Yield Check

An Algorithm for Filtering of Yield Monitor Data

Gregorio C. Simbahan

Achim Dobermann

Department of Agronomy and Horticulture
Institute of Agriculture and Natural Resources

About Yield Check

Yield Check is a software for screening geo-referenced yield monitor data. It detects six common types of erroneous yield monitor values: 1) combine header status up, (2) start-/end-pass delays, (3) grain flow, distance traveled, and grain moisture outliers, (4) values exceeding minimum and maximum biological yield limits, (5) local neighborhood outliers, and (6) short segments and co-located points. Erroneous values are either deleted or flagged for further analysis to identify where they have occurred in the field. The output file can then be used in other commercial mapping software to create a yield map or maps of the different categories of erroneous yield values. **Yield Check** was developed by Gregorio Simbahan and Achim Dobermann at the University of Nebraska-Lincoln. Our intention was to provide a simple utility 'as is', with little flexibility in input and output data formats and a simple graphical user interface. Other commercial software is recommended for pre-processing yield monitor raw data before running the **Yield Check**. Input data must be in a text file that follows the AgLeader advanced export format. Yield Check was written in Visual Basic 6 and we encourage the use of the algorithm in commercial yield mapping software. For further inquires, please contact:

Achim Dobermann
University of Nebraska-Lincoln
Department of Agronomy & Horticulture
P.O. Box 830915, Lincoln, NE 68583-0915, USA
Phone: +1-402-4721501, Fax: +1-402-4727904, E-mail: adobermann2@unl.edu

Suggested citation: Simbahan, G.C., and A. Dobermann. 2004. Yield Check – An algorithm for filtering of yield monitor data. Lincoln, NE: University of Nebraska-Lincoln. http://soilfertility.unl.edu

TERMS OF USE

By accessing or using this software, you agree to these terms of use. If you do not agree to these terms you may not use this software.

All content included in this software, such as text, graphics, logos, button icons, images, as well as any compilation in form of collection, arrangement, and assembly is the exclusive property of University of Nebraska-Lincoln and protected by United States copyright laws. This software may be used as a resource for processing yield map data. Any other use, including the reproduction, modification, distribution, transmission, republication, display or performance, of the content is strictly prohibited.

GRANT. Subject to payment of applicable license fees, UNL hereby grants to you a non-exclusive license to the Software and accompanying documentation (Documentation) on the terms below. This license does not grant you any rights to enhancement or update. You may use the Software on any single computer or copy the Software for archival purposes, provided the archival copy contains all of the original Software's proprietary notices. You may not permit other individuals to use the Software except under the terms listed above; modify, translate, reverse engineer, decompile, disassemble (except to the extent applicable laws specifically prohibit such restriction), or create derivative works based on the Software or Documentation; copy the Software or Documentation (except as specified above); rent, lease, transfer or otherwise transfer rights to the Software or Documentation; or remove any proprietary notices or labels on the Software or Documentation.

DISCLAIMER

LIMITATION OF LIABILITY. UNDER NO CIRCUMSTANCES AND UNDER NO LEGAL THEORY, TORT, CONTRACT, OR OTHERWISE, SHALL THE BOARD OF REGENTS OF THE UNIVERSITY OF NEBRASKA, UNL OR ANY OF ITS AFFILIATES, EMPLOYEES OR OTHER REPRESENTATIVES BE LIABLE TO YOU OR ANY OTHER PERSON FOR ANY INDIRECT, SPECIAL, INCIDENTAL, OR CONSEQUENTIAL DAMAGES OF ANY CHARACTER INCLUDING, WITHOUT LIMITATION, DAMAGES FOR LOSS OF GOODWILL, WORK STOPPAGE, COMPUTER FAILURE OR MALFUNCTION, OR ANY AND ALL OTHER COMMERCIAL DAMAGES OR LOSSES, SPECIFICALLY INCLUDING DAMAGES TO CROPS AND SOIL RESULTING FROM THE APPLICATION OF THE SOFTWARE, EVEN IF UNL SHALL HAVE BEEN INFORMED OF THE POSSIBILITY OF SUCH DAMAGES, OR FOR ANY CLAIM BY ANY OTHER PARTY. THIS LIMITATION OF LIABILITY SHALL NOT APPLY TO LIABILITY FOR DEATH OR PERSONAL INJURY TO THE EXTENT APPLICABLE LAW PROHIBITS SUCH LIMITATION. FURTHERMORE, SOME STATES DO NOT ALLOW THE EXCLUSION OR LIMITATION OF INCIDENTAL OR CONSEQUENTIAL DAMAGES, SO THIS LIMITATION AND EXCLUSION MAY NOT APPLY TO YOU.

Table of Contents

Installation

Yield Check requires about 3 MB of free hard disk space and should run on any PC under Windows 98SE, NT, 2000, or XP. Yield Check does not come with an automatic setup program. Instead, please follow the instructions below for manual installation:

1. Download or copy file **chkyield.zip** to a folder of your choice.
2. Using Windows Explorer, create a new folder **c:\chkyield** in the main (c:) harddrive of your computer.
3. Un-zip all files contained in **chkyield.zip** into the newly created folder **c:\chkyield**. The folder will now contain the following files files:

UNLYieldCheck.exe	the executable program file
Crop.dat	file containing default parameters for different crops (txt file)
Criteria.dat	file containing default filtering criteria (txt file)
Example_in.txt	an example data input file
Example_out1.txt	an example of a basic output data file
Example_out2.txt	an example of a detailed output data file
YieldCheck.pdf	User’s Guide

4. You can run Yield Check by double-clicking on **UNLYieldCheck.exe**
5. Alternatively, use a right-mouse click to create a shortcut, drag it on your desktop, and start the program by double-clicking on the shortcut.

Yield Cleaning Algorithm

Yield data obtained with combine-mounted, geo-referenced grain yield monitors are affected by various systematic and random sources of measured yield variation (Stafford et al., 1996; Doerge, 1999; Arslan and Colvin, 2002b), including (i) naturally occurring yield variation due to climate and soil-landscape features, (ii) management-induced yield variation, and (iii) measurement errors caused by the yield monitoring process itself. Naturally occurring yield variation is often related to more gradual changes in soil-landscape conditions. On the other hand, variation caused by the actual crop management often represents random events that typically occur in small areas, such as planter skips, poor crop establishment, non-uniform fertilizer application, herbicide damage, lodging, or pest damage. Measurement errors include grain flow and other sensor errors (moisture, speed, swath width), errors due to geo-referencing and combine movement, operator errors, and data processing errors (Shearer et al., 1997; Blackmore and Moore, 1999; Arslan and Colvin, 2002b). For most locations within a field, both (ii) and (iii) represent short-distance, random variation that differs from year to year. Such artifacts must be removed from yield monitor raw data to display and properly interpret the major patterns of yield variation as a basis for making site-specific crop management decisions (Ping and Dobermann, 2003). In addition to inappropriate yield monitor calibration, reasons for yield monitor errors can be numerous, but the most obvious are (Adamchuk et al., 2004):

- Varying the crop width which enters the header during harvest
- Changing lag time of the grain as it goes through the threshing mechanism
- Surging grain through the combine grain transport system
- Grain losses from the combine
- Travel speed, grain flow and grain moisture measurement errors
- The inherent 'wandering' error from the GPS

Several data filtering algorithms have been developed and are incorporated into commercially available agricultural GIS products in order to remove or replace data points that fail predefined criteria. **Yield Check** is an algorithm in which the goal is to detect and remove (or flag) six types of errors (Figure 1):

1. Combine header status is up
2. Start- and end-pass delays for both headlands and stop-and-go segments within the field,
3. Frequency distribution outliers of distance traveled, grain flow, and grain moisture
4. User-defined minimum and maximum biological yield limits
5. Small patches or narrow strips with extremely low or high yields that are not closely related to their immediate neighbors
6. Short segments and co-located yield records.

