A protocol for the construction of yield maps from data collected using commercially available grape yield monitors.

Supplement No. 1. February 2005.

Rob Bramley
CSIRO Land and Water and Cooperative Research Centre for Viticulture
Private Bag No. 2, Glen Osmond, SA 5064.

Background

When we produced the original winegrape yield mapping protocol (Bramley and Williams, 2001), the primary source of yield monitor data was the HarvestMaster™ system. Whilst the protocol was, and is, generic with respect to the mathematics of yield map production, the current more common use of the Farmscan™ system may present difficulties for some users of this protocol. In addition, since the publication of the original protocol, we have made a change to our own standard procedures in regard to data trimming. These issues are discussed in this note.

A note for Farmscan™ users

The Farmscan yield monitoring system comes with software (the Farmscan™ Data Manager) for data manipulation, mapping and display. Here, we make no comment on this software other than to point out that the mapping part of this program does not conform to our protocol. However, because the Farmscan system stores yield data in a binary format, users wishing to follow our protocol still need to use the Farmscan™ Data Manager to export data in ASCII or text readable format in order to manipulate it for input into the VESPER kriging software (Minasny et al. 1999) as discussed in the original protocol. The Farmscan™ Data Manager may also be useful for adjustment of recorded yields to match winery weighbridge tonnages, although some users have reported difficulties with this. Users should consult the appropriate Farmscan manual.

Data export from the Farmscan™ Data Manager is easy. With a map shown in the mapping window, select "File…Save as…". We recommend selection of the "Comma delimited UTM (*.csv)" option as this export also converts latitudes and longitudes (decimal degrees) to an Eastings and Northings (m) basis. The resultant exported file can then be opened in a spreadsheet program such as Excel for further manipulation.

Farmscan™ users should take care to note that, as of build 038 of version 3.2 of the Farmscan™ Data Manager software, the exported file contains a column of data with yield in t/ha, unlike previous versions, but the column (column 3) remains labelled with the heading 'rate'. As outlined in the original protocol, interpolation of a yield map requires the raw data to be in units of t/ha. Note that more recent versions of the Farmscan™ Data Manager are due to have more intuitive column headings and data in metric units only.

Data cleaning and trimming

The original protocol stated:

> "*In grain crops, it has often been found necessary to 'despike' the yield data…; this means removing extreme values that appear in the dataset and which may be caused through machine blockage, for example. In our experience with winegrape yield mapping in Coonawarra, Padthaway, Clare Valley and Sunraysia, this has not generally been necessary. However, if you find that your yield maps show erratic and/or extreme variation…, despiking may be useful.*"

We have now changed our view on this, having observed a number of 'spiky' datasets over the last few years (derived from both Farmscan™ and HarvestMaster™ systems), and suggest that the following data pre-treatment be followed to remove possible "spikes" in the data prior to map interpolation using VESPER.

1. Data cleaning

HarvestMaster™ users should follow the guidelines in the original protocol re. cleaning (ie. removal of data records with zero yield or GPS errors). However, Farmscan™ users are advised to run the following cleaning procedure:

(Note: this advice assumes that the yield monitor has been set up to log at 3 second intervals).

a. Delete data records containing GPS errors; ie. those for which the value listed in the GPS column of the data file is not equal to -1.

b. Delete data records for which yield (rate) = 0.

c. Since the yield monitor has been set to log at intervals of 3 seconds, all the values in the 'cycles' column ought to be equal to 3. We recommend deleting data records for which 'cycles' is <3 and also those for which 'cycles' is >4.

d. Finally, given the accuracy of a differentially corrected GPS (+/- 50 cm at best), data records which are closer together than 50 cm should also be deleted. Build 038 of Farmscan™ v.3.2 records distance in inches, so we recommend deleting data records for which distance is <20.

Having run through these steps, yield should be adjusted to match the winery tonnage as per the original protocol.

