Appendix 14 – Herbert Case Study

This study site comprised four sub-blocks on the farm of Brian and Paul Tabone in the Braemeadows region of the Herbert River district. The total intended study area of 26.7 ha (Figure 14.1) was made up of blocks 5-5 (3.54 ha, re-planted in 2009), 5-8 (10.26 ha, re-planted in 2008), 6-1 (6.76 ha, planted post-fallow in 2009) and 6-3 (6.18 ha, re-planted in 2009). All blocks were planted to Q200 except 6-1 which was planted to KQ228. This was ploughed out following the 2012 harvest. Average annual rainfall at this site is 2137 mm.

Figure 14.1 The Herbert study site. From east to west, the four sub-blocks making up the study area were 5-5, 5-8, 6-1 and 6-3; see text for further explanation.

Herbert River growers have a valuable resource in the form of the 'CSR' 1:5,000 soil survey of Wood et al. (2003)[1]. Even though the much higher resolution of this survey by comparison with conventional reconnaissance soil survey is not sufficient for integration with other spatial data for the purposes of Precision Agriculture (see Figure 1 in the main report, for example), it nonetheless provides growers with valuable information about their soils. Thus, Figure 14.2 shows how the study site is dominated by the acidic 'fine grey sandy loam' and 'sand ridge' soils of Wood et al. (2003). These are Kandosols in the terminology of the Australian soil classification (Isbell 1996) and are characterised by shallow fine grey sandy topsoils over coarser yellow sands. Analytical data are presented in Appendix 34; the soils contain little clay, have very low organic matter contents and are quite acidic.

[1] Wood, A., Schroeder, B., Stewart, S., 2003. Soil specific management guidelines for sugarcane production – Soil Reference booklet for the Herbert District. CRC for Sustainable Sugar Production, Townsville.

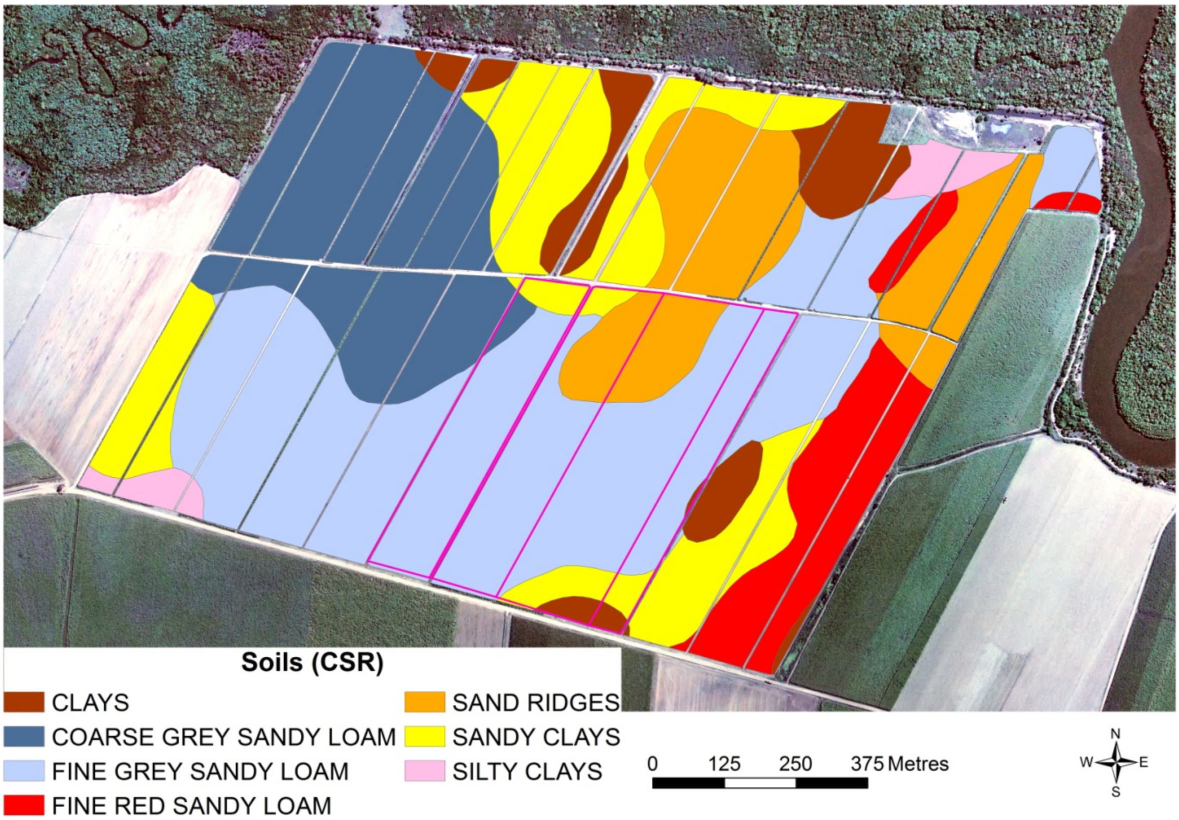

Figure 14.2 Study site soil map derived from the 1:5,000 'CSR' soil survey of Wood et al. (2003).

Whilst a useful resource, the resolution of Figure 14.2 is inconsistent with that of remotely sensed imagery or yield mapping and accordingly, as at the other project study sites, we conducted EM38 soil survey – in this instance using a dual dipole MK2 EM38 sensor (Geonics, Canada) - and also simultaneously surveyed the site using gamma radiometrics. The maps derived from this work are shown in Figures 14.3-14.5.

The soil conductivity maps shown in Figure 14.3 have a number of notable features. First, the values of apparent electrical conductivity (EC_a) are low, as is the range of variation in EC_a across the site. Both of these are reflective of the low clay contents (Appendix 34) and sandy nature of the soils; indeed, when sampling these soils, the coarse, pale coloured and structureless nature of the subsoils with an absence of fine material was striking in its similarity to coarse beach material. On the other hand, the low EC_a values are indicative of the soils not being saline, even at depth, and it is therefore inferred that saltwater incursion from the nearby tidal mangroves is not an issue for these study blocks; whether this is also the case for the blocks immediately to the north is unknown, but would certainly be of interest should the Tabones decide to implement PA in the future.

Also of note, but not unexpected given the shallow topsoil, is the similarity between the 'shallow' (0-75 cm) and 'deep' (0-150 cm) surveys and for this reason, we used the 'shallow' map only for selection of soil sampling locations (Figure 14.4) and for comparison with other

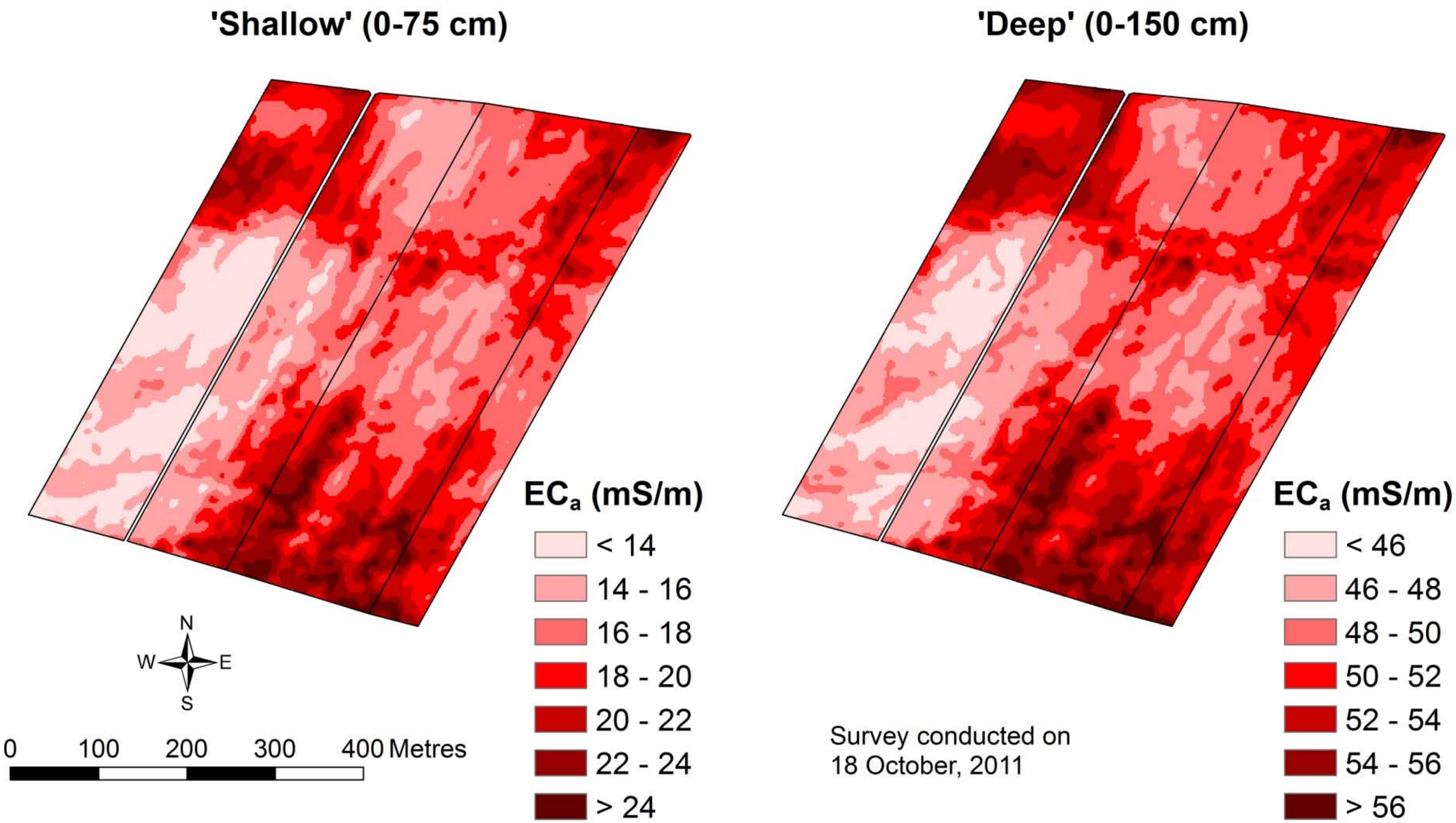

Figure 14.3 Soil conductivity maps derived from EM38 soil survey at the Herbert study site. Since a dual dipole instrument was used for this survey, maps covering both the 'shallow' and 'deep' depth ranges were produced.

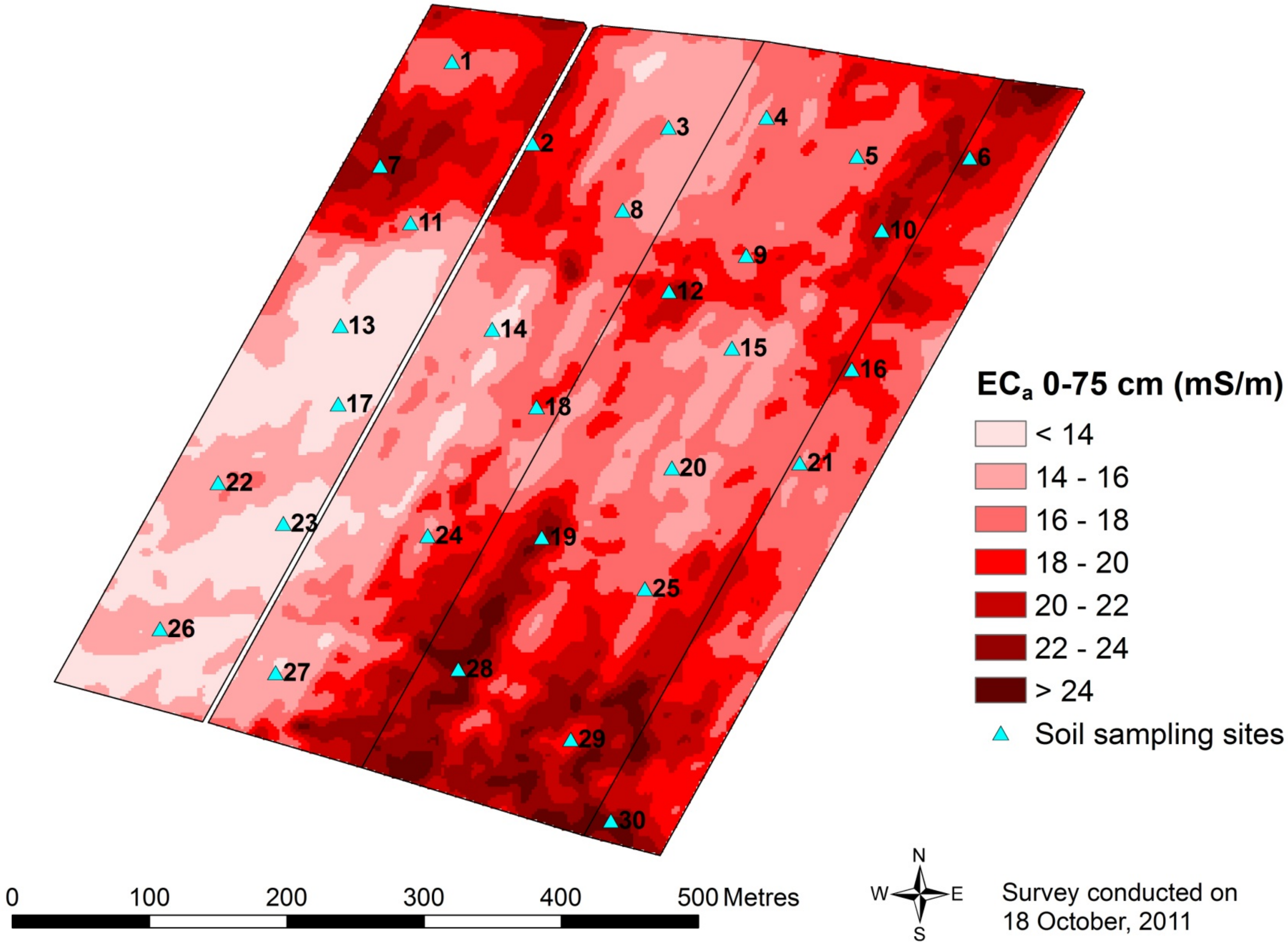

Figure 14.4 Locations chosen for soil sampling on the basis of the variation in the EM38 map.

