Some tips on map production for precision agriculture

Rob Bramley
CSIRO Land and Water / Cooperative Research Centre for Viticulture, Adelaide
E-mail: rob.bramley@csiro.au

It has been apparent for some time that, having decided to adopt some or all of the elements of precision agriculture (PA), one of the main impediments to progress faced by growers is the question of what to do with all the data ? Generally, this means how to turn it into useful maps ?

At the recent SPAA Roseworthy field day, I only saw one yield map in the display booths that had been produced properly. One of the main reasons for this state of affairs is that, in my view (and also the view of many other PA researchers), most of the commercially available software that is either sold in support of PA, or provided with the yield monitor, does not produce the maps properly using robust statistical procedures that result in a *surface* rather than a series of dots. This was the main reason why, in my work in the Viticulture CRCV, the first thing I did was to produce a protocol for the production of winegrape yield maps (see www.crcv.com.au/research/programs/one/CRCVProtocolBkfinal.pdf) – which can also be used for the production of EM38 maps. Clearly, SPAA grapegrowers with yield monitors who want to map their data can follow this protocol. However, the main purpose of this article is to suggest how broadacre cereal producers can use it, with a few modifications, to make robust, and therefore useful, yield maps.

Why bother ?

Figure 1 illustrates the difference between a properly produced yield map (1d), two maps of limited, if any, value (1b and c) and a harvester trace from which these maps were derived (1a).

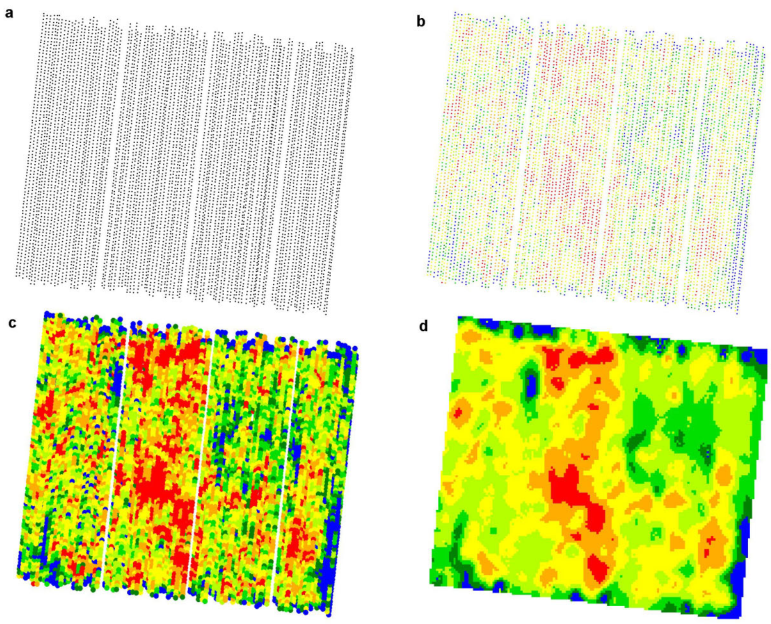

Figure 1. Some possible yield maps (b-d) derived from yield monitor data collected at the points indicated by the harvester trace (a). The only one of these that is useful as a PA decision tool is d.

Fundamentally, a yield monitor measures the instantaneous yield at a large number of geo-referenced points (Figure 1a). We could, if we wished, simply colour code those points according to the measured yield, as shown in Figure 1b. We could also do what much PA software does, and make the dots big enough to give the impression that there is no space between them as in

Figure 1c. However, if we accept that the purpose of producing maps is to use the information shown in them as a basis for decision making, and also accept the conventional wisdom that we need several layers of data (ie yield data for a few years and/or some soil or EM data, maybe some imagery and perhaps some soil or tissue test data) to make these decisions as good as possible, perhaps having used the data to define management zones, then a map such as that shown in Figure 1d is our only real option. This map is an interpolated surface (ie it contains no gaps) projected onto a pre-defined grid.

Using a predefined grid is critical, because if each map for an individual paddock is projected onto the same grid, then we can be sure that each map will contain the same number of pixels and that these pixels will be accurately aligned from one map to the next (Figure 2). This makes the use of clustering for the identification of management zones pretty straightforward. In contrast, reliance on maps like those in Figure 1b and 1c would be dependent on the dots in a map for one year aligning perfectly with those from another – and my guess is that the probability of two harvester traces being identical is close to zero !

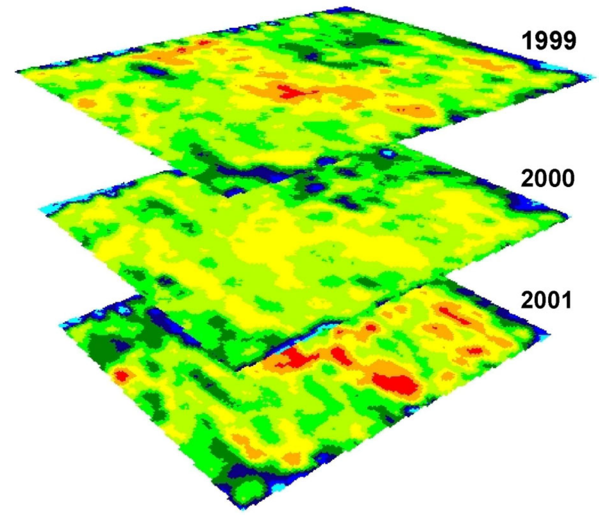

Figure 2. Three years of correctly mapped yield data. Because these maps are all continuous surfaces projected onto the same grid, they are all made up of exactly the same number of pixels, each in exactly the same place. Analysis and clustering of this data is consequently much easier than would be the case with 'dot' maps.

