2. GEOLOGICAL COMPUTER MAPPING: IS IT THE SOLUTION, OR THE PROBLEM?

Christopher Gold
Alberta Environment

INTRODUCTION

I would like to say that almost everything that has just been covered by John Davis is the kind of thing that I've been concerned with over the past few years, and, as I am sure you have, I found his presentation very interesting, indeed.

I frequently find myself in the position of being something of a heretic; continually preaching to the unconverted, trying to say “Hey! geographers have got something to say to you geologists” and then turning around to the geographers and saying “Well look! The geologists have got some really interesting problems.” To make it more difficult, you say “What you really need is a computer” and that throws both of them. Just to reassure you, however, I am really, deep down inside, a geologist.

Why am I here today? I see two very clear reasons. Firstly, I've been involved in a few very specific concerns for several years and I want to share my ideas on them and my feelings about where I think we're at, not answers, but feelings concerning what I've been working towards and what other people have been working towards and where I think we should go. Secondly, I've had a reasonable amount of experience with available computer programs not being written with any concern for my needs. So, although I consider myself at the moment to be on the development end, the real question is: “What do YOU, the end-users, need?” This is what I'd like to emphasize. We all have examples, not merely of programs we needed but didn't get, but perhaps worse, programs we got but didn't need. One example is isometric 3-D block diagram programs. They look nice and are very handy for brochures, but for a geologist they provide very little useful information. He should, after all, know how to rotate the map in the X-Y coordinate system and your octant search will not change. There are many advantages at this Stage for doing this.

TRIANGULATION BASED CONTOURING

Starting with these problems I came up with a triangulation based system for representing the relationship between drill-holes. Figure 1 illustrates some data points and a triangular network which has desirable properties. Once you have created a network, you already have a set of relationships between data points. You can use this for your nearest neighbour searches and a variety of other activities. While I don't think it's appropriate to go into it in great detail, I would point out that this is fundamentally a geographic data base whereby you have relationships between points in space which are not dependent on any coordinate system. You can rotate the map in the X-Y coordinate system and your octant search will not change. There are many advantages at this Stage for doing this.

Figure 1. Triangular network superimposed on data points.

Figure 2 is a very simple contour map. If you wish, you may do some more elaborate mathematics to produce smooth interpolations as illustrated in Figure 3. This involves questions about calculating the gradient of the data points, which in turn involves questions about distance, the zone of influence,
the weighting functions, and a number of things that have been covered earlier. The interesting point, however, is that this does not affect the basic interpretation. All we've done is pretty it up. We now have smooth contours instead of angular contours, but we haven't changed our interpretation. Clearly, this kind of data structure could be used to form the basis of a cross-sectional fence diagram mapping system. As I said, it is a cartographic data structure.

A CONTOURING ALGORITHM CHECK LIST

Considering an Overall View of surface mapping, I wish to offer a check list which might be of assistance to those of you who are looking at mapping programs and systems.

1. Model's surface passes through data point?
2. All available data utilized?
3. Surface affected only by surrounding points?
4. Sample content, adequacy, and isotropy accounted for?
5. Detection/removal of variation caused by small features/errors?
6. Drill-hole data (tops, etc.) validated?
7. Topology of stratigraphic record?
8. Selection of correct model to be utilized?

I wish to cover the points in this check list as it applies to my own, and others', triangulation systems.

Model surface: Does a model surface pass through the data points? I put a check against that because of the triangulation approach that I have been previously talking about. Yes. It does. We've already covered some of the calculations on gridding base techniques. as to how closely they do. The simple-minded triangle does. It has other problems, but it does that.

Are all the available data utilized? Yes, they are. This isn't necessarily the case in a gridding technique. Those of you who have been involved know that the finer the grid you work with, the more it costs, and depending on your data point distribution, you can have several data points within a grid cell, etc., etc.