A detailed description of the algorithm and its application is provided elsewhere (Simbahan et al., 2004). Shifting of raw data to correct for grain flow delay through a combine is not done in **Yield Check.** We recommend using commercial software such as SMS Basic (http://www.agleader.com/sms.htm) to pre-process yield monitor raw files (e.g., .yld files) in order to correct for grain flow delay shift and export the converted data into a text file (in advanced export format**,** see section on **Input Data Format**) for further processing in **Yield Check**.

Primary cleaning involves the deletion of points that represent header status up (**Step 1**) and start- and end-pass delays (**Step 2**). It can be done either in commercial software during pre-processing or within **Yield Check** (see section on **Input Data Format**). **Steps (3) through (6)** are referred to as secondary screening (Figure 1), in which the objective is to further remove both truly erroneous yield records that are caused by combine operation and yield sensing, as well as localized, extreme yield variation due to either measurement error, crop management, or other events that are not related to the general spatial patterns of crop yield variation. The exact causes for such random yield variation are not always known, so the secondary screening process is a combination of statistical tests and empirical decisions.

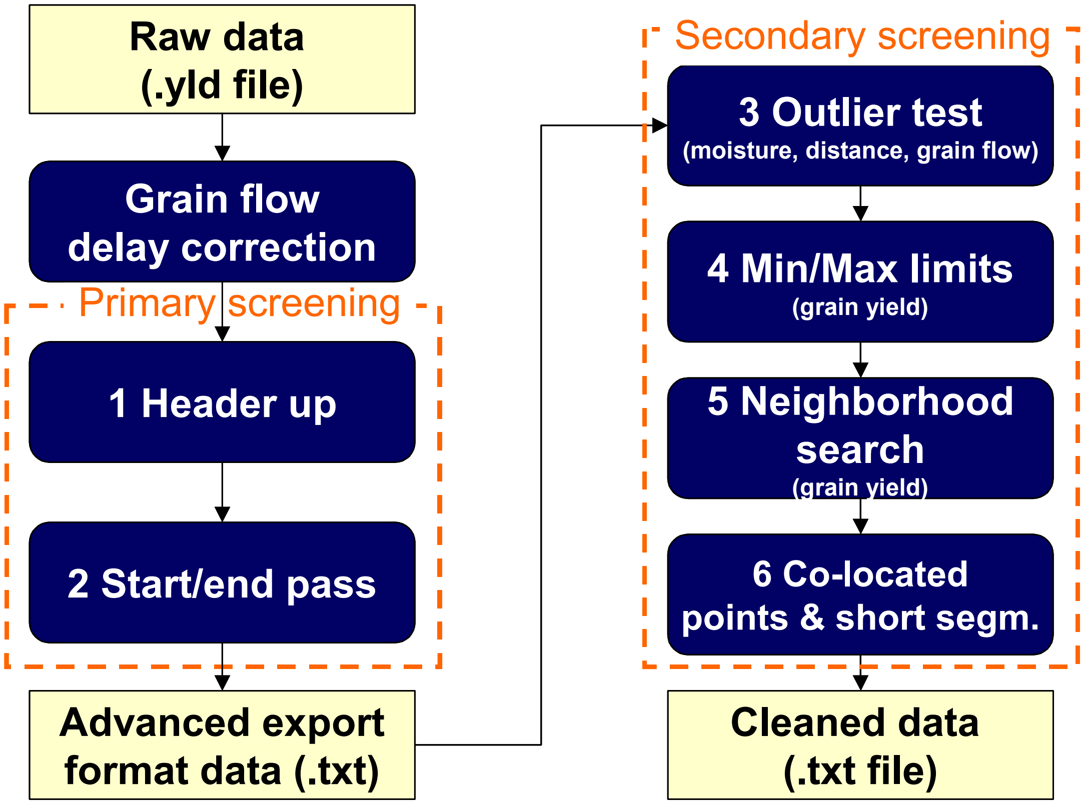


Figure 1. Algorithm for yield data screening used in Yield Check.

Step 2: Values recorded after the header has been lowered but grain flow has not started or has not stabilized yet are referred to as **start-pass delay**. Values at the end of harvest segments, when cutting has stopped but the header has not been raised yet, are referred to as **end-pass delay**. The main decision for the program operator to make is to choose settings for the length of start- and end-pass delays, which may differ among crops and harvest combines due to differences in swath width, harvest speed, and grain flow through a combine. To obtain location-specific settings, grain flow measured during a short time period after start of a new harvest segment or before the end of a harvest pass can be plotted vs. time. Figure 2 shows an example for 15 combine tracks in a field, extracted from yield monitor file. Based on visual inspection, 8 s and 4 s were selected as start- and end-pass delays, respectively.

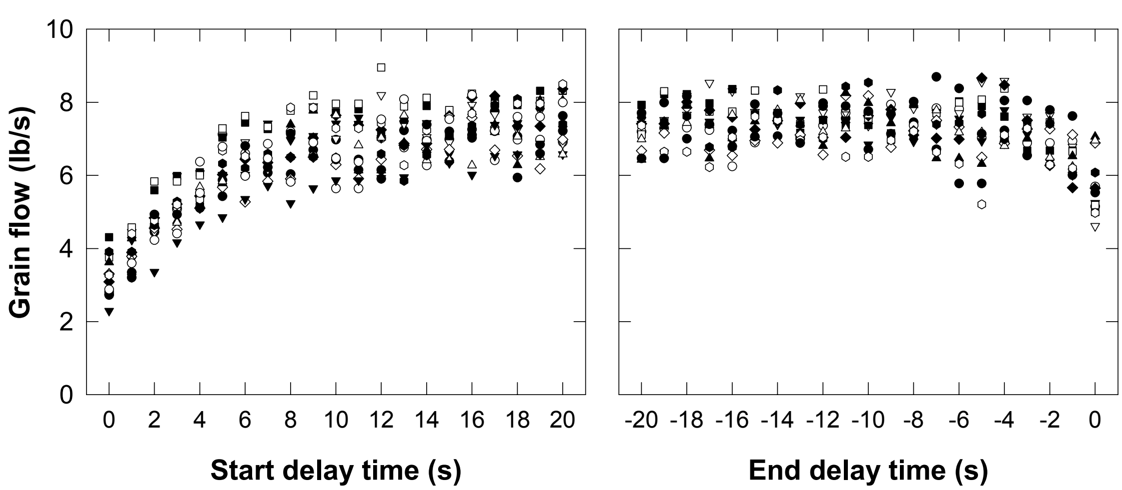


Figure 2. Yield monitor grain flows near start (left) and end (right) of harvest passes.

Step 3: An **outlier test** is performed for the variables **grain flow, grain moisture, and distance traveled**. The main rationale for frequency distribution screening is to eliminate the most extreme values in terms of combine speed, grain flow and moisture that lie outside ranges of optimal yield monitor performance. The global means and standard deviations (SD) are calculated for the whole field and, as a default criterion, yield records for which any one of these variables is outside the mean ± 3 SD range are deleted. The same criterion was used by Kleinjan et al. (2002) as the final step in their cleaning algorithm. In **Yield Check**, the SD value can be changed by the user.

