Processing/Cleaning Corn Silage and Grain Yield Monitor Data for Standardized Yield Maps across Farms, Fields, and Years

Protocol

2-8-2018

Tulsi Kharel, Sheryl Swink, Connor Youngerman, Angel Maresma, Karl Czymmek, and Quirine Ketterings
Nutrient Management Spear Program, Cornell University

Peter Kyveryga
Iowa Soybean Association

John Lory, Theresa A. Musket, and Victoria Hubbard
University of Missouri

Table of Contents

1. Introduction

Yield monitor data, commonly collected by farmers over time can help identify areas of different productivity and within-field yield variability. Linking geospatially recorded long-term yield monitor data with different field measurements and attributes (e.g., soil types, topography, moisture, and soil test values) can aid in identification of factors that influence yield and yield stability. In addition, knowledge of yield limiting factors identified across and within fields will help optimize resource allocation and decision process in precision agriculture and precision conservation.

Calibration of yield monitors during the harvest season is essential for obtaining accurate yield data but even if calibrated properly, the data obtained from the yield monitors still needs to be cleaned. Yield monitor values recorded are estimated based on:

1. Distance (inches or feet), that a combine or chopper travels during data logging time period.
2. Width (inches or feet) harvested during each logging time period.
3. Silage or grain flow (mass) measured by the equipment's flow sensor per logging time period (pounds/second).
4. Moisture content (MC in %) of the harvested mass as measured by a moisture sensor per time period.
5. Logging interval of the yield monitoring system (seconds).

Errors that impact the accuracy of the yield data occur in multiple ways. The distance the combine travels during a time period and the harvest width give the area required for yield calculation. If a combine is not equipped with a harvest swath width sensor, the default used will be the chopper or combine width. This can cause errors when fewer rows are harvested than the width of the chopper or combine. Another source of error is the delay time of grain or silage moving from the chopper or combine head to the flow rate sensor. Flow rate sensors, moisture sensors, and Global Positioning System (GPS) units are located in different places on harvest equipment and since it takes some time for harvested silage or grain to travel to the sensors, adjustments need to be made to the various sensor streams (this is called delay time correction). Each harvest pass (for each sensor) will be affected by this delay correction, independent of whether a new pass starts from one end of the field or from somewhere within the field (in situations where the harvester is paused during harvest). The delay time itself may be variable and related to other factors, such as the speed of the combine or chopper, which may introduce additional sources of error.

Combines and forage choppers are calibrated for a certain velocity range. If the velocities that are recorded fall outside the calibrated range, flow rate and yield values associated with those points are no longer trustworthy and should be removed from the data. Similarly, abrupt changes in velocity affect the flow rate, resulting in erroneous yield calculations for logged data points. Other easily trackable errors are logged data points with zero grain or silage moisture; this may occur as the chopper or combine enters the field or pauses mid-field while the silage or grain flow has not yet reached the moisture sensor.

Last but not least, if the operator does not raise the combine/chopper head after completion of a pass, the pass number will not be updated in the logged dataset. It will require additional time and effort to clean the data. Thus lifting the head of the combine/chopper head while turning in the field is recommended.

Especially for corn silage yield data, use of raw data without proper cleaning can lead to substantial over- and under-prediction of actual yield, depending on the field and harvest conditions.

This manual describes the step-by-step protocol for cleaning individual field datasets and batch processing of harvest data from growers with large numbers of corn silage or grain fields. The protocol developed for whole-farm data cleaning calls for unfiltered or "raw" harvest data files that are downloaded from the yield monitor with corresponding field boundary files (Section 2). These files are read into the Ag Leader Technology Spatial Management System (SMS) software to preview yield maps and reassign any harvest data that might show up in the wrong field. Next, the individual field harvest data are exported in Ag Leader Advanced text file format (Section 3). The yield map files are then imported into Yield Editor (https://www.ars.usda.gov/research/software/download/?softwareid=370) for cleaning (Section 4). Yield Editor is a freely available software developed by the United States Department of Agriculture-Agricultural Research Service (USDA-ARS). The software allows for the use of different 'filters' to remove the errors mentioned above. After cleaning the data in Yield Editor, the cleaned datasets are exported to Microsoft Excel and data points with zero moisture (for grain) and 45% moisture or less (for forage) are deleted. This final step is particularly important for obtaining accurate corn silage yield data (Section 5).

This manual presents an efficient and consistent method for cleaning and handling corn yield monitor datasets for the whole-farm (all corn fields on a farm with corn silage or grain yield monitor data in a given year). It is intended for use by consultants in their own practice and for sharing yield monitor data with researchers working on collaborative research projects. The goal of this manual is to produce quality yield maps for individual fields that can be used to further evaluate areas within a field.
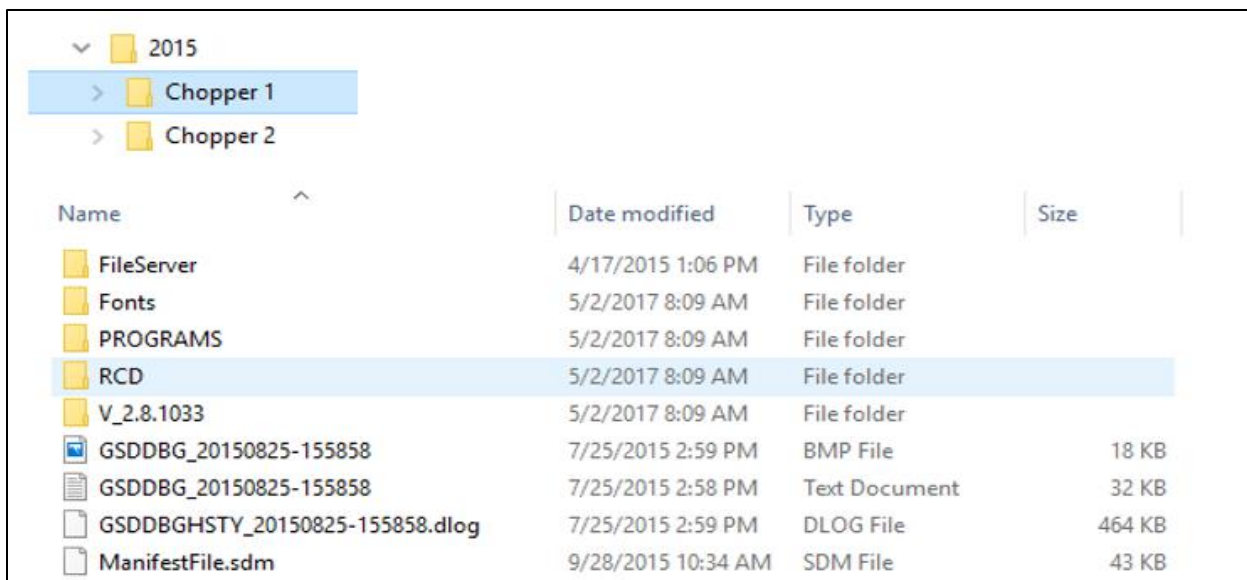
2. Instructions for data sharing

Agricultural crop consultants and agronomists have various means of obtaining yield monitor data from their clients. For producers, consultants, and researchers involved in collaborative projects, yield monitor data can be shared either as raw yield monitor data files or as Ag Leader Advanced format text (*.txt) files exported from the SMS software. To preserve all critical data, the preferred method of data sharing is by downloading raw yield monitor data from the monitor display to a USB drive or, in the case of harvest data already read into SMS (e.g. from earlier seasons), the raw files can be saved to a folder on the computer with the SMS program following the steps described in Section 2.2.

2.1. Exporting original (raw) datasets from yield monitor equipment
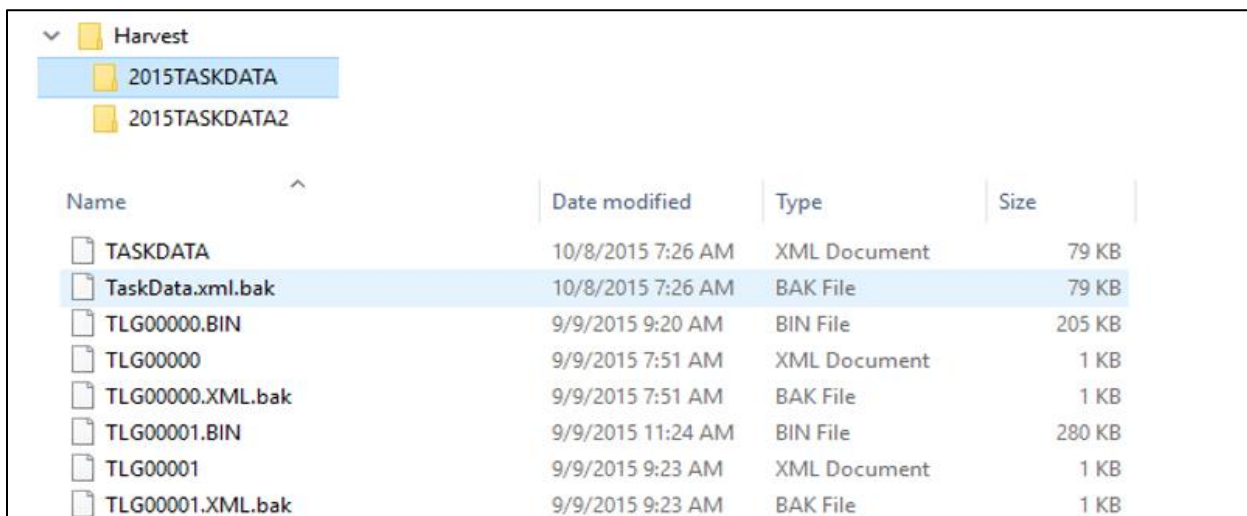
Yield monitors manufactured by different companies have different data file structures. Two examples of raw yield monitor datasets from John Deere and CLAAS equipment are shown in Figures 1 and 2.

For Ag Leader yield monitor systems, copy .agdata files to a USB drive. For other yield monitor systems, make sure to copy all files of the parent folder containing the harvest data.

Figure 1. John Deere GreenStar yield monitor dataset: The 'RCD' folder contains several subfolders and files (with *.fdd, *.fld extensions) that are required to read harvest data. Copy the entire parent folder ("Chopper 1" in this example) to a USB drive.

Figure 2. CLAAS dataset: Multiple files within the task data folder are required to read and process these datasets. Copy folder "Harvest" to a USB drive.

2.2. Saving raw yield monitor data files from SMS and other software

If yield monitor harvest data have already been read into the AgLeader SMS, they can be saved to a folder on the computer and then to a USB drive, or shared by email:

1. In the SMS project that contains the data, go to Tools and click on File Viewer.
2. On the left will be "Archived Files" with the display name, serial number, year, operation, and file names.
3. Highlight the year or operation to be saved and choose "Save Files to…" on the right and save the files in a designated folder onto a computer.

Repeat the above steps until all of the desired files are saved to the folder on a computer. Additionally, raw yield monitor data from APEX software can be exported by:

1. Selecting 'Unload' tab and then 'Manage Archive' will display all the archived dataset in Apex.
2. After selecting the appropriate dataset from the archive, click 'Export'. This brings a browse option where you can select your preferred location/folder to save exported data.

2.3. Obtain field boundary files from collaborators

It is assumed that consultants or growers who use SMS Basic or Advanced software have already established field boundary files and have "frozen" them before reading harvest data files to SMS. Field boundary files for all fields can be exported from SMS as a single shape file. For research projects, the fixed/frozen field boundaries are helpful to facilitate pre-processing of raw yield monitor data. This will allow use of SMS commands to automatically send mislabeled yield loads and splits of loads containing yield data from more than one field to the geospatially correct field boundaries before exporting.