How is the surface affected by adding different points? The concept of the domain comes in here, which has already been mentioned with gridding algorithms. At some distance away you decide to forget it, these points do not contribute to the general information about a specific grid cell or some other point you are interrogating on the Surface of the model. In the triangulation based system it is easy to comprehend that, with the spider's web of triangles. It is a fairly simple system to pull out the nearest
points in all directions and to use these for your various interpolation techniques.

Sample content: This is usually very trivial. What does the data point consist of? Well, X.Y, and Z values. In some cases it also contains slope information. In some of these algorithms, as has been shown with some other examples for gridding techniques, slope information is very valuable. If you have it, it should be used. If you don't have it, which is the norma 1 case. YOU should be able to have some reasonable method of estimating slope.

Sample adequacy: Do you have enough data to find the features you are hunting for? A fundamental problem with most geological type mapping is that you never have enough data. It is fundamental that you should examine your data spacing versus the features you are attempting to identify. If you do not have an adequate data spacing for the smallest features, then you should, if possible, filter the data to smooth it, to remove high frequency components, error components, or other items in the data which will only confuse the final map. If you've only got the level of data collection for a broad regional map with two or three humps and hollows, then you should make do with that. You should realize the limitations of your data.

Sample isotropy: You have no doubt seen examples of seismic traverses. There are many other kinds of traverse data. For example airborne and marine geophysical transverses etc. All of these data are collected along 1ines. Another type of data collected along 1ines. That most people don't realize, is field geology data. I know that in the oil business you are less concerned with this, but in the hard rock business, YOU have these explorationists running around mountains. Rather than crawling up and down the sides of ridges, they usually follow the crests or valleys. Obviously this is traverse type data. Perhaps one of the most important questions to ask of a mapping package is: Will it handle this grossly irregular traverse type data? This is not just irregular spacing, but rather is dense in one direction and sparse in another. I know of various algorithms and mapping packages that are available which quite simply blow up if you provide too fine a grid spacing for the traverse separation which YOU have. I checked off that I me on the check list because, yes, with triangulation techniques, that's not a problem. You simply have a triangulation that sews between the particular traverses. It may not provide a perfect interpretation, but it will at least take into account the data points on both sides of the unknown point under interrogation.

Now we are coming into the items on the check list that are not ticked off, because simple triangulation techniques, as well as others, need additional features to handle these items.

Detection or removal of variation caused by small features or errors (smoothing and filtering): For arbitrarily located data points the kriging approach asks the relevant questions best. The same thing can be done with the triangulation based technique without the mathematical rigor, but still at a useful level. By using the points surrounding a particular data point, the centre of the spider's web if you wish, and using sets of these triples to form triangular plates. these plates can be used to form estimates of the central point, and from this you can calculate means and standard deviations, and estimates of how far off the central point is by comparison with its neighbours. Clearly, if all the points are on a plane you're going to have a zero deviation and a zero difference between the mean and the original point, and a very low standard deviation between all the estimates and the central point. In many cases, however, this will not be the case. The reasons for this may be that the data point is in error and it disagrees strongly with its neighbours; or that you have some kind of break in slope (e.g. faults) with respect to sample density. There are a number of problems with this aspect of contouring that in many cases have not been researched adequately and certainly are not usually taken into account by the user. It is, I think, very important that we start looking at data sampling versus the expected end product. It is usually not considered soon enough.

