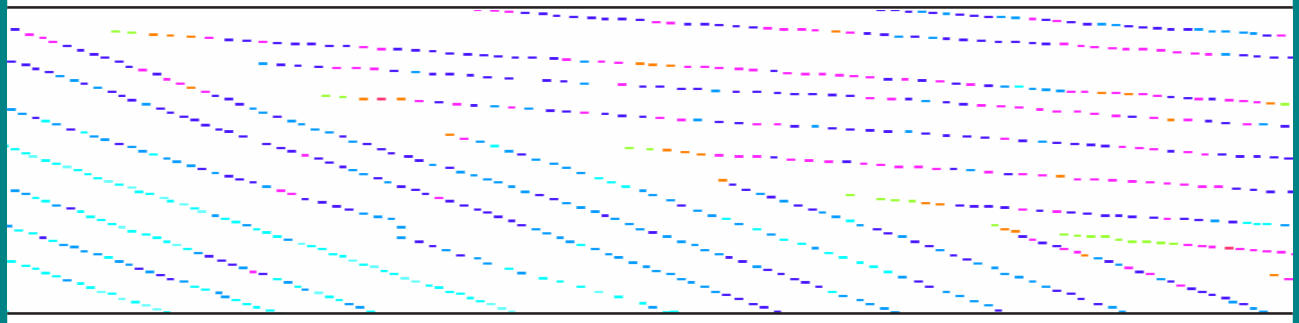
GRDC
Grains Research & Development Corporation



Cleaning Up Yield Data

with JMP software

Raw data files from yield monitor software come in a variety of formats and qualities. A cleaning process based on the distribution of the yield data provides a basic way to clean any data set. More sophisticated methods will eventually be available that work with any data set, but for the present, the process described here should help get data ready for further analysis.

Introduction

The ACPA uses the SAS program *JMP* (www.jmp.com) to calculate crop yield in tons per hectare and trim erroneous data points from yield monitor data files. The full procedure is usually applied to *.txt* files in the AgLeader 'advanced' format but can be adapted to any *.txt* yield file that contains 'mass flow', 'distance travelled', 'cutting width' and 'grain moisture' data columns. The data trimming procedure can also be adapted to *.txt* files where the yield values have already been calculated. In the Procedure section the task is listed in Green followed by the JMP commands in Red. The result and any other requirements for each step are also discussed.

Associated Files

Files used in conjunction with *yield.txt* files:

YieldCalcB.jmp :	for data recorded in metric units (mass flow units = kg/s; length units = mm)
YieldCalcImperial.jmp:	for data recorded in imperial units (mass flow units = lb/s; length units = inches)
JDYieldCalcB.jmp:	for data exported from John Deere Office software (mass flow units = decagrams/s; length units = mm)
DisTrim.jmp:	for data that has an already calculated yield column (yield units = t/ha)

GRDC
Grains Research & Development Corporation

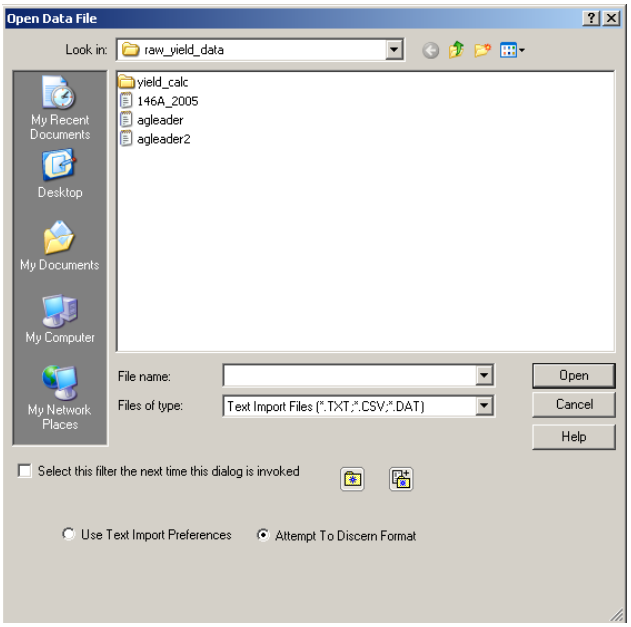
PROCEDURE IN JMP

Calculate Crop Yield and Trim Data

1. Open yield.txt file: *File/Open/Open Data File*

Ensure the 'Files of Type' field reads 'Text Import Files (*.TXT;*.CSV;*.DAT) and the 'Attempt To Discern Format' button is checked. Then navigate to where the yield file is stored and select the desired '.txt' file and press OPEN. This will open a 16 or 17 column data file depending on whether elevation has been recorded.

