Paper No. 991048

Post-Process Filtering Techniques to Improve Yield Map Accuracy

by

Andy D. Beck
Research Assistant

Jody P. Roades
Research Assistant

Stephen W. Searcy
Professor

Texas A&M University
College Station, TX

Written for Presentation at the
1999 ASAE/CSAE-SCGR Annual International Meeting
Sponsored by ASAE and CSAE

Sheraton Toronto
Toronto, Ontario, Canada
July 18-21, 1999

Summary

Yield maps often contain data points that are not correct estimates of the yield in a given area. Often these erroneous points result from the manner in which the harvester is operated. Incorrect yield estimates are a factor of the number of combines used in a field, the shape of the field, and the type of crops being harvested. Incorrect yield maps can lead to inappropriate site-specific management decisions. Therefore, the accuracy of the yield data collected with combine-mounted mapping systems needs to be improved. Post-processing the yield map data with a filter can remove some of the problem data and result in maps that are more accurate.

A filtering technique for exported yield files was developed and tested on ten fields of corn, sorghum and rice. The incorrect yield points were identified with filter functions based on acceptable yield values, acceptable moisture values, appropriate travel distance, sudden surges in yield and overlap of previously harvested ground. Results showed that this filtering algorithm resulted in a higher field average and lower standard deviation than either the unfiltered data or data filtered with maximum and minimum thresholds alone. The filter was successful in eliminating many incorrect yield estimates.

Keywords:

Yield map, filter, grain, precision agriculture

Post-Process Filtering Techniques to Improve Yield Map Accuracy

Andy D. Beck, Jody P. Roades, Stephen W. Searcy

Yield mapping has become the most widely adopted of the precision agriculture technologies. The information provided by yield mapping systems is of such value, that producers want the maps to be true representations of their fields. Unfortunately, errors can enter the data in many forms. The wet basis yield at each data point is calculated using the mass flow rate, time difference between readings (cycle), header swath width, and distance traveled (eq. 1). The mass-flow sensor and moisture sensor on the combine determines the mass flow of grain per second. The swath width is the length of the header unit that is set to harvest the grain. The distance traveled is measured between two GPS positions.

$$Yield_{wet\_basis} = \frac{Flow * Cycle}{Distance * SwathWidth} \tag{1}$$

Converting the wet basis yield to a dry basis with equation 2 allows an equal comparison of data for a specific crop. The crop moisture is measured with the moisture sensor as the grain is harvested, and the payable moisture is the standard payable moisture.

$$Yield_{dry\_basis} = Yield_{wet\_basis} * \frac{100.0 - CropMoisture}{100.0 - PayableMoisture} \tag{2}$$

Potential error sources include the mass flow, moisture content, indicated swath width, and the distance traveled between recorded points. The causes of these errors can include sensor error, machine adjustment and operator practices.

Much research effort has been devoted to determining the sources of error in the grain flow, moisture and position data streams. Higher accuracy systems and proper calibration techniques can improve the overall accuracy of yield maps. However, the most accurate sensors will still provide erroneous results if the operator of the harvester does not follow good practices. Any yield map contains errors in the yield estimates from a variety of sources. Blackmore and Marshall (1996) reported that there are six main groups of error that have been identified and ranked according to their impact on yield maps.

1. Unknown crop width entering the header during harvest
2. Time lag of grain through the threshing mechanism
3. The inherent ‘wandering’ error from the GPS
4. Surging grain through the combine transport system
5. Grain losses from the combine
6. Sensor accuracy and calibration

The effect of an unknown harvest width can manifest itself in several ways in a yield map. If there is a consistent difference between the actual and indicated cutting widths, the area harvested will be overestimated, and the resulting yield proportionately underestimated. This type of error could be nearly constant and affect the entire field in a similar manner. The actual cutting width can also be unknown in instances where the shape of the field results in points that must be harvested. In this case, only a portion of the data is recorded with an incorrect swath width. These situations can not be accurately mapped without an automated sensing of the actual cutting width. As such devices are not commercially available, the best solution for maintaining accurate yield data is to avoid recording data when the full harvest width is not used. This requires proper training of the operators. Unfortunately, the operators frequently record data in situations where the data is in error. Certain crops and conditions seem to create more erroneous data points. Small grains harvested in odd shaped areas typically create a much higher frequency of "bad" data. Rice is an example of a crop harvested in oddly shaped areas (between levees) that result in maps that appear to underestimate the true yield.