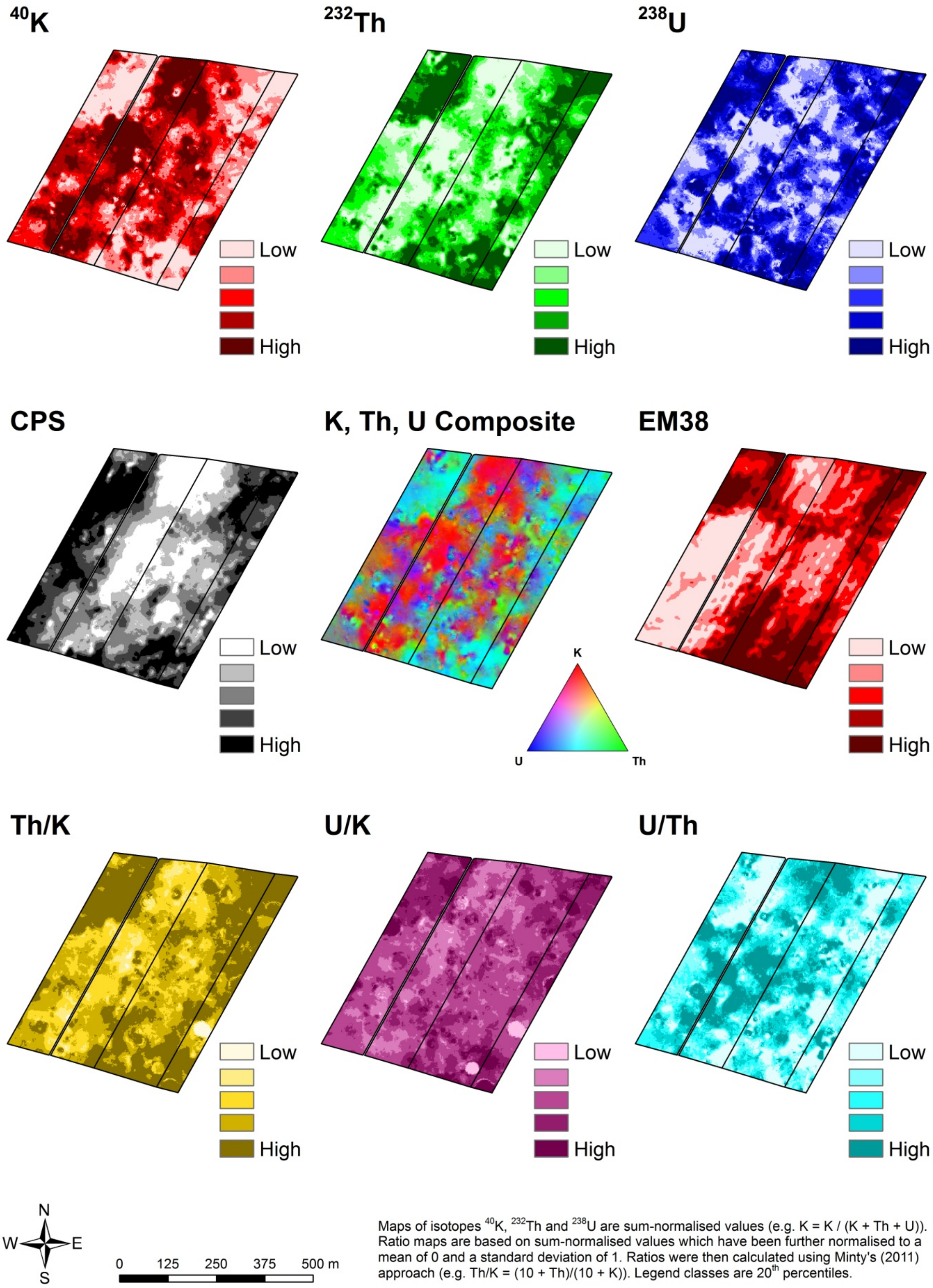

Figure 14.5 Maps derived from gamma radiometric survey of the Herbert study site; the EM38 map (shallow; Figure 14.3, 14.4) is also included for comparison.

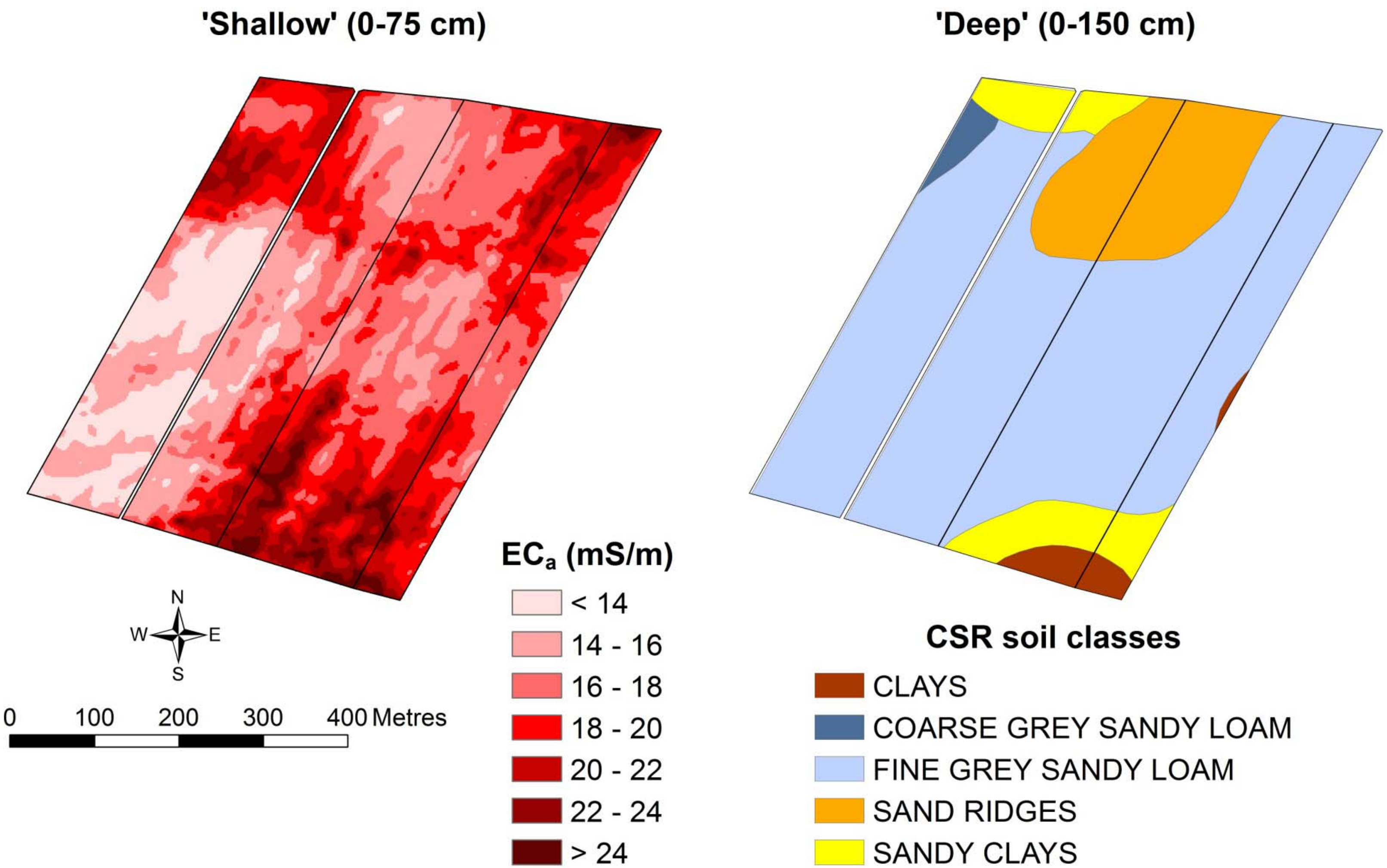

Figure 14.6 Comparison between the EM38 and CSR soil surveys.

soil information. An unusual feature of these maps is the fact that much of the variation is aligned with the row orientation (SSW-NNE) or, in the case of the northern 'boundary' of the low EC_a area in the SW, it runs perpendicular to the rows, and in the case of this 'boundary' runs right across the block. The low EC_a area in the south west is likely an area of 'cut' in the initial 'cut and fill' laser levelling operation used to flatten the block whilst the row-oriented variation may reflect subsequent laser levelling used to promote drainage of surface waters in light of the close proximity to the mangroves to the north and coast to the east.

The patterns of variation in the gamma radiometrics maps (Figure 14.5) are somewhat similar; areas of low K / high Th in the NW correspond to areas of higher EC_a, suggesting that the EM38 survey is primarily responding to variation in the clay content of the upper part of the soil profile. This conclusion is supported by the fact that the areas of higher ECa tend to be where the map of Wood et al. (2003) indicates the presence of clays and sandy clays. Rodrigues et al. (2014) report a reasonable correlation between both clay content and CEC and 'fused' data from the EM and gamma data, whilst the composite gamma map has a spatial structure similar to that seen in the EM38 map. Nonetheless, the range of variation in radiometric counts was low and these soils appear somewhat radiometrically inactive.

Whilst the Tabones use guidance on some of their machinery, we were regrettably unable to download elevation data from their system. As indicated in Appendix 13 (Burdekin) such data are invaluable in gaining understanding of within-paddock variation. The RTK system which enables the sub-decimetre accuracy of machine guidance also delivers accurate elevation data and growers need to ensure that they can source this information and should seek assistance from equipment suppliers as required; experience with PA in a range of cropping scenarios strongly suggests that a digital elevation model (DEM) is almost always a useful spatial data layer for understanding within-block variation. In this case, given the history of land planing, a pre-levelling DEM might be even more valuable. Attempts at accessing such data were not successful.

In light of the lack of elevation data from the machine guidance system, we sought an alternative from the Herbert Resource Information Centre. During the early stage of the project, they had flown the entire region with a lidar sensor with the intention of using this as the basis for a regional digital elevation model. Figure 14.7 shows the lidar data so acquired over the Tabone farm and reveals a surprising alignment to the block boundaries. When an aerial photograph of the site was draped over the DEM derived from the lidar data (Figure 14.8), the reasons for this were clear – the lidar had mapped the surface of the vegetation (i.e. the sugarcane) rather than the land surface.

Clearly, soil variation showing abrupt changes that are either aligned with, or perpendicular to the rows, is not reflective of the continuous variation in soil that is typically seen in less disturbed landscapes. Furthermore, the low range of values in both EC_a and gamma counts and apparent lack of a strong association between variation in yield and soils (see below) suggests that at this site at least, the expected link between soil variation and crop performance has largely been broken by the effects of land forming. Given also the very sandy nature of the soils at this site, it seems possible that, in effect, this is a hydroponic production system in which the role of the soil is simply predominantly one of providing a foundation material for the sugarcane to grow in.

The Tabone harvester had been fitted with a Solinftec (formerly TechAgro) yield monitor as part of the yield mapping work undertaken by Herbert Cane Productivity Services Ltd (HCPSL) prior to commencement of the project. Whilst the data acquired using this sensor presented as a potentially valuable resource, a rigorous evaluation of it led us to have grave concerns as to its utility for agronomic and business decision-making. Appendices 14a and 14b (attached at the end of this Appendix) explain why. Note that these documents formed much of the basis for our yield monitoring and mapping protocols (Appendix 31) and as such, describe work of considerable value to CSE022.

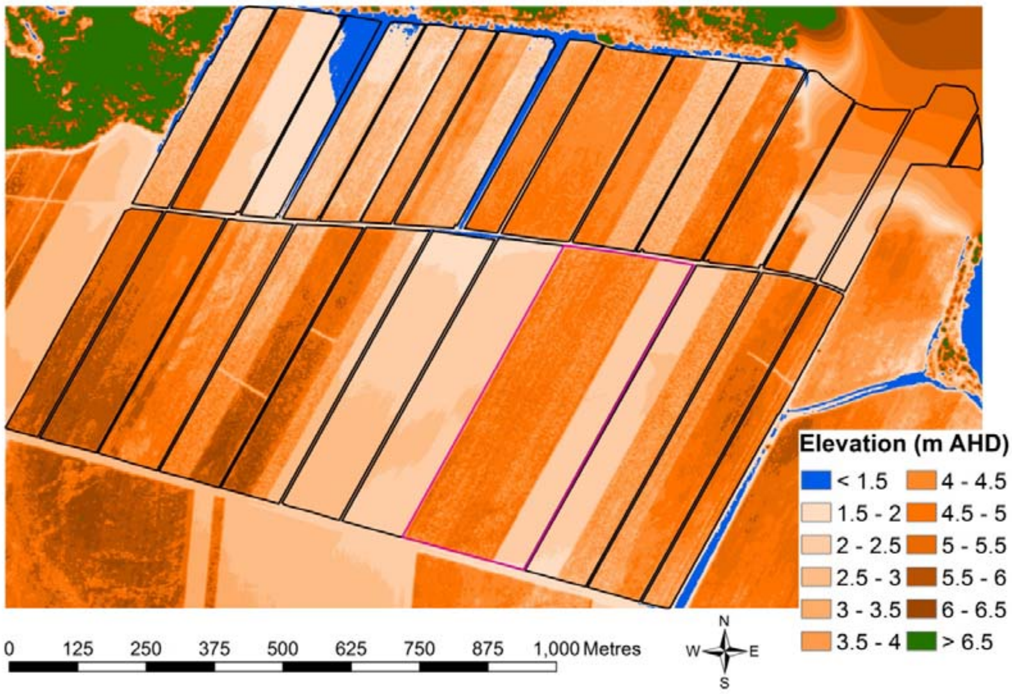

Figure 14.7 Variation in apparent elevation as measured using an airborne lidar sensor. Data provided by the Herbert Resource Information Centre.

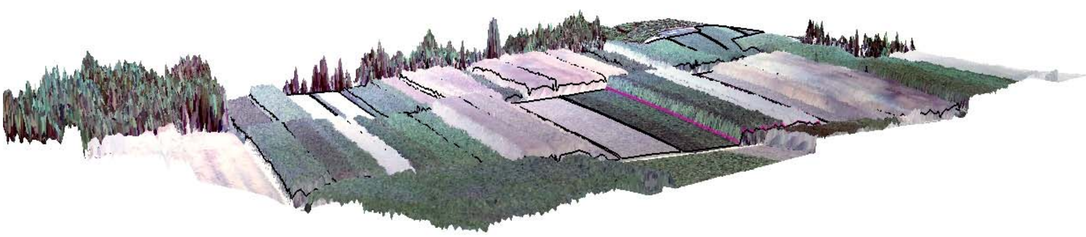

Figure 14.8 Aerial photo of the study site draped over the elevation model shown in Figure 14.7. As can be seen, the lidar has mapped the vegetation surface rather than the land surface. Note that an accurate regional DEM could still be interpolated from these data providing it drew only on data which did reflect the land surface (e.g. roads, bare fallow, etc…).

Initial inspection of the yield maps derived from Solinftec data in 2010 and 2011 alongside remotely sensed imagery (Figures 14.9 and 14.10) obtained in May 2011, tends to lend weight to our concerns at yield map quality – a conclusion which relies on the assumption that the camera never lies. However, this initial interpretation may not be correct. Figure 14.9 compares the 2011 image (cloud cover prevented capture of a useable image in 2010) with the 2011 yield map on the assumption that since sugarcane is a quasi-perennial, its patterns of variation might be expected to be similar between years – as was the case for example at our Burdekin site (Appendix 13). In fact, there is some agreement between the yield map and image in the apparently low-yielding area at the northern end of sub-block 5-8, but otherwise very poor alignment in the patterns of variation in the image with the yield map.

Comparison of the same 2011 image with the Solinftec-derived yield map for 2011 suggests somewhat greater alignment (Figure 14.10), especially in sub-block 6-3, and again, with respect to the low-yielding, albeit striped area to the north of 5-5. However, there is almost no alignment between the image and yield map in 6-1.

When the images for the two years are compared (Figure 14.11) and the underpinning data clustered using *k*-means (not shown), there are few similarities identified between the two years, with differences between the zones identified by *k*-means only differing in mean NDVI values in terms of their second decimal place.