Column 3 = mass flow
Column 5 = measurement time span
Column 6 = distance travelled
Column 7 = cutting width
Column 8 = grain moisture

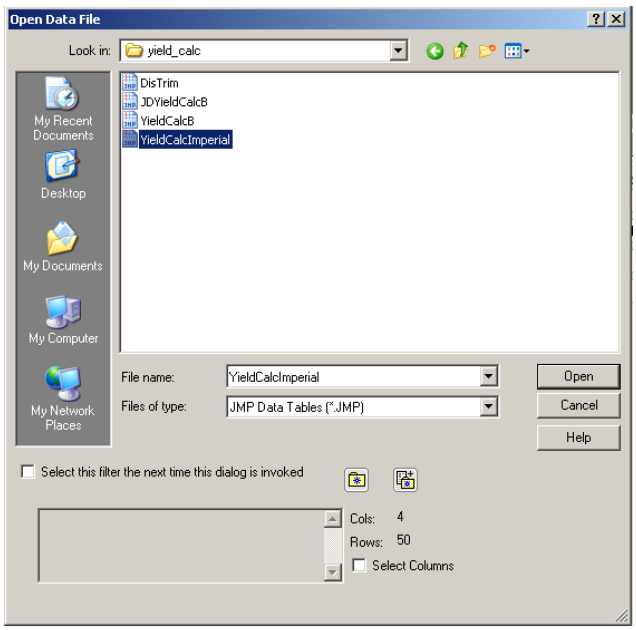


Open Data File window: ensure that the correct 'File of Type' is selected and the 'Attempt to Discern Format' button is checked

2. Open YieldCalc file: File/Open/Open Data File

Navigate to the folder where the YieldCalc.JMP files have been stored. If the yield file is from JDOffice choose the JDYieldCalcB.JMP file. If the file is from another harvester and the meaurement units (imperial or metric) are not known then examine column 7 (cutting width). It should be easy to tell what units are being utilised. Numbers such as 10000 indicate metric measurement of a 10 metre front while 393 indicates imperial measurement of a 10 metre front. Choose the appropriate file and press OPEN.

You will now have 2 spreadsheet files open and the last file opened will be 'active'. Ensure the yield file that you wish to work on is made active by clicking on the file spreadsheet. An 'active' file or window is denoted by a blue strip along the top.



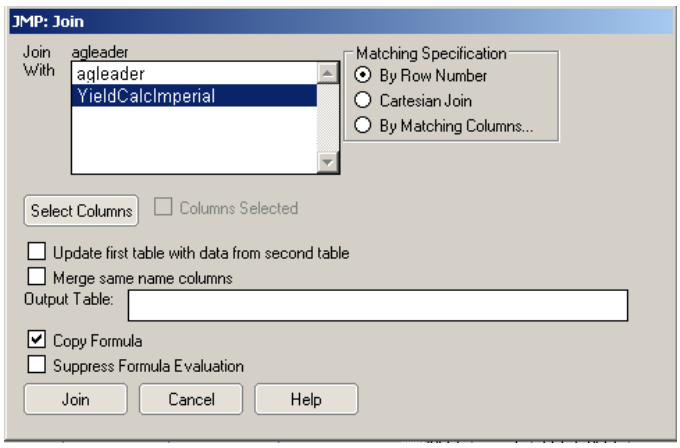
Open Data File window: ensure that the 'File of Type' field reads 'JMP Data Tables (*.JMP)

3. Join the Yield.txt file with the YieldCalc file: Tables/Join

Select the appropriate YieldCalc file from the list, ensue the 'Suppress Formula Evaluation' box is UNTICKED and press JOIN.

This will produce a new spreadsheet file where the columns of the YiledCalc file will be added onto the end of the Yield file and the last 4 columns will be 'yield, trim1, yield_t_ha, trim2'.

.