Step 4: The user must provide an estimate of the expected **minimum and maximum yield limits**, which should represent the biologically possible yield range. Although many erroneous data outside this range are likely to be screened out in step (3), the additional step was added as a machine-independent cross check, for which limits can be set based on expert knowledge about crop yield potential. The value for the minimum possible yield should be set to a small number, slightly larger than zero, unless there is evidence that areas with true zero yields occur in a field and are not caused by the yield monitoring operation itself. The value for the maximum possible yield should represent crop yield potential, defined as the yield of a crop cultivar when grown in environments to which it is adapted, with nutrients and water non-limiting, and pests and diseases effectively controlled (Evans, 1993). Information from high-yield trials or crop simulation models can be used for setting site- and season-specific upper yield limits. For example, upper limits used in our studies were usually set to 350 bu/acre (22 Mg ha^{-1}) for maize (15.5% moisture content) and 105 bu/acre (7 Mg ha^{-1}) for soybean (13% moisture). For both crops, the upper values represented maximum yields achieved in yield contests or simulated by crop simulation models for the climatic conditions in Eastern Nebraska and other areas of the Corn Belt (Duvick and Cassman, 1999; Specht et al., 1999; Dobermann et al., 2003a).

Step 5: Local yield extremes that occur in **small patches or narrow strips** with little relationship to neighboring yield records are detected and deleted in step (5). This may include remaining yield monitor or combine operation errors, but also true short-distance yield variation due to random events affecting crop growth. The assumption underlying this screening step is that such local outliers represent random events that are limited in scope and unlikely to occur at the same location in succeeding years. Such data should be filtered out for integrating yield maps over time to obtain maps of spatially varying yield performance (Dobermann et al., 2003b). As shown in Figure 3, following the movement of the combine through the field a floating local neighborhood test is performed for each yield monitor record.
At each location x_j, yield z^* is predicted by inverse distance weighted interpolation with a power of two from all yield values z measured within a local neighborhood surrounding the sampled location (Shepard, 1968):

$$z^*(x_j) = \frac{\sum_{i=1}^{n} z(x_i)\, d_{ij}^{-2}}{\sum_{i=1}^{n} d_{ij}^{-2}} \qquad [1]$$

where x_i are the locations of measured yield data points within the local neighborhood, and d_{ij} is the distance from each point to x_j. The local neighborhood (Figure 3) includes the three preceding and three succeeding yield records in the same track as well as yield records within a band perpendicular to the tangent of the path of traveled. The user can specify the size of this band. For most applications, the band should cross two adjacent harvest passes on both the left and right sides of the path of travel (**Y)** and its width (**X)** should be equivalent to about one swath width of the combine. This definition of the local neighborhood assigns more weight to the neighboring harvest passes because those are likely to have yield monitor and operation errors that are independent from those affecting data points that precede or succeed x_j in the same combine track.
The confidence interval of the estimate z^* is obtained. A default value of 95% or ±2 SD is used as confidence interval, but, depending on the intended use of a yield map, this value can be changed. For example, if there is evidence that certain management-induced yield patterns are small and permanent (e.g., compacted travel paths), the test criteria for the local neighborhood test could be relaxed. If the measured yield z at location x_j is outside this interval, the yield value is considered a spatially uncorrelated outlier and discarded. The rationale for this definition is that yield at any location x_j is likely to be spatially

correlated to its immediate neighbors, irrespective of the direction of the combine movement. If that is not the case, a random event must have caused an unusually high or low yield at location x_j, either due to yield monitor error or due to specific crop management events that occur in very small patches. The former may include sudden changes in speed or grain flow (Arslan and Colvin, 2002a; Arslan and Colvin, 2002b), whereas the latter may be caused by planter skips, poor crop establishment, non-uniform fertilizer application, herbicide damage, lodging, pest damage, etc. A super-block search algorithm was implemented in step (5) to facilitate a fast, floating local neighborhood search. In this, the whole field is divided into 9 super blocks to narrow the search radius for identifying neighboring points around a test location x_j.

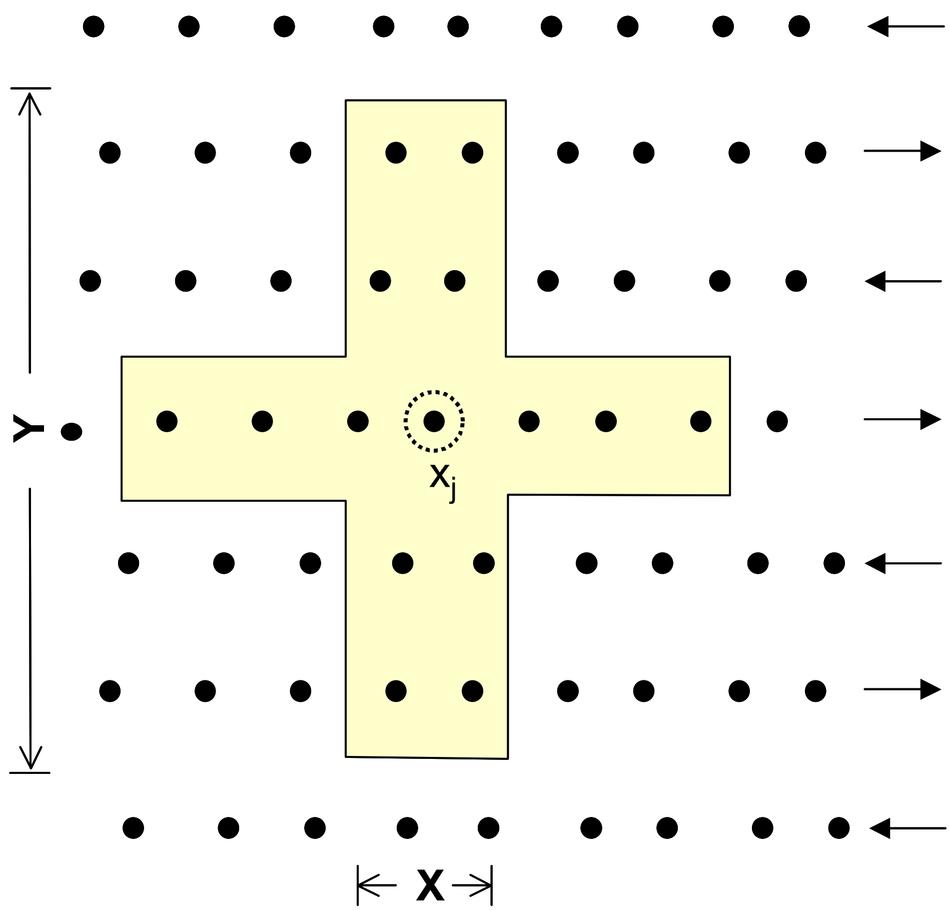


Figure 3. Determination of local neighborhood outliers (Step 5). The dotted circle shows the yield data point x_j being tested. An estimated yield is calculated by inverse distance interpolation from the yield points located within the local neighborhood and compared with the actual yield. Arrows indicate the direction of the different combine passes in this example. The combine used had a 20 ft swath width (6 m). Therefore, settings chosen for the size of the search window were 30 m across (Y, to include two passes on each side) and 6 m along (X) the direction of travel.

Step 6: removes short segments due to combine stop-and-go events and points that were recorded with the same geographical coordinates, Short segments are considered unreliable because most data points in them are affected by start- or end-pass delays. The default value In **Yield Check** is that segments with less than 12 yield monitor points are identified as short segments and deleted, but this criterion can be changed. Co-located data points can be caused by GPS error or overlapping harvest passes. Their removal is necessary to avoid difficulties with kriging algorithms used for interpolating yield maps.