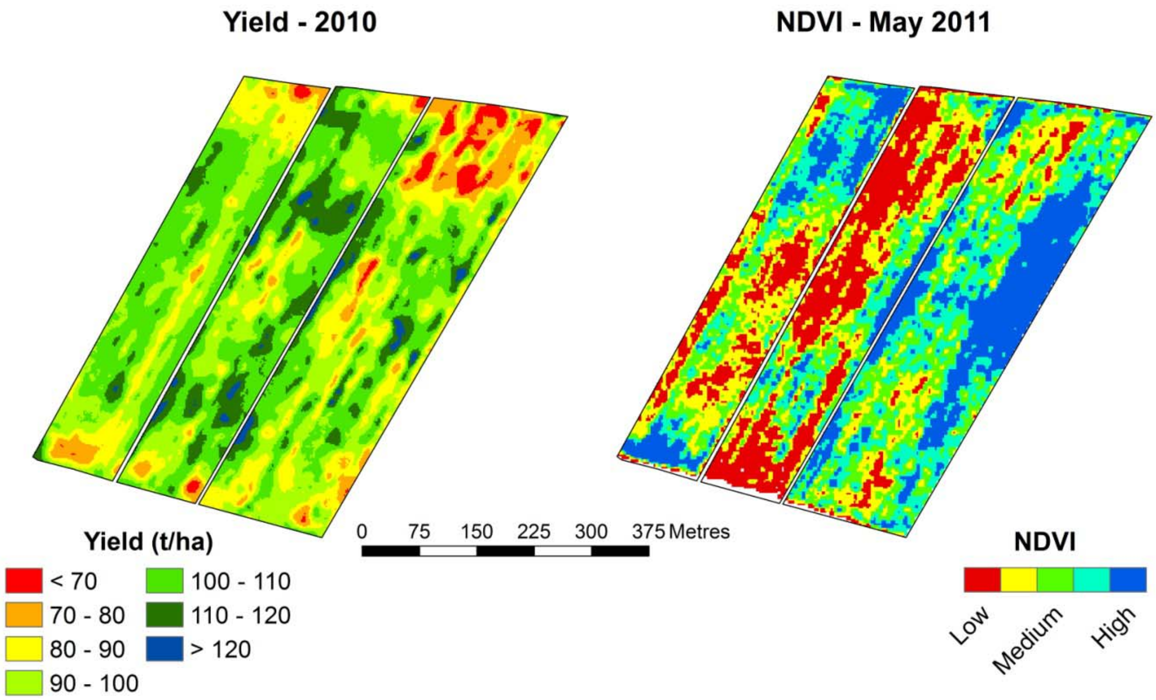

Figure 14.9 Comparison between the yield map derived from the Solinftec yield monitor in 2010 and remotely sensed imagery (Ikonos satellite) obtained in May, 2011.

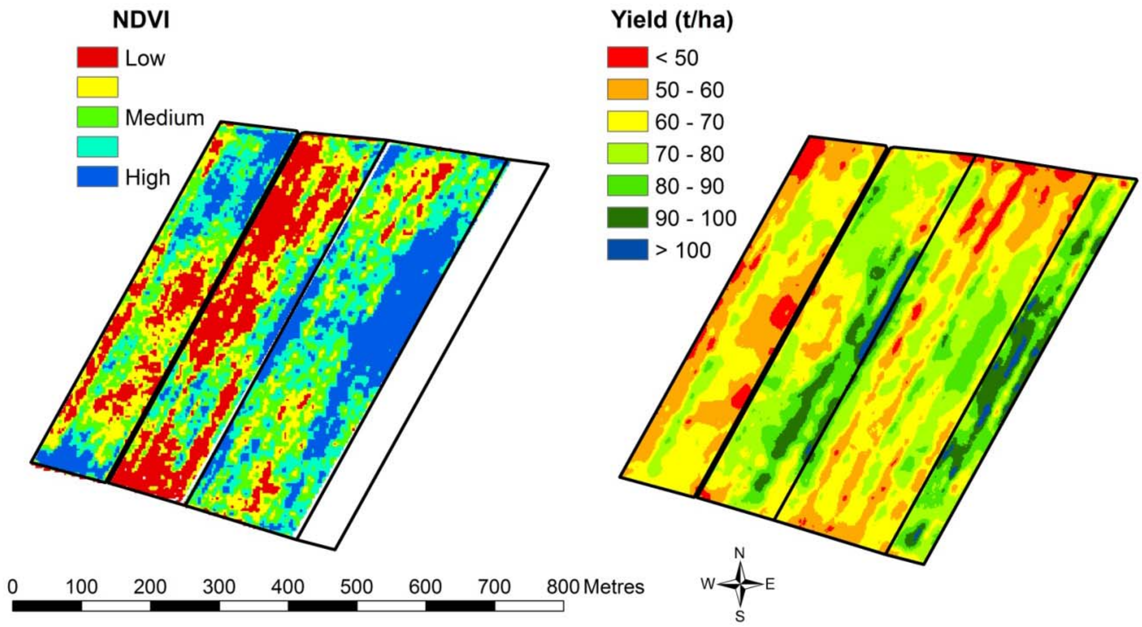

Figure 14.10 Comparison between the yield map derived from the Solinftec yield monitor in 2011 and remotely sensed imagery (Ikonos satellite) obtained in May of the same year.

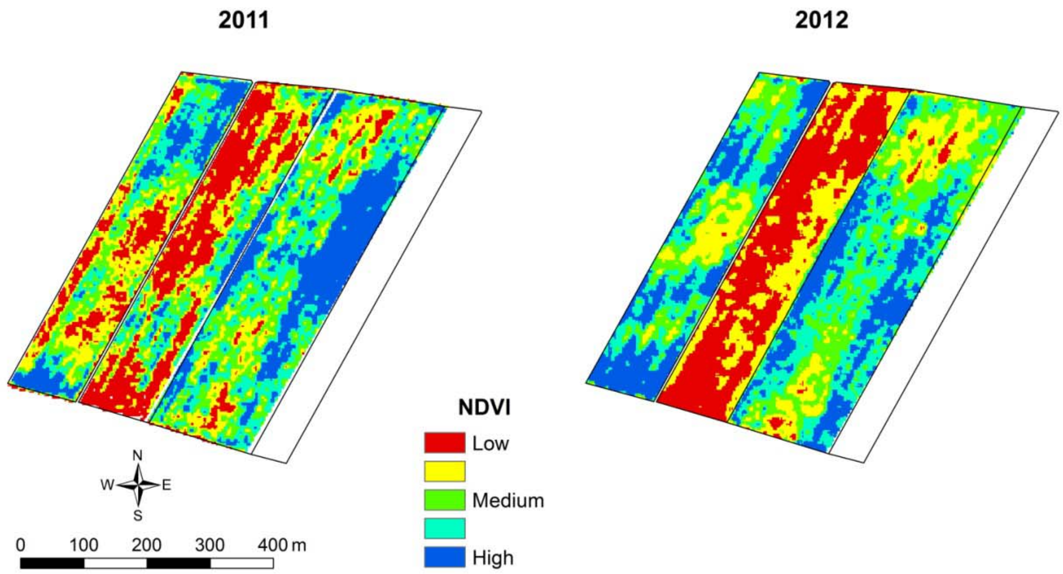

Figure 14.11 Comparison between remotely sensed imagery (Ikonos satellite) obtained in May 2011 and April 2012.

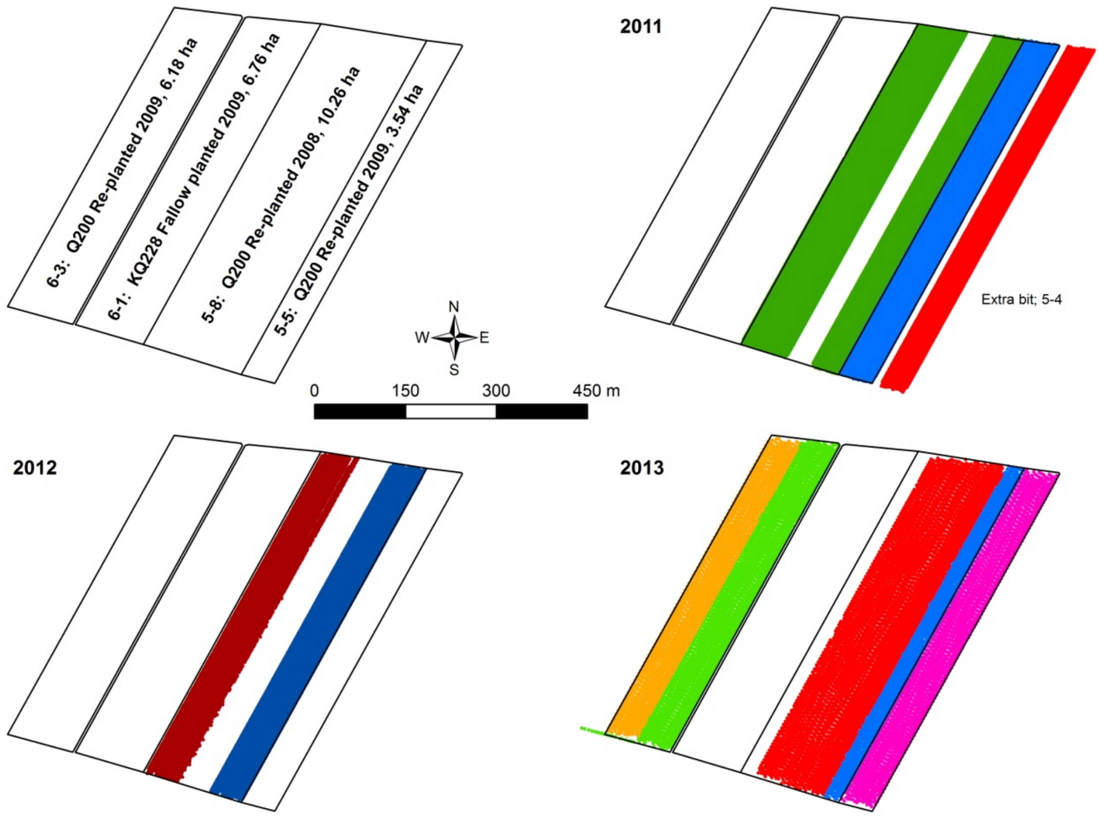

Figure 14.12 Harvester trace in 2011-13, as recorded through our own yield sensing; note that block 6-1 did not contain a crop in 2013.

Given our lack of confidence in the TechAgro/Solinftec data, it has been most regrettable that capture of yield data using our own yield sensors has been 'patchy' (Figure 14.12) – a consequence of the combined effects of equipment failure and some early season harvesting for which we were not sufficiently well prepared. As a consequence of this, our yield maps derived from the Elevator Pressure sensor in 2011 are confined to sub-blocks 5-5 and 5-8 (Figure 14.13), to 5-8 only in 2012 (Figure 14.14) and to 5-5, 5-8 and 6-3 in 2013 (Figure 14.15). In producing these yield maps, following the careful data cleaning, trimming, calibration and display procedures outlined in Appendices 31-33, it apparent that the range of yield variation is much lower than was seen at our Burdekin site, or indeed when the maps were produced from the Solinftec sensor (Figures 14.9 and 14.10). This observation suggests that perhaps our greatest limitation in data interpretation at this site is that it is simply not very variable. Indeed, when the yield data shown in Figures 14.13-14.15 are clustered with the remotely sensed imagery (Figure 14.11) and EM38 soil data (Figure 14.4), for sub-block 5-8, it was not possible to delineate zones for which differential targeted management might be warranted whether all of the data, or just selected layers were included in the analysis. This Herbert case study therefore highlights an important truism with respect to PA; just because PA clearly has potential to deliver advantage at some sites, does not mean that it will deliver such advantages ubiquitously. Whilst it is possible and even likely that over the entire area farmed by the Tabone family (Figures 14.2 and 14.2), there is good reason to consider differential management, within the 26.7 ha selected for this study, and on the basis of the data we collected during it, we have not been able to

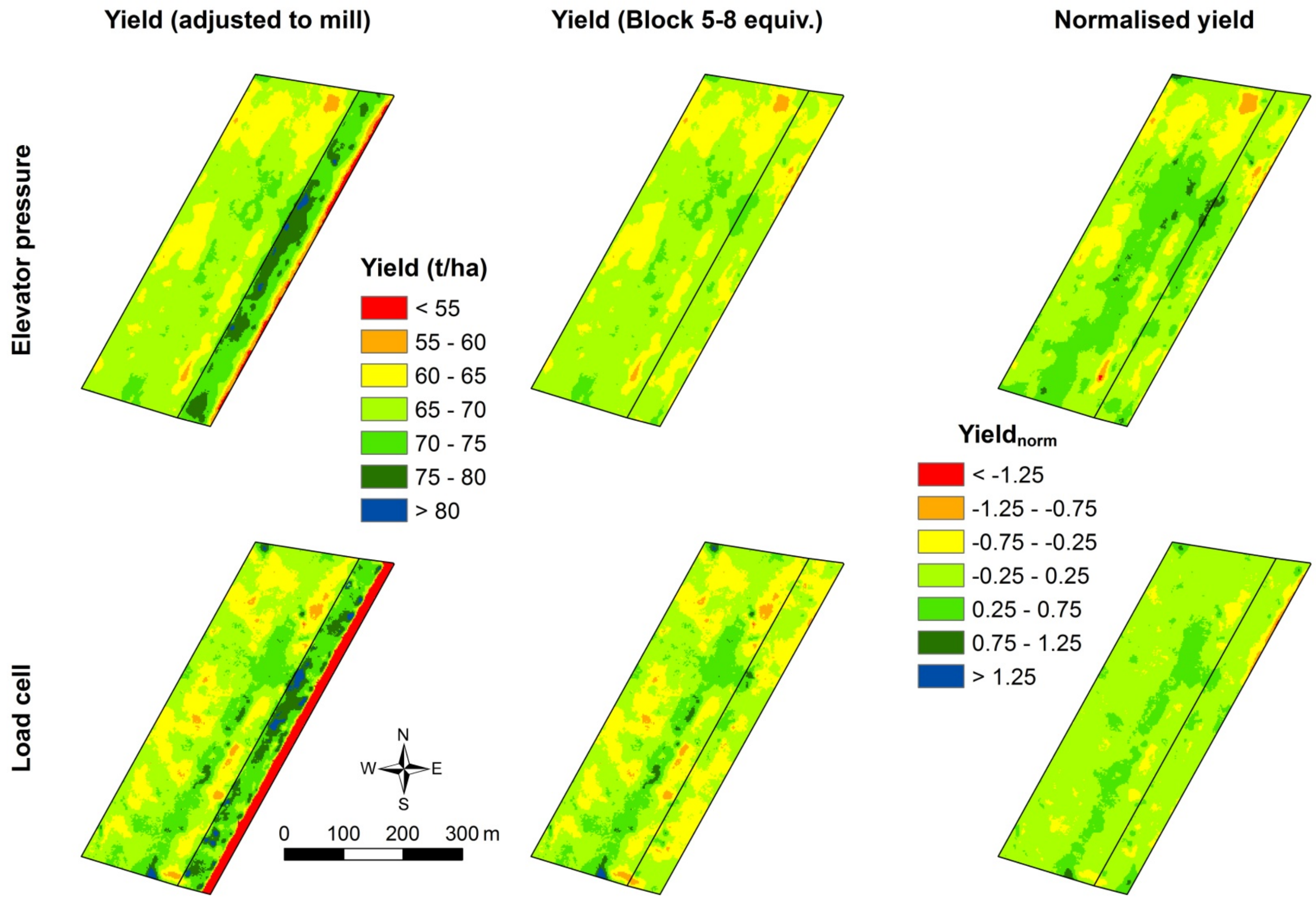

Figure 14.13 Yield maps obtained in 2011 for sub-blocks 5-5 and 5-8 using the elevator pressure sensor (see Appendix 26). Here the data have been mapped either 'as delivered' with the means adjusted to match those recorded at the mill on a per harvest event basis, with the means adjusted to the mean yield in sub-block 5-8, or in terms of normalised yield (mean of zero, standard deviation of 1). See Appendix 33 for further explanation.