From grapes to grain

The grape yield mapping protocol explains the map production process and some of the thinking behind it; this article should therefore be read in conjunction with the grape yield mapping protocol. In summary, we use the VESPER program to 'block krig' the yield (and EM) data using a local variogram with the results projected onto a 2 m grid (ie pixels of $4m^2$). You can download the VESPER program as shareware from the website of the Australian Centre of Precision Agriculture (ACPA) at the University of Sydney (www.usyd.edu.au/su/agric/acpa/pag.htm).

Grapes are typically grown along rows that are about 3 m wide. Discussion with vineyard managers suggested that the smallest area that they would consider for targeted management (ie smallest size of management zone) would be 3 rows wide. For this reason, when block kriging, we use 10 m blocks (ie blocks that are about 3 rows wide); note that 'block kriging' results in a smoother mapped surface than 'point kriging' because it involves some local averaging of the kriged values for points within the block (in this case in blocks of 10 m). Following similar logic, I would recommend that for grain yield maps, a block size of not less than 3 times the width of your header be used. Some people may argue that their seeder or sprayer provides a more appropriate basis for determining the size of their minimum management zone (that is fine !).

Whatever, a block size of 25-50 m would be appropriate for kriging grain yield maps; if in doubt, it is probably better to go larger than smaller.

Our 2 m grid was chosen because it (a) is typically smaller than the width of a single row; but (b) is larger than the smallest pixel size that can be justified on the basis of the accuracy of a differential gps (about 1 m); and (c) results in a reasonable map resolution for vineyards larger than about 3 ha. In vineyards smaller than about 3 ha, maps projected onto a 2 m grid appear quite 'pixelated'. In this situation, a smaller grid size (ie pixel size) cures the problem, but is difficult to justify on the basis of gps accuracy. Again, using similar logic, my suggestion is that for grains yield mapping, a grid size of 5-10 m is used.

Let's say that a typical vineyard is 7 ha. This means that on a 2 m grid, its yield map will comprise 17500 pixels. Since this protocol works well, it seems reasonable to aim for a similar number of pixels in a grains yield map. I'll assume that a typical SPAA members' paddock is 80 ha. This means that each pixel needs to cover about 46 m^2, equivalent to a pixel size of about 6.8 m. Note however, that a gps without differential correction is accurate to about 6 m at best, so where differential correction is not used, I would recommend that maps are projected onto a 10 m grid. Where differential correction is being used, smaller grids are fine. However a 'row' in a cropping situation is equivalent to a header width. Since kriging is an interpolation method (ie it fills in the gaps between un-sampled points to create a surface), I think there is sense in projecting kriged yield maps onto grids that have a smaller size than the row width – otherwise you are not filling in a gap. So for data collected with a differentially corrected gps, a grid size of 5-8 m seems appropriate; if in doubt, go for larger rather than smaller, but once you have created the grid, stick to it !

(VESPER tip: The authors of VESPER know what they are doing, so as a general rule, accept the default settings. For yield and EM maps, set VESPER to run with a local variogram, fitting an exponential model with the weighting set to unity. Some of the other settings depend on the points discussed in this article and in the grape yield mapping protocol).

Remember – all this stuff about grids is important because you want to make sure that every map for any one paddock is projected onto the same grid. Ideally, this grid will be defined on the basis of the paddock boundaries and can be generated in a GIS.

Datums and data trimming

Remember too, that prior to mapping, you need to convert the gps data from the WGS84 standard datum to one appropriate to your local situation. At the same time, you will need to convert from decimal degrees to eastings and northings (m) – how many decimal degrees wide is your header ?!! Since there is only a 2 cm difference between the new GDA94 Australian datum and WGS84, the conversion to eastings and northings is the important bit; it is not unreasonable to ignore transformation to GDA. To do the eastings / northings conversion properly, you need to know which UTM zone you are in. As a rough guide, Port Lincoln is in zone 53, as is Port Broughton. However, Clare is in zone 54, as are Mildura and Naracoorte; Wagga Wagga is in zone 55.

For a variety of reasons, grain yield monitor data tends to be quite 'spiky'. The 'spikes' (ie aberrant high values) should be removed before mapping. The question is: what is a high value ? The following is recommended as a means of sorting this problem out.