DRILL-HOLE VALIDATION

Let's move to a partly related topic -- drill-hole validation. I've done a great deal of work attempting to pin down errors in data from drill-holes. It's a major problem when you have certain of your expected sequence of formation tops missing, due to erosion or non-deposition or whatever, e.g. even the fact that it wasn't logged in the first place. Incomplete logging is more common than many would like to admit, and many garbage maps have been produced because the supposed top of a particular formation was non-existent. It was inadequately labelled in the first place and the topology of the situation was not considered. There is no problem when you have continuous blankets with no edges, but where you have zero isopaths you have problems.
Figure 4 represents an example where the topology is not usually taken into account. Figure 5 represents a graph of its adjacent geological units. We have units 2, 3, and 4, which were originally stratigraphically continuous. Unit 1 is on top of a erosional surf ace. Several drill-holes penetrate various combinations of these. The little triangles in the diagram represent labels given to a particular contact when it is described in drilling, or when it is interpreted from a formation pick. The problems should be obvious. On the left hand side, drill-hole A. there is no problem. On Drill-hole B. we have a problem where a pick was not made. Why was that pick not made? In this particular case. because the data simply were not recorded. Again. not a uncomon occurrence. In D. E. and F we have situations where we have different labelling conventions. Is this the top of unit 2 (let's Call it Cretaceous) or is this the top of unit 3? What is it? In many cases you may run into a direct conflict. The conventions may not even be formalized. but they certainly need to be defined and taken into consideration for the computer mapping program. In many cases nobody will have been told, and different interpretations will be used by different people who are coding it. The fundamental problems are evident. It is a major item on the checklist.

One last point: What do you mean by an isopach? In conventional terms you will normally take the thickness difference and try to contour that. This frequently causes problems. Whenever YOU get the jaggies around the zero thickness isopach on the contour map you are ignoring the fact that a zero thickness on the map is not necessarily a zero thickness. It is missing data. A continuing problem is that people code thickness directly. They insert the difference values into their contouring program and wonder why everything goes crazy at the isopach edge. What you really should be doing is defining a model of an isopach. I feel that a surface model is a representation of one surface. If you wish to define this surface, and your drill-hole data has been coded in a valid fashion to do so. then by all means do it. An isopach has two components: the current top, and the original bottom. What you should be coding is not the difference between the top and the bottom, but an estimation of the top surface. If your mapping system can select valid data points, you should indeed have a negative value just beyond the intersection of the two surfaces. Your mapping system is then simply programmed to take the positive values only, and contour them. You then avoid all the problems that go into conventional isopach mapping.
I will summarize by using an industrial case history which is listed in point-form as follows.

Problems:

Data coding errors hard to find
1. Hand contouring allowed discretionary smoothing.
2. Incorrect handling of data where a particular contact was absent in a particular hole.
3. The required contact was contoured even where known to be absent.

Commentary:

Data coding errors were hard to find; I've already mentioned this in terms of filtering. You may use various techniques to identify a particular point out of kilter with respect to its neighbours. If all the neighbours put their thumbs down on it, you should look at it. There are ways of automatically flagging it. Somebody will have to make the judgement as to whether this is genuine, or an error.

2. Hand contouring allows discretionary smoothing. You know that there is something funny going on or your errors are on a particular scale. Automated techniques don't usually permit this. I've discussed some possible directions using filtering, but it is not a final subject. It is a problem that should be examined further, particularly when you are looking at arbitrarily located data, not predefined grids.

3. Incorrect handling of data where a particular contact was absent in a particular hole.

4. The required contact was contoured even where it was known to be absent: I've mentioned this respecting isopaths. If you think of an isopach map as a difference map, a map of the difference in elevation between the top surface and the bottom surface, each independently calculated as opposed to having zero plugged in where it is missing, then you may need to handle isopachs correctly. You may even use the zero isopach as a trigger to decide that the formation top contour map may be suppressed where the isopach values are zero or negative.

Some final conclusions on mapping methods are listed in point-form as follows.

Single surface methods:

1. Control led smoothing (or none) to catch data errors or high frequency filtering.

Multiple surface comparisons:

Primarily for where a contact is absent in a drill hole.