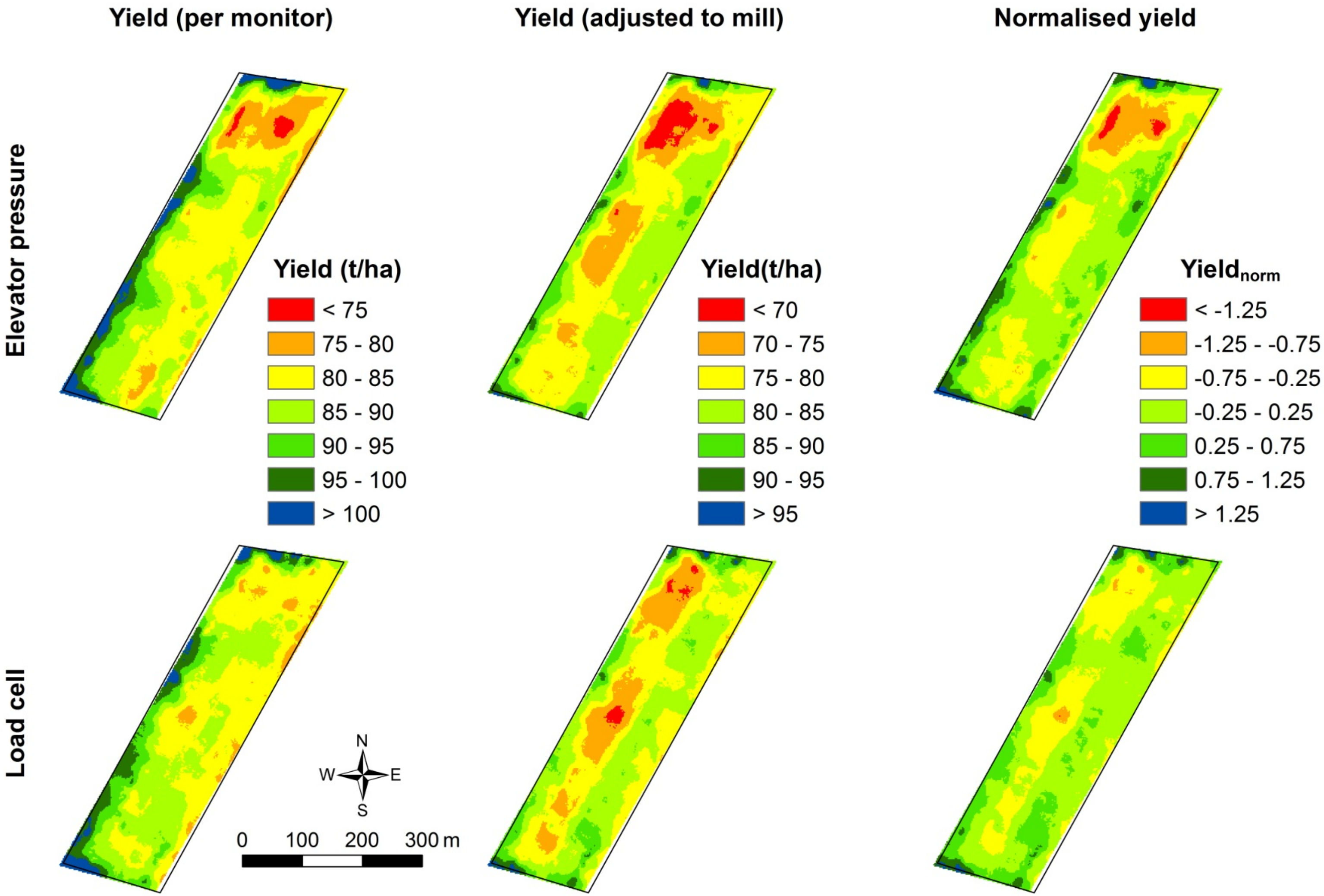

Figure 14.14 Yield maps obtained in 2012 for sub-block 5-8 using the elevator pressure sensor. See Figure 14.13 for further explanation of map formats.

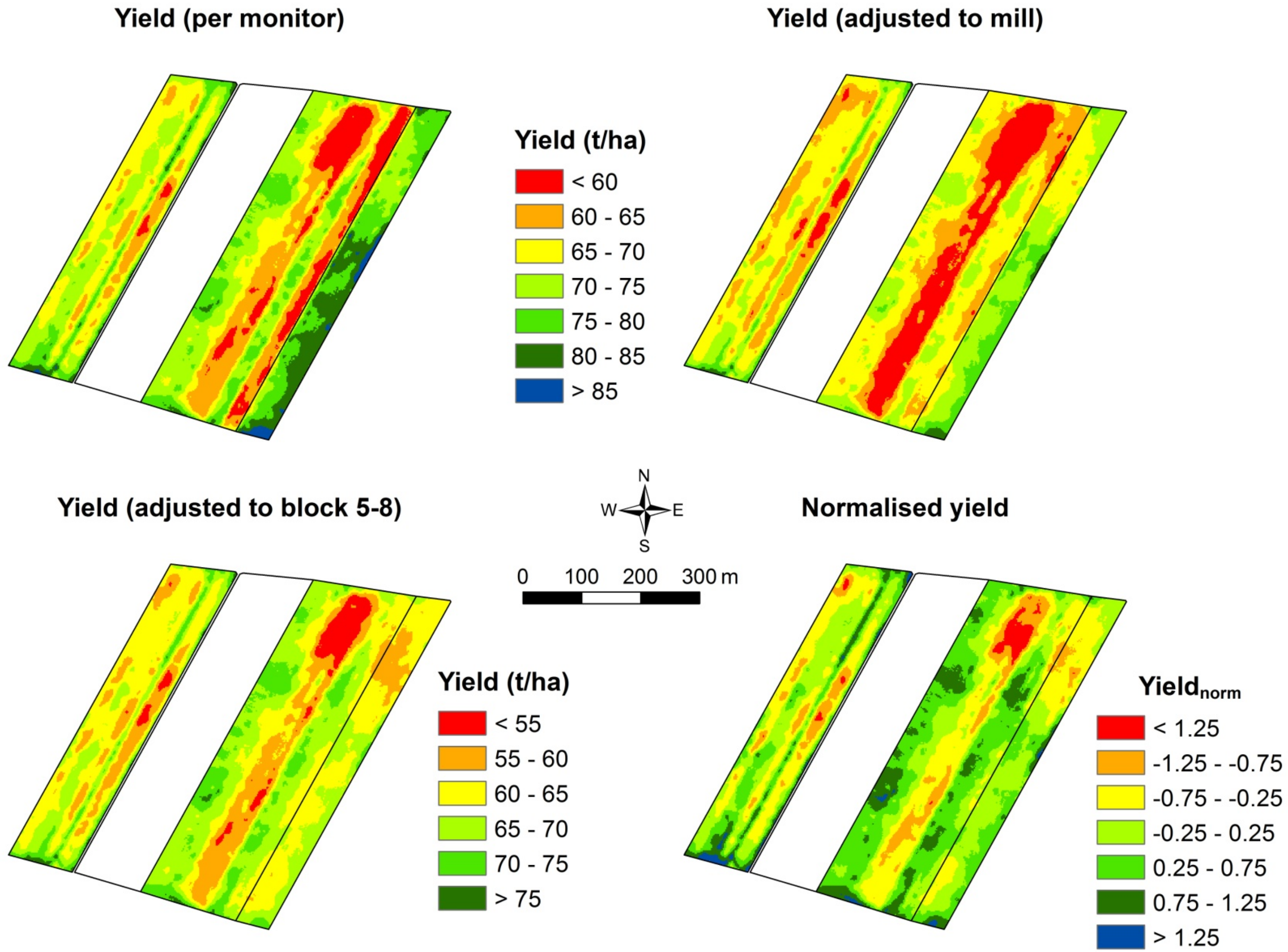

Figure 14.15 Yield maps obtained in 2013 for sub-blocks 5-5, 5-8 and 6-3 using the elevator pressure sensor. See Figure 14.13 for further explanation of map formats.

establish a basis for targeting management. Whether this is due, in part, to the effects of laser levelling on coarse sandy soils is unclear. However, within the 26.7 ha study area, conventional uniform management does appear to be the optimal strategy. Issues such as management of soil acidity and the build up and retention of soil organic matter will arguably therefore deliver greater benefit:cost than attempts to adopt technologies such as yield mapping or variable rate fertilization.

Acknowledgments

We are indebted to Brian and Paul Tabone for allowing us to conduct this work on their farm. Their assistance and support, along with that of their harvester crew has been greatly appreciated. We are also grateful to Lawrence Di Bella and Mike Sefton (HCPSL) and Raymond De Lai (HRIC) for their assistance with providing and understanding Mill records and other background data, to Dr Andrew Robson (University of New England; formerly QDAFF) for providing the remotely sensed imagery used in this study, and to Santiago Marrero (Solinftec) for providing the data derived from the Solinftec/TechAgro yield monitor in 2009-11.

Appendix 14a: Analysis of TechAgro yield monitor data for 3 blocks of the Tabone farm, season 2010 (May 2011)

Rob Bramley
CSIRO Ecosystem Sciences, Waite Campus, PMB 2, Glen Osmond, SA5064

Background

SRDC-funded project CSE022 seeks to establish the basis for the adoption of Precision Agriculture (PA) in the Australian sugar industry. A key aspect of this work is to quantify and understand the nature and extent of yield variation at the within-block scale. In turn, this is highly dependent on both access to robust yield data at high spatial resolution, and also its appropriate mapping. Recent work by Jensen et al (2010) has questioned the efficacy of current commercially available yield monitoring technology. Notwithstanding the findings of that work, the potential for improvements to the yield monitoring systems studied by Jensen et al (2010) and the likely future commercial availability of additional yield monitoring systems, the existing availability of yield monitor data nevertheless represents a potentially valuable resource to CSE022. The present study therefore sought to consider alternative approaches to yield monitor data processing, especially in regard to data pre-processing prior to mapping. It was undertaken with a view to informing refinement of a draft mapping protocol for sugarcane yield maps (Bramley – in preparation), and also to providing key data for progressing CSE022 at one of our core study sites - the farm of Brian Tabone in the Herbert River district (Macknade Mill area).

Study site and data availability

The CSE022 study site on the Tabone farm (5226A) comprises 3 sub-blocks (6-3, 6-1 and 5-8) of approximately 6.2, 6.6 and 10.2 ha (22.9 ha in total). These were planted to either Q200 or KQ228 in either 2008 or 2009 such that at the time of harvest in 2010, block 6-1 was under a plant crop of KQ228, 6-3 was under a replant crop of Q200, whilst 5-8 was 1st ratoon Q200. Given the likelihood of differing yield potential between varieties and crop ages, the study area therefore represented a useful co-location of mixed crop age and variety as is typical of many Australian sugar farms.

Yield monitor data were collected over parts of the Tabone farm during the 2010 harvest season and these data were made available for the present work by Herbert Cane Productivity Services Limited (Mike Sefton) and TechAgro Australia, now Solinftec (Santiago Marrero). Fortunately, yield monitor data were collected over the entire CSE022 study site (Figures 1-3).

Data evaluation

The TechAgro yield monitor operates on the basis of sensing the roller opening during harvest and uses a proprietary algorithm, along with measurement at the Mill of the tonnage of cane delivered, geo-location using GPS, and knowledge of the cane row spacing, to provide geo-referenced estimates of yield expressed in t/ha.

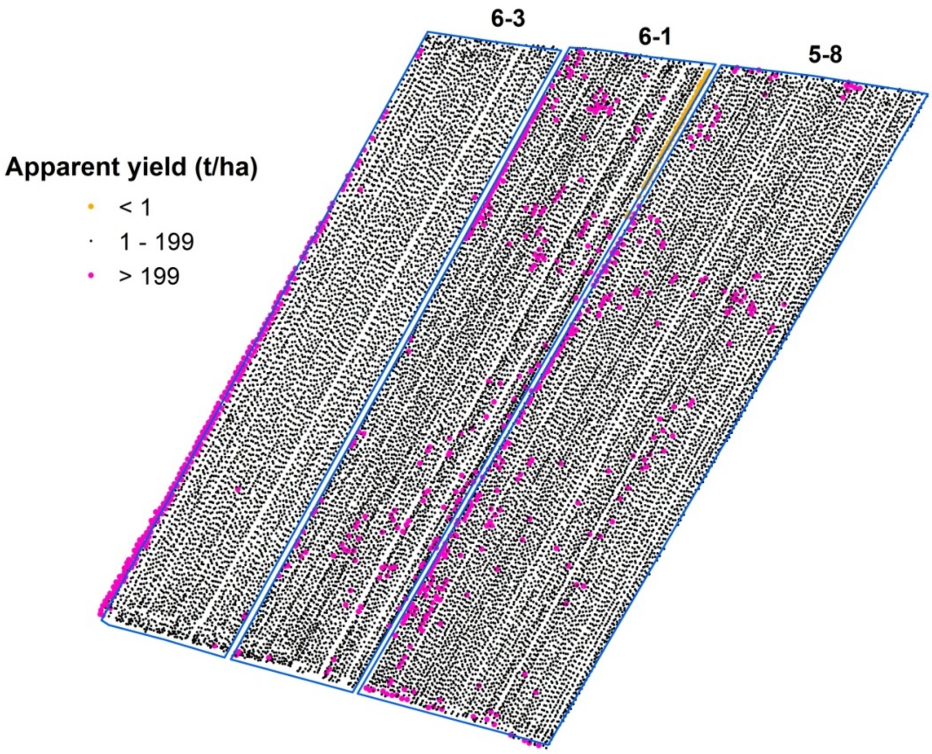

Figure 1. Yield monitor trace for the 3 study blocks. Also shown are the locations of data records for which yield was recorded as either > 200 t/ha or < 1 t/ha.

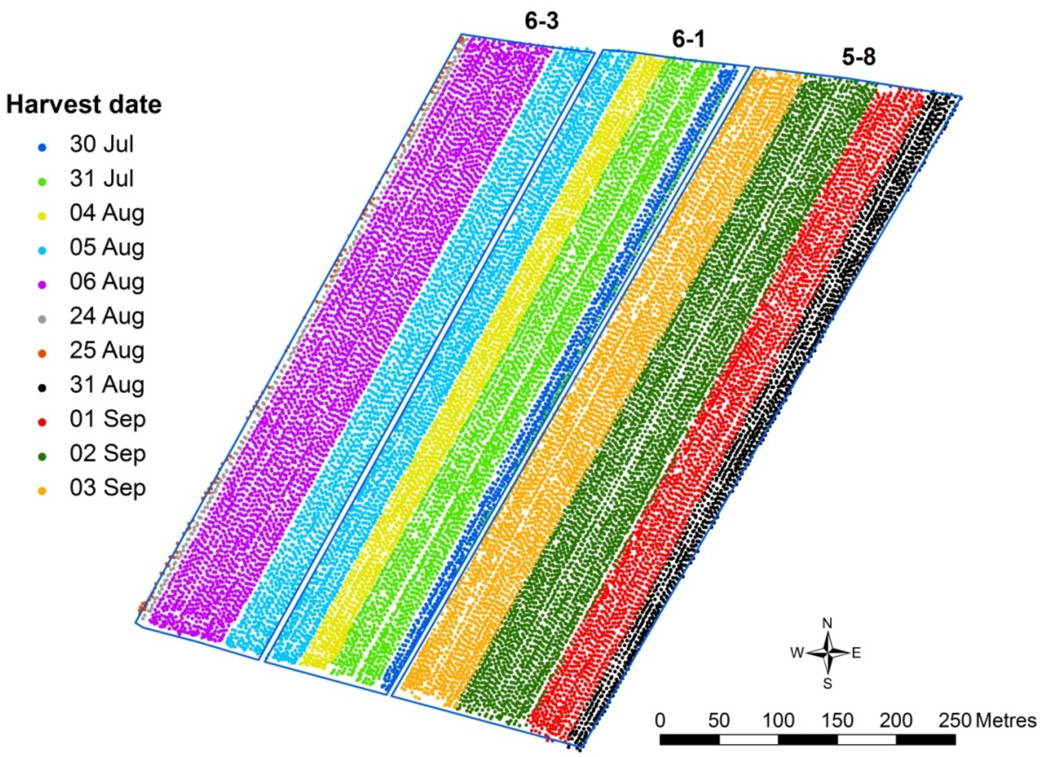

Figure 2. Date of harvest over the 3 study blocks; points for which yield was recorded as being outside the range 1-199 t/ha (Figure 1) have been deleted, as have data records that were closer together than 1 m.