A related problem occurs when the operators do not lift the header to disable data recording during turns. This results in two problems, statistically and visually. Statistically, the points recorded during turns add to the total area harvested, but not to the harvested mass. Consequently, the calculated yield is low. Visual interpretations of the maps can result in errors because the maps are printed in a "first recorded, first displayed" order. This means that the second or third data points (the erroneous ones) will obscure the initial, "correct" data in the final map. Observation of yield data files from Texas producers shows this phenomenon to be commonly occurring. While this type of incorrect data can be avoided (Searcy, 1998), most combine operators do not try to disable recording when the mapping system assumptions are invalid.

Upper and lower threshold limits are commonly use to eliminate data points presumed to be erroneous. The upper limits can be used to identify those points that are abnormally high due to surging grain flow or sudden reductions in the travel speed. The lower limit can be selected to eliminate points that might result from partial header width harvesting or recording while turning, but it is often difficult to separate true low yield points from those that result from poor operating procedures. An alternative approach would be to utilize the spatial relationships of the recorded points to identify those points with a high probability of being incorrect.

The objective of this research was to develop a filtering technique that could be used to post-process yield data files and eliminate the majority of erroneous points. This filter should be capable of performing correctly in situations where the mapping harvester was operated in tandem with non-mapping machines.

Filtering Technique Development

To improve the accuracy of mapped yield estimates, errors must be eliminated, corrected or ignored. Calibration of sensors and operator training are attempts to prevent or minimize the occurrence of errors. Correcting yield data requires some means of identifying those points that contain error, and using some knowledge of common harvest conditions to determine how to correct the error. Rands (1995) developed an expert filter system to identify and attempt to correct errors when possible. Those data that could not be corrected were deleted. Han et al. (1997) developed a bitmap technique that attempted to correct the swath width for combines. The system required that the entire field is harvested with a mapping harvester, and that the positions be determined with a high accuracy. While this technique was successful, it is not applicable to most commercial situations. The DGPS receivers currently used by commercial yield mapping systems are not accurate enough to determine swath width. In commercial practice, several combines maybe used in the same field simultaneously, and often only a portion of the machines have mapping capability. Blackmore and Marshall (1996) suggested a technique that they called "potential mapping" to correct for swath width uncertainty. Unfortunately, this technique would not work well for fields with multiple harvesters if some were not mapping yield.

The need for an improved filtering technique became obvious when conducting research with cooperating producers. For these studies, the crops were harvested by the producers or their employees. The yield data files were provided to us for evaluation of treatments that had been applied to the fields. On many fields, incorrect data was obvious from the driving patterns of the operators. To ensure that the research results would be as accurate as possible, a new filtering technique was developed for cleaning up the yield data files.

The approach used for the filter was to identify and eliminate erroneous data. Yield maps represent very intensive descriptions of a field. In most cases, the loss of a few points will not affect the information content for a producer. The yield patterns will still appear with fewer points displayed. In many situations, the mapping combine worked in tandem with up to three non-mapping combines. The patterns followed by combines was uncertain, so few assumptions could be made about the yields in unmapped regions or the harvesting order of the combines. The filtering techniques described here are based solely on the data from the mapping combine.

The accuracy and calibration of the mass flow sensor and the pattern of movement of grain through the combine can result in unreasonable flow rates. If an unreasonable flow rate and/or moisture reading is measured, the data point is discarded. The filter was intended to remove five different types of error; unrealistic yield or moisture, inappropriate distances, yield surges, turnarounds and overlaps. The techniques used to identify each of these will be discussed individually.

<u>Unrealistic yield or moisture</u> Removing values that fall above or below reasonable values is a common filtering technique. Many software packages use upper and lower thresholds as the sole method of filtering data. Selection of the optimum threshold values may be difficult without knowledge of the field and its productivity for the year in question. Appropriate values for one year may be completely inappropriate for other years. For this reason, the upper and lower yield thresholds were made operator selectable. The default values were in standard English units (1 and 300 bu/ac respectively for low and high thresholds), and the user needed to determine the appropriate conversion for crops yields normally given in mass units. The minimum was slightly greater than zero on the assumption that any area of the field being harvested should have small amounts of grain entering the combine. The default values were used for all fields in the analysis reported here. The resulting yield thresholds for the different crops were the following.