3. Importing/reading yield monitor data

Yield monitor data come in various formats. A software that can read a wide variety of yield monitor data formats is required for processing raw yield monitor data from farms with different equipment or systems. The Ag Leader Advanced format data structure is required by the Yield Editor software for cleaning. Ag Leader's SMS is a widely used and very comprehensive commercial software package that works well for this purpose. An unregistered trial version of SMS (free download) with limited functionality can read and process data from a wide number of monitors. In addition, SMS software contains functions for reassigning incorrectly labeled harvest data to the geospatially correct field and for splitting loads harvested from more than one field to the correct field locations. Once all harvest data are assigned to their correct geospatial field boundaries in SMS, it is easy to export the yield data for each field in the Ag Leader advanced file format required by the Yield Editor cleaning software. A list of equipment from which SMS software can import raw yield monitor data is shown in Figure 3.
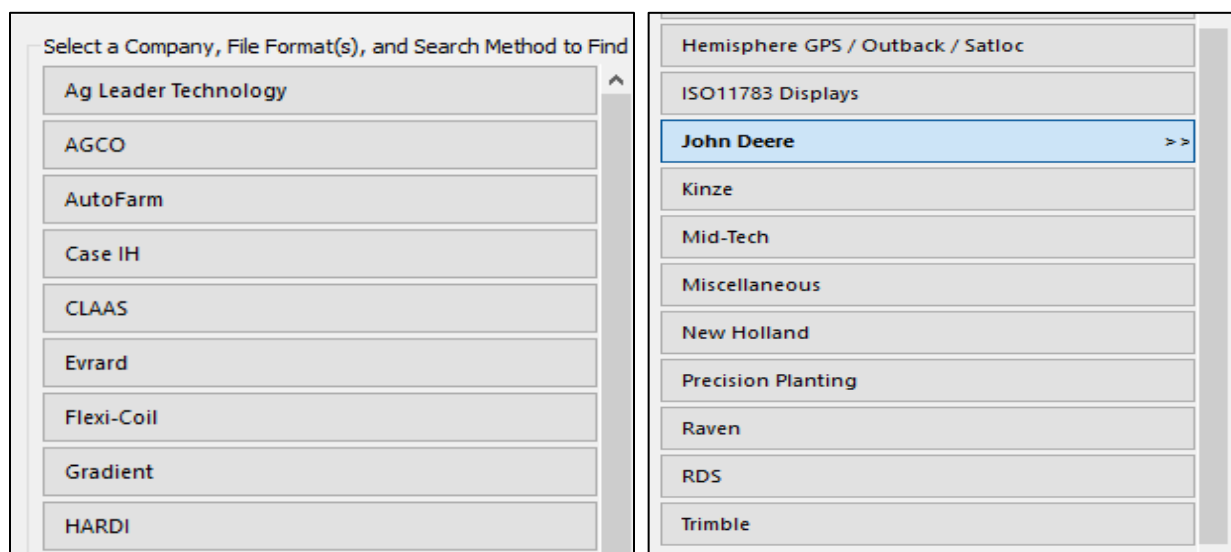
Figure 3. List of equipment from which SMS software can import raw yield monitor data.

3.1. Setting project and boundary files in SMS

The following steps are adapted from the SMS manual. These steps describe how to set up a project along with the farm and grower information:

- To set up a 'Project' in SMS, click 'File → Projects→ Add New Project → Create New Project → Create Empty Project' (SMS Basic users are limited to 5 projects, but many growers, farms, and fields can be included in one project).
- Give a name to the project.
- If the desired project is already established, go to File menu, then open the Project to which you want to add a grower, farm, or field.
- In the pull down Tools menu, select Management Item Editor.
- Select Grower button (if adding new grower) on left of window, click on Add button (right side of window), then name the grower in the pop-up window.
- To add a farm, select Farm button on left side of window, click on Add button, then use the pull down menu of growers to select which grower the farm belongs to.
- To add a field, select Field button on left, click on the Add button, then use the pull down menus for Grower and Farm and select the grower and farm the field belongs to if they are not already showing by default.
- Note: Project, Grower, and Farm do not appear in the management tree until data has been loaded for the new Project.
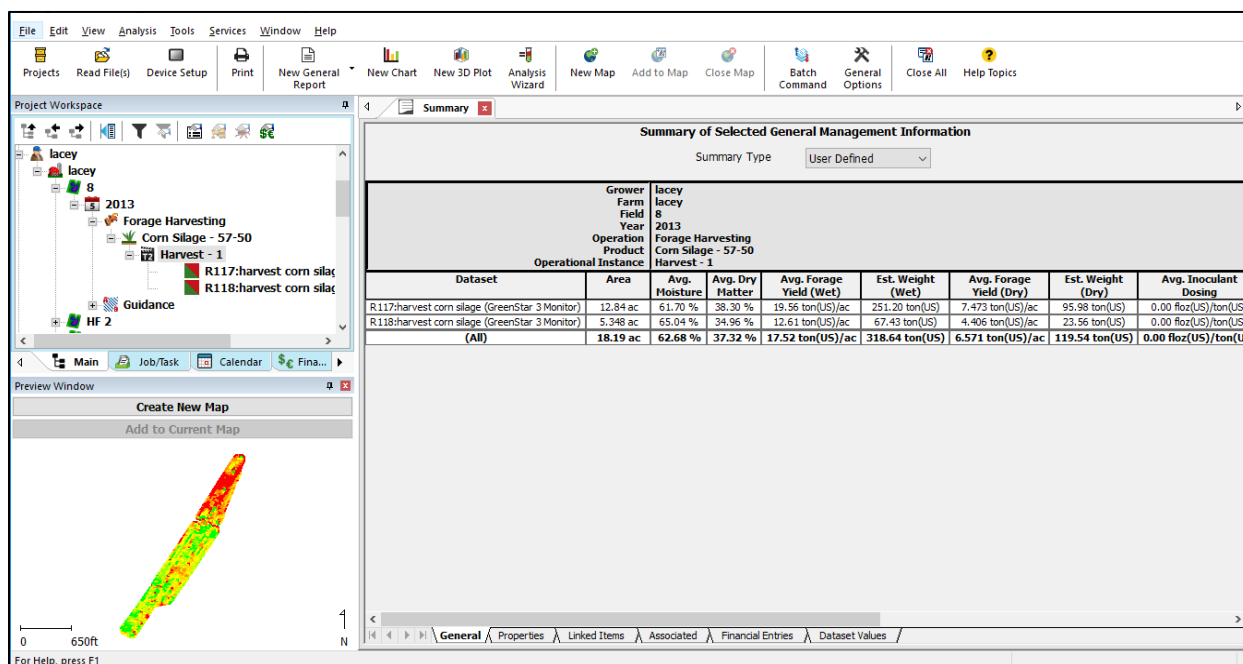
Figure 4. SMS data organization: Grower, Farm, Field, Year, Operation, Product, Operational Instance and dataset as shown in summary of selected general management information.

If a set of frozen or fixed boundaries ("shape file" set) for all the grower or farm's fields is obtained from a collaborating agricultural consultant or grower, add the yield monitor dataset's grower, farm, and at least one field to the project and bring that shape file of frozen field boundaries into SMS **before** reading in the first raw yield monitor harvest dataset from that grower. This will allow users to sort harvest load data into geo-referenced fixed field boundaries in case any fields were mislabeled at harvest.

To import a set of frozen or fixed field boundaries for a farm:

After obtaining a set of frozen field boundaries exported as a shape file set (must contain .shp, .shx, and .dbf extension files, not just .shp file), from the farmer or agricultural consultant and setting up everything for the farm as described in Section 3.1 of this manual:

- Open the appropriate Project and Click on the Read Files pulldown menu in the main menu at top of window.
- Select "Import a File from a Generic Source" then click on "Start Importing Generic Files at bottom of window.
- Select "Shape File" in next window that appears and then click on the "Select File(s) to Import" at bottom of window.
- Browse to the fixed border shape file you want to import in the window that appears. Select it and click on "Open".
- In the "Projection" window that appears, scroll down and adjust "Datum" selection to WGS 84 if that is not already showing. For "Projection", scroll down the list and select Projection used by equipment – e.g., "18N" (for most of New York; maps of U.S. UTM zones can be found by Googling "United States UTM Zones") or "Lat Lon" if that was used. Hit Finish. If everything is correct, the set of fields will be seen in the Import Preview window.
- If okay, click on "Next".
- In the "Select Import Type" window that opens next, set Import Data Type as "Boundary" from that pull down list.
- Check box next to "Set as Frozen Field Boundary", then click "Next".
- In the "Assign Columns to Import" window, first check that the sample row data in the lower table looks correct for the farm or field. If it looks good, click on "FIELD" in the File Column list (that row should then be highlighted in black), then click "Assign column" button.
- In Columns Settings: FIELD window that pops up, click on "Use as Management" button. Then for Management type, choose "Field." Click "Okay" – returns to previous window which should now show "Field" under Column Mapping, if so, click on "Next".
- In the Import Clipping Options window that opens, select "No Clipping – Import Entire File." Click on "Next".
- In Management Selection window, choose the correct grower in Select grower pulldown list. If everything has previously been set up, it may self-populate the Farm, Field, Year, Operation, etc., fields. Check to see that these are correct (if any one of the fields for the farm is selected, all fixed field boundaries in the boundary file will be imported and set up as individual field listings).
- Click on "Next" – will then see Import Review Window – this will indicate the Number of records to be imported (# of frozen field boundaries in the import shape file).
- Click on "Finish".
- Wait while all the data is imported. When done, the new grower and, if expanded, the farm and all the fields with imported frozen boundaries will appear in the Management Tree.

3.2. Import raw yield monitor dataset

In the Management Tree, click on appropriate grower or farm for which data are being imported. Once its name is highlighted in the management tree:

- Click on "Read File(s)" in main menu bar.
- Select "Read File(s) from a Supported Field Display or Monitor," then click on "Start Reading Field Display Files…" at bottom of the pop-up window.
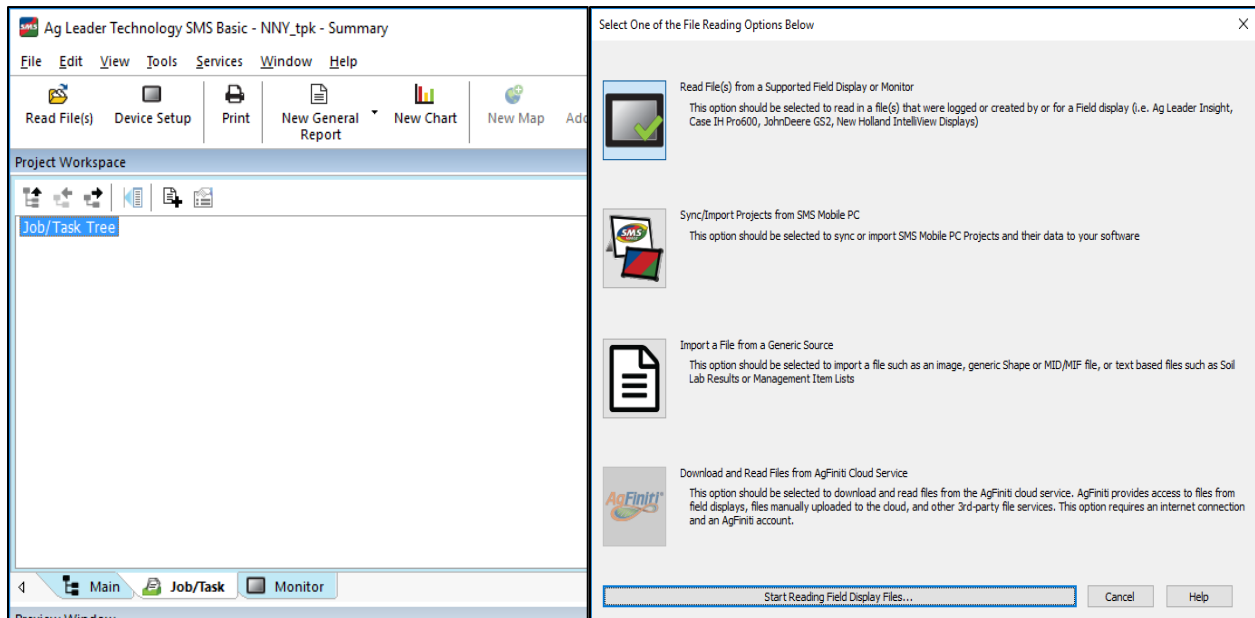
Figure 5: File reading options in Ag Leader SMS software.