One of the areas of concern with data obtained from the TechAgro yield monitor, as identified in CSE022 hitherto, has been the use of an arbitrary upper tolerance for yield of 200 t/ha. That is, any value for yield calculated from the roller opening that is > 200 t/ha is assigned a value of 200 t/ha. On the face of it, this seems reasonable. However, it is problematic when it comes to the question of calibrating the roller opening against Mill tonnage because questions arise as to whether these data should be regarded as aberrant and excluded before conversion of the roller opening data to cane tonnages, or simply regarded as legitimate high yield values and included in the calibration against Mill tonnage.

Figure 1 shows that with the exception of parts of the westernmost rows of each sub-block, the locations of yields > 200 t /ha were generally not co-located in easily identifiable areas. We therefore considered it reasonable to regard these data as aberrant and so excluded them from further analysis. Also excluded were a small number of 'zero' yields which were confined to the northern end of the easternmost row in sub-block 6-1. The result was a dataset corresponding to yield monitor data collected over several days, with 0.1 < yield > 199.9; indeed, each sub-block contained data collected over a (different) four day period (Figure 2).

One option for mapping these data is to display them as a colour-coded yield trace on the assumption that no correction for between-day differences in calibration is required (Figure 3). However, good PA practice requires the data to be interpolated to a smooth surface. In this study, map interpolation followed the sugarcane yield mapping protocol (Bramley – in preparation) which depends on the use of local block kriging of yield monitor data onto a grid of 2 m pixels (ie 4 m^2) using exponential variograms, a block size of 10 m (ie 100 m^2) and data cloud of 100 points, to interpolate robust sugarcane yield maps.

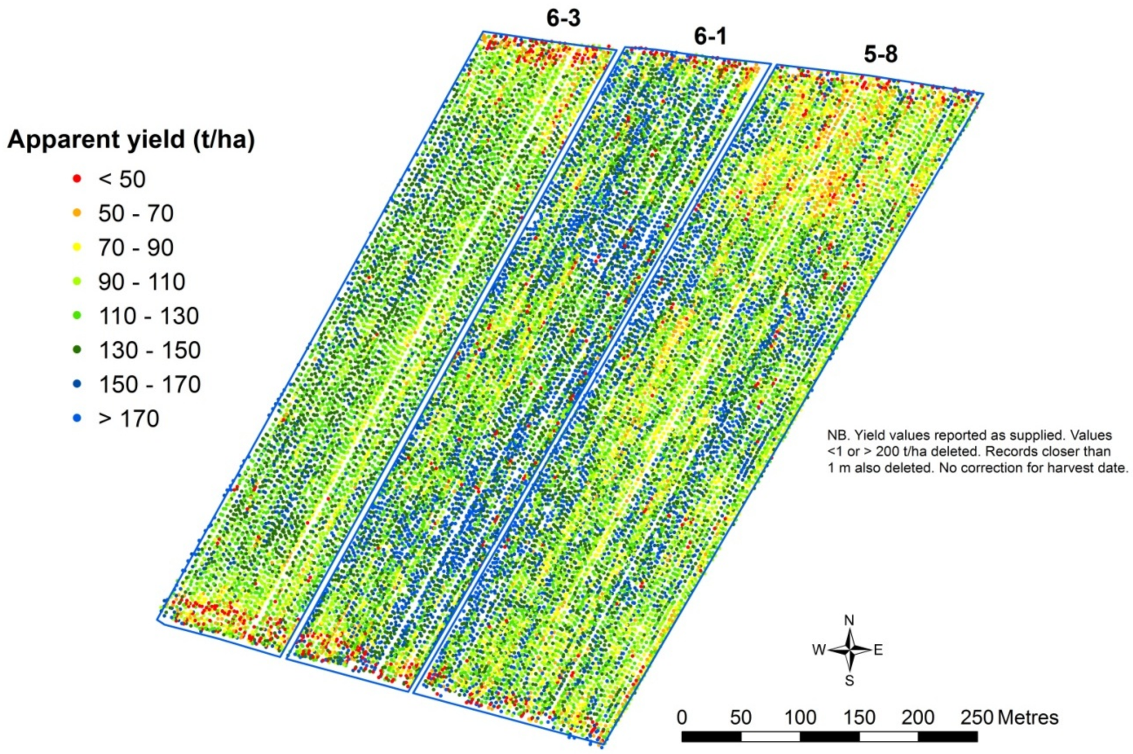

Figure 3. Yield monitor trace, colour coded for apparent yield, after removal of data records with 0 < yield > 200 t/ha.

Figure 3 takes no account of any between-sub-block differences in yield potential which may be caused by varietal or crop age differences, nor does it include any adjustment for between-day differences in either yield monitor performance or its calibration against the Mill tonnage, and nor does it consider the possibility that further data cleaning and trimming of additional aberrant or outlier values might be necessary. The primary purpose of the study was to explore these issues.

Mill data

Table 1 summarises the Mill data corresponding to the harvest events depicted in Figure 2. The first 6 columns were provided by HCPSL and the last 4 are calculated from the others. Of note is the remarkable stability of mean yield (t/ha) in each sub-block on different days in spite of the wide variation in the total tonnes and area harvested on each day. This is consistent with the common observation in PA that the full range of variation in a whole paddock may be observed in a single row. However the small range of variation in the daily means is unusual. Also of note is an apparent between-block difference in yield of ~5 t/ha which presumably reflects differences in yield potential due to age and/or variety.

Table 1. Summary of Mill data corresponding to the harvest of the study site.

LINKCODE	DATE_CUT	HA HARV	NETT WT	CCS	RAKEID	TICKET NO	t/day	ha/day	t/ha	Mean t/ha
5226A-6-1	30/07/2010	0.80	82.23	13.20	122347301	21	82.23	0.80	102.79	102.34
5226A-6-1	31/07/2010	0.98	100.11	12.90	122351501	24	296.15	2.90	102.12	
5226A-6-1	31/07/2010	0.92	93.63	12.90	122359901	25				
5226A-6-1	31/07/2010	1.00	102.41	12.50	122360001	25				
5226A-6-1	4/08/2010	1.43	146.10	12.80	123349401	26	146.10	1.43	102.17	
5226A-6-1	5/08/2010	1.56	159.58	12.90	123350401	27	159.58	1.56	102.29	
5226A-6-3	5/08/2010	0.68	66.29	15.05	123350301	28	155.58	1.59	97.85	97.15
5226A-6-3	5/08/2010	0.91	89.29	15.10	123357801	29				
5226A-6-3	6/08/2010	0.90	88.38	15.10	123370201	30	367.43	3.75	97.98	
5226A-6-3	6/08/2010	0.98	96.04	15.30	123370301	30				
5226A-6-3	6/08/2010	0.88	86.11	15.00	123377301	31				
5226A-6-3	6/08/2010	0.99	96.90	15.15	123377401	31				
5226A-6-3	24/08/2010	0.23	22.01	15.80	126330801	32	22.01	0.23	95.70	
5226A-6-3	25/08/2010	0.47	45.63	15.10	126342301	33	45.63	0.47	97.09	
5226A-5-8	31/08/2010	0.61	56.94	14.75	127310401	42	128.41	1.37	93.73	93.62
5226A-5-8	31/08/2010	0.76	71.47	15.00	127316701	43				
5226A-5-8	1/09/2010	1.20	112.56	14.90	127317901	44	219.89	2.35	93.57	
5226A-5-8	1/09/2010	1.15	107.33	14.40	127320901	46				
5226A-5-8	2/09/2010	1.62	151.25	15.40	127327301	47	327.71	3.51	93.36	
5226A-5-8	2/09/2010	0.89	83.19	15.70	127336401	48				
5226A-5-8	2/09/2010	1.00	93.27	15.60	127336501	48				
5226A-5-8	3/09/2010	1.46	136.21	15.10	127342901	49	291.73	3.11	93.80	
5226A-5-8	3/09/2010	1.44	134.43	15.20	127344001	50				
5226A-5-8	3/09/2010	0.21	21.09	15.20	127349401	51				

In light of the apparent data stability in Table 1, and the suggestion of spatial structure in the data seen in Figure 3, the yield data (Figures 2 and 3) were mapped on a per-block basis with no account taken for between-day or between block differences, with no additional data cleaning beyond that used to create Figure 3 and with no adjustment to Mill tonnage. The result is shown in Figure 4A.

Whilst these data were mapped on a per-block basis (ie separate mapping runs for each block), there is some evidence of 'striping' with the maps aligned with the row orientation, between block discontinuities and of edge effects at the ends of rows; note that given that land is continuous, one might expect greater pattern continuity across the boundaries of adjacent blocks.

By convention, given the wealth of data provided by a yield monitor, and as a means of cleaning out all possible outlier data, yield monitor data are typically trimmed such that all data used for mapping lie within +/- 3 standard deviations of the mean (Bramley and Williams, 2001; Taylor et al., 2007). Typically, this results in ~2% of the data being discarded. Thus, the data were trimmed on a per block basis (no adjustment between days) and remapped. The result of this was to reduce the apparent edge effect at row ends (Figure 4B).

The process of data trimming based on +/- 3 standard deviations from the mean relies on calculation of the mean yield recorded by the yield monitor. Table 2 lists these mean values whether calculated on a per block or harvest event (ie daily) basis. As can be seen, in contrast to the Mill data, the yield monitor data suggests considerable variation between harvest events. The data were therefore

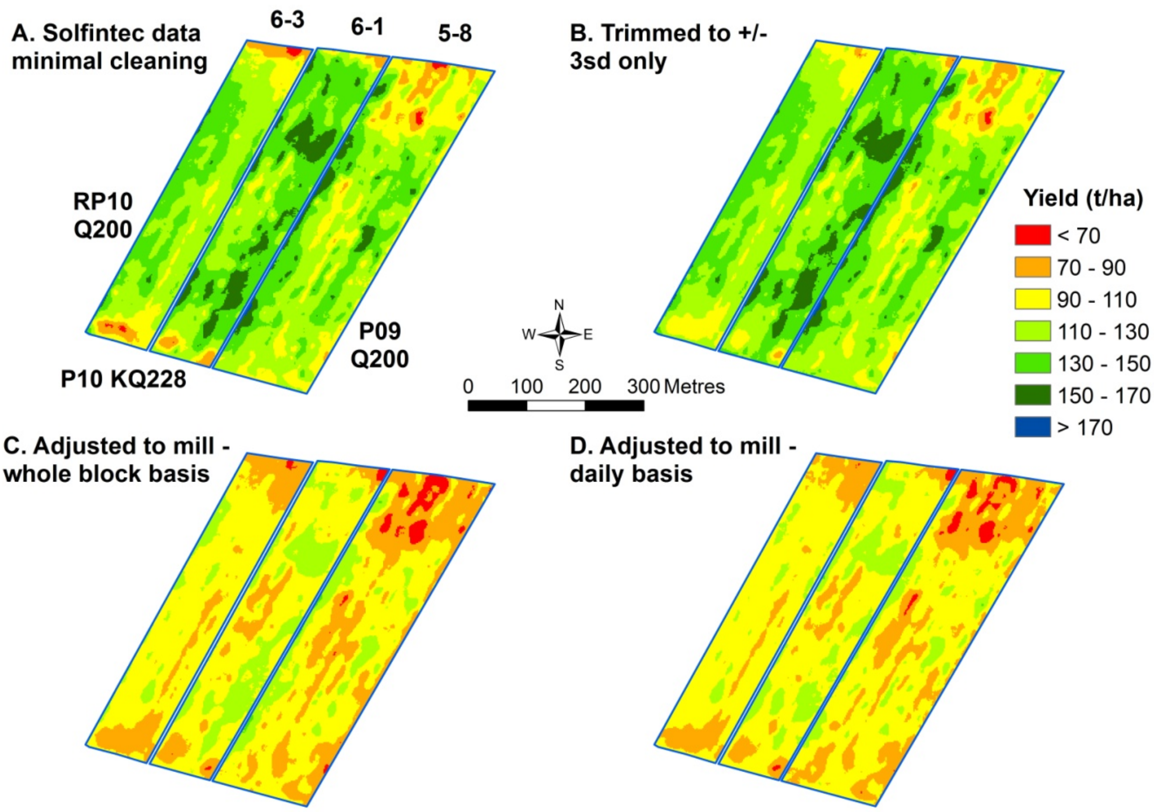

Figure 4. Yield maps derived from the Tabone yield monitor with different data pre-treatment prior to mapping – see text for further explanation. Note that these maps have been produced on a per block basis and not for the study area as a single whole.

Table 2. Daily summary of Mill and Yield monitor data and the effect of data trimming to within +/- 3 standard deviations (σ) of the mean (μ). Note that CV and n denote coefficient of variation and number of data records.

Block	Date	Mill μ	Yield monitor – pre trim				Yield monitor – clean			
			μ	σ	CV %	n	μ	σ	CV %	n
6-1	All	102.3	133.5	34.3	26	6897	136.1	30.4	22	6729
	July 30	102.8	130.3	37.1	28	916	130.3	37.1	28	916
	July 31	102.1	139.0	32.6	23	2690	142.2	27.4	19	2608
	Aug 04	102.2	130.9	33.3	25	1659	132.8	30.4	23	1629
	Aug 05	102.3	128.9	35.2	27	1632	131.3	31.8	24	1595
6-3	All	97.2	119.8	29.5	25	5674	122.8	24.5	20	5469
	Aug 05	97.9	118.5	30.9	26	1648	118.0	25.7	22	1597
	Aug 06	98.0	120.5	28.5	24	3734	124.1	22.8	18	3586
	Aug 24	95.7	123.9	25.5	21	204	128.7	16.3	13	191
	Aug 25	97.1	165.0	36.1	22	88				
5-8	All	93.6	119.3	34.1	29	10584	120.2	32.6	27	10488
	Aug 31	93.7	113.5	31.6	28	1437	114.9	29.5	26	1417
	Sep 01	93.6	123.2	34.0	28	2355	124.6	31.7	25	2323
	Sep 02	93.4	113.8	31.8	28	3880	114.7	30.6	27	3848
	Sep 03	93.8	126.3	36.4	29	2912	127.1	35.0	28	2890

adjusted to the mill tonnage by multiplying each record by a correction factor calculated as the mean Mill tonnage (t/ha) divided by mean yield monitor value (t/ha). Because the mean yields recorded by the yield monitor are greater than those measured at the Mill (Table 2), the effect of adjusting the yield monitor data to Mill tonnages is to greatly reduce apparent yield (Figure 4), whether the adjustment is done on a sub-block (Figure 4C) or harvest event (Figure 4D) basis. However, comparison of Figures 4C and D indicates that whether the adjustment of yield monitor data is done on a daily or whole block basis has no bearing on the spatial structure of the data – this, presumably due to the remarkable stability of mean yield between different sections of the same sub-block (see above). Figures 5A and B show the same maps with the legend adjusted to a more appropriate range.

Producing yield maps using normalised yield data is often useful. In the case of data obtained in the same year, it enables the effects of between block differences due to different varieties, crop ages, harvest dates or even species to be accommodated. It can therefore assist in yield map interpretation, given that it is often easier to interpret the map from one paddock by having access to those for the surrounding paddocks (NB since land is continuous, if its variation is the primary driver of yield variation, then the patterns of yield variation should extend across block boundaries). In the case of data obtained for the same paddock in different seasons, normalisation can be used to remove the effects of between-year differences in yield potential due to differing seasonal conditions. In the case of the present study, data normalisation was thought likely to be useful in helping to accommodate both the effects of between-variety and crop differences in yield potential, and also the day-to-day variation in calibration between the yield monitor and mill data.

