A methodology to assess the accuracy and reliability of yield-monitor data

TA Jensen¹, D Gobbett², RGV Bramley² and AG Garmendia¹

¹National Centre for Engineering in Agriculture, University of Southern Queensland, Toowoomba, Australia; troy.jensen@usq.edu.au
²CSIRO Agriculture and Food, Waite Campus, Adelaide, Australia

Abstract  The ability to accrue full benefits from the adoption of precision agriculture (PA) technologies depends on having complete confidence in the data layers on which decisions are based. Previous research has shown that yield data can reliably show spatial patterns in within-block yield, but the process of filtering, manipulating and attributing yield-sensor data to sugar mill records for the block from which it was derived can have considerable bearing on the utility and reliability of the resultant yield maps. This paper reports on the production of data-handling tools to evaluate, clean and prepare the collected sensor data (harvester position/speed, chopper pressure/speed, feed roller separation, elevator pressure/speed and weigh pad) for input into a yield-mapping protocol for rigorous map generation. The tools developed use Python scripts have been developed using freely available Python libraries and Excel macros. The data manipulation steps include aggregating data packets, clipping to block boundaries, identifying individual harvest events, aligning with mill records, predicting yield and kriging the data. Applying the filtering protocol in an automatic fashion vastly reduced the total time required for the task, while ensuring that all harvest events were processed consistently and resulting in improved confidence in the resultant maps. The block mean yields derived from each yield map generated compares well with the mill tonnages for the individual blocks. In comparing the mill tonnage with the sensor-derived average yields for each block, all sensor yields are within the 95% confidence interval of the mill tonnage for all except one block. This is well within the level of accuracy in commercial yield sensors sold for other crops such as grains or wine grapes. The time to process the data has also been greatly reduced - from weeks to hours. The procedures and tools reported on in this paper have gone a long way towards automating the analysis of yield monitor data. They have enabled the datasets to be treated in a consistent and regimented fashion, with only limited manual input. This has improved confidence in the data on which the derived yield maps are based.

Key words  Precision agriculture, yield mapping, harvest event, protocol, sugarcane

INTRODUCTION

Yield-monitoring equipment has been available for cereal grains since the late 1980s (Searcy et al. 1989) and the first commercial yield map that was published was from a German canola crop in 1989 (Haneklaus et al. 1991; Schnug et al. 1991). Monitoring yield variation in sugarcane production systems has also been addressed by numerous researchers (Cox et al. 1998; Bramley and Quabba 2002; Magalhães and Cerri 2007; Price et al. 2011). However, yield-monitoring equipment is still not readily commercially available as an accessory on current-model sugarcane harvesters. Various attempts have been made at delivering a commercial yield monitor to Australian sugarcane producers but an evaluation of those available in 2008 (Jensen et al. 2010) found that they were strongly influenced by the presentation (i.e. lodged versus erect, etc...) of cane to the harvester. In turn, this led to doubt about the reliability of yield maps derived from them. Rather than using these devices, the same team undertook an evaluation of the fundamental yield-monitoring concepts that had been considered to that time (Jensen et al. 2013). Each concept was shown able to produce reliable yield maps, but the method of data manipulation, filtering and calibration to mill tonnages was seen to have a potentially major impact on the reliability of the resulting map. As a part of this work, and in order to produce maps in a consistent fashion, a yield-monitoring protocol specifically for sugar was produced (Bramley and Jensen 2014).

The protocol requires that the data be processed following a series of data-cleaning steps in order to remove aberrant values and obtain a consistent yield map. This process has, up until now, been done manually for every harvest event and takes considerable time to complete. Harvest events have to be identified using GIS software, clipped from a data file that might contain several days/weeks/months of data, and then the protocol has to be applied for each harvest event.
Invariably, as data refinements are made (such as changing sensor filtering thresholds), these steps may have to be repeated. To expedite these repetitive steps/processes, tools have been developed to make the process more streamlined. A brief description of the tools/protocol is given below. The current process requires manual intervention at several steps to input data. It is hoped that these will be fully automated and incorporated into a final app/software package.