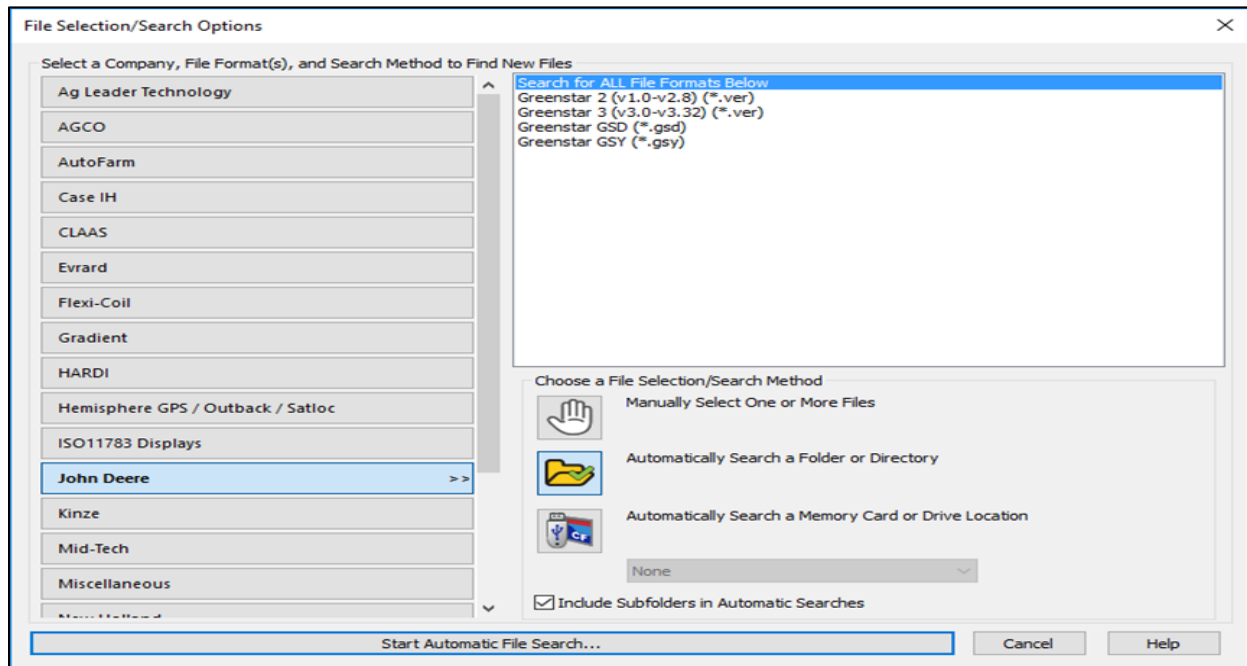Figure 6. SMS file selection window.

- Select Equipment type/brand (e.g., CLAAS or John Deere) from the list on the left side of the window, then select "Search for ALL File Formats Below" in the top right box and choose a File Selection/Search Method ("Automatically Search a Folder or Directory" is a good one in most situations). Click on "Start Automatic File Search…" at the bottom of window.

- Navigate to the file folder with the raw yield monitor dataset(s) for download. It could be a folder with files from more than one chopper/combine or harvest day. Once you have clicked on the folder you want to select, click "OK".
- The selected data files will appear in the search window. Click on "Edit" to see details identifying the files selected and to make any adjustments. Under "File Processing Options" in the lower part of this window, "Spatial Sort Data into Fields" and "Automatically Merge/Split Data" should be turned "ON" for growers or farms that already have frozen field boundaries in the project. To change the ON/OFF setting, click the "Edit…" button to open File Processing Options window and check or uncheck the boxes for "Spatial Sort Loads into Fields while Processing Files" and "Automatically Merge/Split Data" ("Merge All Sensor Data into a Single Dataset and Operation" can be selected). "Secondary File Copy" to create a zipped backup of raw files probably only needs to be enabled if downloading directly from a harvester or monitor chip. After selecting desired File Processing Options, click "OK".
- Turn the "Auto-Create Management Items and Skip Processing Settings" button to off and then click "Start Processing Selected Files…".
- In next window "Set Management Items for New Data", under "Set Grower Name" select the existing grower name that you created during project setup. Under "Set Farm Name" select the farm you created during project setup. Click "OK".
- In next pop-up window, select "Processing Settings" tab. It will display the default processing settings applied by the yield monitor system. An example is shown in Figure 7 where the minimum (min) (0 ton/acre) and maximum (max) (200 ton/acre) filters for corn silage yield (labelled 'P' on the Physical Product and 'Forage Harvesting' under Mode for this dataset) were applied to import data. The range (0 to 200 ton/acre) is good enough to import most of the data. The boxes under "Operation Settings" for manual swath and manual moisture should be unchecked. Under "Equipment Settings", make sure to keep zero values for the Start Delay, Stop Delay and Flow Shift to make sure no data points will be removed during import. Click on the dropdown menu for both 'Physical Product' and 'Mode' and repeat the process for each product if more than one product or Mode is listed there. Click "Next" to process and read raw yield monitor data with these settings.

Figure 7. Processing settings applied to SMS during raw yield monitor data import.

3.3. Check individual field and load harvest files for misplaced, extra or missing areas

When the raw yield monitor data have been imported, all farm fields for which there were harvest data will now have a set of load files (Dataset) listed under the harvest year and operation (e.g., "Forage Harvesting") subfolder in the Management Tree. Fields that are not included in the harvest data set (e.g., planted to a different crop or not harvested in that data set) will not be shown for that year.

Expand fields from a given farm in the Management Tree to identify those that have new harvest data and check preview window maps of each load file to find any with data from other fields:

- Click on the Farm name (barn with silo image one) in Management Tree.
- In the menu bar just above the Management tree, click on arrow pointing to the right to expand all farm fields and see if new harvest year data was downloaded.
- For each field with newly downloaded harvest data, click on the operational instance (Harvest – 1 in the example above) to see a preview map of combined load files and note if there appears to be data from more than one field. Click on "Boundary" or "Borders" in the Management Tree for that field to compare with the shape of the harvest data preview map.
- If more than one load file (dataset) is shown under operational instance, click on each file to see which part of the field is in that load. If there is more than one field, identify which load contains the extra data.
- Make a note of any load files that appear to have data from more than one field. These load files will need to be split according to the data's geo-spatial information (see Figure 8).
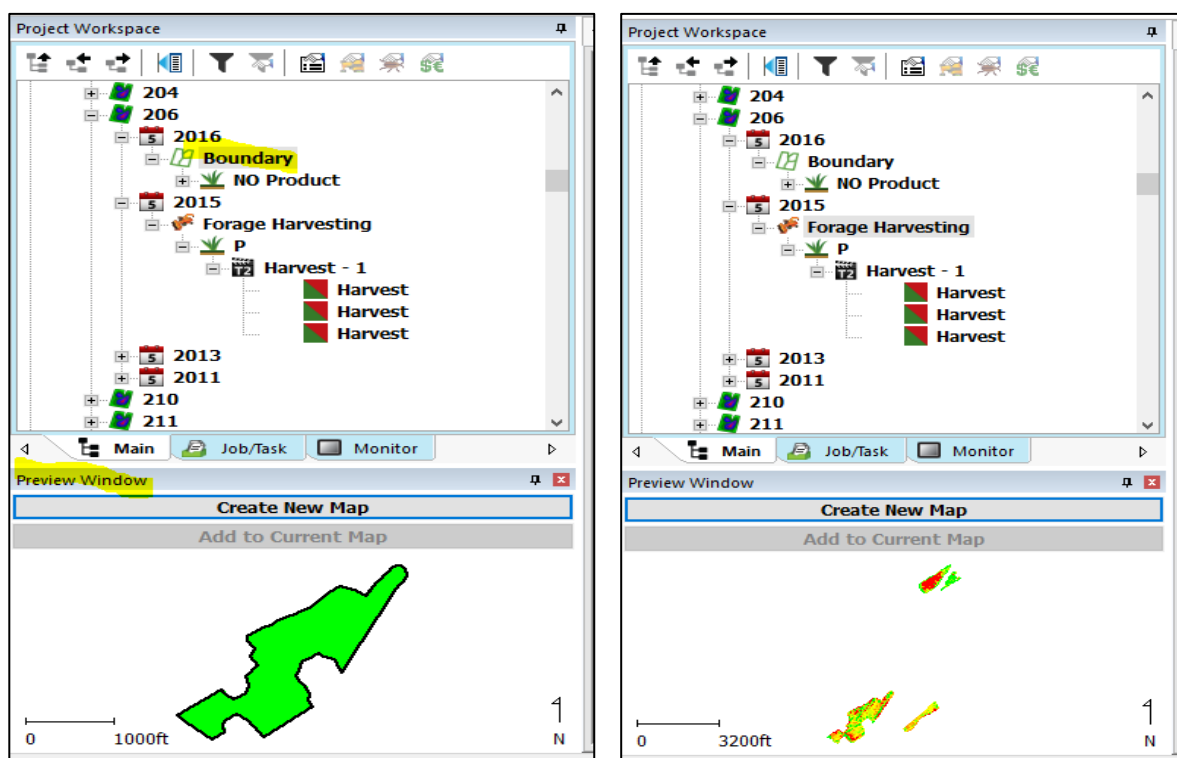

Figure 8. Forage harvest in 2015 from Field 206 contains data from more than one field. Compare the field boundary for this field (2016 in first picture) to harvest data (2015 in second picture) in the preview window.

Split loads that contain data from more than one field:

- In the Management Tree, select the load file with data from more than one field. Only one load can be split at a time. Click on the right mouse button. A menu of options will appear. Select the Split Loads / Region option.
- The software will check the data contained in the load and compare its location to the boundaries that you have frozen or set. If a load partially falls within a boundary, the points that are within the boundary will be split and a new load with the same name will be added to the field that corresponds to the data's geospatial information.
- Once split is completed, the software will generate a report that identifies what field load was split and which fields received parts of that load's data.
- Any loads that were split, even the original load, are now spatially-based only and do not have monitor summaries. These files now cannot be calibrated by new calibration files. Note: If these files containing data for the split load are re-processed, the load will be added back into the system like it was before it was split out.
- By default, the save as screen will save the split-data results report (showing which fields have split data) to the Documents>Ag Leader Technology>SMS>Export folder. When doing the Save as, name the text file using the farm code and field ID code.
- If any load files do not match the field they were sent to during the download, use the 'Sort Load into Field' command to send the data to the correct geo-referenced field location.

Split and sort load can be done in batch mode if the registered version of the SMS advanced software is available. This will save time when working with a large number of fields and farms. This process will split and sort each harvest dataset if they were outside of the fixed field boundary. At the end of this process it will generate a report for all the fields. This report contains information about the data points moved during this process and where they moved. Save as option can be used to save this report within appropriate folder for later review.

- From 'Tool' dropdown menu click 'Batch Command Utility'.
- In next window, 'Batch Command Utility', select Split Loads' and click 'OK'.
- In the 'Select Data Filters' window that appears next, select 'ALL' for all the filter variables and click 'Next'.
- Click 'Finish'

3.4. Exporting Ag Leader Advanced format text files

Exporting raw yield monitor data can be done individually for each field or in batches for all fields.