- Rice – 0.05 Mg/ha (45 lbs/ac) and 15.0 Mg/ha (13 400 lbs/ac)
- Corn and sorghum– 0.06 Mg/ha (56 lbs/ac) and 20.0 Mg/ha (18 000 lbs/ac)

The moisture reading thresholds were also user selectable. The default values were 0 and 30%. The normal harvesting moisture for the grain is 14-17%, but the moisture content as high as 28% was experienced in some fields.

<u>Inappropriate distance</u> The distance traveled between two sequential data points is determined by the GPS receiver, a radar sensor or a wheel rotation sensor, depending on the manufacturer of the yield mapping system. Errors can result from apparent jumps in the GPS locations or actions by the combine operator. The ‘wandering’ error of the GPS may contribute to the inaccurate travel distance estimates for the machine on those systems that obtain distance from the GPS receiver. Excessive wheel slippage could also generate similar errors. Operator actions, such as stopping the harvester with the header down, can generate points when the machine is not moving. For this filter, data points were removed for zero distance traveled and speeds greater than 4.9 m/s (16 ft/s) between GPS positions. The maximum travel speed was selected as twice the normal travel velocity of a combine during harvest.

<u>Yield surges</u> Sudden yield changes can result from mechanical effects, but most often are only apparent surges. An apparent surge can result if the operator quickly reduces forward speed at a time when a maximum grain flow is occurring. This phenomenon is observed most frequently at turns in the field. Erroneous surges were removed with a statistical identifier based on moving average mean and standard deviation. The mean and standard deviation of a moving window of yield values is calculated and

compared to the current point. Yield values were considered valid when the difference between the mean and the value was less than a specified multiple of the standard deviation. The length of the statistical window and the standard deviation multiplier were user specified. The default values were 25 points and 3 standard deviations.

Turnarounds Combine operation during turns frequently creates yield errors. This is especially true when the sides of the harvest area create an acute angle. A series of data points with partial swath width will be created as the combine cuts to the end of a side of the field and starts back onto the adjacent side. Figure 1 shows the machine path at a turn and indicates those points which are likely to be recorded with less than a full header width. Turns are detected by the change of pass number. Each time the header is raised, a new pass number is generated. A user specified number of points before and after the pass number change are compared with each other to calculate the distance between each pair of points. For any pair that has a distance less than a user specified percentage of the swath width, the point after the pass number change is eliminated. Turns can only be detected with this filter if the header is raised enough to stop recording temporarily. In cases where data recording is never stopped during the turn, this filtering method has no effect. The default number of points to be compared is 25 and the default swath width percentage is 70.

Overlaps Data recorded while traveling over previously harvested areas is the primary justification for development of this filter. Rice, sorghum and wheat are broadcast crops that are harvested with platform combine headers. During harvest, it is common for the machine have acute angle turns or narrow lands. In rice fields, the normal path of the harvester is along the levees. As the combine enters the field, only a partial header width will be cut. Fields with several levees will prevent the combine from cutting a complete header swath width as a section of the field is completed. This harvest pass could produce totally erroneous yield estimates. Similar situations occur when the combine operators do not pick up the header during turns. This lack of action leaves the data recording on while the machine is traveling over ground previously harvested. Figure 2 shows a sorghum field with many instances of recording data while driving over previously harvested areas.