2. The two-surface model of a geological unit means that an isopach map is a positive difference map.

3. In principle, missing contacts are classifiable as not logged/not deposited/eroded out.

4. Features should be in mapping package, but contact naming may be too inconsistent.

They are broken into two categories. Single surface methods and multiple surface methods. In single surface methods, we're talking about controlled smoothing, or even none. That is another dig at gridding techniques where, for all of their advantages, you cannot get a zero smoothing situation. This controlled smoothing can be used to catch data errors or perform high frequency filtering. Although the filtering methods are still somewhat non-rigorous and primitive. Multiple surface comparisons are primarily intended for where a contact is absent in the drill log. If you do not mind having missing contacts, then there is no problem. The two surfaces model of the geological unit means that the isopach map is a positive difference map. In principle, missing contacts are classified as not logged, not deposited, or eroded out. In practice it isn't that easy. These features should be in a mapping package, but even when they are, if the contact naming is too poorly done on the original data, then you have no real means of success. You're going to have to go back to hand editing and there is no alternative.

Processing Costs

I would add one more point about mapping techniques, referring back to the triangulation methods. Some of the costs that have been associated with various methods have been discussed. My own experience on the triangulation approach is that the computing costs can be broken into two parts: input and output. For input of even up to a few thousand data points, the cost of building a triangular network is almost linear with the number of points. To give you some idea; if reading (X,Y,Z) values for one data point using standard FORTRAN 1/O takes one unit of time, it only takes one and a half units of time with my system to insert that point into a triangulation network of up to a few thousand points. If you were up to a million points, it would take longer. But for a few thousand points the input of the data into a mesh, into a network, and into a data structure, is almost linear. From there on you are processing individual triangles. There is never any searching of the whole network. You just select each triangle in turn, or you select each node in turn, or you look at all the surrounding points, to calculate the
slopes. and do most everything else. Computation time is linear with the number of triangles, and there are twice as many triangles as there are data points. Ignoring boundaries. We therefore have output which is also linear. This kind of system can therefore be considered linear in input and output, which is an attraction when you start accounting for computer time.

CONCLUSIONS

In conclusion, the preceding points are topics which I've been addressing in my recent work. They are not necessarily solved, and most of them need further refinement. There are two additional steps required. Firstly, full problem definition, i.e. proper definition of a smoothing function which handles a particular wavelength of feature and suppresses features with shorter wavelength: and secondly, they need implementation, acceptance, and usage, and this is where industry understanding of the problems and attempts to use them would be appreciated. Software suppliers are often badly at fault for not listening to the user's needs. The geologist or the manager is also at fault for not insisting on products that meet specifications. However, this is relatively difficult to achieve, since he must speak two languages: that of a geologist, and that of a computer scientist. In part that is the reason for this particular problem.

Now I have a series of quotes I wish to throw out to you. I'll run through them fairly quickly without making any additional comments. They are comments (perhaps cynical) that I put together from talking to industry people about some of the fundamental problems with computer aided mapping, and perhaps you would like to respond to them during the discussion.

1 Current mapping packages are useful for reservoir work where information has previously been carefully validated. True or false?

2. The biggest problem in formation mapping is the choice of picks from downhole logs.

3. Regional studies do not benefit from computer contouring since, in addition to simple errors, formation picks will be inconsistent over large distances.

4. When mapping lithofacies or drill stem tests, formation picks are again the limiting factor.

5. Second derivative maps, isometric diagrams, etc., are almost never used. Currently, isopach maps showing differential compaction, etc., are the primary mapping tool. True or false?

6. Any mapping tool must be simple enough to be understood thoroughly by the geologist. If not, he will do it by hand anyway!

7. The value of posted maps is underestimated. Leave the contouring and editing to be done by hand. All you want is a database. True or false?

8. The use of posted symbols for drill stem test results should be examined further.

9. Here's my favorite. Most plays, in Western Canada anyway, are generated by dreaming in front of hand constructed cross sections. Who even needs the map?

10. Most new plays in North America, excluding Alaska, now come from new interpretation of old data (i.e. different facies combinations).

11. And a final one: Interactive computer graphics for data displays or mapping is still rare, primarily due to current management structures in industry. True or false?

This would be a good place to stop, and I am looking forward to your response.