Exporting Individual fields:
Once all harvest data are assigned to the correct fields, select the operational instance ("Harvest – 1" in the above example) of the desired field in the Management Tree, right click, and from the bottom of the menu that appears, choose "Export…":

- For Export Method select "Export to a Generic File Format" and click on Start Generic File Export Process…
- For Generic File Format to Export to… Choose "Generic" and "Ag Leader Advanced Export (*.txt) and click on "Export to Selected File Format…".
- In the Save As screen, name the file appropriately.

Exporting multiple fields using batch export (option is only available with the registered SMS version):
- From 'Tool' dropdown menu click 'Batch Command Utility'.
- In next window 'Batch Command Utility' select 'Export' and click 'OK'.
- From 'Select Data Filters' window that appears, choose the desired filter variables. For example, all the fields that contain corn silage (or corn grain) harvests were exported separately for each year from a grower/farm. Select grower, farm, year, operation, and product in 'Select Data Filters' window.
- Select 'Year' from the dropdown menu for 'Merge Filter Results Datasets by...' and click 'Next'.
- In next window 'Edit Data Filter Results', click 'Finish'.
- In the next window 'Select an Export Method...' select 'Export to a Generic File Format' and click 'Start Generic File Export Process...'.
- In the next window, choose "Generic" and "Ag Leader Advanced Export (*.txt) and click on 'Export to Selected File Format...'.
- In the 'Browse for Folder' window, select or make a new folder to save exported dataset and click 'OK'. Once done, all the exported dataset will be in the specified folder.

Using MS Excel or a similar program, open and check Ag Leader Advanced Export file(s) for the harvest data to be cleaned:
- Double check to make sure that the exported data content matches the Ag Leader Advanced Format column order and units (Table 1) and that there are no issues. Some problems were discovered in this required data structure when shapefiles were used to read into SMS. We found a pass (transect) numbering problem where each data point was assigned a new pass number in a dataset exported from a KRONE forage harvester. Single pass number for entire dataset and missing GPS time are some other examples of what might be seen.
- If the column info as shown in first 10 columns of Table 1 are correct, these files are ready to run in Yield Editor. If something looks suspicious such as some units do not make sense or pass number looks odd, revisit the SMS dataset and correct the units there before re-exporting in Ag Leader Advanced format.

Table 1. Required column sequence for Yield Editor (Ag Leader Advanced Format).

Data in column order	Description
1. Longitude	Decimal degrees
2. Latitude	Decimal degrees
3. Flow	Pounds per second
4. GPS Time	Seconds
5. Logged Interval	Seconds
6. Distance	Inches
7. Swath	Inches
8. Moisture	Percent (wet basis)
9. Head Status	1 = harvesting, 0 = not harvesting
10. Pass Number	Generally +1 for each header up/down
11. Any other information	

Note: The file must have 11 columns at a minimum. The content of columns 11 and higher will not affect the reading and processing of the data but the 11th column is needed.
- Check the coding in the Header Status column to determine if it is different from the ones for Yield Editor (1=down, harvesting; 0=up, not harvesting).

- By default, any odd number in the header column means 'header down or harvesting' and an even number means 'header up'. We found '33' both in several CLAAS and John Deere datasets. Yield Editor accepts these values as 1.

4. Clean yield data using Yield Editor

The following protocol uses Yield Editor, freely available software developed by USDA-ARS, to clean raw yield monitor data (https://www.ars.usda.gov/research/software/download/?softwareid=370). As explained in the introduction, raw yield monitor data should be cleaned before use in analyses. Multiple errors that can affect accuracy of yield monitor data are overlap of harvest passes, delay that occurs as material travels from header to sensor, abrupt changes in machine speed, and ramping effects at the beginning and end of passes. Yield Editor was developed to address the errors listed above. A manual that comes with the software ('Open .pdf Manual' under 'Load/Import File' tab of Yield Editor) should help for better understanding this tool and the cleaning process. Yield Editor facilitates cleaning of yield monitor data using different 'filters.' The descriptions of the filters here are from Sudduth et al. (2012; courtesy of USDA ARS Cropping Systems and Water Quality Unit, Columbia MO):

- **Grain flow delay (DELAY)** – This corrects for the transport time between the location where the crop is harvested and the location where flow rate is sensed. Because delay time can be affected by design of harvesting equipment, speed, ground slope, load, and other factors, it is best to determine the DELAY parameter value for the harvest conditions within a particular field.
- **Start pass delay (START) and End pass delay (END)** – The START filter removes observations when entering each transect and the END filter removes data when exiting each transect. These filters eliminate points with low and unreliable yield estimates that are obtained as the harvester fills at the beginning and then empties at the end of each transect. If the START and END delay parameters are too small, the points with erroneously low yields that remain in the dataset will deflate the yield estimate and exaggerate the yield standard deviation.
- **Maximum velocity (MAXV)** – This filter removes data points collected at velocities greater than the specified limit. It is particularly useful for removing partial swaths and observations where the operator may have kept the separator running and the header down without harvesting.
- **Minimum velocity (MINV)** – The MINV filter removes points collected at velocities less than the specified limit. It assists in removing unrealistically high instantaneous spikes in the calculated yield that can occur as velocity approaches zero.
- **Smooth velocity (SMV)** – This filter eliminates data where rapid velocity changes have occurred. The SMV parameter represents an allowable ratio of velocities from one point to the next point along a transect.
- **Minimum swath (MINS)** – This filter removes data points with swath width readings smaller than the minimum such as in point rows or in partial swaths. The effectiveness of the MINS filter depends on the operator manually adjusting the swath indicator on the yield monitor; an action that the operator may not be willing or remember to do.
- **Maximum yield (MAXY) and minimum yield (MINY)** – These filters set yield thresholds above (MAXY) and below (MINY) which points will be deleted.
- **Standard deviation of yield (STDY)** – The STDY filter removes yield data that are more than a certain number of standard deviations from the field mean.

- **Position (POS)** – This filter removes positional "flyers," consisting of single points or data segments that lie outside the boundaries of the field of interest. If data points are within the field, but exhibit positional error, they can be deleted using the MAN filter described below.
- **Manual (MAN)** – Due to the complex nature of the problems that can occur with equipment and sensors in the field, yield data will often include some data points that are clearly in error, yet are not easily captured by any of the above filters. The manual filter allows the user to select individual points, transects or regions for removal, addressing the case where other filters fail to recognize errors that are visual.

4.1. Initial test cleaning to determine farm specific settings

In most cases, filter values used for a particular crop, year and harvester combination appear to be very similar across fields within a farm. To find the most common filter values for the specific crop (such as corn silage) within a farm and particular year, clean about 10 fields interactively (use Yield Editor's Automated Yield Cleaning Editor (AYCE) followed by Filter Selections to adjust AYCE results) or manually. Note the impact of AYCE "Bitmap based overlap filter" and "Localized STD filter" on the farm's data – did this just remove visually obvious errors or too many points throughout the field? Record filter values of each field after a thorough cleaning. Compare the filter values for each of the cleaned fields (around 10 fields) and decide if values are close enough to select a common one for each filter to use in batch processing or initial import configuration for that grower/farm's fields.

For initial testing, select large fields with large spatial variation in yield. Finding appropriate filter values in smaller fields or a field with low spatial variability is more difficult.

If more than one type/brand of chopper or combine was used across the farm, filter values will likely be different for each of the harvesting systems. An ideal condition for using filters is when a single person operates the harvester and uses the same equipment (harvester) at a similar speed across all fields.

4.1.1. Data import and cleaning

- Start Yield Editor (Version 2.0.7, USDA-ARS) software. Under "Load/Import File" tab, select the type of crop from the table "select grain type and properties." Silage corn is not in the table of crops when Yield Editor is downloaded. To add it and its properties, click on 'Edit table?' and enter following information in the space under "Selected Grain Type and Properties":
  > Crop type: Corn Silage, Density: 2000, Market Moisture: 65
- Click on 'add' and this new information will be added to the table under "Select Grain Type and Properties."
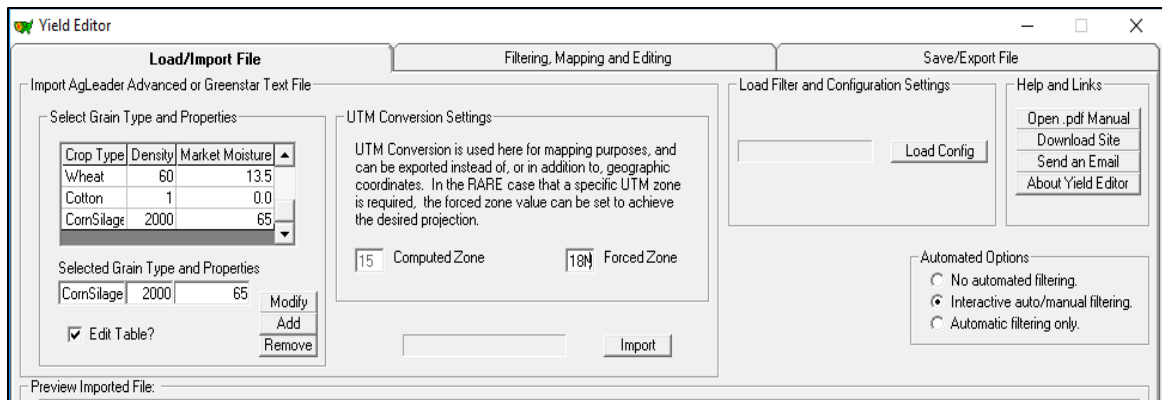
Figure 9. Example of adding corn silage information.

- Go to the tab 'Save/Export File'. Check all items as shown in Figure 10 before saving as default configuration. Save the settings as default under 'save filter and configuration settings'. Check 'save as default configuration?' and click on 'Save Config'. This will ensure that all the data files will have the same settings applied to them.
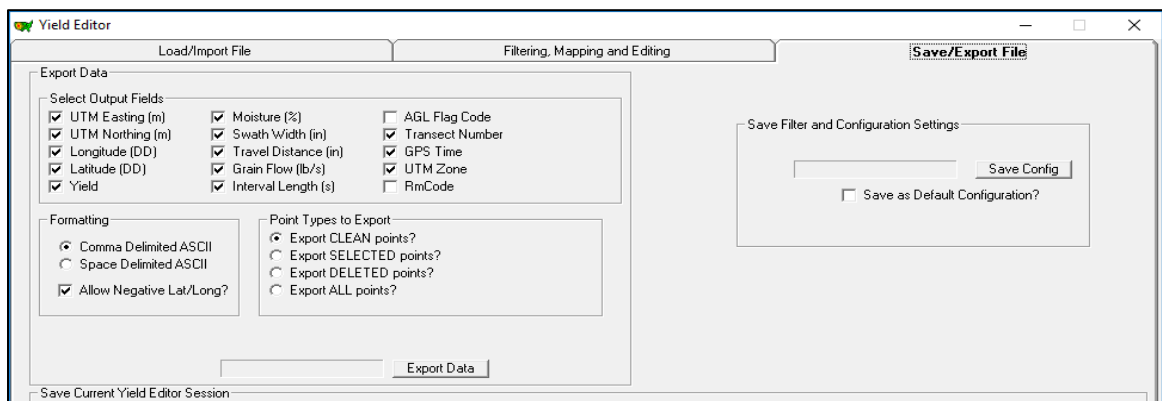

Figure 10. Data columns that will be exported from Yield Editor are selected.