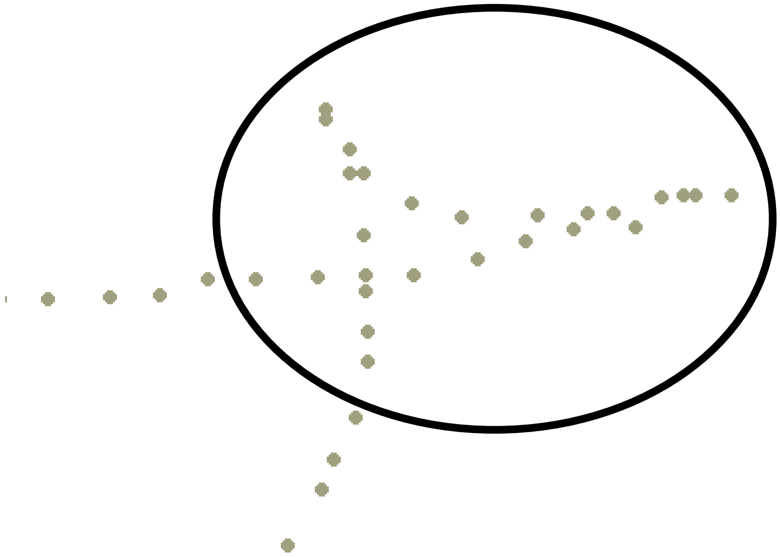

Figure 1 – Points recorded during a turn. The points within the ellipse were recorded with a partial swath width or traveling over previously harvested ground. These are the points that the turnaround filter attempts to eliminate.

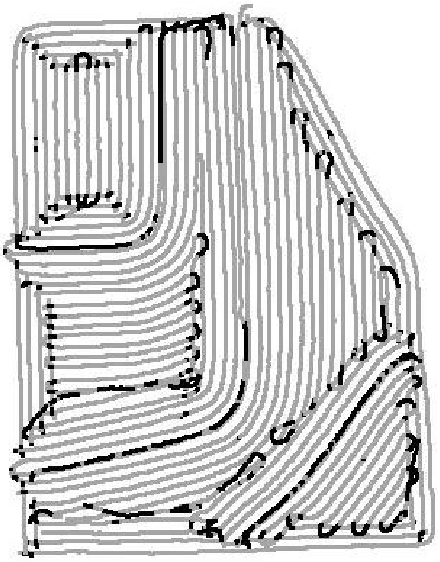

Figure 2 – A sorghum field with overlapping yield points highlighted.

The procedure used to correct incomplete header widths is the overlap filter. This filter places a bitmap grid over the entire field in a north-south, east-west configuration. The size of the square grid cells is user specified as a percentage of the header swath width. Because of the bitmap orientation, passes traveling at 45 degrees to an ordinate direction are closer than a full swath width apart in the N-S and E-W directions. When the diagonal of a square cell equals the swath width, the length of both sides of the cell is 70.7% of the swath width. For this reason, the default value for the grid width is 70 percent.

As the data points are sequentially processed, each point is assigned to a cell location. If the cell is empty, meaning that area has not yet been harvested, then this data point is assumed to be harvesting a full swath width and the data is valid. To accommodate a slow travel speed, sequential yield points are considered valid within the same cell. After consecutive data points have left a cell, all subsequent data points falling into that cell are discarded.

The entire algorithm is named the Texas Agricultural Experiment Station (TAES) filter. Its intended use is to remove as many erroneous yield data points as possible without relying on assumptions about the accuracy of the GPS locations, the number of machines operating in the field or the patterns of harvest. The program is written in Borland C++ 5.0 for the EasyWin format. EasyWin is a windows type DOS program. The input to the TAES filter is an exported yield file in either JD-MAP® or Ag Leader® Advanced formats. The user has the option of filtering with the default values or entering alternative values. The filter produces three or four output files, depending on the users selection of the original input format or the Ag Leader® Basic format. The Basic format is provided as an output option for those users who wish to save the calculated dry basis yield. One output file is created for all yield points considered valid, another for all points considered invalid and a third file with all points is created if the Basic format is chosen. The final file contains the latitude, longitude, a point identification number and the pass/fail status for each of the filter criteria. This status file was produced to allow analysis of the performance of the filter elements.

Since the TAES filter examines every point in the file, the time required for filtering is proportional to the number of data points in the field. A field with 33 000 points takes about three minutes when processed on a Pentium II 266 MHz computer. Memory requirements are modest, and normally not a problem on a Windows operating system.