**MATERIALS AND METHODS**

We used data collected during 2014 from the Bundaberg region of Queensland, Australia. The range of sensors tested and details of the associated logging equipment are given in Jensen et al. (2013); these were installed on a 1997 Austoft 7000 harvester. A recent addition to the system described by Jensen et al. (2013) is the use of a 3G serial modem (Maxon model EM-770W) to upload data packets every 15 minutes to a server for storage and future access. The previous method was to collect several weeks of data onto a CompactFlash (CF) card in the data logger and swap cards for data download on an occasional basis. Since the associated data tables can approach several hundred megabytes in size, such a mode of data collection made manipulation more difficult. The 15-minute data files from the modem are never larger than 200 kb.

The following principles have guided the development of the current tools:

- The tools should not require the use of costly software licenses, such as for proprietary GIS software;
- Processing should adhere to the Bramley and Jensen (2013) yield mapping protocol; and
- Tools should process large volumes of yield data with minimal user intervention.

The yield data processing is broken into five main steps summarised in Table 1. Rather than using proprietary GIS functions, the Python scripts have been developed using freely available Python libraries, such as:

- fiona: For reading and writing spatial data files (https://pypi.python.org/pypi/Fiona/).
- pandas: High-performance, easy-to-use data structures and data analysis tools (http://pandas.pydata.org/).

<table>
<thead>
<tr>
<th>Description</th>
<th>Implementation</th>
<th>Notes</th>
</tr>
</thead>
<tbody>
<tr>
<td>Block data preparation</td>
<td>Python script</td>
<td>Separate the cane block dataset (multiple blocks in one Shapefile) to single block per Shapefile. Generate GDA94 projected Shapefile and Vesper block grid to be used for kriging. This is dependent on consistent formatting and attributes in the cane block boundary Shapefiles, and may need to be revised or enhanced to accommodate differences between cane block spatial datasets.</td>
</tr>
<tr>
<td>Merge 15 minute yield data files</td>
<td>Python script</td>
<td>This processing merges the multiple &quot;.dat&quot; files created by the yield monitor, into a single file for each day of harvest. Some preliminary data cleaning is performed as part of this processing, including converting format of longitude and latitude coordinates, removing invalid coordinates and null sensor readings. Large numbers of .dat files are generated (i.e. Bundaberg 2014: 900 files). The script tracks which files have been processed so that additional files can be added (e.g. as received from the harvester) and merged without repro cessing all files.</td>
</tr>
<tr>
<td>Split daily yield files by block</td>
<td>Manual (using ArcGIS)</td>
<td>Using the individual block boundaries, separate the daily files of yield data into separate harvest events. This has been done manually to date, but an implementation of this as a Python script is envisaged. As part of this processing, the an automated tool would output a list of output a list of block name, date of harvest, and area harvested which can later be matched to mill data on a per event basis.</td>
</tr>
<tr>
<td>Clean, filter and adjust event yield data</td>
<td>Excel macro</td>
<td>This processing step, implemented as an Excel tool is used to read in the merged data files, apply the cleaning and filtering rules (e.g. elevator and ground speed, 3 second intervals, trim to +/- 3 standard deviations and adjust yields to the mill tonnage. This phase outputs a separate x, y, yield file for each sensor and event.</td>
</tr>
<tr>
<td>Mapping block yield</td>
<td>Python script</td>
<td>Merge the cleaned/adjusted sensor event files belonging to the same block and year. Use the filename to match the yield data to a block boundary. Krig the block data with Vesper (Whelan et al. 2002) and convert to TIF/raster for mapping. Calculate the mean block yield and 95% confidence interval from the kriged dataset.</td>
</tr>
</tbody>
</table>
We define a harvest event as the area harvested in a mapped cane block on a single day. A single block may be harvested over several days, sometimes a number of days apart and, thus, the data underpinning the yield map for a single block may comprise several harvest events. A harvester may also cut cane from multiple blocks in a single day. In generating maps of cane yield, it is best that the data from harvest events are pre-processed separately, and yield-monitor tonnages are calibrated to mill tonnages prior to combining with data from other harvest events on the same block.

RESULTS AND DISCUSSION

The yield monitor data from 11 'test' harvest events (detailed in Table 2) covers nine cane blocks from the Hubert farm (Fig. 1) in Bundaberg Queensland, and have been processed into yield maps. The Hubert farm is the focus of this analysis as it was one of the key sites used during our earlier work (Bramley et al. 2014). A yield map has been generated for each of five harvester-mounted sensors: chopper pressure 1 and 2 (CP1, CP2), roller opening (RO), elevator pressure (EP1) and elevator load cell (Cell).