- Under the 'Load/Import File' tab, click Import, browse to the folder and select the *.csv or *.txt data file in AgLeader Advanced format for the field to be cleaned and click on "Open".
- Confirm Import Settings window pops up asking "Continue import with current file format, grain and projection settings?" Confirm that the settings match the crop type being imported (e.g., silage or corn grain) then click on "OK".
- In the 'Preview Import File' box under the 'Load/Import File' tab, some of the rows and columns of the dataset are displayed.
- Check the "UTM Conversion Settings" tab. Add information about appropriate UTM zone of the imported field only if it does not agree with the computed one. The software calculates the UTM zone based on the first values of Latitude and Longitude in the dataset. If a field falls on the border of two zones, manual correction may be necessary.
- Click on the 'Filtering, Mapping and Editing' tab. It will display a map of data points. Look for the 'Yield statistics' table (lower left in window). Pay attention to the number of points (N) for raw and clean data. If some data are already removed during import, look at the 'Filter Selection' section. What filter(s) removed these points? Uncheck everything under 'Filter Selection' section and click 'Apply Filters'. This will bring back all the data points.
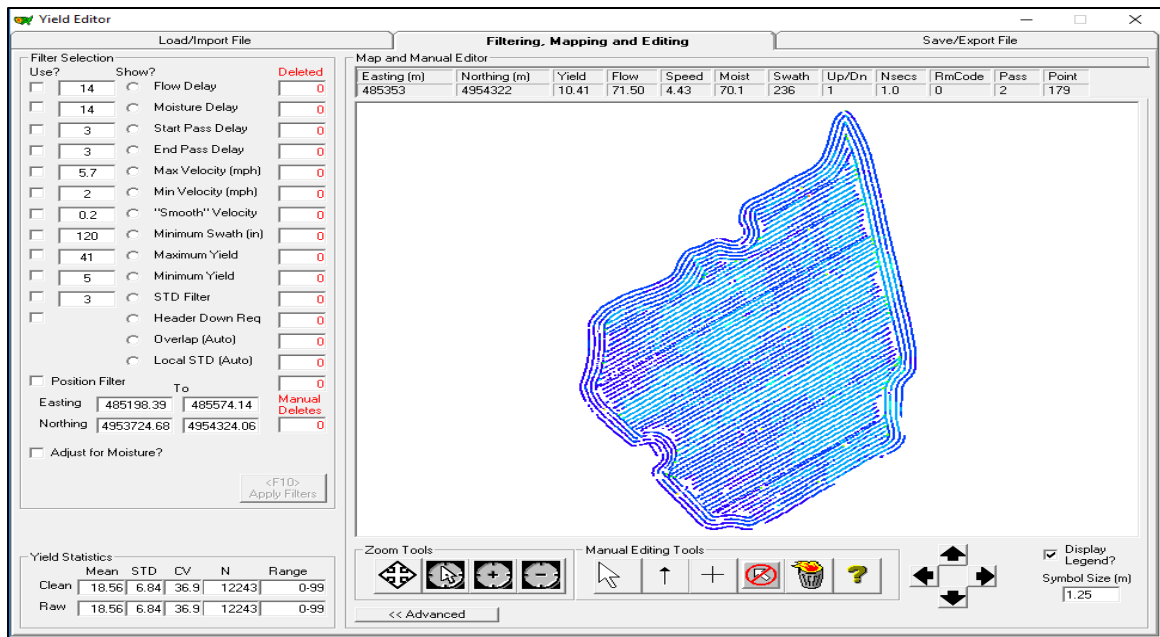
Figure 11. Uncheck every filter and hit 'Apply Filters' to bring back all the data points to the display.

4.1.1.1 Run automated AYCE on a number of fields

Follow these steps for at least 10 fields that came from the same crop, year, and harvester combination:

- Once all the data points appear in the map window, go to a separate window called "AYCE – Automated Yield Cleaning Expert". Select all of the checkboxes in AYCE window as shown in Figure 12 and hit 'Run Auto Filters' to run AYCE with default settings.
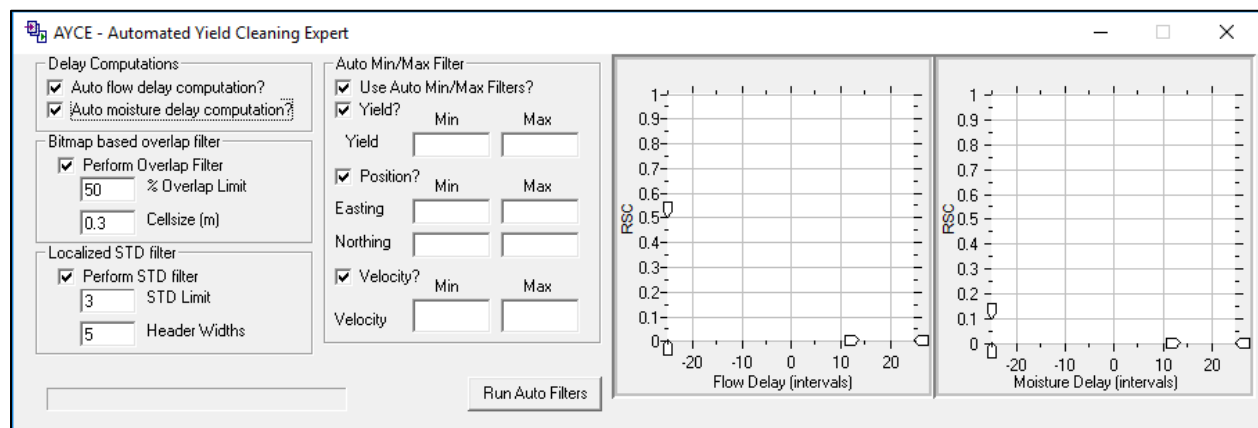
Figure 12. Default settings for AYCE.

The sequence of filters applied when running AYCE is shown in Figure 13. AYCE first determines the position of the field, next it calculates the flow and moisture delay values for the field, then it determines min and max values for yield and velocity. Next, it computes swath (pass) overlap based on given parameters (% overlap limit and cell size) and removes any pass points that show greater overlap with the previous pass. Finally, it calculates local standard deviation and removes data points with greater than the

parameter supplied for this filter. To decide the min and max yield values, the software extends ¼ (0.25) of the inter quantile range to the upper (99th) and lower (10th) quantiles of the dataset (Sudduth et al., 2012). To decide the min and max velocity values, the software extends 0.18 of the inter quantile range to the upper (99th) and lower (8th) quantiles of the dataset (Scott Drummond, USDA-ARS, personal communication).

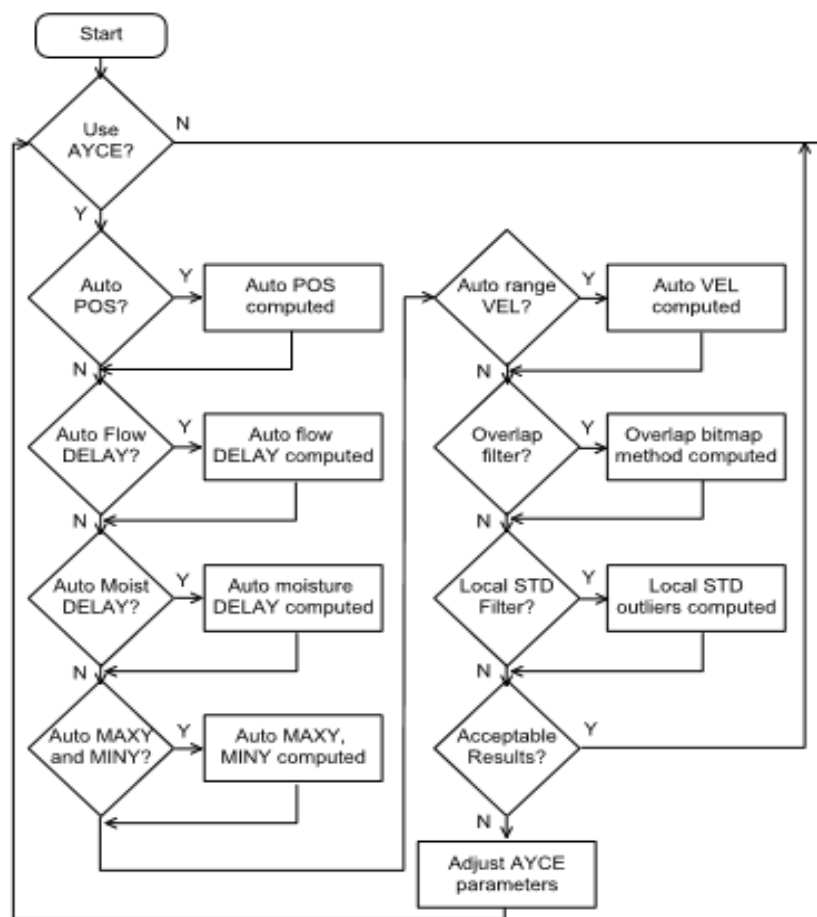Figure 13. Sequence of filters applied during AYCE (Source: Yield Editor 2.0 Manual, courtesy of USDA-ARS).

To choose the delay interval values, the phase correlation method (Lee et al., 2012) is used, where a random noise is added onto the dataset and iterated 10 times to estimate delay value. The 'relative spatial consistency' or RSC obtained with this method are plotted relative to the flow or moisture delay intervals (Figure 14). The flow or moisture delay interval is chosen based on the maximum RSC value. The overlap filter uses the bitmap method (Han et al., 1997) to identify if the current harvest point is overlapped with a previously harvested pass. It counts the number of smaller grid cells that were harvested in the previous pass relative to the total number of grids within a current harvest polygon (each time step) to decide overlap status. The local standard deviation filter uses a square grid method to calculate the mean and standard deviation within each grid, and removes yield values greater than specified standard deviation (Sudduth et al., 2012). Grid size is determined by the number of header widths specified by the user.

An example of AYCE output is shown below in Figure 14. The flow delay and moisture delay calculated by AYCE are 3 and 5 seconds, respectively, which have been applied to the dataset (the flow delay checkbox is selected in 'Filter Selection' and the number of points removed is in red under 'Deleted'). The number of moisture delay points removed will not appear until the "Adjust for Moisture" and "Sensor Based?" box is checked. The AYCE looks for the RSC of different delay intervals and finds one with the highest consistency. If the RSC values were not consistent during 10 repetitions, those charts are marked with yellow (warning: low consistency) or red (no consistency) backgrounds.
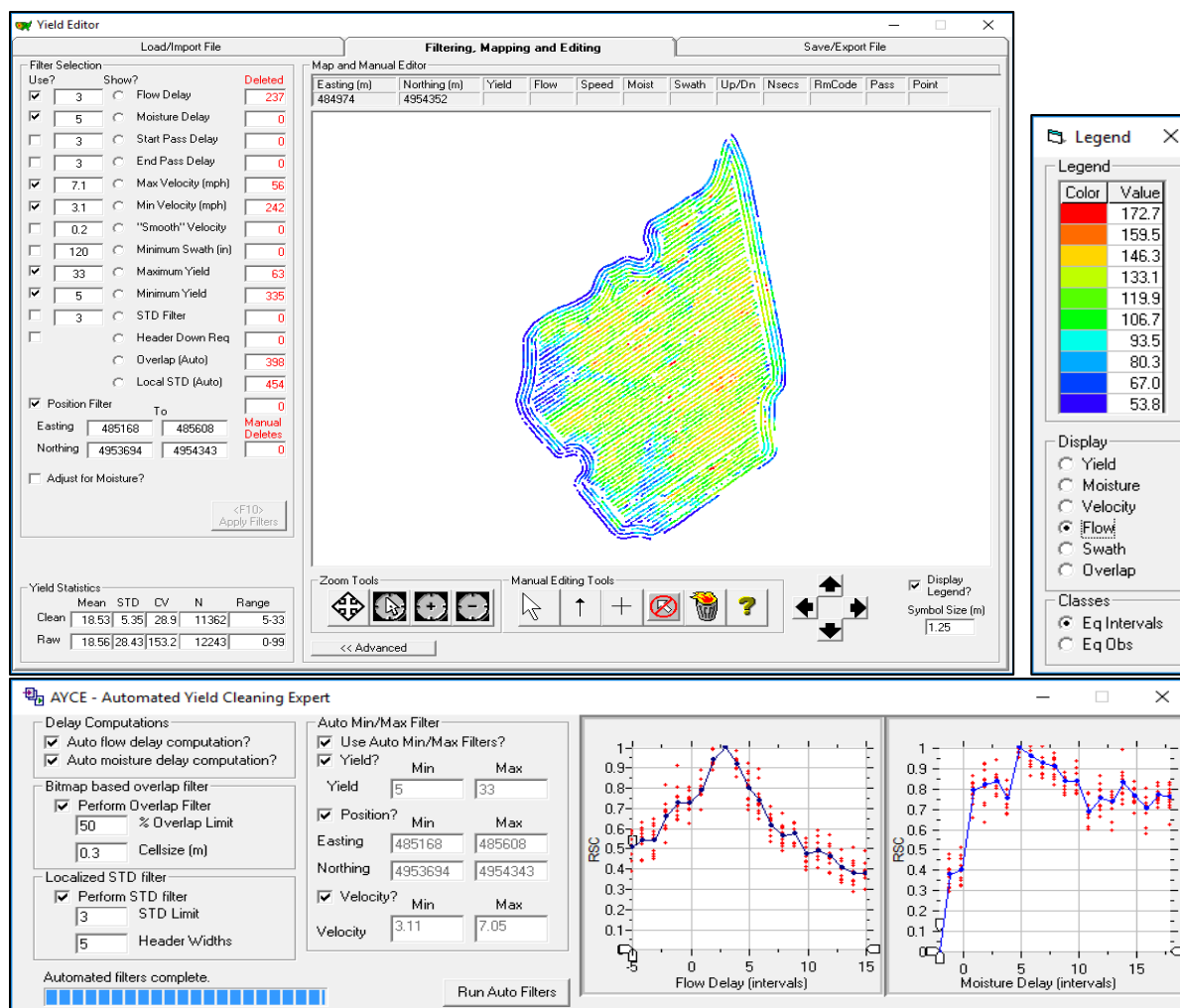
Figure 14. Example of AYCE output.