Fields analyzed

For this study, yield data from 10 rice, sorghum, and corn fields were processed and analyzed for the impact of the various filters mentioned above. The rice and sorghum fields were harvested in 1998, and the corn data was from the 1997 growing season. Six of the fields were harvested with one combine, while the other four fields had two combines harvesting in tandem. Table 1 contains information describing the fields. The data was recorded with John Deere, Ag Leader and Case AFS yield mapping systems. Each of the 10 fields was exported from the manufacturer's software package in the advanced format.

The range of the yields before any filtering was greater than expected in the rice and sorghum fields. The maximum yields varied from 71.3 Mg/ha (375% above maximum expected yield) at one rice farm to 278 Mg/ha (1750% above) at another rice farm to 388 Mg/ha (1840% above) at a sorghum field. With the three corn fields, the maximum calculated yield was 25.9 Mg/ha (29.5% above). No difference was seen between the maximum yields when one or two combines harvested the field. The corn fields were under center pivot irrigation systems, and the rows were straight across the entire field. Therefore, there were no turns within the fields. The rice and sorghum fields were generally odd shaped polygons, and were harvested in rectangular patterns or following the levees that existed in the fields. Consequently, many turns were made within the field, resulting in a high portion of suspect yield data points.

Results and Discussion

The TAES filter was applied to each of the fields, and the resulting output files were analyzed for the impact of the various filtering techniques. A simple threshold filter was also used for comparison. The actual field area and harvested mass were requested from the cooperating producers, but for a variety of reasons, many of the values provided were questionable. Consequently, only three fields had actual field average yields that could be compared to the original and filtered values. Table 2 shows the field average results for the filtering methods. An unexpected result of the filter comparison was that the average yield for the TAES filter was greater than that from the threshold filter for all ten fields. Since the TAES filter

Table 1. Field descriptions

Crop	Field ID	Area (ha)	No. of combines	Maximum yield (Mg/ha)
Sorghum	A	10.1	1	81.4
Sorghum	B	50.6	1	388.4
Corn	A	52.6	1	27.9
Corn	B	53.8	1	25.9
Corn	C	50.6	2	27.7
Rice	A	21.0	1	270.8
Rice	B	47.3	2	297.4
Rice	C	23.1	1	266.3
Rice	D	30.8	2	64.3
Rice	E	29.1	2	74.8

included thresholding, this means that the data points removed by the other filter elements must have been lower values. This is reasonable, since the additional points removed were expected to be primarily points recorded during partial header width operation or turns.

Filtering greatly reduced the variability of the yield data as indicated by the standard deviation of the entire field. Most of the reduction in variation was accomplished with the threshold filter, but for all ten fields, a slight further reduction was achieved with the TAES filter.

Table 2. Field average yields with and without filtering

		Average yield (Mg/ha)		
Crop	Field ID	Unfiltered	Threshold filter	TAES filter (70% Cell Size)
Sorghum	A	2.23	2.16	2.22
Sorghum	B	5.38	4.96	5.02
Corn	A	13.09	13.06	13.15
Corn	B	12.23	12.21	12.39
Corn	C	14.42	14.39	14.52
Rice	A	6.30	5.28	5.64
Rice	B	6.93	6.52	6.78
Rice	C	6.00	5.64	6.06
Rice	D	8.29	8.26	8.34
Rice	E	8.39	8.35	8.39

The different filtering elements had varying effects for the ten data sets processed for this study. In many cases, individual points would be identified for elimination by multiple filter elements. The minimum and maximum thresholds eliminated from 0.4 to 15 percent of the data points. For the fields with more data points eliminated by the threshold, there seems to be an increased incidence of driving over previously harvested ground while recording yield data. The moisture data were reasonable in all the fields so no data was removed with the moisture filter. However, several fields show multiple data points with a travel distance of 0 cm (0 in.). Few data points were removed because of an excessive distance

Table 3. Effect of filtering on yield variation.

		Standard Deviation (Mg/ha)		
Crop	Field ID	Unfiltered	Threshold filter	TAES filter
Sorghum	A	1.90	0.85	0.75
Sorghum	B	8.76	1.61	1.44
Corn	A	3.02	2.98	2.88
Corn	B	3.76	3.75	3.53
Corn	C	2.17	2.13	1.84
Rice	A	6.60	2.44	2.14
Rice	B	7.41	2.43	2.14
Rice	C	5.59	2.47	2.00
Rice	D	1.82	1.53	1.39
Rice	E	2.04	1.63	1.54

traveled. The percent of points removed as a result of an invalid distance traveled is up to 2.9% in one rice field. This error can be contributed to GPS, mapping software, and operator error. The distance filter did not remove any points in many of the fields tested. The yield surge filter removed between 1.1 and 2.0% of the data in all the fields.