Table 2. Summary of 11 harvest events processed.

<table>
<thead>
<tr>
<th>Block</th>
<th>Number of events</th>
<th>Area (ha)</th>
<th>Harvest date</th>
<th>Variety</th>
<th>Ratoon class</th>
<th>Notes</th>
</tr>
</thead>
<tbody>
<tr>
<td>15A</td>
<td>1</td>
<td>3.33</td>
<td>9 Oct 2014</td>
<td>KQ228</td>
<td>3R</td>
<td>Fertiliser trial</td>
</tr>
<tr>
<td>20A</td>
<td>1</td>
<td>2.38</td>
<td>2 Dec 2014</td>
<td>KQ228</td>
<td>3R</td>
<td></td>
</tr>
<tr>
<td>20B</td>
<td>1</td>
<td>3.07</td>
<td>2 Dec 2014</td>
<td>KQ228</td>
<td>3R</td>
<td></td>
</tr>
<tr>
<td>20C</td>
<td>1</td>
<td>1.58</td>
<td>2 Dec 2014</td>
<td>KQ228</td>
<td>3R</td>
<td></td>
</tr>
<tr>
<td>5A</td>
<td>1</td>
<td>2.70</td>
<td>3 Dec 2014</td>
<td>Q138</td>
<td>Old</td>
<td>NW end of W side of boundary does not match yield data well</td>
</tr>
<tr>
<td>26B</td>
<td>1</td>
<td>3.50</td>
<td>19 Nov 2014</td>
<td>KQ228</td>
<td>1R</td>
<td></td>
</tr>
<tr>
<td>26C</td>
<td>2</td>
<td>3.24</td>
<td>18-19 Nov 2014</td>
<td>KQ228</td>
<td>1R</td>
<td>Coverage of block 26C in 2 events</td>
</tr>
<tr>
<td>27A</td>
<td>2</td>
<td>1.09</td>
<td>17 Nov 2014</td>
<td>KQ228</td>
<td>1R</td>
<td></td>
</tr>
<tr>
<td>27B</td>
<td>1</td>
<td>1.05</td>
<td>17 Nov 2014</td>
<td>KQ228</td>
<td>2R</td>
<td>Yield data from a narrow strip of 1-2 rows are missing down W side of block</td>
</tr>
</tbody>
</table>

Fig. 1. Maps showing the blocks used for the yield mapping.
Applying the filtering protocol in an automatic fashion vastly reduced the total time required for the task while ensuring that all harvest events were processed consistently. Future improvements include the possibility to process sensors individually (in case a sensor presents faulty values and the tool is unable to follow the protocol, other sensor values can still be processed) and extension to regions other than Bundaberg (which might present slight differences in the data format and/or sensors present).

Figure 2 shows the yield maps generated for blocks 20A-C (Fig. 1, Table 2) using data collected using a load cell yield sensor. Since all of these blocks were third ratoon, variety KQ228<sup>3R</sup>, each map has been classified using the same legend on the assumption that yield potential for this variety and age was approximately the same in each block. However, it is evident both from the maps (Fig. 2) and also Figure 6 that the mean yield in each of the sub-blocks varies. It is worth emphasising here that each of the harvest events have been mapped separately – which is one reason for apparent discontinuities in the legend categories across sub-block boundaries.

Arguably, since the sub-blocks in Figure 2 are all of the same variety and crop age, they would have been better mapped together as a single map. However, in the absence of an electronic consignment system, and/or a more sophisticated means of block/crop class discrimination than reliance on standard mill data output, the degree to which our mapping process can be automated is confined to the sub-block and harvest event.

Figure 3 gives a similar presentation for blocks 5A, 26B,C and 27A,B using the roller-opening yield sensor. However, because of differences in mean block yield, a number of different legends have been used. Block 5A contains a plant crop of variety Q138 so it is unsurprising that its mean yield (65.1 t/ha) should be different to the other blocks which were planted to variety KQ228<sup>3R</sup>. Block 26B is first ratoon KQ228<sup>3R</sup> (mean yield of 69.1 t/ha); 27A is also first ratoon KQ228<sup>3R</sup> yet its mean yield is considerably higher (115.7 t/ha). In contrast, blocks 26C and 27B have similar mean yields (95.0 and 95.4 t/ha, respectively) even though 26C is a first ratoon and 27B is a second ratoon (both KQ228<sup>3R</sup>). Whilst the maps for the 3 sub-
blocks comprising 26B highlight the potential merit of being able to interpolate maps across sub-block boundaries where appropriate (see above discussion of Fig. 2), the presentation in Figure 3 is clearly confusing; the utility of the yield maps for any individual block or group of blocks is constrained by the apparent need to view them independently of neighbouring blocks. This is counter-intuitive given that patterns of yield variation tend to be driven by variation in the land (soil, topography) underlying the block as detailed by Bramley et al. (2014) and as many other studies in a range of crops have demonstrated (Bramley 2009; Bramley and Trengove 2013).