Running Yield Check

Steps Involved in Running Yield Check

1. Prepare input data file(s) so that they strictly adhere to the AgLeader advanced export text file format.
2. Start the program by double-clicking on **UNLYieldCheck.exe** or double-clicking on the desktop shortcut
3. **Crop Tab:** Select/edit site name, crop, min./max. yield limits (Step 4).
4. **Files Tab:** Select input file, specify output file and its format.
5. **Check Criteria Tab:** choose re-scaling option, specify start-/end-pass-delays (Step 2), specify standard deviation for the outlier test (Step 3), specify local neighborhood size and test criterion (Step 5), and specify criterion for short segments (Step 6).
6. Click on button [**Run**].
7. Review/map the results. Use **Cleaned data** output files to create an interpolated yield map in other software. Use **Detailed output** files to analyze data points discarded by cleaning steps.

Input Data Format

Yield monitors generate a binary file (e.g., a .yld file) which requires the yield monitor's proprietary software or other commercial or public software to convert to a ASCii text file. **Yield Check** requires that the input data are stored in a text file that strictly adheres to the **AgLeader Advanced** export format. Each yield record is separated by a carriage return, i.e., it occupies one single line in the file. Such a data file looks as follows (see **example_in.txt** in c:\chkyield):

```
-96.467896,41.168594,18.97,969288847,2,174,240,11.3,33,1,980415,"F13:2.10NC","L1: ","CORN",7,0,1178.5
-96.467949,41.168594,22.31,969288849,2,176,240,11.1,33,1,980415,"F13:2.10NC","L1: ","CORN",7,0,1178.5
-96.468002,41.168594,23.15,969288851,2,174,240,11.0,33,1,980415,"F13:2.10NC","L1: ","CORN",7,0,1178.5
-96.468056,41.168598,23.87,969288853,2,174,240,11.1,33,1,980415,"F13:2.10NC","L1: ","CORN",7,0,1178.5
-96.468102,41.168598,22.80,969288855,2,171,240,10.9,33,1,980415,"F13:2.10NC","L1: ","CORN",7,0,1178.8
```

Each line (yield record) contains 17 values (columns), each separated by a comma:
Column 1: Longitude (decimal degrees), WGS 84 Lon
Column 2: Latitude (decimal degrees), WGS 84 Lon
Column 3: Grain flow rate (lb/sec)
Column 4: GPS time (seconds since GPS start time)
Column 5: Logging interval (seconds)
Column 6: Distance traveled since last reading (inches)
Column 7: Swath width (inches)
Column 8: Moisture (%)
Column 9: Header status (1-UP, 33-DOWN)
Column 10: Pass no.
Column 11: Serial no.
Column 12: Field ID
Column 13: Load ID
Column 14: Grain type (crop)
Column 15: GPS status
Column 16: PDOP
Column 17: Altitude (feet)

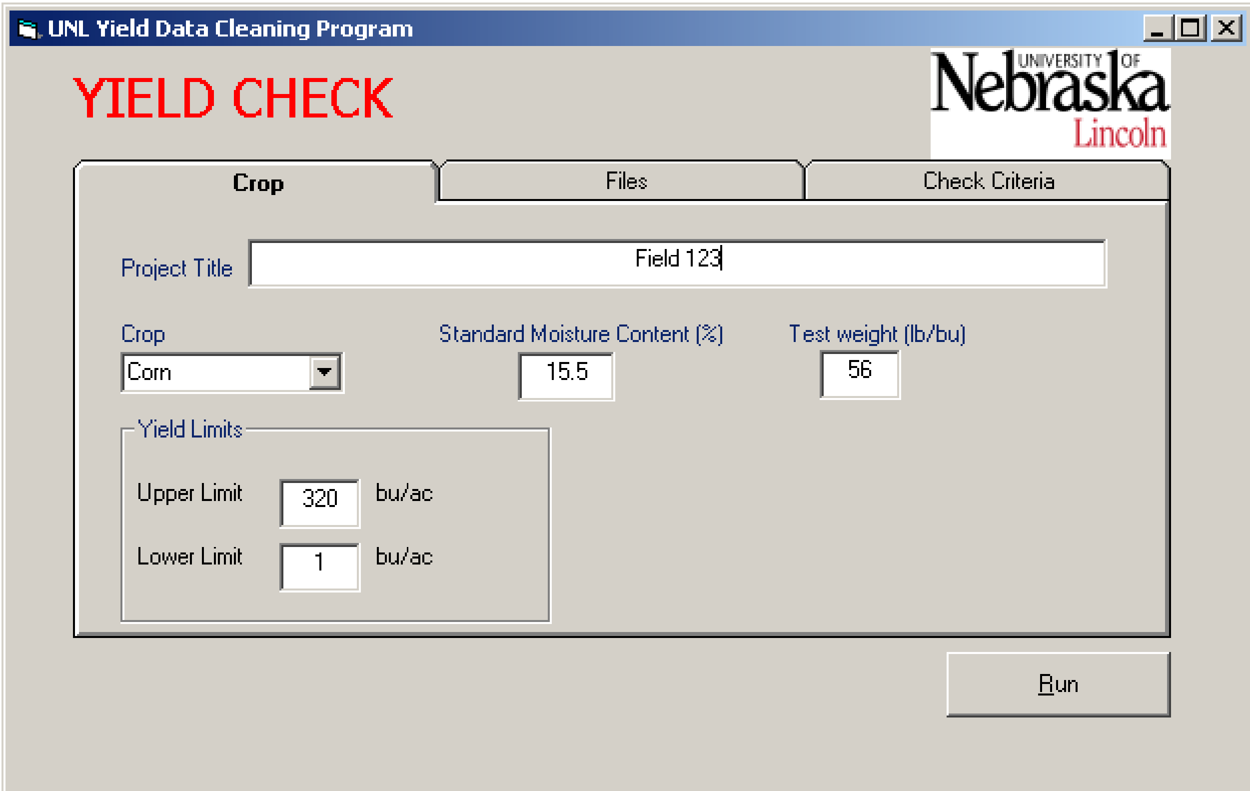
Of those, columns 1 through 11 are used in the various filtering steps of **Yield Check**. Examples, of software packages that allow pre-processing of binary yield monitor raw data files (.yld files) and exporting the data in Ag Leader Advanced export format include:

SMS Basic (http://www.agleader.com/sms.htm)
JDOffice 1.3. (http://www.deere.com/en_US/ag/servicesupport/ams/JDOffice.html
SSToolbox (http://www.sstsoftware.com/toolbox.htm)
Easi Suite (http://www.mapshots.com/products/es)

When processing the yield monitor raw data files, check the settings used in your software. Ideally, only use such software to combine .yld files into one yield monitor file per field, perform the grain flow delay correction (Figure 1), and export the resulting file in the **Ag Leader Advanced Export** format for further filtering by **Yield Check**. Turn off other filters in SMS Basic (or similar software) because **Yield Check** will perform those tasks. Specifically, set the values for specifying start- and end-pass-delays in SMS Basic (or other software) to zero and also turn off filtering for lower and upper yield limits and header status up.

If it is not possible or not desirable to turn off filtering for header status up, start-/end-pass delays, and min/max yield limits, use your commercial software for performing those filtering steps (Steps 1, 2 and 4, in Figure 1) as part of the processing of the yield monitor raw data. However, in this, case, when running the exported file through **Yield Check,** you must set the start- and end-pass-delay settings in **Yield Check** to zero (see **Check Criteria Tab**) and also make sure that the upper and lower yield limits specified on the **Crop Tab** are similar to those used in the pre-processing. If that is not done correctly, more data points will be deleted at the beginning and end of each harvest pass (segment).