The critical filtering technique applied to the fields was the overlap filter. In order to select the optimum cell size, widths of 50, 70 and 100 percent of the indicated swath width were examined. Table 4 shows the number and percentage of points identified for elimination for overlap. The percent data removed is nearly doubled when changing the cell size from 70 to 100% of the swath width. Even with two combines harvesting a field, about 3% of the points are removed with this overlap filter with 100% of the swath width. The percent of data removed increases to 10.8-22.9% when only one combine is harvests a field. Using the 70% of the swath width will reduce the percent of data removed from the single-combined fields to 4.2-11.6%. This change is apparent in areas where the combine is harvesting an arc through the field, or traveling in any direction other than horizontal or vertical. The yield average is not changed by more than 2% between the 70 and 100% bitmaps. As expected, there is not as much change between the 50 and 70% cell sizes as there is with the 70 and 100% cell sizes. The fewest points are removed with the densest bitmap (50% of swath width). Figures 3, 4 and 5 show the results of using a 100, 70 and 50 percent cell size for a field with a curves travel path. The increased number of "good" points eliminated with the 100 percent cell width is obvious.

Table 4. Points eliminated because of overlap with previously harvested ground.

			50% Cell Size		70% Cell Size		100% Cell Size	
Crop	Field ID	Total points	Number	Percent	Number	Percent	Number	Percent
Sorghum	A	13867	1138	8.2%	1613	11.6%	2685	19.4%
Sorghum	B	43986	1738	4.0%	2008	4.6%	4769	10.8%
Corn	A	33535	1633	4.9%	2678	8.0%	5561	16.6%
Corn	B	34114	1422	4.2%	2525	7.4%	5467	16.0%
Corn	C	13292	90	0.7%	131	1.0%	436	3.3%
Rice	A	24288	2108	8.7%	2661	11.0%	4819	19.8%
Rice	B	34064	1625	4.8%	1937	5.7%	3628	10.7%
Rice	C	32135	2760	8.6%	3509	10.9%	7372	22.9%
Rice	D	10414	59	0.6%	106	1.0%	710	6.8%
Rice	E	11828	25	0.2%	53	0.4%	88	0.7%

The overlap filter removed yield points that are too close to previous data points. This function is intended to identify partial swath width situation, but these can not be distinguished from the 'wandering' of the GPS locations. In the corn fields, there were no instances of partial swath operation, but large numbers of points were removed. This was due to the error in GPS location. This is a limitation of the current implementation of the overlap filter. It currently does not examine any characteristics associated with the data points besides position. If assumptions are made about the harvesting patterns, some of these points could be corrected and retained. However, for this filter, we did not want to make assumptions about the harvest operations. The filter also can not always identify partial swath data. When two combines harvest in tandem, the filter cannot detect where the non-mapping combine has traveled.

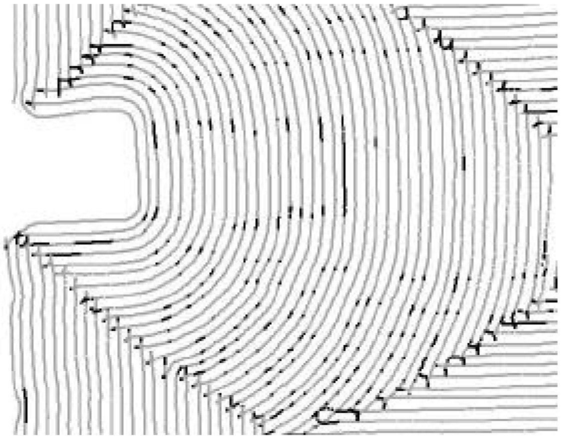

Figure 3 – Data points eliminated by the 100% overlap filter. Dark points are identified for removal.

