some issues when removing polygons that fell under the specified area cutoff. Because the elimination must maintain topological consistency, the polygons are not just removed, but absorbed into a neighboring unit. The two options that dictate this absorption are 1) the neighboring polygon that shares the longest border, or 2) the neighboring polygon that has the largest area. In this exercise, I chose the longest border option, but either of these solutions is prone to geologically inconsistent results. A more sensible geologic solution would incorporate attribute information, so that the unit would be absorbed by the most lithologically similar neighboring polygon, regardless of border length or area. In the hypothetical case where two neighboring polygons share similar lithologies but are surrounded by polygons that were significantly different in lithology, the two polygons would combine into one (the one which was bigger of the two), if the resulting merged unit was above the area cutoff. Software limitations precluded the use of this kind of intelligent grouping approach.

Looking toward future applications of attribute-based generalization in geologic maps, I would like to incorporate a more quantitative approach to my feature selection rules. Minimum polygon areas, and line simplification parameters (point remove, bend simplify) were created using qualitative methods based on my own aesthetic biases. Part of my reasoning for a qualitative approach is that there are no set rules for the minimum geologic unit sizes for 1:24,000 maps (McLaughlin, personal communication, 2007). Decisions are largely left in the hands of the geologist to represent those features that are deemed significant, often exaggerating their size in the process. However, this stance is not defensible and lacks quantitative rigor required for widespread adoption, so future work would focus on rule creation for line simplification and polygon area cutoffs through the scales.

To conclude, I feel that the use of attribute generalization in geologic maps is perhaps the only way to maintain the lithological relationships and significance of potentially hazardous features through a broad range of scales. Faults and geologic units are difficult to generalize en masse, and in order to facilitate the creation of small scale maps from 1:24,000 data, additional time must be invested in coding these necessary fault and polygon attributes. Additional work must also be done to optimize how current ArcGIS-based tools operate in the context of geologic maps, particularly how faults and polygons can be processed together to better maintain their interrelationships throughout the generalization process. If I can point to a breakthrough, it would be the use of the polygon buffer and centerline approach to fault network simplification at small scales. The results at 1:250,000 show promise in how geologic mapping agencies can generalize these complex features with minimal effort.

Acknowledgements

I would like to acknowledge the guidance of Cindy Brewer, my fellow cartography seminar graduate students, Charlie Frye at ESRI for his input and assistance, and Barbara Buttenfield and Lynn Usery for their additional comments and suggestions.

References

Aagaard, B., Zoback, M.L., and Brocher, T., 1906 San Francisco Earthquake Ground Motion Simulations. USGS Website: http://earthquake.usgs.gov/regional/nca/1906/simulations/index.php

Bader, M. and Weibel, R. (1997): Detecting and Resolving Size and Proximity Conflicts in the Generalization of Polygonal Maps. *Proceedings of the 19th International Cartographic Conference*, Stockholm, 23-27 June 1997.

Bonham-Carter, G.F. and Broome, J. (1998): Tools for Effective Use of Geological Map Data: A Topic for GeoComputation Research? *Proceedings of the 3rd International Conference on GeoComputation,* 17-19 September 1998.

Downs, T.C. and Mackaness, W.A. (2002): An Integrated Approach to the Generalization of Geological Maps. *The Cartographic Journal*, 13(2): 137-152.

ET Geowizards tool website: http://www.ian-ko.com/ET_GeoWizards/gw_main.htm (Nov. 13, 2007)

Graymer, R.W., Moring, B.C., Saucedo, G.J., Wentworth, C.M., Brabb, E.E. and Knudsen, K.L. (2006): Geologic Map of the San Francisco Bay Region: U.S. Geological Survey Scientific Investigations Map 2918 (http://pubs.usgs.gov/sim/2006/2918/)

Graymer, R.W., Bryant W., **McCabe, C.A**., Hecker S., and Prentice, C.S. (2006): Map of Quaternary-active Faults in the San Francisco Bay Region: U.S. Geological Survey Scientific Investigations Map 2919 (http://pubs.usgs.gov/sim/2006/2919/)

Langenheim, V.E., Roberts, C.W., **McCabe, C.A.**, McPhee, D.K., Tilden, J.E., and Jachens, R.C. (2006): Preliminary isostatic gravity map of the Sonoma Volcanic Field and vicinity, Sonoma and Napa Counties, California: U.S. Geological Survey Open-File Report 2006-1056 (http://pubs.usgs.gov/of/2006/1056/).

McLaughlin, R.J. Sarna-Wojcicki, A.M., Langenheim, V.E., McPhee, D.K., **McCabe, C.A.,** Valin, Z. (2007): Geologic and Geophysical Framework of the Santa Rosa 7.5' Quadrangle, Sonoma County, California. U.S. Geological Survey Scientific Investigations Map XXXX (publication in review).

McPhee, D.K., Langenheim, V.E., Hartzell, S., McLaughlin, R.J., Aagaard, B.T., Jachens, R.C., and **McCabe, C.A.** (2007): Basin Structure beneath the Santa Rosa Plain, northern California: Implications for damage caused by the 1969 Santa Rosa and 1906 San Francisco earthquakes, *Bulletin of the Seismological Society of America* (*in press).*

Steiniger, S. and Weibel, R. (2005): A Conceptual Framework for Automated Generalization and its Application to Geologic and Soil Maps. *Proceedings of 22nd International Cartographic Conference*, La Coruna , 11-16 July 2005.