Crop Tab



Enter a **Project Title**, which usually is the name of the field to be mapped. This title will appear in the first line of the output file header.

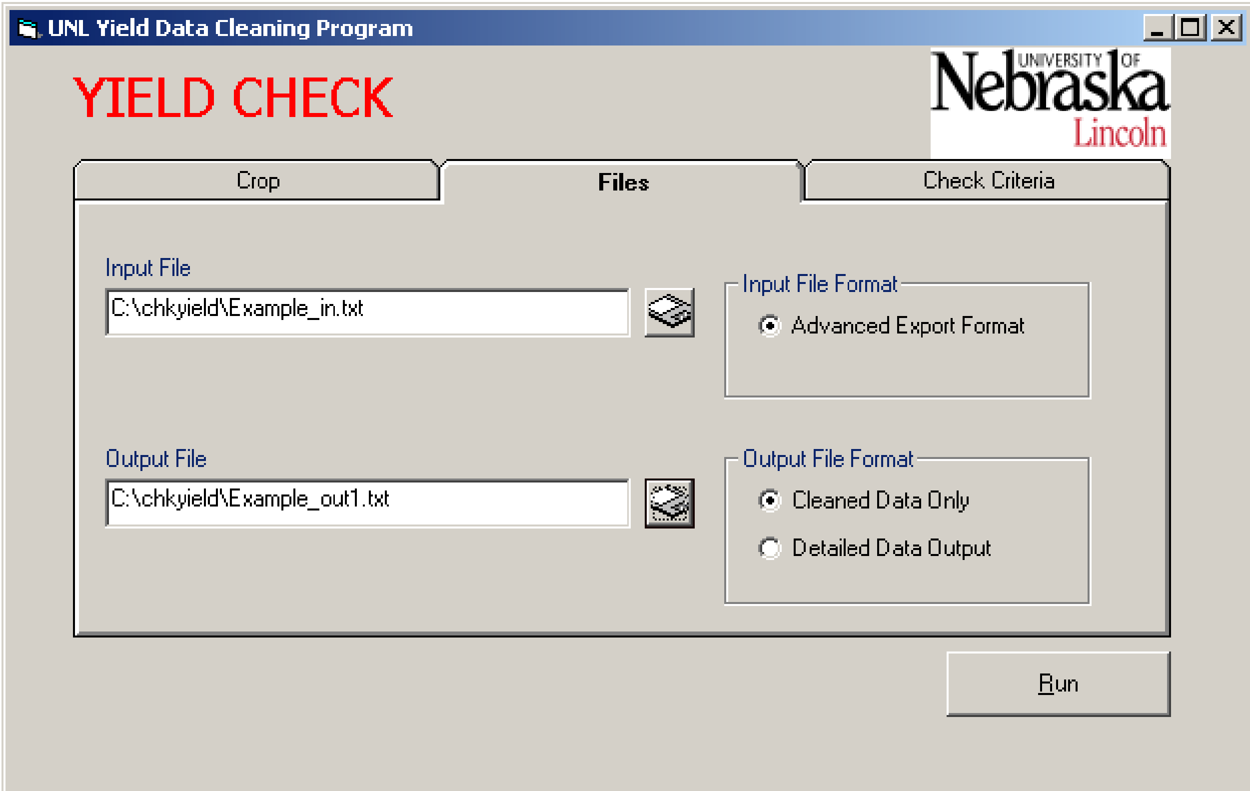
Choose a **Crop** and enter/edit the standard moisture content and test weight for it. **Yield Check** includes default values for corn, soybean, wheat, and sorghum. You can enter any other values, but to make your choices permanently available, you must edit the file **crop.dat** found in C:\chkyield. Use any text editor to edit/add entries in this file. Each line contains the default values for one crop, with 5 columns separated by commas:

```
"Crop","MC","UpperLimit","LowerLimit","BushelWeight"
"Corn",15.5,320.0,1,56
"Soybean",13.5,120,1,60
"Wheat",12.0,160,1,60
"Sorghum",13.0,200,1,56
```

Crop	Crop name
MC	Standard moisture content
UpperLimit	Default upper yield limit used in Step 3
LowerLimit	Default lower yield limit used in Step 3
BushelWeight	Default test weight for this crop (lb/bushel)

Enter upper and lower **Yield Limits** for your crop/field, i.e., trimming limits used in **Step 4** (Figure 1).

Files Tab



Specify location and name of your input file (see **Input Data Format** specifications). All input files must be in **Ag Leader Advanced Export** format. No batch processing is possible in **Yield Check**.

Select the output format and specify location and name of your output file. You may choose between two types of output files:

Output File Format: Cleaned Data Only
This is a simple text file in **GSLIB** format (Deutsch and Journel, 1998). It only contains yield records remaining, i.e., all erroneous values (Figure 1) have been removed. Output is minimized and mainly used for creating a yield map, i.e., the **Clean Data** output file can be easily imported into other software for statistical analysis and/or mapping. **Yield Check** automatically converts Lat/Long coordinates from the input file into UTM coordinates, calculates grain yield from the variables contained in the input file (grain flow, distance, swath width), and adjusts the yield according to the standard moisture content, based on the actual moisture measured. The final yield is then exported to the output data file.

The **Cleaned Data** output file contains a 6-line header comprising the project title (Line 1), the number of variables (Line 2, =4), and the names of the four variables exported (Lines 3-6). Following the header, each lines contains one data record of four variables:

1 Easting	UTM coordinates in east-west direction, m
2 Northing	UTM coordinates in south-north direction, m
3 Yield	bu/acre, adjusted to the standard moisture content entered
4 Moisture	Actual moisture content recorded in the input file

The **Cleaned Data** output format is shown in **Example_out1.txt** in c:\chkyield:

```
Example field no. 1
4
Easting
Northing
Yield
Moisture
712482  4560178   76.15  36.4
712479  4560177   77.06  37.1
712222  4560165   54.23  36.4
712220  4560165   55.97  35.8
712216  4560165   49.98  36.3
…..
```

Output File Format: Detailed Data Output
This is a detailed output file in **GSLIB** text format (Deutsch and Journel, 1998). It contains **all yield records** remaining, i.e., all accepted and erroneous records, but with a column that contains a specific data flag for each record. **Yield Check** automatically converts Lat/Long coordinates from the input file into UTM coordinates, calculates grain yield from the variables contained in the input file (grain flow, distance, swath width), and adjusts the yield according to the standard moisture content, based on the actual moisture measured. The final yield is then exported to the output data file, along with 11 other variables. The **Detailed Clean Data** output file can be easily imported into other software for statistical analysis and/or mapping. This may involve creating of yield maps (use all data with error flag = 0) or mapping the field locations of erroneous values filtered out in the various screening steps (error flag = 1 to 12).
The **Clean Data** output file contains a 12-line header comprising the project title (Line 1), the number of variables (Line 2, = 10), and the names of the 10 variables exported (Lines 3-12). Following the header, each lines contains one data record for the 10 variables, with variables (columns) separated by a space:

1 Easting	UTM coordinates in east-west direction, m			
2 Northing	UTM coordinates in south-north direction, m			
3 Yield	bu/acre, adjusted to the standard moisture content entered			
4 Moisture	Actual moisture content recorded in the input file			
5 VID	Error flag:	0	no error, good yield record	
		1	Error due to header is up (HStatus = 1)	**Step 1**
		2	Start-pass delay error	**Step 2**
		3	End-pass delay error	**Step 2**
		4	Grain flow outlier	**Step 3**
		5	Distance traveled outlier (speed)	**Step 3**
		6	Grain moisture outlier	**Step 3**
		7	Yield < Minimum yield limit	**Step 4**
		8	Yield > Maximum yield limit	**Step 4**
		9	yield > Neighborhood average + n SD	**Step 5**
		10	yield < Neighborhood average - n SD	**Step 5**
		11	Short segment error	**Step 6**
		12	Co-located point (both points removed)	**Step 6**
6 Long	Longitude (decimal degrees), WGS 84 Lon			
7 Lat	Latitude (decimal degrees), WGS 84 Lon			
8 Grainflow	Grain flow rate (lb/sec)			
9 DTrav	Distance traveled since last reading (inches)			
10 Load	Load ID			