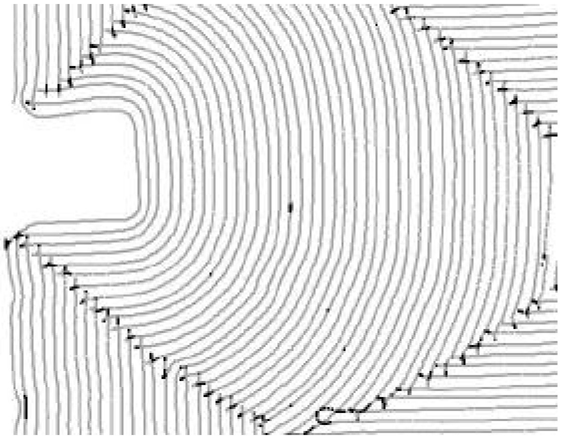

Figure 4 – Data points eliminated by the 70% overlap filter. Dark points are identified for removal.

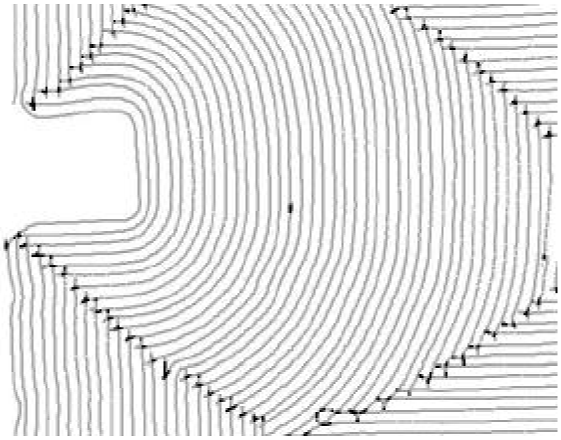

Figure 5 – Data points eliminated by the 50% overlap filter. Dark points are identified for removal.

Summary

Overall, the TAES filter worked as designed. The filter was able to remove many of suspect yield points while not eliminating too many good points. Several observations resulted from the operation of this filter.

- Most visibly erroneous data is removed.

The effects of poor operating procedures and erroneous data were generally removed from the filtered yield maps. This allowed a better understanding of the actual yield trends that existed in the fields.

- Non-mapping combines in the field cause fewer errors to be detected.

The TAES filter was developed for use in situations where the operation of non-mapping combines affected the data files. Since the operation of these machines was uncertain, no assumptions were made about where the non-mapping combines might have passed through the field. As a result, fewer of the low yield points that probably resulted from partial swath situations could be identified and removed.

- The TAES filter works better than a threshold only filter in crops with high yield variability.

Threshold filters allowed more erroneous yield points to remain in the valid data file. The difference was primarily overlapped points that were greater than the minimum threshold. That minimum could be increased, but at the cost of possibly incorrectly eliminating data from low yield areas.

- Good operating practices minimize the need to filter yield data.

Some operators used better yield mapping practices while operating the combine. Those files had many fewer points eliminated than files from operators who did not consider the data recording function of the combine.

References

Blackmore, B. S. and C. J. Marshall. 1996. Yield Mapping; Errors and Algorithms. *Presented at the 3rd International Conference on Precision Agriculture*, Silsoe College, Cranfield University.

Han, S., S. M. Schneider, S. L. Rawlins, R. G. Evans. 1996. A Bitmap Method for Determining Effective Combine Cut Width in Yield Mapping. *Transactions of the ASAE* 40(2):485-490.

Rands, M. 1995. The development of an expert filter to improve the quality of yield mapping data. *Unpublished MSc. Thesis*, Silsoe College, Cranfield University.

Searcy, S. W. 1998. Operating Practices Can Improve Yield Maps. *Factsheet*, Texas Agricultural Extension Service, Texas A&M University System.

Acknowledgements

This research was supported by the Texas Rice Research Foundation, Rice Belt Warehouse, Cargill, Inc. and the Texas Agricultural Experiment Station. Thanks go to the cooperating producers who generously shared their yield data to support research activities.

Filter Availability

The TAES filter can be downloaded from the Precision Agriculture Laboratory of the Texas A&M Agricultural Engineering Department at the following URL. http://www.agen.tamu.edu/txprecag/