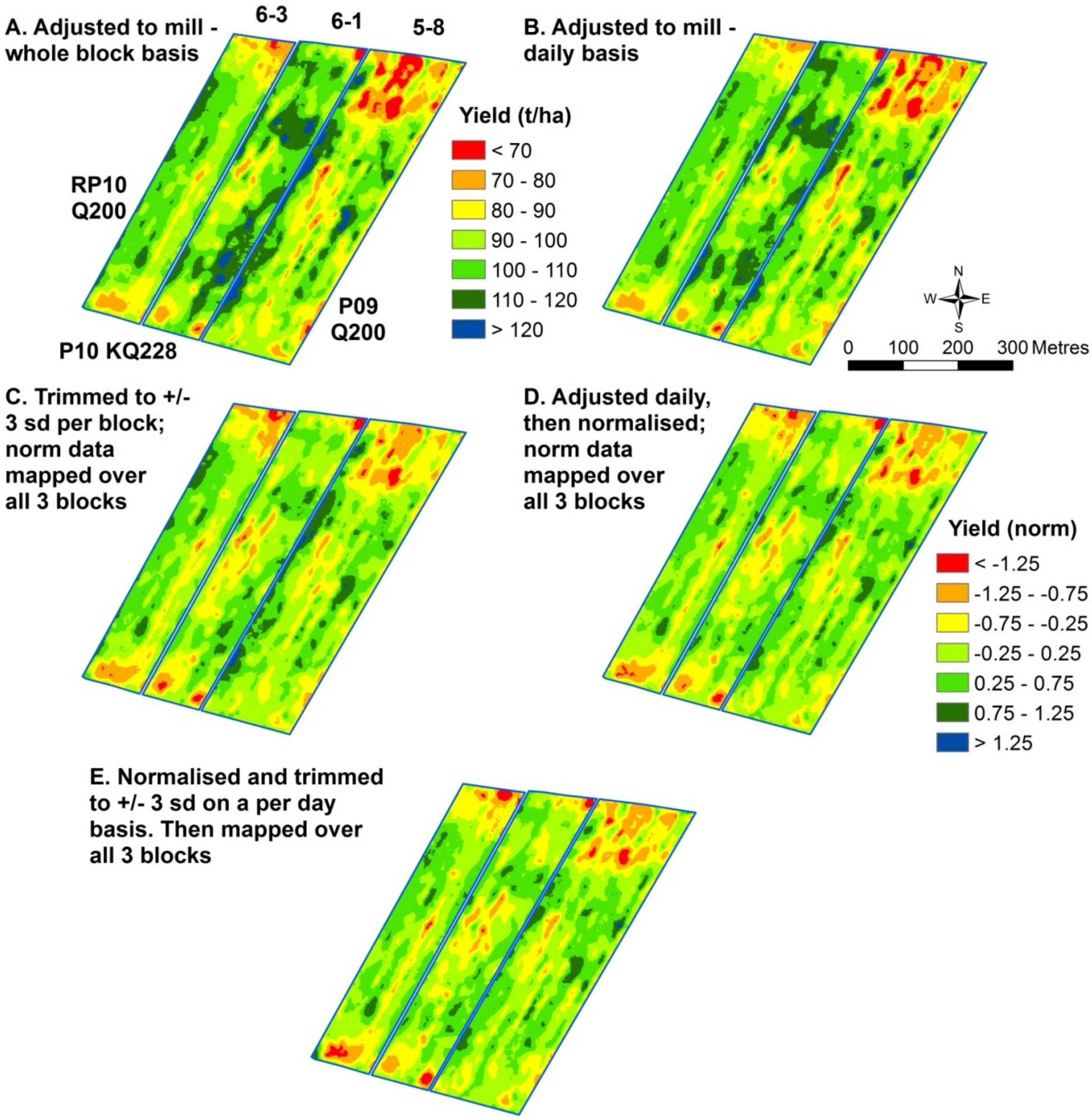

Figure 5. Yield maps derived from the Tabone yield monitor following different approaches to data pre-treatment and normalisation prior to mapping – see text for further explanation. Note that A and B are the same as Figures 4C and D; whereas these maps have units of t/ha and have been mapped on a per sub-block basis, C-E show normalised data. As a consequence, these have units of standard deviations, and have been mapped for the study area as a single whole.

Figures 5C-E show normalised maps which have been interpolated over all 3 sub-blocks together following different approaches to data trimming, adjustment to the Mill and/or normalisation. Thus Figure 5C derives from exactly the same dataset as Figure 4B, but with data normalised on a per block basis but then kriged to a single grid covering all 3 blocks; Figure 4B derived from yield data (t/ha) kriged to separate grids for each block. As indicated, one reason for using this approach is to explore the continuity of patterns of variation across block boundaries. It is therefore of interest that comparison of Figures 5C and 4B suggests little improvement in delineation of such patterns. In

Figure 5D, the data for each sub-block have been adjusted to the Mill on a daily basis, normalised and trimmed on a daily basis for each sub-block, before the normalised data have been mapped over the entire study area. In Figure 5E, the Mill adjustment step has been omitted. As can be seen, the spatial structure (ie patterns) in Figures 5C-E are extremely similar. Note that in all cases, there is little difference in across-boundary pattern continuity compared to the maps shown in Figure 4.

Finally, Figures 6A and B again show the same maps as Figures 5A and B along with a third variant (Figure 6C) aimed at maximising the opportunities for smoothing out between-block and between-harvest event differences. Thus, in Figure 6C, the data for each sub-block have been normalised and trimmed on a daily basis. Yield (t/ha) has then been calculated from these normalised data using the mean yield derived for the block from the mill tonnage and block area and a calculated mean value for CV (from which σ can be derived knowing μ). The latter is a weighted mean value for CV with the weights derived from the number of data values obtained in each day of harvesting. Since CV = $100\sigma/\mu$, this approach forces the data to conform to a single common distribution for each day/sub-block. As Figure 6C shows, using this approach nevertheless results in marked discontinuities remaining at the sub-block boundaries. This is unusual by comparison with yield maps for crops such as cereals and winegrapes.

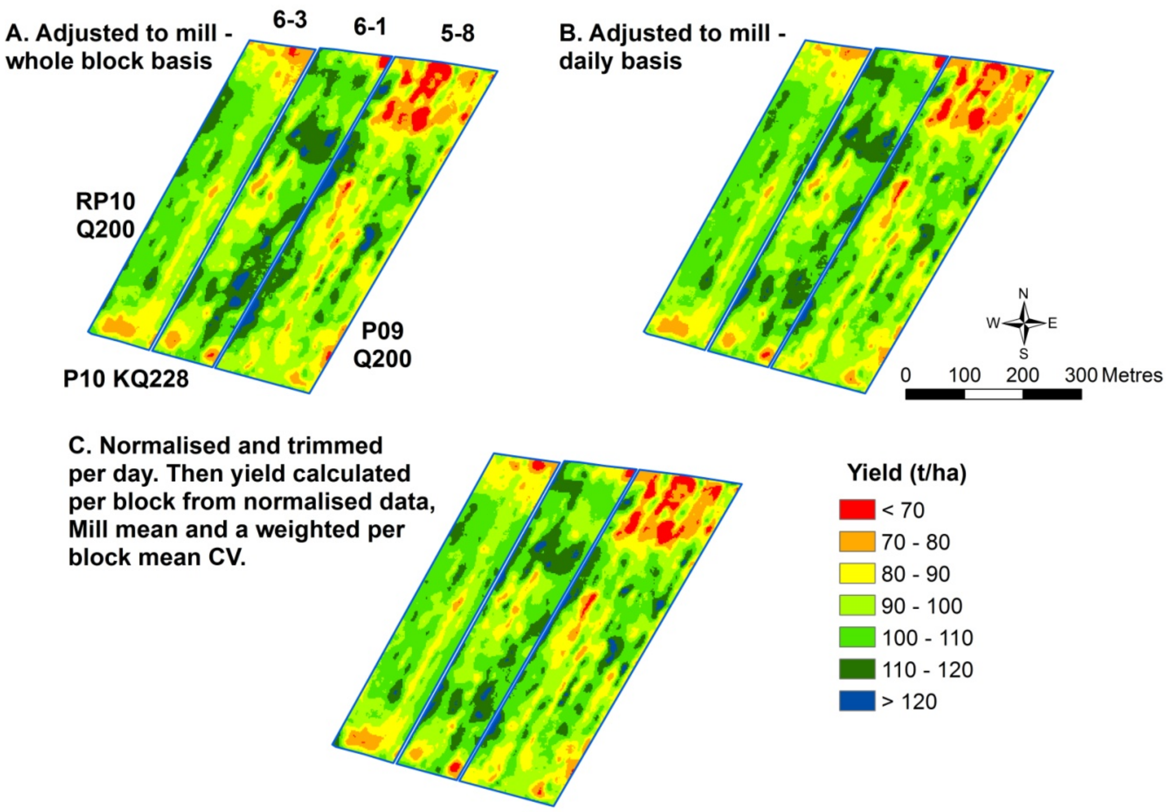

Figure 6. Yield maps derived from the Tabone yield monitor following different approaches to data pre-treatment and normalisation prior to mapping – see text for further explanation. Note C is arguably the smoothest possible variant of the maps (Figures 4-6) given that the data were logged at 3 second intervals and have been mapped to a 2 m grid.

Discussion

The obvious weakness in the above analysis is that it has relied on TechAgro yield data (t/ha) that have already been through the proprietary algorithm that converts roller opening to yield. Thus, each variant of the cleaning and trimming process has used the same starting point of the TechAgro data, as supplied, with only values outside the range of 0.1-199.9 t/ha omitted. Much better may have been to have started with the raw sensor data ? This is especially the case given that 200 t/ha is a somewhat arbitrary cut-off – why would it not be 199, or 196, or perhaps 170 (or whatever figure makes sense based on grower or research experience ?) ? It was noticeable when doing these analyses that when cleaning and trimming using the +/- 3 sd rule, it was almost never necessary to trim off high values; those removed were almost always low values, where 'low' could easily mean as much as 40 t/ha. In other words, based on the statistical approach, which forces the data to approximately conform to a normal distribution, it was the tail at the low end causing skewness, not at the high end. This probably suggests that the roller opening sensor is disposed to record high values rather than low – a suggestion which is consistent with the idea (Jensen et al. 2010; Troy Jensen – pers. comm.) that the TechAgro unit is susceptible to the effects of harvester speed, crop presentation (erect cane much better than a tangled mess) and whatever else might cause high pour rates. Since the algorithm which converts roller opening to yield is proprietary, we cannot readily investigate the merits of data cleaning and trimming prior to conversion to a t/ha basis without the assistance of Solinftec. Nevertheless, this is an issue which they may wish to explore themselves and I encourage them to do so.

The foregoing raises another issue which is a cause for concern. The data as supplied were dependent to some extent on the Mill data recorded for each block as these are required for conversion of the sensor value to yield (t/ha). The initial cleaning (ie removal of zeros and values > 200 t/ha) predominantly involved removal of high aberrant values which ought to reduce the mean for the dataset. Yet in every case, the mean yield monitor yield was considerably higher than the mean yield calculated from Mill data and the block area. This suggests to me that there is a problem with the initial conversion of sensor data to t/ha based on Mill data; one would have expected much closer agreement between the mean yields whether derived from the Mill or yield monitor. Overall, I would much prefer a methodology in which:

- first, the Mill tonnage and harvested area are used to identify the mean yield for a particular harvest event.
- second, this Mill derived value is matched to the mean sensor value (after trimming of aberrant sensor readings, not aberrant yields, with the aberrant sensor values appropriately identified, perhaps using the +/- 3 sd rule ?) and used as the basis for then converting the other sensor values to a t/ha basis.

The above comments notwithstanding, we should not assume that all the error in this process is attributable to the TechAgro yield monitor. Frankly, I am rather suspicious of the mill data since the finding that the estimates of average yield for each harvest event within the same block are so similar seems too good to be true; I would have expected a much greater range of values. Thus, Table 1 suggests that the mean yield estimated from different harvest events within the same block only varies by about +/- 0.5 t/ha (ie roughly 0.5 %); I would have expected this to vary by at least 5%.

In spite of all of the above, I think that the analysis presented here gives us reason to have greater confidence in the TechAgro data than we may have had previously. The fact that the spatial structure in the various maps is so similar is encouraging and it enables us to ask some potentially important questions:

First, if we think we can be reasonably confident that these maps are telling us something useful, what does that say about the striping in the maps ? What management practices are causing these (bearing in mind that straight lines are highly unusual in most landscapes – that is, the striping is not reflecting variation in some key underlying attribute of the block which is causing such yield variation) ?

Second, why is it that even when normalised data are mapped (eg Figure 5E), marked discontinuities in the patterns of variation across block boundaries are seen ? What might be the aspect of management that causes this ?

Finally, and most importantly, what do Brian and Paul Tabone think of these yield maps ? They know the land better than anyone, so should have a feel for whether the maps make sense.

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.

Jensen, T., Baillie, C., Bramley, R., DiBella, L., Whiteing, C. and Davis, R. 2010. Assessment of sugarcane yield monitoring technology for precision agriculture. Proceedings of the Australian Society of Sugar Cane Technologists, 32nd Conference, Bundaberg. 410-423.

Taylor, J.A., McBratney A.B., Whelan B.M. 2007. Establishing management classes for broadacre agricultural production. *Agronomy Journal* **99**, 1366-1376.

Appendix 14b: Yield mapping and forensic science – the case of the Tabone farm 2010-2011 (May 2012)

Rob Bramley
CSIRO Ecosystem Sciences, Waite Campus, PMB 2, Glen Osmond

The analysis of yield mapping at the Tabone farm in 2010 was described in a previous document. That work highlighted concerns as to the sugarcane yield mapping process using data derived from the TechAgro (now Solinftec) yield monitor relating to:

- Uncertainty as to the reliability of Mill data for the purposes of yield monitor calibration. In particular, the lack of day-to-day variation in apparent mean yields did not seem believable;
- The application of an arbitrary maximum yield of 200 t/ha to yield monitor data (now removed);
- Incorrect specification of the row spacing; and
- A consequent lack of confidence in yield monitor calibration.

The present document continues this analysis using 2011 data for the same study site (5226), albeit with block 5-5 added to the 3 blocks used in 2010 – 5-8, 6-1 and 6-3.