**Fig. 3.** Yield maps derived from the roller opening sensor in blocks 5A, 26B and C and 27A and B. Note the different legends that apply to different blocks.
Experience suggests that most farmers who adopt precision agriculture are interested in the identification of zones for differential treatment, rather than pursuing continuous variable rate application of inputs such as fertilizers or soil amendments. In light of this, the actual yield values underpinning a yield map, whilst important to the consideration of issues such as identification of potential yield, are somewhat less important to the identification of patterns of variation; a ‘low’, ‘medium’, ‘high’ classification is often adequate. With this in mind, the issues highlighted by Figure 3 were addressed by adjusting data on a sub-block or harvester event basis to a common mean value, a procedure detailed in Bramley et al. (2014). This adjustment has been implemented in Figure 4 for blocks 26 and 27 in an attempt to remove the confusion apparent in Figure 3; the maps for all blocks have been adjusted to a mean yield of 95.0 t/ha, the mean yield in block 26C. Whilst the result of this adjustment (Fig. 4) allows for a less confusing presentation by comparison with the unadjusted maps (Fig. 3), it is evident that in this section of the Hubert farm, other aspects of management in individual sub-blocks might have a greater bearing on yield variation in this mapped area than inherent variation in the underlying land; the contrast between the western-most and eastern-most sub-blocks of 26C and the rest of the area highlight this. Nonetheless, in further development of the map automation tool, we will be considering how to refine the processing and for what minimum area of interest, so that the results are those which will deliver greatest utility to the end-user.

![Yield maps derived from the roller opening sensor in blocks 26B and C and 27A and B when the mean yield in each block is adjusted to 95.0 t/ha, the mean yield for 26C. Each sub-block is planted to KQ228® and is first ratoon except for 27B which is second ratoon.](image)

Finally, Figure 5 shows the maps produced from all sensor options in block 15A. Note the southern side of the eastern-most corner of this block is a trial site that is part of Sugar Research Australia funded project (2014/045 - Boosting NUE in sugarcane through temporal and spatial management options). This N x K trial, was yield monitored along with the remainder of the block. Each of the blocks in the trial was four rows wide by 10 m long with each of the centre two rows being cut into a weigh truck so plot yields could be calculated; this necessitated stopping and starting the harvester every 10 m. It is instructive that the trial does not show up readily in any of the maps that were generated using the automated tool. The ‘fly out’ map in the centre of the graphic was specifically generated for only the trial area using non-standard protocols. On further investigation, because of the stopping and starting of the harvester, much of the data was excluded from the analysis using the current tool due to filtering of data 3 s either side of a break in harvesting in order to allow errors due to slow harvester and belt speeds to be removed. This filter resulted in only one valid data point being evident in each
of the 10 m row sections. Careful consideration should therefore be given to using the current tool in the analysis of trial data. It should also be noted that this block was suffering badly from soldier fly infestations and, following this harvest, was ploughed out. Soldier flies kill the stool, creating gaps in the row that may then be occupied by Guinea grass. Guinea grass appears as biomass to the roller opening sensor. However, by the time the biomass has progressed to the elevator, the primary extractor has removed most of the material, which is, therefore, not reflected by either the elevator pressure or load cell sensors. This may explain some of the differences in performance between the sensors.

**Fig. 5.** Yield maps for block 15A derived from different yield-sensing options.

The block mean yields derived from each yield map compare well with the mill tonnages (Fig. 6) for the individual blocks, including the blocks 26C and 27A for which maps were generated from two separate harvest events. In all blocks, the differences between sensor maps are within the 95% confidence interval of the maps. Comparing the mill tonnage with the sensor-derived average yields for each block, all sensor yields are also within the 95% confidence interval of the mill tonnage, except for map for 27A where the mill-reported block yield falls outside the 95% confidence range of the maps for sensors CP1, EP1 and RO1 but these sensor yields are still within 6.0% of the mill yield recorded for the block. This is well within the level of accuracy in commercial yield sensors sold for other crops such as grains or wine grapes.