Join JMP files window: highlight the file to join with the yield file and ensure the 'Suppress Formula Evaluation is unticked

umn10	Column11	Column12	Column13	Column14	Column15	Column16	Column17	yield	trim1	yield_t_ha	trim2
1	960446	F5:COMETB	L0:44 012911	WHEAT	7	0	1011.4	8.91660275	5	8.91660275	0
1	960446	F5:COMETB	L0:44 012911	WHEAT	7	0	1011.1	7.50215454	5	7.50215454	5
1	960446	F5:COMETB	L0:44 012911	WHEAT	7	0	1011.4	7.10019052	5	7.10019052	5
1	960446	F5:COMETB	L0:44 012911	WHEAT	7	0	1011.8	5.52338244	5	5.52338244	5
1	960446	F5:COMETB	L0:44 012911	WHEAT	7	0	1012.1	5.42670455	5	5.42670455	5
1	960446	F5:COMETB	L0:44 012911	WHEAT	7	0	1012.1	5.66147932	5	5.66147932	5
1	960446	F5:COMETB	L0:44 012911	WHEAT	7	0	1012.8	5.48553787	5	5.48553787	5
1	960446	F5:COMETB	L0:44 012911	WHEAT	7	0	1011.8	5.9195876	5	5.9195876	5
1	960446	F5:COMETB	L0:44 012911	WHEAT	7	0	1011.4	5.8504535	5	5.8504535	5
1	960446	F5:COMETB	L0:44 012911	WHEAT	7	0	1009.8	5.49686105	5	5.49686105	5
1	960446	F5:COMETB	L0:44 012911	WHEAT	7	0	1009.5	5.43745582	5	5.43745582	5
1	960446	F5:COMETB	L0:44 012911	WHEAT	7	0	1009.5	5.33353441	5	5.33353441	5
1	960446	F5:COMETB	L0:44 012911	WHEAT	7	0	1008.5	5.58973439	5	5.58973439	5
1	960446	F5:COMETB	L0:44 012911	WHEAT	7	0	1008.5	5.49999533	5	5.49999533	5

Joined data file showing the four added columns

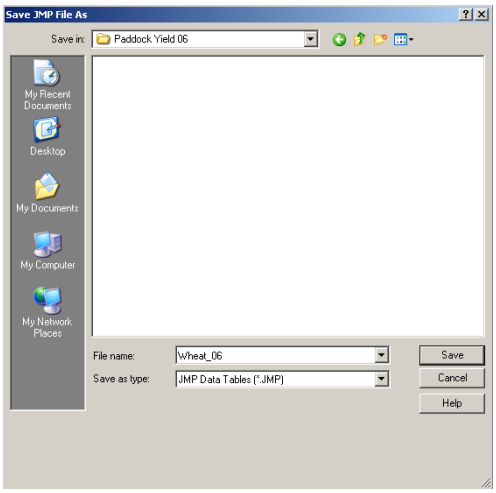
The four added columns are:

‘yield’:	has the yield in t/ha calculated from the data in Columns 3,5,6,7 and 8.
‘trim1’:	is an indicator column that identifies values in the ‘yield’ column that are either 0 or >10. Such values are shown as 0 in this column and all others are shown as 5.
‘yield_t_ha’	yield values identified by ‘5’ in the previous column are carried across to this column. Those identified by 0 in ‘trim 1’ are left blank.
‘trim2’	uses the identifier concept again, but here the 0 represents ‘yield_t_ha’ values that are either greater or less then the mean ‘yield_t_ha’ value +/- 2.5 standard deviations. An indicator 5 identifies yield values that fall within these distribution limits.

In effect the process calculates the yield values from the data, identifies where the yield is 0 t/ha or greater than 10 t/ha and discards them, then identifies where the remaining yield values fall outside the data limits of the mean yield +/- 2.5 standard deviations.

4. Save the file: File/SaveAs

Navigate to the appropriate storage location for the yield data, give it a relevant name and press SAVE



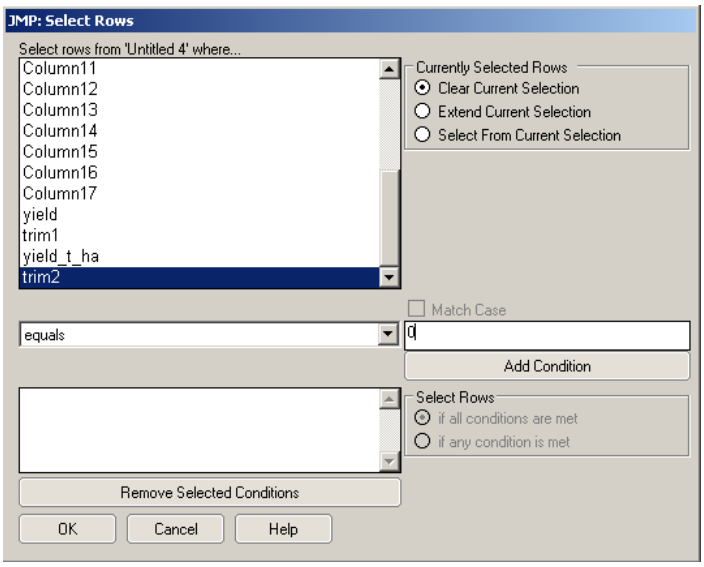
Save the joined file as a JMP file in the appropraite location