4.1.1.2 Check impact of AYCE bitmap-based overlap filter and localized STD filter

The default values for the bitmap-based overlap filter are 50% overlap and 0.3 m cell size. Inspect the location and number of data points removed by this filter ("Overlap" in the "Filter Selection" panel). This filter is affected by the header width and accuracy of the GPS system used. It works well and removes overlapped passes effectively where GPS systems are accurate. However, it can remove too many scattered data points (instead of a whole pass) from some datasets. Consider the accuracy of the GPS – for example, a WAAS system can be 5-15 ft off. For very poor DGPS corrections on very small headers, you may need to turn the overlap filter off completely. For a wide swath and an RTK system, you may want to be far more aggressive (smaller cell size and higher overlap% than default settings).

Identify the appropriate percent overlap and cell size value for the dataset by changing and observing the deleted points. Look to see if the narrower swath (partial, overlapped) passes were removed by using the new filter value or not. To make changes to the %Overlap, Cellsize, STD Limit, or Header Width values, the raw data must be re-imported. After the value adjustments are made, then rerun "Run Auto Filters". The Local STD filter selection cannot be undone once it has already been applied.

4.1.1.3 Flow delay – evaluate and adjust

Choose 'Flow' in the AYCE output map legend. It will display the flow delay values for the entire field.
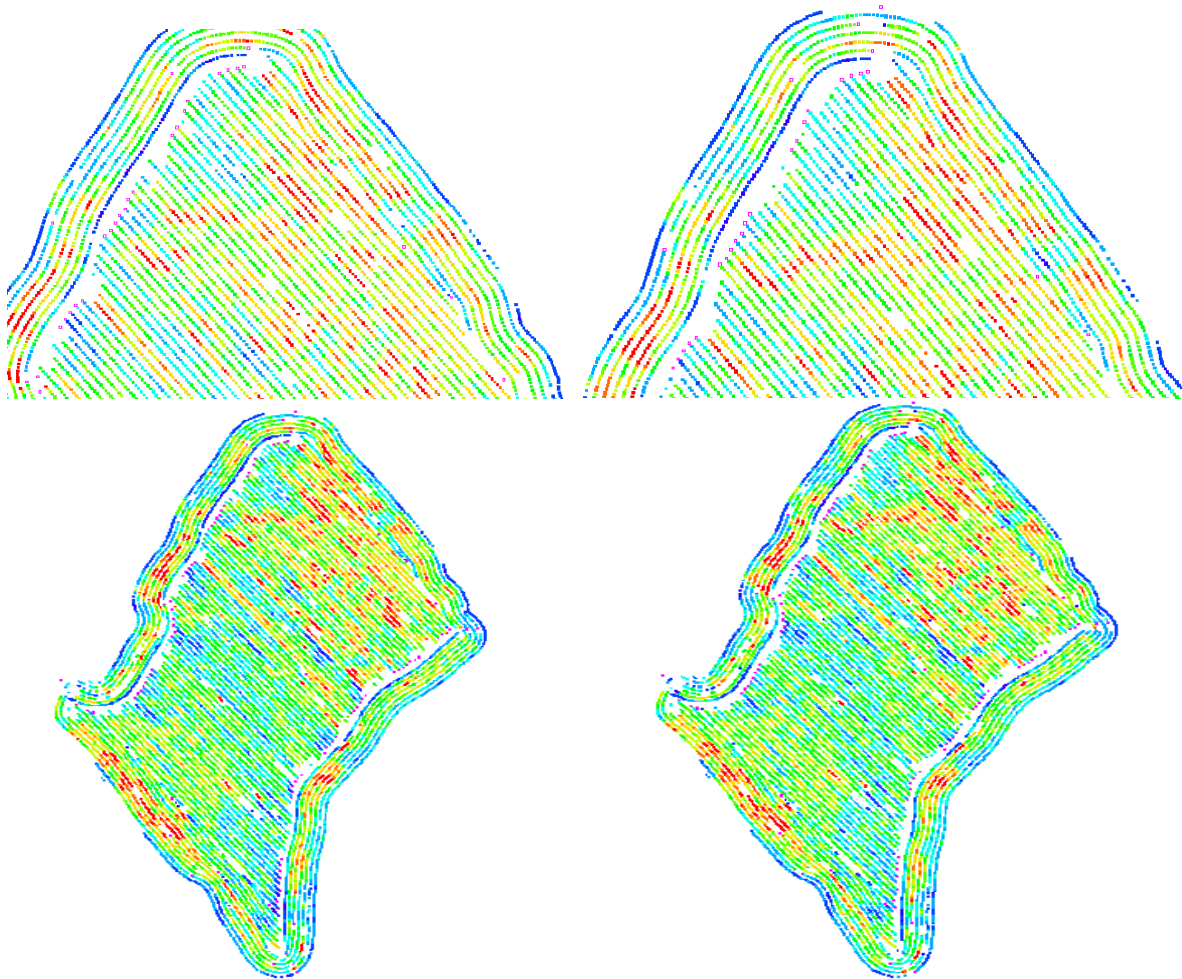

Figure 15. Map showing flow delay of 0 sec (left top and bottom) and flow delay of 4 sec applied (right top and bottom).

Displayed values will be colored based on the 10 classes it computes for two choices; either in an equal interval basis or an equal observation basis (see selection options at bottom of Legend display). An Equal interval map is useful to identify spatial trends as well as individual data points with relatively large differences. An equal observation map adds contrast to the uniform field where values (yield, flow, moisture, velocity, etc.) are relatively similar. Once the flow map is displayed, go to 'Filter Selection' and choose 'Show?' for either 'Start Pass Delay' or 'End Pass Delay' (some values need to be applied for these two filters). This will display the deleted points in pink for your choice (either start pass delay or end pass delay). This will help to reference the direction of each pass/transect. Inspect two opposite pass/transects and decide if the flow values (colors) match on both transects. Changing the 'Flow Delay' value and inspecting the flow value match on these two passes is one way to decide appropriate 'Flow Delay' value.

4.1.1.4 Moisture delay – evaluate and adjust

Choose 'Moisture' in the AYCE output map legend. It will display the moisture values for the entire field. Follow a similar approach as in flow delay value to decide the appropriate moisture delay value for batch processing settings.

4.1.1.5 Determine start and end pass delay intervals

These filters remove data points equal to the selected filter value for "start pass delay" and "end pass delay" at the beginning and end of each pass/transect. Look at the displayed 'flow' and 'yield' map and color differences in a few points at the beginning and end of each pass. If two opposite passes (beginning and ending) show different patterns in the map (difference in color) at the edge of the fields, then these points should be removed. Decide the filter value based on number of points showing each distinct color pattern. Use "Bounding Box Zoom" in the "Zoom Tools" selection panel to inspect data points (Figure 16).
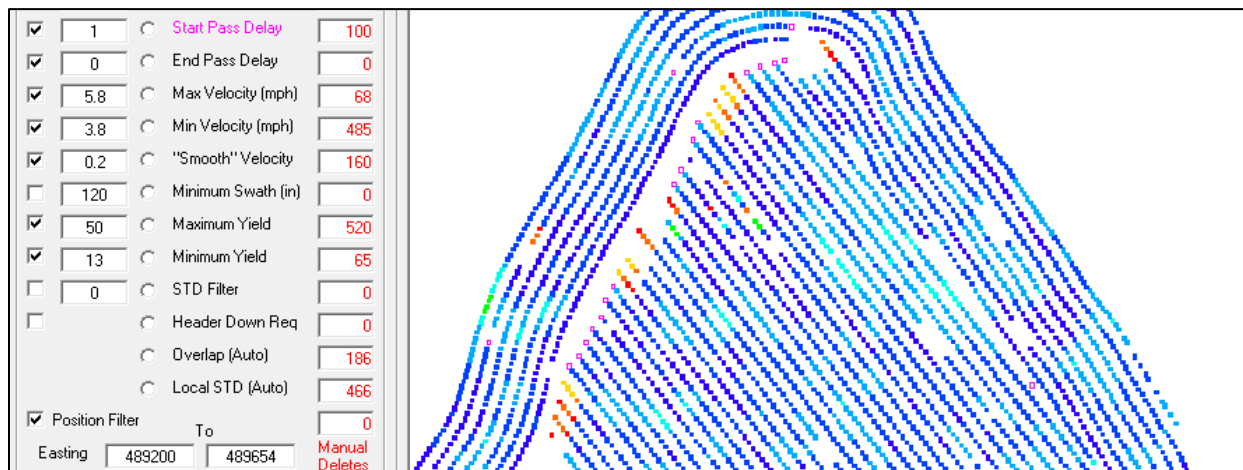

Figure 16. Example to decide the end pass delay: 'Start Pass Delay' filter value 1 is applied and displayed in flow map with 'pink' color. Very high flow at the end of each pass (red colored points) indicates an end pass delay of 4 is appropriate for this field, because there are a total of 4 red points. Observe with yield map (for calculated yield) if similar pattern exist for the start and end pass.