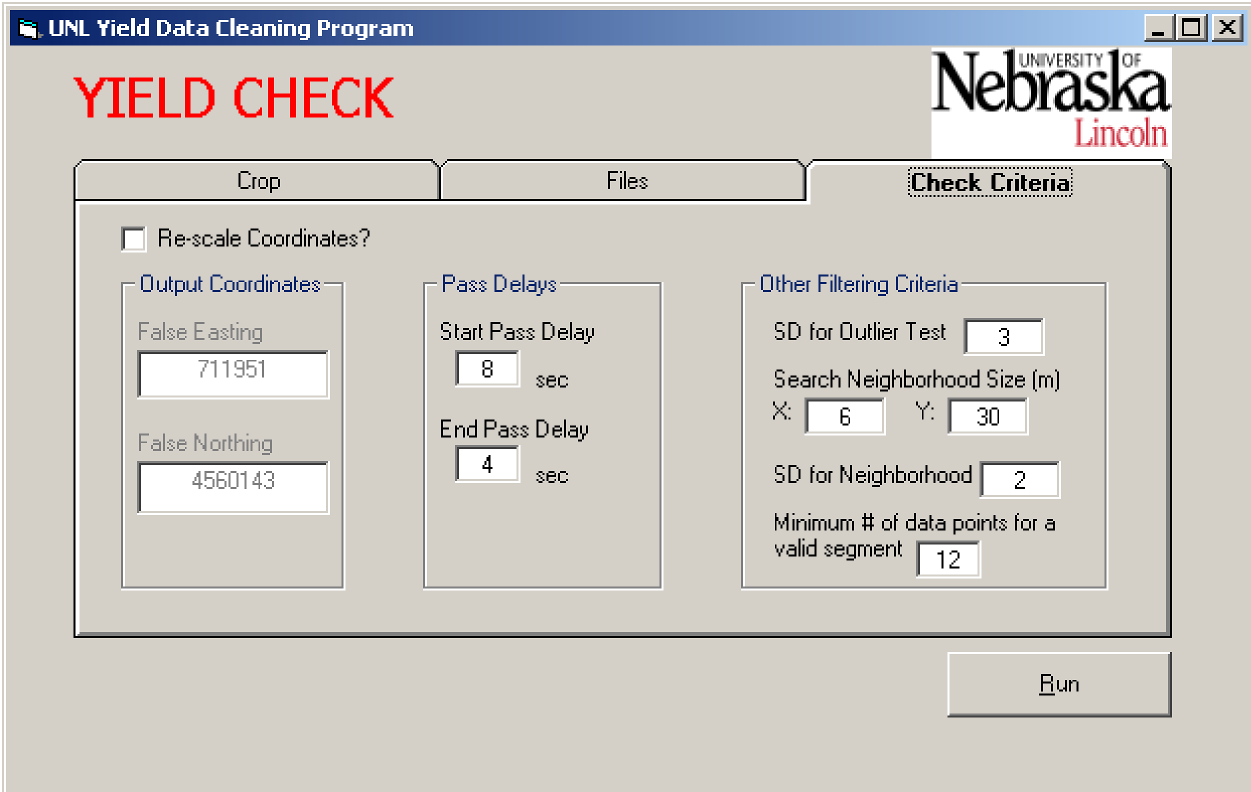
The **Detailed Data** output format is shown in **Example_out2.txt** in c:\chkyield:

```
Field 123
10
Easting
Northing
Yield
Moisture
VID
Long
Lat
GrainFlow
Distance
LoadNo
712525  4560176     0.00 8.4 2 -96.46677 41.16507 0 19 'L1:'
712525  4560176     0.00 8.4 2 -96.46677 41.16507 0 6 'L1:'
712525  4560176    13.97 8.4 2 -96.46677 41.16508 0.29 21 'L1:'
712525  4560178    84.45 8.4 2 -96.46677 41.16510 4.34 52 'L1:'
712524  4560180   119.98 8.4 2 -96.46678 41.16512 9.13 77 'L1:'
712524  4560182   137.43 8.4 2 -96.46678 41.16513 10.73 79 'L1:'
712524  4560184   156.28 22.6 5 -96.46678 41.16515 11.15 61 'L1:'
712524  4560185   129.51 26 5 -96.46678 41.16516 11.09 70 'L1:'
712524  4560187   149.15 23.9 5 -96.46679 41.16518 12.42 70 'L1:'
712523  4560190   125.49 24.2 5 -96.46679 41.16520 11.24 75 'L1:'
712523  4560191   101.15 24.9 5 -96.46679 41.16521 9.51 78 'L1:'
712523  4560194   121.38 24.6 5 -96.46680 41.16524 10.93 75 'L1:'
712523  4560196   114.54 24.9 4 -96.46680 41.16526 12.15 88 'L1:'
712522  4560197    89.51 26.5 4 -96.46680 41.16526 9.15 83 'L1:'
712522  4560197   130.72 25.3 3 -96.46680 41.16526 6.02 38 'L1:'
712522  4560197   126.31 29 3 -96.46680 41.16527 3.06 19 'L1:'
….
```

For example, the first record in this file reads as follows:
712525 4560176   0.00 8.4 2 -96.46677 41.16507 0 19 'L1:'

712525	Easting
4560176	Northing
0.00	Yield
8.4	Moisture
2	Start-pass delay error (combine is moving into the field, but header is already down)
-96.46677	Long
41.16507	Lat
0	Grain flow is still zero (combine is moving into the field)
19	Distance traveled
'L1:'	Load 1

Check Criteria Tab



Re-scale Coordinates

For most applications, it is recommended to leave this option unchecked (default). Check this option only if you wish to re-set the Easting and Nothing coordinates to start at 0, 0 m or any other relative location. To do this, enter the desired offset values in **False Easting** and **False Northing.** The program automatically displays the minimum Easting and Northing values found in the input data file (converted from the Long/Lat coordinates). Thus, if those are not changed and the Re-scale Coordinates check box is marked, Easting and Northing values in the output file will start at 0.

Pass Delays

Specify the values for Start-pass delay (seconds) and End-pass delay (seconds).

Other Filtering Criteria

SD for outlier test: Standard deviation multiplier **n** used in **Step 3** to identify outliers for grain flow, distance traveled, and grain moisture. For each variable, the global mean and the standard deviation (SD) is calculated for the whole field. Yield records for which any one of these variables is outside the mean ± **n** x SD range are deleted. The default **n** value is 3 SD, but it can be changed here. Only make this change if you have good reason to do so.

Search neighborhood size: Size of the search window used in **Step 5** (Figure 1) to perform the local neighborhood outlier test. See Figure 3 for definition of the window.

X Window size (m) in driving direction, should be about one swath width of the combine used (Default: 6 m)

Y Window size (m) across the driving direction, should be about 4 to 5 times the swath width of the combine used (Default: 30 m)

SD for neighborhood:	Standard deviation multiplier used to perform the local neighborhood outlier test of grain yield (**Step 5**). The default value is 3 SD, but it can be changed here.
Minimum # of data points for a valid segment:	Number of data points needed to have a valid segment (**Step 6**). The default value is 12, i.e., any segments with less than 12 yield records between header up/down or stop-and-go events will be deleted here (or flagged as error). Note that this filtering is done after **Steps 1 and 2**, i.e., in **Step 6** we screen for short segments remaining after header up and start-/end-pass delays have already been removed.