Mill data

The Mill data relevant to the study site are presented in Table 1. As was the case in 2010, a striking feature of these data is the lack of variation in the daily mean yield. Whilst it is common in other crops, and also consistent with our observations at other sugar sites, for the range of yield variation in a single row to approximate that for the entire field, it is simply not believable that the range of variation in daily mean harvested yield should be so small (<1 t/ha – equivalent to < 1% in all cases). Also of note are the

Table 1. Summary of harvest data for the Tabone site sourced from Macknade Mill (via Mike Sefton, HCPSL)

LinkCode	Date_Cut	Ha_Harv	Nett_Wt	CCS	RakeId	Ticket_No	t/day	ha/day	t/ha	Mean t/ha	Ha_Pdock	V_Class	Total area	My area	Mean yld my area
5226A-6-1	28/07/2011	0.83	60.63	12.8	122361001	1	60.63	0.83	73.05	73.20	6.81	KQ228-1R	6.81	6.76	73.75
5226A-6-1	29/07/2011	1.51	110.75	12.8	122362701	2	221.75	3.03	73.18		6.81	KQ228-1R			
5226A-6-1	29/07/2011	1.52	111	12.1	122365001	3					6.81	KQ228-1R			
5226A-6-1	30/07/2011	1.59	116.4	12.1	122378101	4	116.4	1.59	73.21		6.81	KQ228-1R			
5226A-6-1	31/07/2011	1.36	99.79	12.25	123300101	6	99.79	1.36	73.38		6.81	KQ228-1R			
5226A-5-5	18/09/2011	0.54	43.8	15.6	130304901	35	43.8	0.54	81.11	81.25	3.08	Q200-1R	3.08	3.54	78.22
5226A-5-5	19/09/2011	0.58	47.16	14.45	130320501	37	206.71	2.54	81.38		3.08	Q200-1R			
5226A-5-5	19/09/2011	0.55	44.69	15.2	130320601	36					3.08	Q200-1R			
5226A-5-5	19/09/2011	1.41	114.86	15.45	130315101	38					3.08	Q200-1R			
5226A-5-8	20/09/2011	0.41	26.4	14.8	130322801	41	203.79	3.15	64.70	64.76	10.46	Q200-2R	10.46	10.26	66.03
5226A-5-8	20/09/2011	0.54	35.24	14.9	130322901	40					10.46	Q200-2R			
5226A-5-8	20/09/2011	1.06	68.4	15.25	130323001	39					10.46	Q200-2R			
5226A-5-8	20/09/2011	1.14	73.75	14.8	130326101	42					10.46	Q200-2R			
5226A-5-8	21/09/2011	0.92	59.61	14.9	130340701	45	161.54	2.5	64.62		10.46	Q200-2R			
5226A-5-8	21/09/2011	0.91	58.73	15.2	130340801	44					10.46	Q200-2R			
5226A-5-8	21/09/2011	0.67	43.2	14.9	130340901	43					10.46	Q200-2R			
5226A-5-8	22/09/2011	0.98	63.48	14.9	130349401	47	139.68	2.16	64.67		10.46	Q200-2R			
5226A-5-8	22/09/2011	0.71	45.94	14.9	130349501	46					10.46	Q200-2R			
5226A-5-8	22/09/2011	0.47	30.26	14.8	130344501	48					10.46	Q200-2R			
5226A-5-8	23/09/2011	0.78	50.67	15.2	130350701	51	172.43	2.65	65.07		10.46	Q200-2R			
5226A-5-8	23/09/2011	0.76	49.33	14.9	130350801	50					10.46	Q200-2R			
5226A-5-8	23/09/2011	0.84	54.12	15.3	130351901	49					10.46	Q200-2R			
5226A-5-8	23/09/2011	0.27	18.31	15.3	130354801	52					10.46	Q200-2R			
5226A-6-3	15/10/2011	2.29	138.84	13.4	133378201	65	138.84	2.29	60.63	60.55	6.25	Q200-1R	6.25	6.18	61.24
5226A-6-3	16/10/2011	0.31	18.9	14.2	134301001	71	152.52	2.52	60.52		6.25	Q200-1R			
5226A-6-3	16/10/2011	2.21	133.62	14.2	134303501	72					6.25	Q200-1R			
5226A-6-3	17/10/2011	1.44	87.1	13.8	134313201	73	87.1	1.44	60.49		6.25	Q200-1R			

small differences between the total Mill estimated harvested area, and the block areas derived from accurate field survey. Admittedly, these are small and within the bounds of measurement error, although this is not the case for block 5-5.

Yield monitor data

As in 2010, yield monitor data were supplied already processed by Solinftec. The specifics of data pre-treatment are not known, but it was apparent that the data ‘as-supplied’ had been trimmed to within +/- 3 standard deviations of the mean on a per block basis. The data were however, supplied per block and not per harvest event. Figure 1 shows a trace of harvesting activity at the study site displayed on a daily basis as determined by the date stamp in the yield monitor data files (Note that in these files, date is incorrectly labelled as time and vice-versa). Figure 2 then shows the yield map obtained from these data. This map has two striking features. First, the mean yield (pale green colour in the legend) is clearly considerably higher than is inferred from the Mill data (Table 1). Second, the map shows a marked ‘striping’ and in general, the patterns of variation in this map are aligned with the row orientation.

In regard to the elevated mean yield seen in Figure 2, it quickly became apparent that, as was the case in 2010, the yield monitor data had been processed on the basis of an incorrect specification of the row spacing; the row spacing on the Tabone farm is 2.0 m and <u>not</u> 1.5 m as was specified in the 2010 yield monitor data. Applying the appropriate correction resulted in a map that better matched the block means (Figure 3). Nonetheless, the presence of this error in the dataset raises concerns as to the process followed for calibration of yield monitor data to the Mill data. Our understanding amongst the CSE022 project team is that the Mill data are critical to the conversion of yield monitor sensor data to a t/ha basis and the fact that this calibration is apparently insensitive to the correct specification of the row spacing is of some concern and raises questions as to the confidence that can be placed in the yield monitor data.

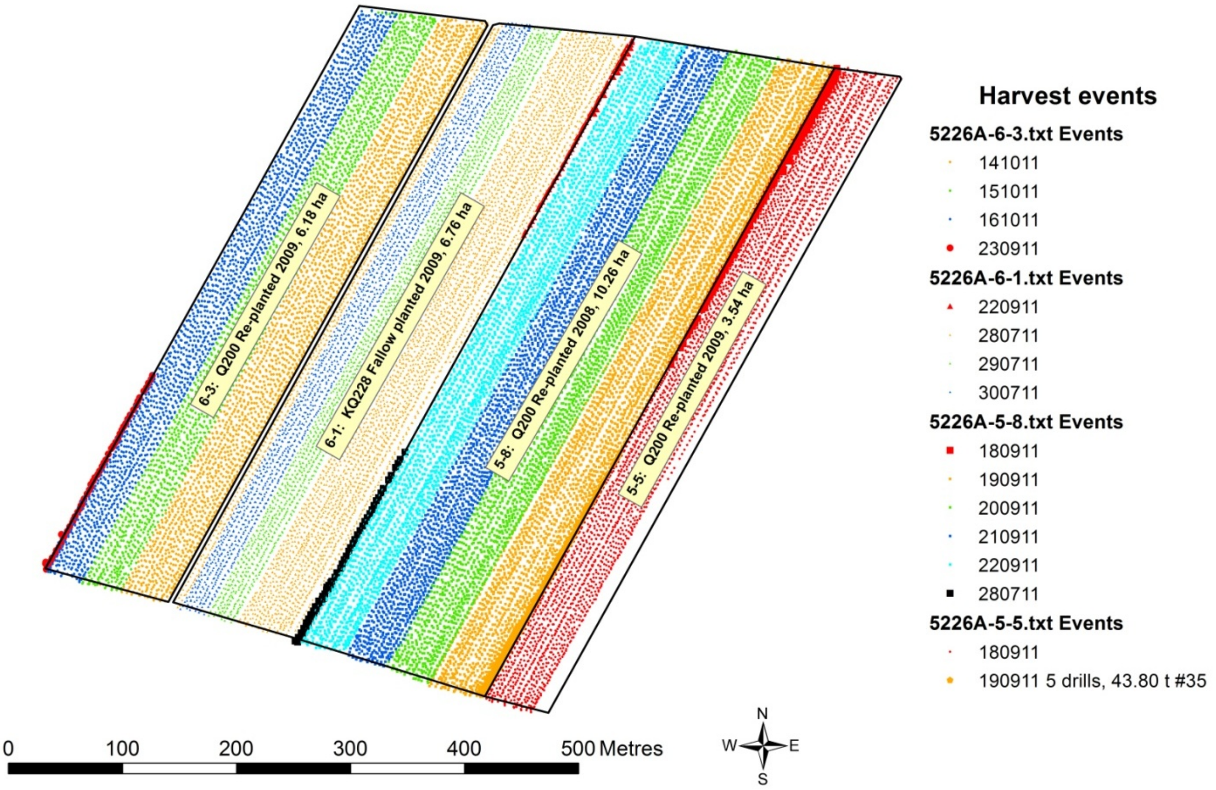

Figure 1. Yield monitor trace for the Tabone study site, 2011.

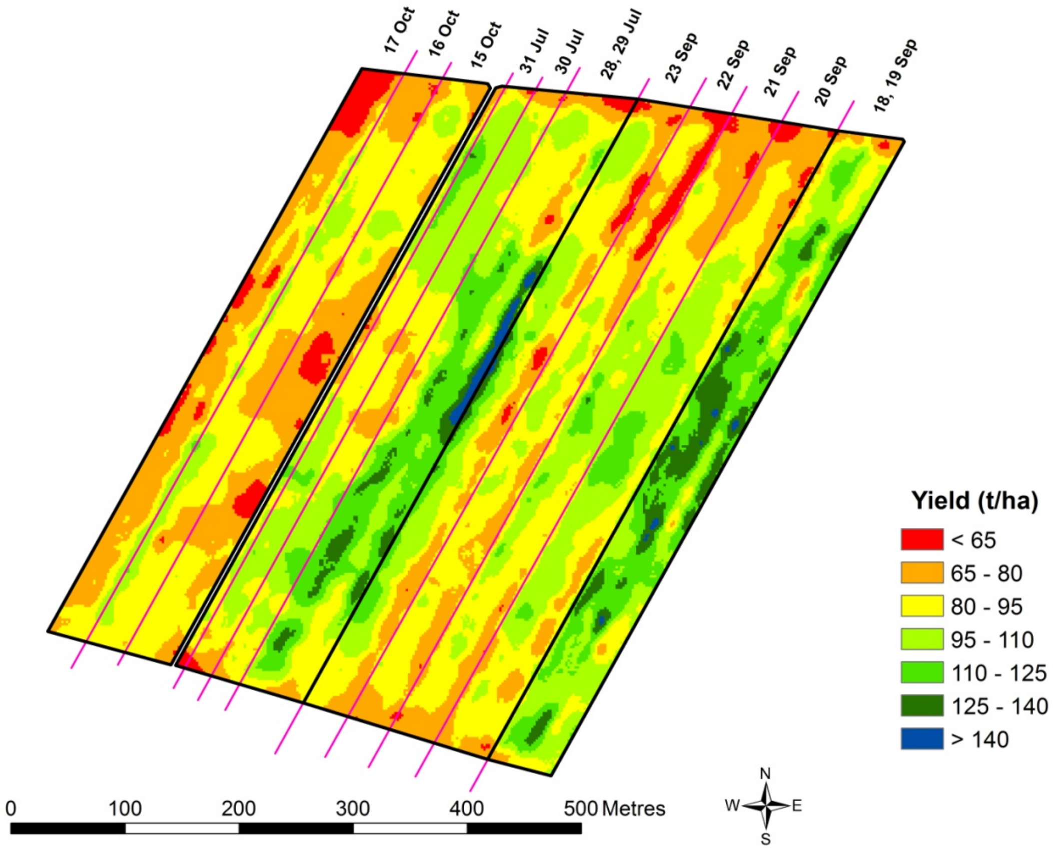

Figure 2. Yield map derived from yield monitor data, as delivered. Also shown are delineations between the harvest dates identified in Figure 1.

As is apparent in Figure 2, some of the striping in the yield map appears to be related to block boundaries; the 5-5 to 5-8 boundary is particularly strong and may simply reflect a crop age effect which is also reflected in the Mill data shown in Table 1. However, a core component of the striping within blocks appears to be related to the different harvest events; Figure 2 shows the boundaries between these as identified by the data shown in Figure 1. Yet more striping appears to be unrelated to either block boundaries or harvest events. However, comparison of the harvest dates specified by the Mill (Table 1) with those listed in the yield monitor files suggested a significant discrepancy.

Date and time in the yield monitor files is reported on a UTC basis (ie Greenwich mean time), yet for obvious reasons, the Mill data are recorded on a local basis, that is, GMT + 10 hours. After appropriate reformatting in MS Excel, we applied the appropriate date and time correction to the yield monitor data so that these could be examined on a local Queensland time basis. The result is shown overlain on Figure 3 (the row-spacing adjusted yield map). Comparison of the harvest date delineation shown in Figures 1 and 2 with that shown in Figure 3 shows some discrepancy. Examination of the Mill data (Table 1) in light of Figure 3, suggests that some small harvest events have been mis-attributed on both a date and block basis. This may explain at least a part of the reason why the mean harvest event yields recalculated from Mill data are of concern.

Figure 3 also indicates a closer alignment of striping in the yield map with the adjusted harvest event delineation suggesting that a day-to-day variation in the yield monitor calibration is not being properly accounted for in the whole-of-block matching of yield monitor data to Mill data. This is especially apparent when Figures 2 and 3 are compared, with marked differences visible in block 6-3, for example, and with some of the striping in block 6-1 aligning with harvest events in Figure 3 that were not delineated in Figure 2. These differences raise a serious question about the appropriateness of applying the same calibration between Mill and yield monitor on a per block basis, rather than a per day basis as is

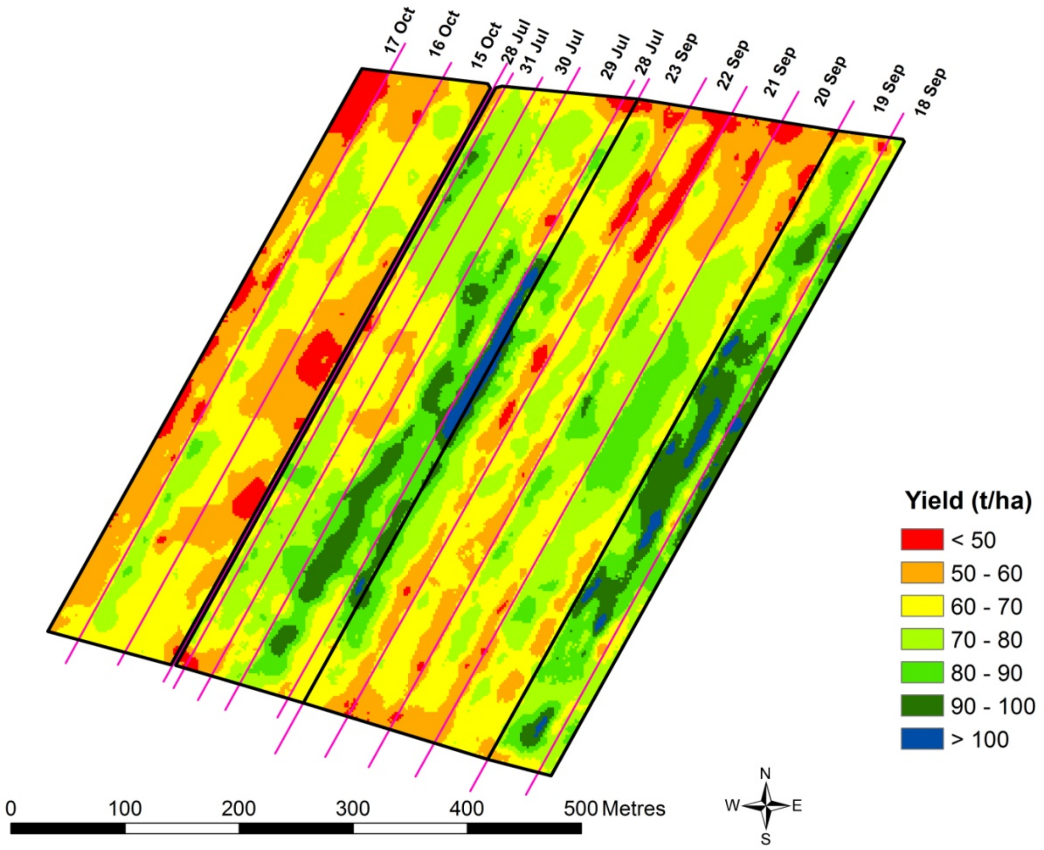

Figure 3. Yield map obtained when the yield monitor data are adjusted on the basis of the correct row spacing being 2 m and not 1.5 m. Also shown are the delineations between harvest events following correction of the date and time stamp in the yield monitor file to local Qld time.

typically done in grape yield monitoring, for example, and as was recommended following the 2010 analysis of data from this site.