**CONCLUSIONS**

The procedures and tools we report have gone a long way towards automating the analysis of yield-monitor data. They have enabled datasets to be treated in a consistent and regimented fashion, with only limited manual input. As this work is ongoing, we plan to fully automate data processing, as well as making the format sufficiently generic so that other organisations/institutions can utilise this tool to analyse their datasets.

In the sample blocks analysed, the mean yields derived from each yield map fall within the 95% confidence interval of the mill tonnage, in all but one of the datasets. There were, however, some discrepancies among some of the yield maps derived for the different sensors when an unusually high number of stoppages occurred due to the cutting of a fertiliser trial (block 15A). The yield-mapping protocol and tool is, therefore, being reviewed for its suitability for use in plot-based trials.
Fig. 6. Block mean yields (t/ha) and 95% confidence intervals (t/ha) derived from five different sensors compared with mill tonnages.

ACKNOWLEDGEMENTS

We gratefully acknowledge the financial support from Sugar Research Australia (SRA) for project CSE022 (A collaborative approach to Precision Agriculture RDE for the Australian Sugar Industry) and for the continuing funding for project 2014/028 (Product and profit – Delivering precision to users of Precision Agriculture in the Australian Sugar Industry – Yield Monitoring). A special mention goes to the Hubert Family in Bundaberg for their continued perseverance with researchers asking them to interrupt their usual commercial practice.

REFERENCES


Metodología para evaluar la exactitud y fiabilidad de los datos de monitores de rendimiento

Resumen: La habilidad de obtener el máximo beneficio de las tecnologías de Agricultura de Precisión (AP), utilizadas para la toma de decisiones, depende de tener completa confianza en las diferentes capas de información. Previas investigaciones han demostrado que la información recogida por monitores de rendimiento es capaz de mostrar tendencias espaciales del rendimiento a nivel de parcela, pero el proceso de filtrado, manipulación y emparejamiento de la información sensorial a los datos del ingenio para la parcela de la que proceden puede tener un considerable efecto en la utilidad y fiabilidad de los mapas de rendimiento resultantes. Este artículo presenta la producción de herramientas para evaluar, limpiar y preparar la información sensorial recogida (posición y velocidad de la cosechadora, velocidad y presión de las cuchillas, ángulo de los rodillos, velocidad y presión del elevador y sensor de peso) para su uso en un protocolo de mapeo que genere mapas de forma rigurosa. Estas herramientas han sido desarrolladas utilizando librerías gratuitas de Python y macros de Excel. Los pasos para manipular los datos incluyen la agregación de paquetes de datos, recortar los datos a los límites de parcela, identificar eventos de cosecha individuales, emparejar con información del ingenio, predicción del rendimiento y krigaje de los datos. La aplicación del protocolo de forma automática redujo enormemente el tiempo requerido para esta tarea, además de garantizar que todos los eventos de cosecha fueran tratados de manera consistente y ofreciendo mayor confianza en la calidad de los mapas resultantes. El rendimiento medio por parcela, derivado de cada mapa de rendimiento, se corresponde adecuadamente con el tonelaje del ingenio para cada parcela individual. Comparando el tonelaje del ingenio con el rendimiento medio derivado de los sensores, todos los rendimientos de los sensores se encuentran dentro del 95% del intervalo de confianza, excepto en una parcela. Esto se encuentra dentro del nivel de exactitud de monitores de rendimiento comerciales utilizados en otros cultivos como cereales o vid. El tiempo requerido para su procesamiento también se ha reducido – de semanas a horas. Los procedimientos y herramientas presentados en este artículo han avanzado el proceso de automatización del análisis de la información recogida por los monitores de rendimiento. Posibilitan que los datos sean tratados de manera consistente y regimentada, con tan solo un reducido aporte manual. Esto ha mejorado la confianza en la información utilizada para generar mapas de rendimiento.

Palabras clave. Agricultura de precisión, mapa de rendimiento, evento de cosecha, protocolo, caña de azúcar