4.1.2. Exporting cleaned data, saving settings and session files

- Once cleaning is done, click on the "Save/Export File" tab in Yield Editor. Figure 10 shows all the parameters that should be checked for data export. Click on 'Export Data,' browse to folder, type the name of the exported file and save as CSV (*.csv) file. Caution: If any delays were applied to flow and/or moisture in Yield Editor, the values for flow and/or moisture in the exported datasheet will not correspond to the yield and location of the row because Yield Editor calculates yields involving delays in flow or moisture by simply reading back or forward in time (rows) for the number of intervals. It does not actually move the moisture or flow column values back or forward to match the row with the corresponding yield calculation, time point, and GPS coordinates. Hence, to avoid confusion, export yield and corresponding latitude and longitude information only. The next step is to save the session file as a *.yes file. These session files will be helpful if more editing or cleaning is needed later. Click 'Save Session', name it, and navigate to the desired folder, and click on Save (Figure 17).
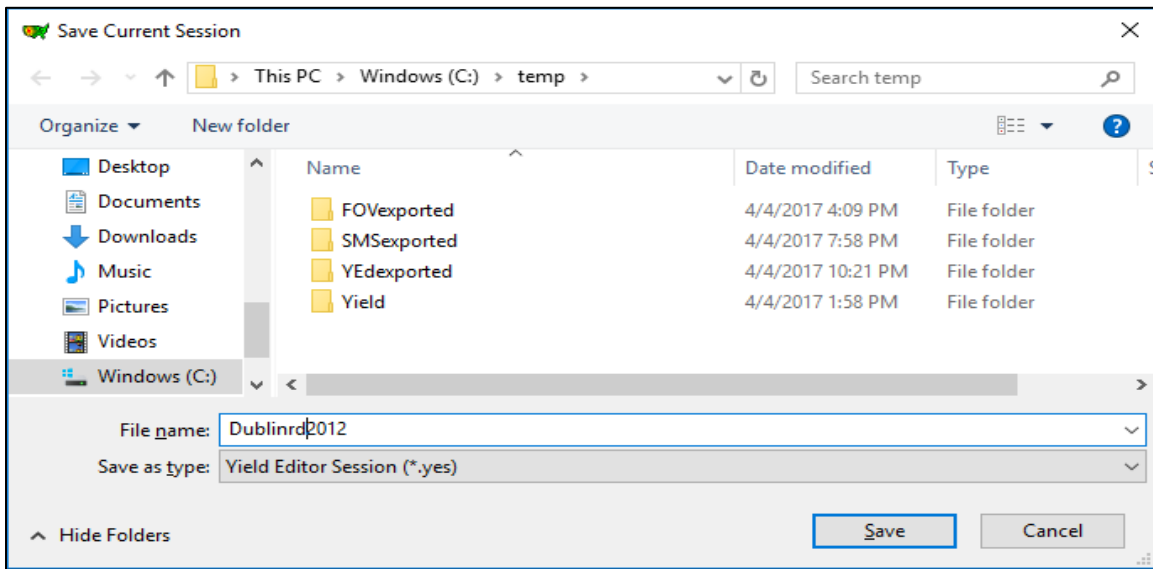
Figure 17. Session files saved as *.yes file extension.

- After cleaning about 10 fields with the approach outlined so far, compare the resulting filter values across these fields. *Choose the most representative value for each of the 4 filters (flow delay, moisture delay, start pass, and end pass) and 0.2 for Smoothing.* Save these as a setting file.
- Settings and configurations applied to these fields can be saved and applied to other fields quickly. To apply those setting parameters to a new field while it is imported, save these settings as default under 'Save Filter and Configuration Settings.' Check 'Save as Default Configuration?' and click on 'Save Config.'
- To save these settings for later use, uncheck 'Save as Default Configuration' and click on 'Save Config'. In the pop up window, browse to the folder where these files need to be saved, give a name under 'filename' and save file as a *.set (configuration settings) file (Figure 18).
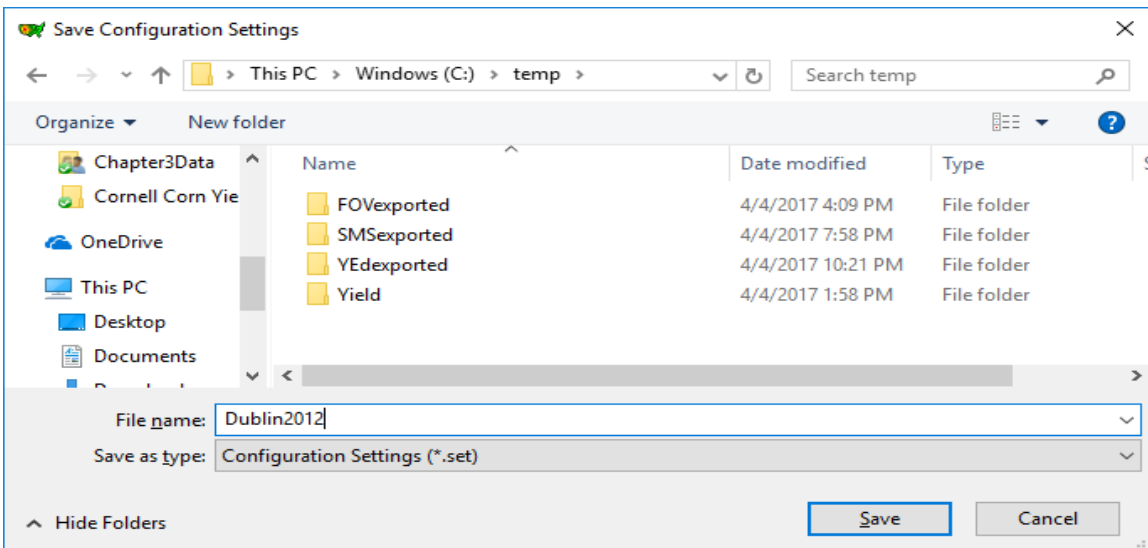

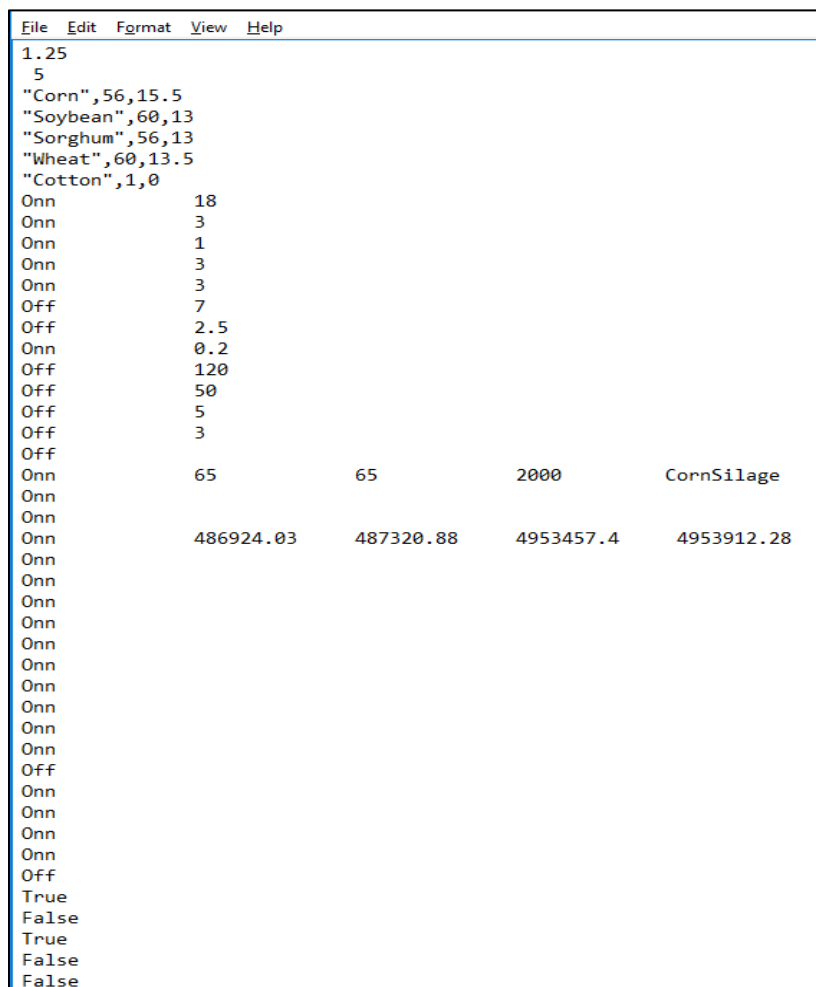Figure 18. Setting files saved as *.set file extension.

4.2. Processing individual fields with farm specific settings

4.2.1 AYCE method

Use the setting file(s) saved in initial testing (Section 4.1) above, that have been determined to work well for an individual farm. Using AYCE and saved settings file, run Yield Editor to clean the rest of the fields in the farm. Save cleaned file and session file with the appropriate label.

4.2.2. Batch processing

Batch processing is a powerful tool to speed up the cleaning process. If there are multiple fields to process, at least 10 individual fields should be interactively cleaned from the same operating settings (same combine harvester, crop type, and year) to derive reliable filter setting values for the whole farm (as described in Section 4.1). Once these setting values are saved as a settings (*.set) file (as described in Section 4.2), this file can be used during batch processing of other fields harvested the same operating settings (equipment, crop, year). The Yield Editor 2.0 manual describes the batch processing details under the "Using AYCE in Automated Mode" and "Batch Mode Language Details" section.

```
File  Edit  Format  View  Help
1.25
 5
"Corn",56,15.5
"Soybean",60,13
"Sorghum",56,13
"Wheat",60,13.5
"Cotton",1,0
Onn              18
Onn              3
Onn              1
Onn              3
Onn              3
Off              7
Off              2.5
Onn              0.2
Off              120
Off              50
Off              5
Off              3
Off
Onn              65          65          2000         CornSilage
Onn
Onn
Onn              486924.03   487320.88   4953457.4    4953912.28
Onn
Onn
Onn
Onn
Onn
Onn
Onn
Onn
Onn
Onn
Off
Onn
Onn
Onn
Onn
Off
True
False
True
False
False
```

Figure 19. Example of filter setting values on the saved configuration settings (*.set) file.

Example of batch processing:

- One folder "BatchFileExample" is created inside "C:\Users" for this example. Name this folder appropriately to identify the farm and batch processing information.
- Three folders named 'input', 'output' and 'session' are created inside "C:\Users\BatchFileExample" folder.
- All input data (*.tx or *.csv) should go inside the folder 'input'.
- Save a copy of default settings file (*.set) (that was derived from 10 individual fields) under 'BatchFileExample" folder and name it "Config".
- To see what this settings file looks like, right click the file and open it with notepad. An example of saved settings file for 'corn silage' is shown in Figure 19.

Information on the settings values is shown in Figure 20. A repeated run by changing each value one at a time and their effect on the output is used to produce the table as shown below.

Row position	Settings file value	Description
1	1.25	legend symbol size
2	5	Number of rows in table
3	corn	Table itself
4	soybean	Table itself
5	sorghum	Table itself
6	wheat	Table itself
7	cotton	Table itself
8	Onn	UTM zone
9	Onn	Flow delay
10	Onn	Moisture delay
11	Onn	Start pass delay
12	Onn	End pass delay
13	Off	Max velocity
14	Off	Min Velocity
15	Onn	Smooth velocity
16	Off	Minimum Swath
17	Off	Max yield
18	Off	Min yield
19	Off	STD filter
20	Off	Header down required
21	Onn	, Adjust for Moisture?
22	Onn	Sensor based?
23	Onn	Expand Dry?
24	Onn	Position filter
25	Onn	Export table Latitude(DD),
26	Onn	Longitude (DD),
27	Onn	UTM easting
28	Onn	UTM Northing
29	Onn	Yield
30	Onn	Moisture
31	Onn	Swath Width

32	Onn	Travel Distance
33	Onn	Grain Flow
34	Onn	Interval Length
35	Off	AGL Flag code
36	Onn	Allow Negative Lat/Long?
37	Onn	Transect Number
38	Onn	GPS Time
39	Onn	UTM Zone
40	Off	RmCode
41	True	Formatting: Comma Delimited ASCII
42	False	Formatting: Space Delimited ASCII
43	True	Export Clean points?
44	False	Export Deleted points?
45	False	Export All points?

Figure 20. Each row of the saved setting (*.set) file and description of those rows.