Program-internal default values for **Pass Delays** and **Other Filtering Criteria** are stored in a parameter file called **criteria.dat** found in C:\chkyield. Edit this ASCii text file to make permanent changes of the default values so that they better represent your crop and combine conditions. Use any text editor to edit/add entries in this file:

```
"StartPassDelay","EndPassDelay","SearchX","SearchY","SDOutlier","SDNeighbor","MinSegment"
8,4,6,30,3,2,12
```

The entries in this file are:

8	StartDelay	Default start-pass delay used for this crop/combine (seconds)
4	EndDelay	Default end-pass delay used for this crop/combine (seconds)
6	SearchX	Window size (m) in driving direction (**X)**
30	SearchY	Window size (m) across driving direction (**Y)**
3	SDOutlier	SD multiplier for general outlier test of grain flow, distance, and moisture
2	SDNeighbor	SD multiplier for local neighborhood outlier test of grain yield
12	MinSegment	Number of data points needed to have a valid segment

Example

Yield monitor data were collected from an irrigated corn field in Nebraska. using a calibrated Ag Leader™ PF3000 yield monitor with elevator mounted moisture sensor (Ag Leader® Technology, Inc., Ames, IA) and a differential Global Positioning System (GPS) receiver. The combine harvested with a swath width of 6.1 m (eight rows) and yields were recorded at 2 s logging intervals. Grain yields were expressed in Mg/ha with moisture content 15.5% (Conversion: Mg/ha x 1000 / 62.78 = bu/acre).

In this case, 16.4% (4992) of the original yield data were removed, which greatly improved the frequency distribution of grain yield. The frequency distribution of grain yield computed from the original yield monitor data was negatively skewed, including many zero values, but also some extreme yields, which exceeded the known biological yield limit of about 22 Mg/ha (350 bu/acre) for this site. Data removal mainly occurred in the non-irrigated, northwest field corner and near headlands, but also around stop-and-go segments within the field as well as locations dispersed throughout the entire area (Figure 4).

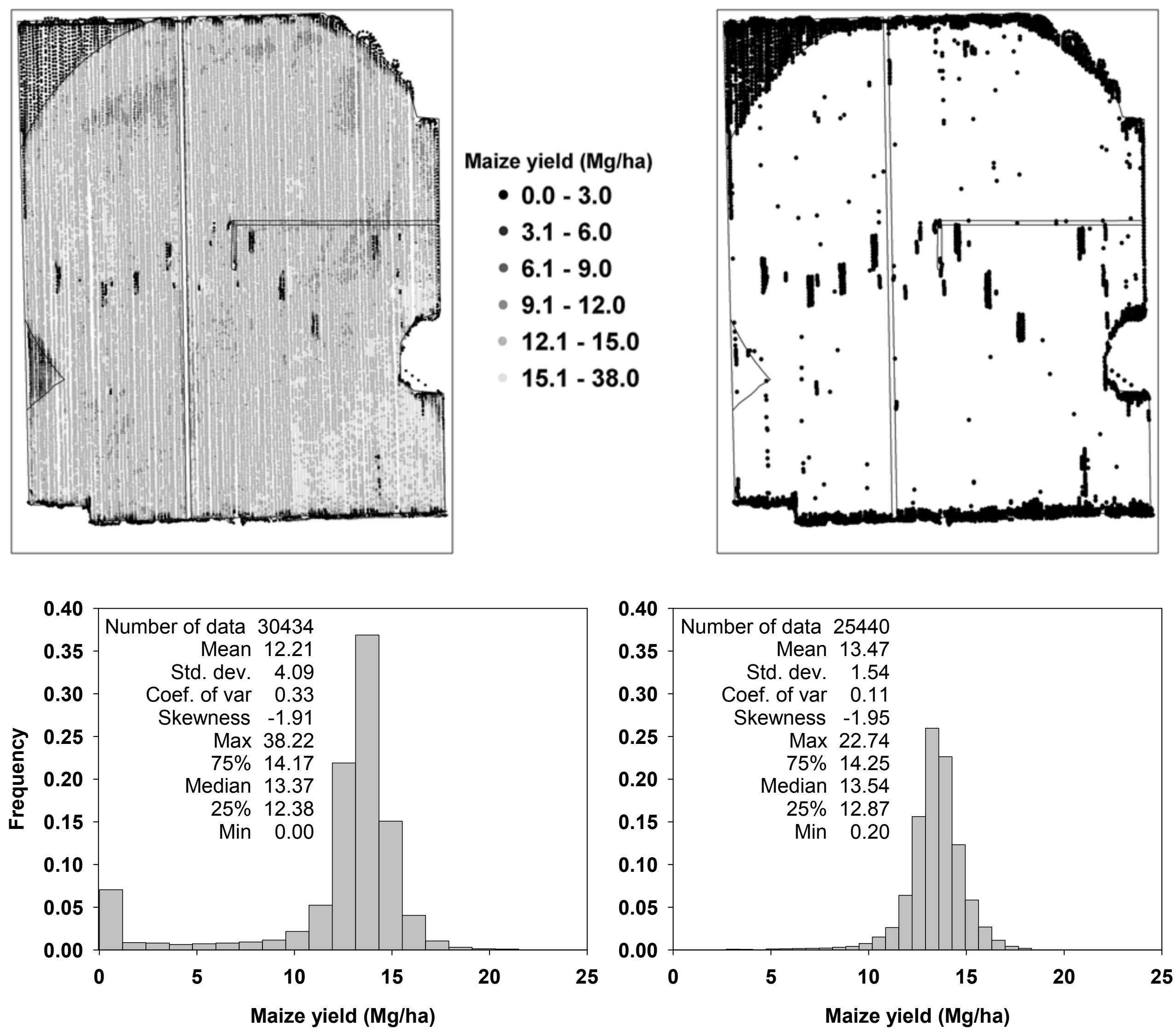


Figure 4. Yield monitor raw data (top left), all data points removed by the yield screening algorithm (top right), and the frequency distribution of corn grain yield before (bottom left) and after screening (bottom right). Yield was expressed in Mg/ha (Mg/ha x 1000 / 62.78 = bu/acre).