GRDC Grains Research & Development Corporation

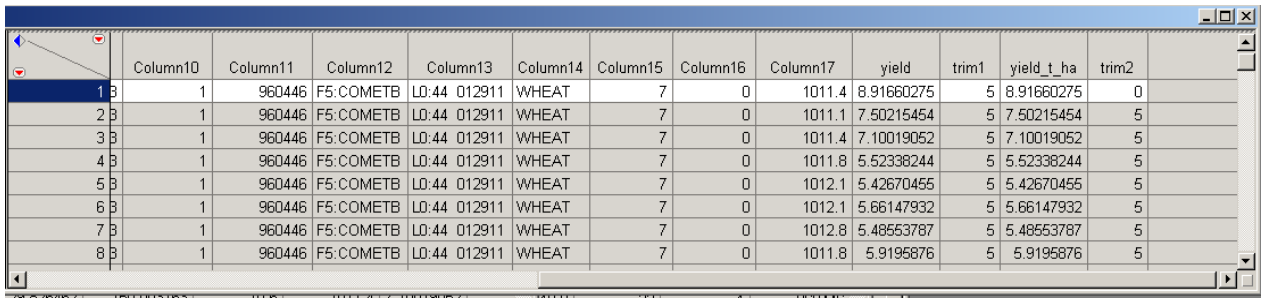
5. Examine which points will be trimmed Rows/Row Selection/Select Where

Highlight the 'trim2' column from the list offered, select 'equals' from the pull down menu and type '0' into the vacant box next to the pull-down menu arrow. Press OK.

This will highlight all the rows in the spreadsheet that fulfill this criteria. These are the data points that will be trimmed.



Select the points in the 'trim2' column that equal '0'