Note: If the yield monitor has been logging at intervals of less than 3 seconds, then the data density will be too great for implementation of the mapping procedure outlined in the original protocol. In light of a range of issues which relate mainly to GPS accuracy, it is recommended that data records be deleted so that the remaining data have an effective logging interval of 3 seconds. Thus, for example, if logging was set to 1 second, then 2/3 of the data need deleting. An easy way to do this is to import the data to Excel and to sort them on the basis of time. Then in a new column called 'marker' assign values of 1, 2 and 3 to the first 3 data records. Then copy this sequence to the end of the data file. Then re-sort the data file on the basis of the 'marker' values and then do one of the following:

a. delete all the data records for which 'marker' = 2 or 3.
b. delete all the data records for which 'marker' = 1 or 2.
c. delete all the data records for which 'marker' = 1 or 3.

The result will be a data set that can then be used for processing as per the original protocol.

2. Data trimming

The purpose of data trimming is to remove aberrant values from the data file. This is done on a statistical basis by normalising the data so that it has a mean of zero and standard deviation of 1, and then removing records for which the normalised yield is either greater than +3 or less than –3; that is, data with yield values more than 3 standard deviations from the mean. This is a common pre-treatment for yield monitor data (eg Pringle et al. 2003), and in our experience is effective at removing aberrant values that are artefacts of the harvesting process (eg. occasional blockages of the discharge belt and high values at row ends when the discharge belt is switched on or off).

Data normalisation can easily be done in a spreadsheet program such as Microsoft Excel. First the mean (ie average) yield (t/ha) and standard deviation of yield are calculated. In Excel, the AVERAGE and STDEV commands can be used. Thus if the yield data are in column X of your worksheet and lie between rows 2 and 5000, the mean yield could be calculated in cell X5002 using the formula:

=AVERAGE(X2:X5000).

Similarly, the standard deviation is calculated by inserting the formula:

=STDEV(X2:X5000)

into cell X5003.

Normalised values can the be calculate in a new column using the formula:

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

Again, using Excel as an example, if you wanted to place the normalised data into column Y of the worksheet, the following formula would be entered into cell Y2 and copied down the length of the column:

=(X2-X5002)/X5003

Yield records for which the value of normalised yield is either less than -3 (ie more negative than -3) or greater than +3 are then discarded. The remaining data are then used for yield map production as described in our original protocol.

In addition to maps of yield (t/ha) produced for a single year, maps of normalised yield are also often useful. This is especially so in the context of analysis of multi-year data because by mapping normalised yield, users can examine yield variability over a few seasons independently of any seasonal effects due, for example, to differences in annual rainfall or the number of growing

season degree days. Normalised maps are also of value when data are not available for calibration of the yield monitor against winery tonnages

For normalised maps, following the initial data trimming as described above, we find it useful to use an iterative procedure in which the data are re-normalised based on new values of the mean and standard deviation for the trimmed data sets. The data are re-trimmed until all normalised values (N_i) conform to the rule: $-3 < N_i < +3$.

Note that for a normally distributed data set, 99% of the data lie within 3 standard deviations of the mean. Thus, for maps of actual yield, approximately 1% of the data are discarded in the trimming process, whilst for the production of maps of normalised yield, between 2 and 5% of the data are trimmed. The iterative process described ensures that for a multi year analysis, the data for each individual year follow similar distributions. At the Coonawarra research site shown in the original protocol, a yield monitor data file typically contains more than 11000 georeferenced yield records, so the need to 'de-spike' the data poses no threat to the viability of the kriging process in terms of its data requirements.

Following cleaning and trimming, yield map production can proceed as per the steps outlined in the orginal protocol.

References

Bramley, R.G.V. and Williams, S.K. 2001. A protocol for the construction of yield maps from data collected using commercially available grape yield monitors. www.crcv.com.au/CRCVProtocolBkfinal.pdf Cooperative Research Centre for Viticulture, Adelaide.

Minasny, B., McBratney, A.B. and Whelan, B.M. 1999. VESPER version 1.2. Australian Centre for Precision Agriculture, The University of Sydney. (www.usyd.edu.au/su/agric.acpa).

Pringle, M.J., McBratney, A.B., Whelan, B.M. and Taylor, J.A. 2003. A preliminary approach to assessing the opportunity for site-specific crop management in a field, using yield monitor data. *Agricultural Systems* **76**, 273-292.

Readers may also find other useful information at:

http://www.clw.csiro.au/staff/BramleyR/publications.html

http://www.crcv.com.au/research/programs/one/project1.1.1.asp