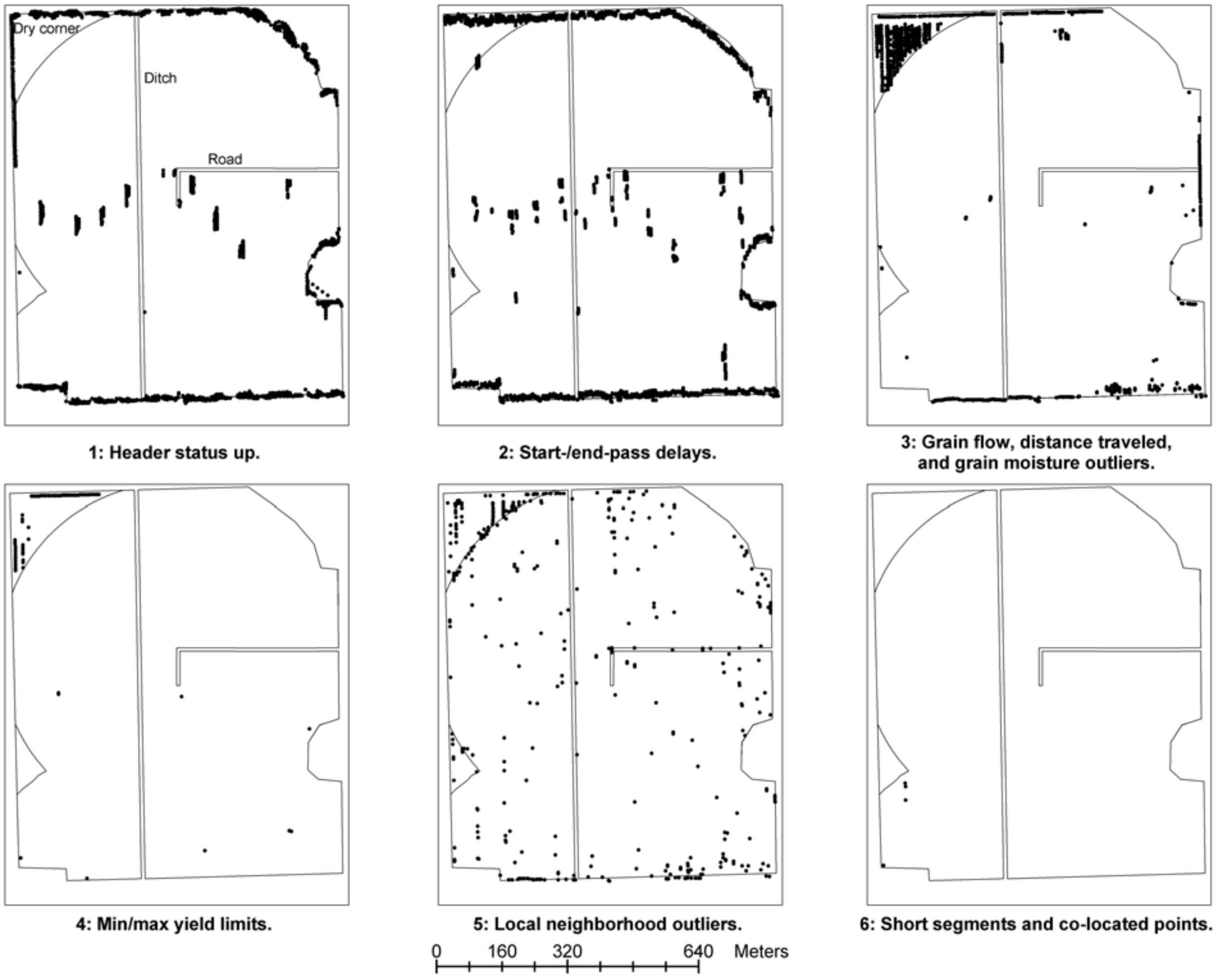


Figure 5. Locations of erroneous yield data points removed in the six screening steps.

Stepwise removal of yield monitor data indicated that 71% of all data removal took place in the first two steps of data screening, whereas the remaining 29% were removed in steps three through six of the screening algorithm (Figure 5). These proportions were similar to observations made at other sites for both corn and soybean (Simbahan et al., 2004). Data removal in the first two steps mostly included zero or very low yields, but few extremely high values were also removed. Erroneous yield points due to header-up status (1941) and start-/end-pass delays (1581) in the yield monitor operation were mostly removed in the headland areas, but also included stop-and-go locations inside the fields.

Step 3 removed 1007 yield data points, mostly located in the northwest field corner and around field edges. Most of these locations were outliers in the grain flow outlier test because corn in that non-irrigated corner suffered from severe drought, resulting in nearly complete crop failure. Step 4 removed data points in the same area as well as some other locations scattered throughout the field based on the empirically defined yield limits. Because many raw data values that would cause outliers in the computed grain yield were removed in the preceding step (3), only 86 additional points were deleted in step 4.

Step 5 removed 371 yield points that were identified as local outliers within the moving local neighborhood (Figure 3). Such outliers included most of the remaining yield points in the dry northwest corner as well as locations that were widely dispersed across the field. The latter included locations at which spikes or sudden drops in yield occurred due to localized management problems or sudden shifts in combine speed. Step 6 deleted six points that had repeated records of yield for the same locations.

References

Adamchuk, V.I., A. Dobermann, and J.L. Ping. 2004. Listening to the story told by yield maps. EC 04-704. http://ianrpubs.unl.edu/fieldcrops/ec704.pdf. University of Nebraska Cooperative Extension, Lincoln,NE.

Arslan, S., and T.S. Colvin. 2002b. Grain yield mapping: yield sensing, yield reconstruction, and errors. Prec. Agric 3:135-154.

Arslan, S., and T.S. Colvin. 2002a. An evaluation of the response of yield monitors and combines to varying yields. Prec. Agric 3:107-122.

Blackmore, B.S., and M. Moore. 1999. Remedial correction of yield map data. Prec. Agric 1:53-66.

Deutsch, C.V., and A.G. Journel. 1998. GSLIB. Geostatistical software library and user's guide. 2 ed. Oxford Univ. Press, New York,Oxford.

Dobermann, A., T. Arkebauer, K.G. Cassman, R.A. Drijber, J.L. Lindquist, J.E. Specht, D.T. Walters, H. Yang, D. Miller, D.L. Binder, G. Teichmeier, R.B. Ferguson, and C.S. Wortmann. 2003a. Understanding corn yield potential in different environments. p. 67-82. *In* L.S. Murphy (ed.) Fluid focus: the third decade. Proceedings of the 2003 Fluid Forum, Vol. 20. Fluid Fertilizer Foundation, Manhattan,KS.

Dobermann, A., J.L. Ping, V.I. Adamchuk, G.C. Simbahan, and R.B. Ferguson. 2003b. Classification of crop yield variability in irrigated production fields. Agron. J. 95:1105-1120.

Doerge, T.A. 1999. Yield map interpretation. J. Prod. Agric. 12:54-61.

Duvick, D.N., and K.G. Cassman. 1999. Post-green revolution trends in yield potential of temperate maize in the North-Central United States. Crop Sci. 39:1622-1630.

Evans, L.T. 1993. Crop evolution, adaptation, and yield. Cambridge University Press, Cambridge, UK.

Kleinjan, J., J. Chang, J. Wilson, D. Humburg, G. Carlson, D. Clay, and D. Long. 2002. Cleaning yield data. http://plantsci.sdstate.edu/precisionfarm/Publications.htm. South Dakota State University, Brookings,SD.

Ping, J.L., and A. Dobermann. 2003. Creating spatially contiguous yield classes for site-specific management. Agron. J. 95:1121-1131.

Shearer, S.A., S.G. Higgins, S.G. McNeill, G.A. Watkins, R.I. Barnhisle, J.C. Doyle, J.H. Leach, and J.P. Fulton. 1997. Data filtering and correction techniques for generating yield maps from multiple-combine harvesting systems. ASAE paper no. 97-1034. St. Joseph, MI, USA.

Shepard, D. 1968. A two-dimensional function for irregularly spaced data. p. 517-524. *In* Proc. 23rd ACM Nat. Conf. Brandon/Systems Press, Princeton, NJ.

Simbahan, G.C., A. Dobermann, and J.L. Ping. 2004. Screening yield monitor data improves grain yield maps. Agron. J. 96:1091-1102.

Specht, J.E., D.J. Hume, and S.V. Kumudini. 1999. Soybean yield potential - a genetic and physiological perspective. Crop Sci. 39:1560-1570.

Stafford, J.V., B. Ambler, R.M. Lark, and J. Catt. 1996. Mapping and interpreting the yield variation in cereal crops. Computers Electronics Agric. 14:101-119.