If the default overlap (50%) is not suitable and you need to change this value for the batch processing, locate the 'autoset.cfg' file within 'c:\program files(x86)\Yield_Editor' as shown in Figure 21. Open and edit 'autoset' file with text editor such as Notepad. Change default value 50 under 'Percent of area necessary to keep point' to the desired overlap value you want and save.
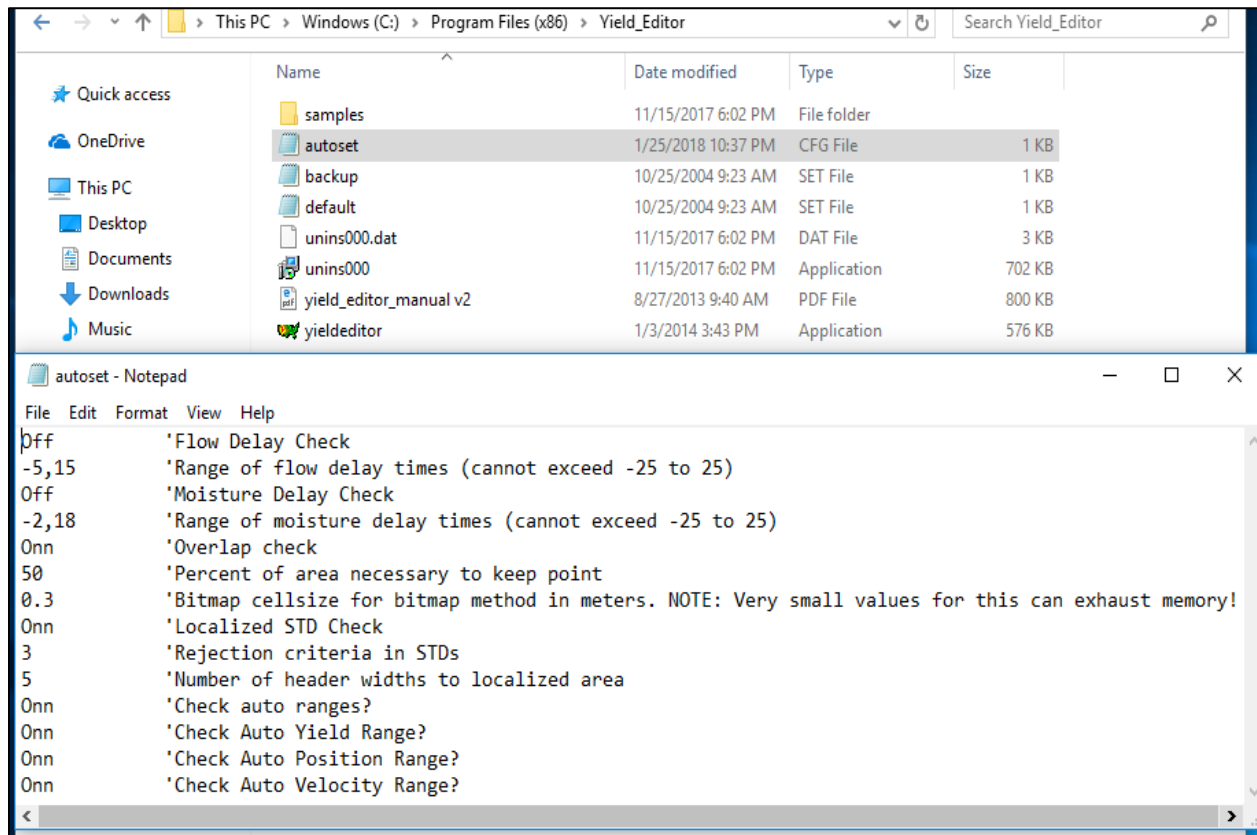

Figure 21. Location of the 'autoset.cfg' file to modify default overlap settings.

Box 1 below shows the batch processing code. Copy the entire contents of Box 1 and paste it into a blank notepad. Change the address path of each variable (the piece after the = in the first 5 "set" lines) so it is correct on the computer being used. Note: none of the folders in the pathway to the "BatchFileExample" folder can have spaces in their names or the batch will not run; delete the spaces or replace spaces with underscores (_). Also make sure that the "yieldEditorPath" is correct; it should be pointing to the location where the "yieldeditor.exe" file can be found. Save the notepad file as a batch file (add batch extension (.bat); see Figure 22). In the example below (Box 1), an Excel CSV file was the input file (inside "inputfolder"), hence "*.csv" was used to identify the file structure. If you have a text file (SMS exported Ag Leader advanced text file system), change "*.csv" to "*.txt" while the rest of the code stays same.

Box 1. Batch file code.

```
:: Batch processing example

@echo off

setlocal EnableDelayedExpansion

set "configFile=C:\Users\BatchFileExample\Config.set"

set "inputfolder=C:\Users\BatchFileExample\input"

set "outputfolder=C:\Users\BatchFileExample\output"

set "sessionfolder=C:\Users\BatchFileExample\session"

set "yieldEditorPath=C:\Program Files (x86)\Yield_Editor\yieldeditor.exe"

FOR %%f IN (%inputfolder%\*.csv) DO (

 "%yieldEditorPath%"  /fin="%configFile%" /csvin="%inputfolder%\%%~nxf"
/csvout="%outputfolder%\%%~nxf" /auto="y" /hide="y"
/yesout="%sessionfolder%\%%~nf%.yes")
```
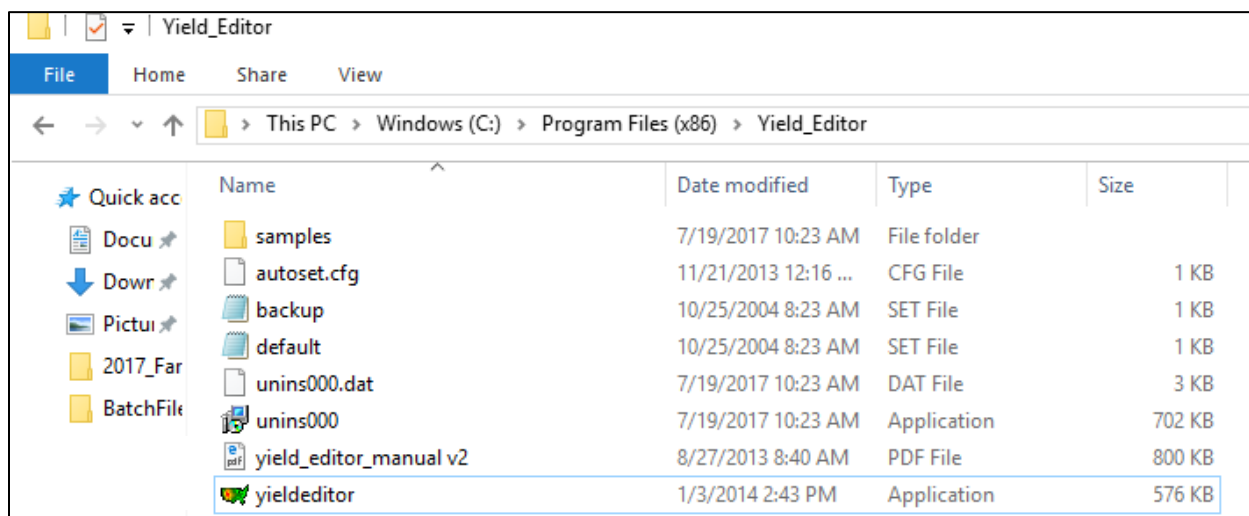

Figure 22. Verify the correct address for "yieldeditor.exe".

Ensure the notepad file with the Box 1 contents is saved as a *.bat file (Figure 23).
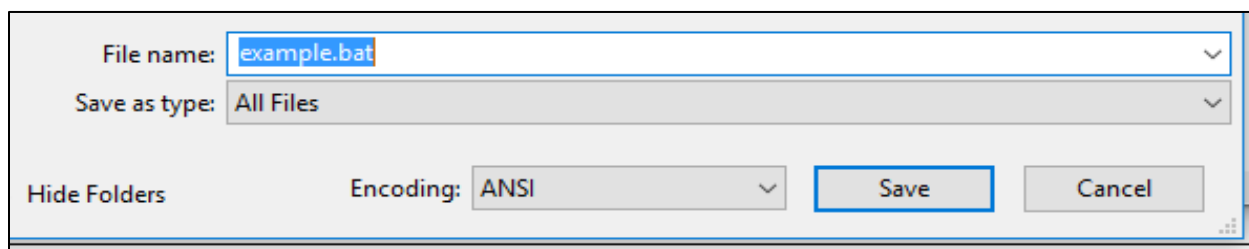
| File name: | example.bat | ⌄ |
| Save as type: | All Files | ⌄ |

| Hide Folders | Encoding: ANSI ⌄ | Save | Cancel |

Figure 23. Save the Box 1 contents as a *.bat file.

- Save this batch file inside the "BatchFileExample" folder.
- Go to the "BatchFileExample" folder and find the "example.bat" file.
- Double click the example.bat file to run these settings in batch mode.
- If the code does not run:
  - check for errors in paths, addresses, and folder names for case sensitivity, spaces, underscores, and slashes.
  - check for unmatched " " and ( ).
- The code in Box 1 calls yield editor ("%yieldEditorPath%"), loads the settings file (/fin="%configFile%"), takes each "*.csv" file from the input folder (/csvin="%inputfolder%\%%~nxf"), saves the output as a *.csv file in the output folder (/csvout="%outputfolder%\%%~nxf"), calls AYCE processing (/auto="y"), hides batch processing details (/hide="y"), and saves the session file as *.yes in the session folder (/yesout="%sessionfolder%\%%~nf%.yes").
- Once the processing is done, the cleaned (*.csv) files can be found in the output folder and the session files are saved in the session folder.
- To visually inspect and/or adjust filters as the batch processing takes place, change /auto="y" to /auto="p" in the FOR string of the Box 1 code. By changing to "p" this sets the processing to the interactive mode. This means that each field will pop up in Yield Editor and the filters can be adjusted as necessary. When the desired adjustments are achieved, close the Yield Editor window. A new window will appear for each subsequent field.

5. Moisture filter

Our silage data analyses showed that some yield values calculated based on a 0% moisture value (due to the moisture delay correction) had remained in the cleaned output. These points can be deleted by opening the saved session files and selecting individual points in the displayed moisture map in Yield Editor.

For automation purpose:

Deleting data points with moisture content <1 % for corn grain and moisture content < 46 % for corn silage helped to remove yield values calculated with erroneous (low) moisture content. This improved the yield estimation between manual and auto cleaned output and resulted in cleaner and more accurate maps. If too many zero moisture data points are in the dataset, affecting speed of producing a cleaned yield map:

- Open the SMS exported data .csv file in MS Excel and use the Data sort function to select and remove those points. Resort the file based on the GPS time column to put the data back in harvest sequence order.
- Batch process these files using the same settings derived from original dataset (10 fields).

6. References

- Han, S., S.M. Schneider, S.L. Rawlins, and R.G. Evans. 1997. A bitmap method for determining effective combine cut width in yield mapping. Trans. ASAE 40(2): 485-490.
- Lee, D. H., K.A. Sudduth, S.T. Drummond, S.O. Chung, and D.B. Myers. 2012. Automated yield map delay identification using phase correlation methodology. Trans. ASABE 55(3):743-752.
- Sudduth, K.A., S.T. Drummond, and D.B. Myers. 2012. Yield Editor 2.0: Software for automated removal of yield map errors. Paper No. 121338243. ASABE Ann. Intl. Mtg. Dallas, TX. 7/29/12. (http://extension.missouri.edu/sare/documents/ASABEYieldEditor2012.pdf)
- Sudduth, K.A. and S.T. Drummond. 2007. Yield Editor: Software for removing errors from crop yield maps. Agron. J. 99:1471–1482 (https://www.ars.usda.gov/ARSUserFiles/50701000/cswq-0372-sudduth.pdf.