Bring the yield monitor data into a program such as Excel, Access or Lotus. The data should then be *normalised* as follows. Calculate the average yield (t/ha) and standard deviation for the whole dataset. Then for each individual record, do the following calculation:

Normalised yield = (Yield – average yield)/standard deviation

Then sort the data on the basis of normalised yield and delete all records for which normalised yield is either less than -3 or greater than +3. Use the remaining data (actual yield) for map production.

Note that it is often useful to produce maps of normalised yield too because these enable maps for different years to be compared without the confounding effect of the yield potential being different for different years due to climatic differences. If you wish to do this, then the above procedure should be repeated iteratively (ie. recalculate normalised yield using the trimmed data) until the remaining data all lie between -3 and +3. The basis for this is that if the data are 'normally distributed', then 99% of them will lie between -3 and +3. Trimming once should therefore remove about 1% of the data in the raw yield monitor file. In my experience with winegrapes, it is generally necessary to trim 3 or 4 times with the result that about 2.5-5% of the data are removed. Don't worry – you have plenty of data and can afford to lose a bit !

GIS, Standard legends and Consultants

Of course, once you have interpolated a yield map, you want to be able to look at it and start to interpret it. First up, therefore, the map needs a legend.

At the Roseworthy meeting, there was a fair bit of talk about legends and standards. At present, there are no standards for yield mapping, and certainly none for the legends. On the basis that red can variously mean 'stop' or that 'you're going bust !', and that green means 'go', it seems sensible for red to infer low yields and green to infer better ones. My experience is that most growers like a legend to have about 8 colours, so I generally standardize on: red (lowest), orange, yellow, light green, bright green, dark green, dark blue, bright blue (Figure 3). If you want additional colours, then browns and light blues or purples can be added at either end. I assign my colours to equally-sized categories of yield and ensure that the average yield for the paddock lies within the range of values covered by light green (or occasionally the bright green). I would not wish to infer that this is the only or best colour scheme for a yield map, but it works. If SPAA members were to standardise on this colour scheme, then a lot of the concerns expressed at Roseworthy would disappear ! Note that I do not favour the idea of a standard legend with respect to actual yield (or EM38) values, because these would end up being restrictive in that what is appropriate for one year, paddock or region might not be so for another. And since the object of all this is to see within-paddock variation, there is little point in reducing the apparent magnitude of this by imposing a general legend – in other words, if you are averaging 1.5 t/ha, why use a legend based around an average of 4 t/ha ?! (The same comments can be made for EM38 maps for which I use a monochrome red legend like that shown in Figure 3).

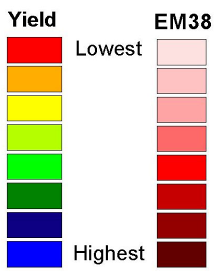

Figure 3. Suggested legends for yield and EM38 maps. These have proved useful in viticultural work and may provide the basis for the development of a SPAA standard.

In addition to inspection, most people want to be able to analyse their maps and see how they relate to other data layers. This requires the use of a geographical information system (GIS). By any measure, such software is expensive and would have a high degree of redundancy with respect to the needs of many PA practitioners. I would imagine that the majority of farmers have better things to do than spend hours at the PC struggling with the vagaries of GIS, whether these be other farm tasks, playing with the kids or watching the footy. Furthermore, whilst yield mapping is not difficult, it is certainly made easier by having a good understanding of spatial statistics. So I would recommend that all but the most tenacious readers of this article get someone else to do their map production for them. The consultants amongst the SPAA membership may not thank me for saying so, but it probably either falls to them to get up to speed with the computational and geostatistical requirements of PA, or to form some alliances with appropriately skilled partners. The obvious additional benefit of this approach is that one set of software (ie owned by the consultant) could then service many farms. Further, it would then fall to the consultant to carry out the essential task of data management.

Beware the real-time snake-oil merchants !

It is probably beyond the scope of this article to explain why kriging is the preferred method of map interpolation or how it works. Suffice to say that it looks at the data within a region, and uses this to estimate values at unsampled points based on a relationship called a variogram which describes the variance amongst the data values as a function of how far apart they are.

Several manufacturers have suggested that real-time yield map production is the cure of all ills. I do not agree ! Map production in real-time can only be done using data that have already been collected (the bit of the paddock you have already harvested) whilst for any point within a paddock, kriging requires information about the bits you are yet to harvest as well those already harvested. So it is impossible to produce a yield map properly in real time !

Further Info

Anyone interested in recent PA research may wish to navigate to the ACPA website (www.usyd.edu.au/su/agric/acpa/pag.htm) which has some good stuff in addition to the VESPER software. Details of my work can be found at www.clw.csiro.au/staff/BramleyR and www.crcv.com.au/research/programs/one/project1.1.1.asp (these sites do not entirely relate to grapes).

Feedback

If SPAA members are indeed concerned about standardisation in mapping, then maybe they would like to agree on some standards for SPAA to adopt. Do you agree with those suggested here ? Would you like to suggest modifications or alternatives ? Feedback please to either Rob Bramley, Rohan Rainbow or Malcolm Sargent.

CSIRO
CSIRO LAND and WATER

COOPERATIVE RESEARCH CENTRE *for* VITICULTURE