In light of these results, a laborious process was followed in which the yield monitor data (corrected row spacing) were treated on a per-harvest event basis, as follows:

The data were first sorted on the basis of the date and time in Queensland. Knowing the row spacing and the coordinates of each data point, the distance between consecutive data points was calculated, as was the harvested area corresponding to each point and thus the actual tonnage of cane cut. Thus, the total tonnage of cane cut and area from whence it came could be calculated for each harvest event. Data values which were more than 20 m from the previous data record were deleted. 20 m in 3 seconds equates to 6.7 m/sec or 24 km/h. It is recognised that this is much faster than typical harvester speed, but it provided an initial useful filter for removing data points that were far apart and to which an abnormally large area was being ascribed. After removal of these aberrant values, the yield data (t/ha) were normalised (mean of zero, standard deviation of 1) and values outside of +/- 3 standard deviations of the mean were removed. From the remaining data, approximate totals for the tonnage and area harvested could be calculated on a daily basis along with a mean yield. The tonnage and area data were a useful check on the Mill data and highlighted the fact that there were 3 small harvest events on Sept 19 (block 5-8), Sept 23 (6-1) and Sept 24 (6-3) which were seemingly not reflected at all in the Mill data.

For all harvest events, there was a small discrepancy between the Mill average yield and that calculated from yield monitor data, even after applying the row spacing adjustment. Generally this was < 3 t/ha, although on a couple of occasions (16, 17 Oct) it was around 6 t/ha and on 18 September was approximately 10 t/ha. In light of uncertainty over the areas and total tonnages harvested on a harvest event basis (see above and also Table 1, Figures 1-3), it was not possible to accurately recalibrate the yield monitor data on a per harvest event basis. It is envisaged that an adjustment of the calibration between the sensor and Mill in terms of both slope and bias is required. However, using the mean block yields (Table 1), a bias correction, only, was made to each harvest event dataset; that is, for each harvest event, the yield monitor data were adjusted on the basis of the ratio of the mean block yield (Table 1) to the mean yield calculated from yield monitor data. The map derived from these data is shown in Figure 4. As can be seen, whilst there is some apparent smoothing out of the apparent yield variation by comparison with Figure 3, most notably in blocks 6-3 and 6-1, the map is still characterised by a striping which, to at least some extent, aligns with the delineation between harvest events.

Overall, the analysis is summarised by Figure 5. Even when the data are handled in such a way as to remove as many artefacts of the harvesting process as possible (Figure 5c and d), the maps retain a strong striping. This may be associated with some aspect of agronomic management (eg a group of rows which missed fertilizer or spray), but the evidence seems strong that a major part of this striping is associated with one or more of a combination of Mill consignment errors, some other mis-management of Mill data (the small range of variation in Table 1) or the failure to calibrate the yield monitor data on a per harvest event basis for slope in addition to bias.

Discussion

In general, it is unusual to take more than a couple of hours to construct a robust yield map for either grapes or broadacre crops, even when the harvest of a vineyard or paddock comprises several harvest events. The analysis reported here took over 3 days. Clearly, that sort of time is not going to be available in a commercial sugarcane scenario, and even if it is, I have no doubt that it will not be tolerated. Automation of the data management and manipulation is an obvious solution to this problem and is

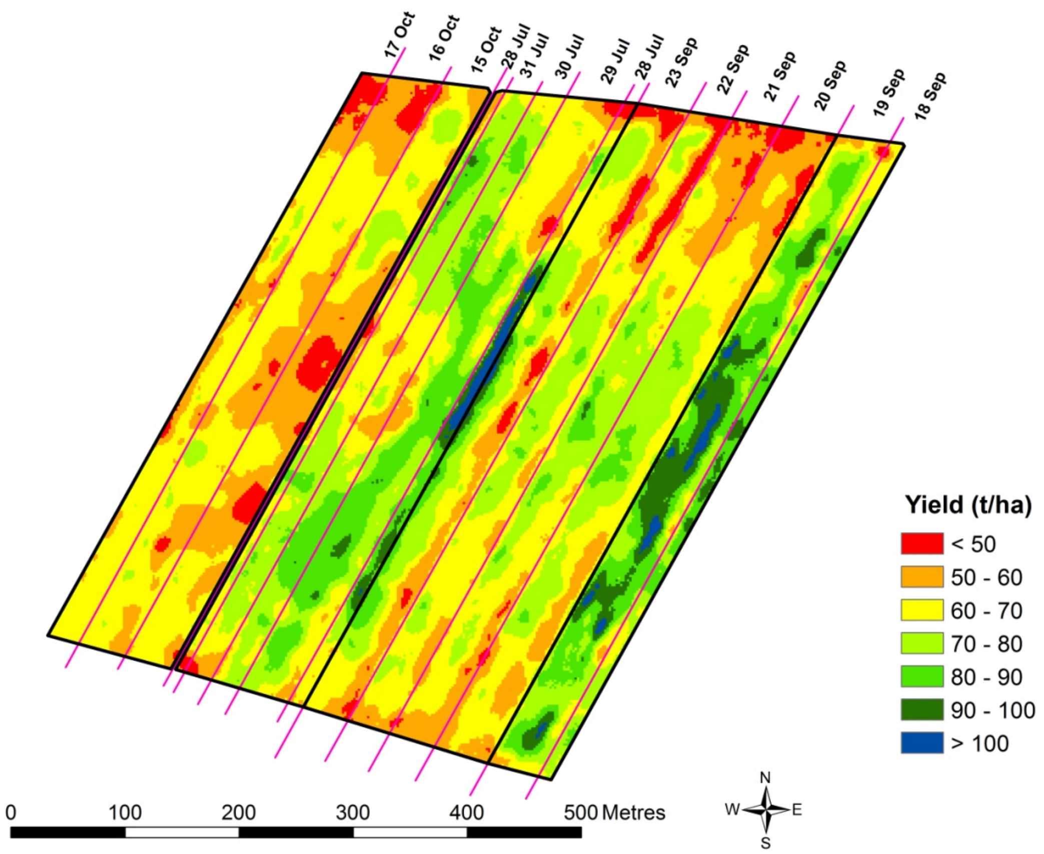

Figure 4. Yield map obtained following adjustment to correct row spacing and bias correction of yield monitor data to Mill average block yields on a per-harvest event basis.

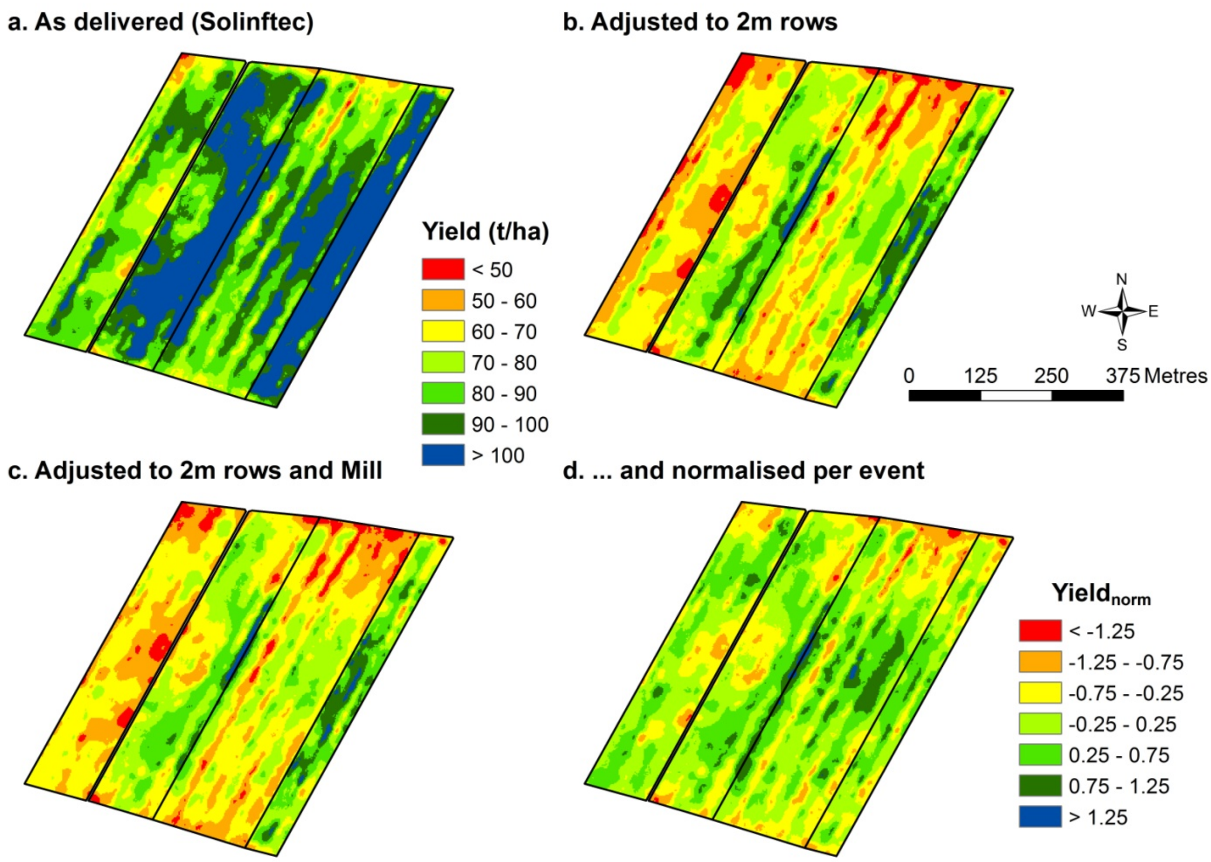

Figure 5. Yield maps produced following various approaches to yield monitor data treatment. Maps (c) and (d) reflect the maximum possible adjustment of yield monitor data to Mill records that is possible in the absence of access to Solinftec's proprietary algorithms.

clearly one which Solinftec have sensibly pursued. However, such automated processing very evidently becomes a hindrance rather than a help when it contains flaws, especially if these are not readily detected. Further to a similar conclusion which was drawn following the 2010 data analysis, I am now firmly of the view that the matching and calibration of yield monitor and Mill data needs a complete re-think if the yield monitoring of sugarcane is to be both cost effective and with minimal error. Similarly, I am of the view that given the apparently high likelihood of consignment and other errors in Mill data, sugarcane yield mapping is going to have an uncertain future unless it is accompanied by a more robust consignment process, most likely electronic consignment. One possible alternative, which nonetheless would still be best implemented with appropriate electronic matching of sequentially harvested bins to yield monitor data, would be to use portable weighbridges at cane railway sidings, such as is used for example, by Treasury Wine Estates at its Padthaway vineyard. As things stand at present in the Herbert, and in my opinion based on over 15 years experience with yield mapping, there is too much error associated with the present modus operandi for the resulting yield maps to be considered a reliable basis for decision making. The work carried out by Troy Jensen as part of CSE022 strongly suggests that sensing sugarcane yield is not difficult; an array of sensing options appear to do an acceptable job including that used by Solinftec. However, based on the Jensen work, the present analysis and also the corresponding analysis from 2010, it is clear that the limiting steps in the application of yield mapping to sugarcane are all in the domain of data management and handling; somewhat ironically, one would think that this ought to be the comparatively easy part !

The above notwithstanding, it was of interest to examine the similarity between the yield maps obtained for this site in 2010 and 2011. Figure 6 shows the results of clustering normalised yield maps for the two years using *k*-means; two and three cluster solutions are shown. Of note is the fact that for both maps, the 95% confidence interval, derived from the median kriging variance, is 0.61 standard deviations. Experience suggests that this is a large number and also indicates that, without careful examination of the

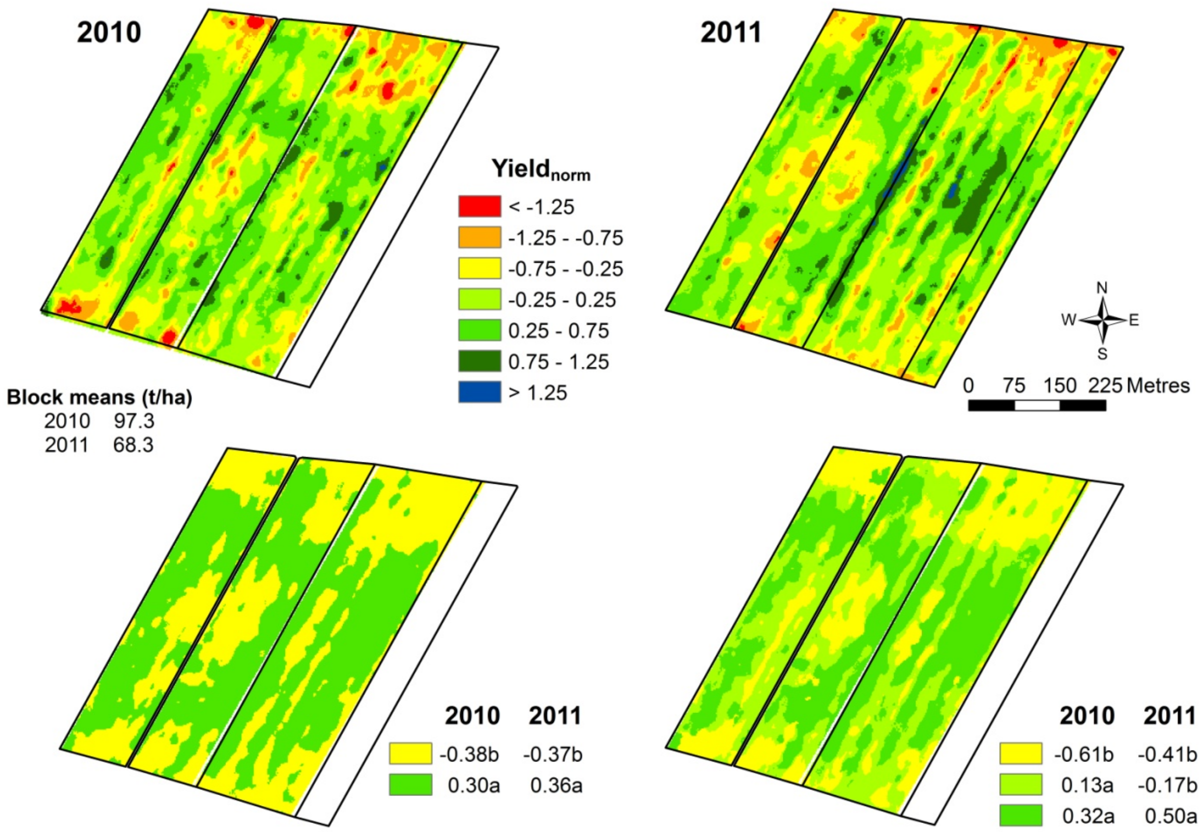

Figure 6. Cluster analysis of yield maps obtained in 2010 and 2011. Zone means followed by different letters in the same year are significantly different.

kriged data underlying the map, one cannot be confident that adjacent different legend categories in the base yield maps are indeed different. This high value of the confidence interval is almost certainly due, at least in part, to the uncertainty in the yield maps caused by the striping. Indeed, it is interesting that the lack of confidence expressed in the data for 2011 (above) and for 2010 in the corresponding report produced last year, is matched by the lack of statistical confidence in the maps identified by the median kriging variance. That said, and somewhat surprisingly, the two cluster solution identifies zones which are consistently significantly different across both years, even though there is a mean yield difference between the two years of approximately 30 t/ha. However, when the analysis is extended to 3 zones, some between zone differences are seen to be not significant. The striping in the yield maps is also seen to be impacting on the three zone analysis. The degree to which agronomic confidence can be attached to the two zone map shown in Figure 6 is something that Brian Tabone will need to ponder. For now, it is enough to point out that the maps in Figure 6 do not bear a strong resemblance to soil maps derived from either EM38 or gamma radiometrics. Whether properly calibrated maps would bear such a resemblance is not known.