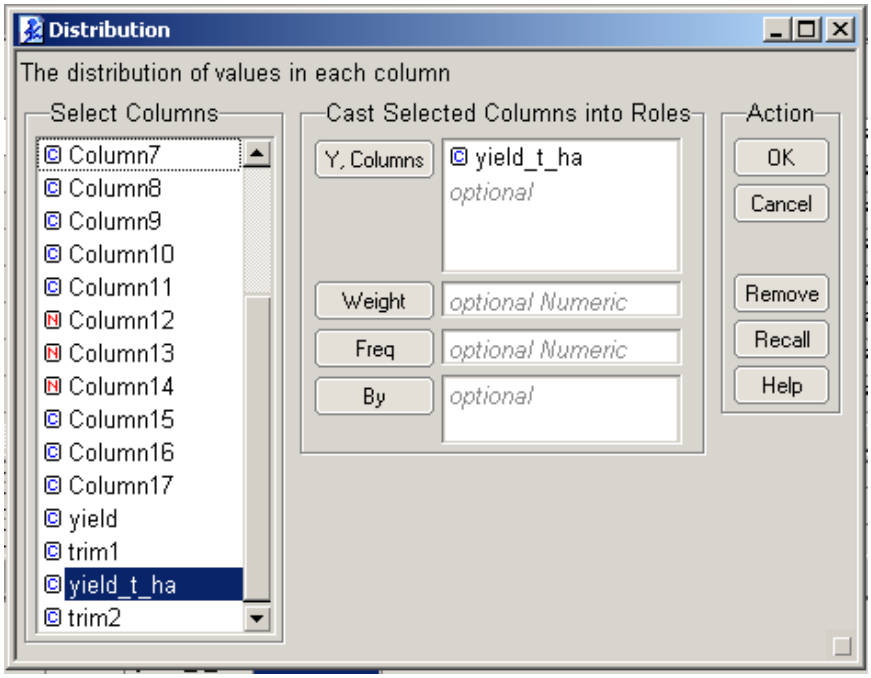


	Column10	Column11	Column12	Column13	Column14	Column15	Column16	Column17	yield	trim1	yield_t_ha	trim2
1	3	1	960446	F5:COMETB	L0:44 012911	WHEAT	7	0	1011.4	8.91660275	5	8.91660275	0
2	3	1	960446	F5:COMETB	L0:44 012911	WHEAT	7	0	1011.1	7.50215454	5	7.50215454	5
3	3	1	960446	F5:COMETB	L0:44 012911	WHEAT	7	0	1011.4	7.10019052	5	7.10019052	5
4	3	1	960446	F5:COMETB	L0:44 012911	WHEAT	7	0	1011.8	5.52338244	5	5.52338244	5
5	3	1	960446	F5:COMETB	L0:44 012911	WHEAT	7	0	1012.1	5.42670455	5	5.42670455	5
6	3	1	960446	F5:COMETB	L0:44 012911	WHEAT	7	0	1012.1	5.66147932	5	5.66147932	5
7	3	1	960446	F5:COMETB	L0:44 012911	WHEAT	7	0	1012.8	5.48553787	5	5.48553787	5
8	3	1	960446	F5:COMETB	L0:44 012911	WHEAT	7	0	1011.8	5.9195876	5	5.9195876	5

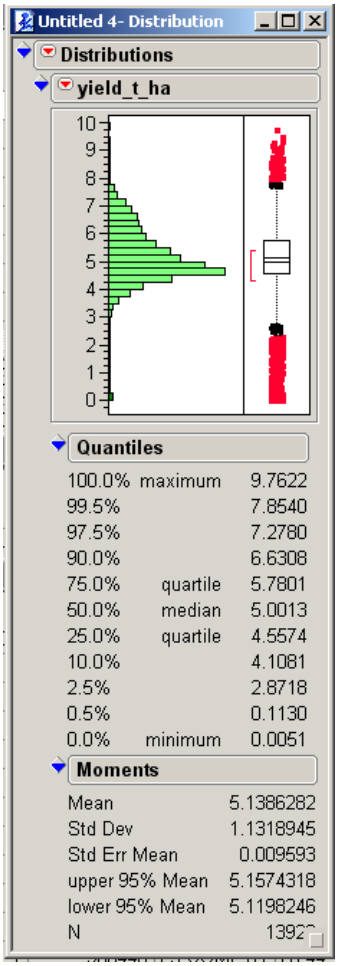
The rows where 'trim2' equal '0' are highlighted blue in the spreadsheet

To see where these data points fit in the distribution: Rows/Colours/Select a colour. This will highlight the points with a coloured dot in the far left column. To see the distribution histogram, the commands are: Analyse/Distribution. Then scroll down and with a single click select the column name 'yield_t_ha' and press the 'Y,Columns' button to cast the 'yield_t_ha' column into the adjacent box.

Then press OK. A histogram displaying the data distribution will be displayed and the upper and lower trim points will be highlighted in red.

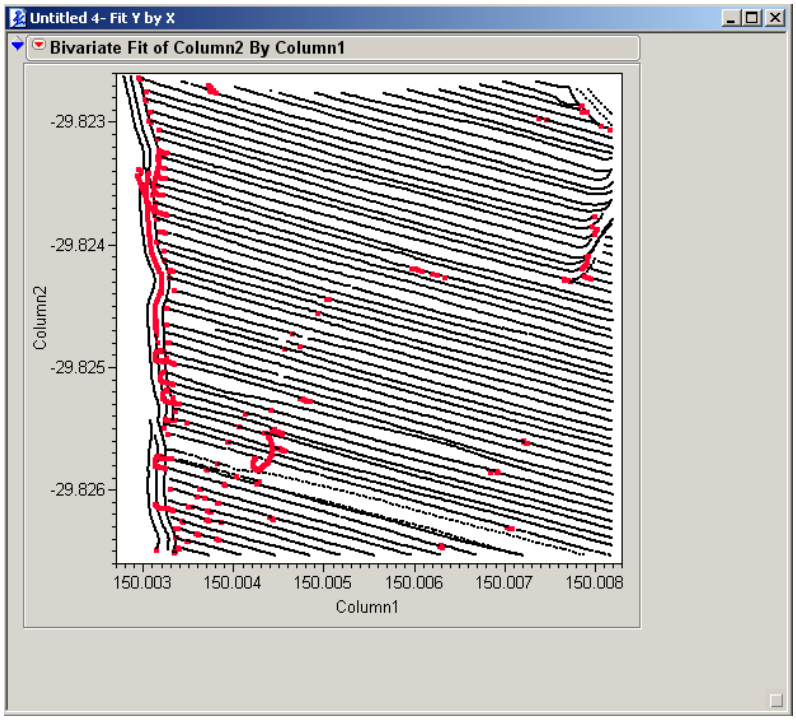


Examine the distribution of the 'yield_t_ha' data in a histogram



The data histogram with the points to be trimmed identified in red

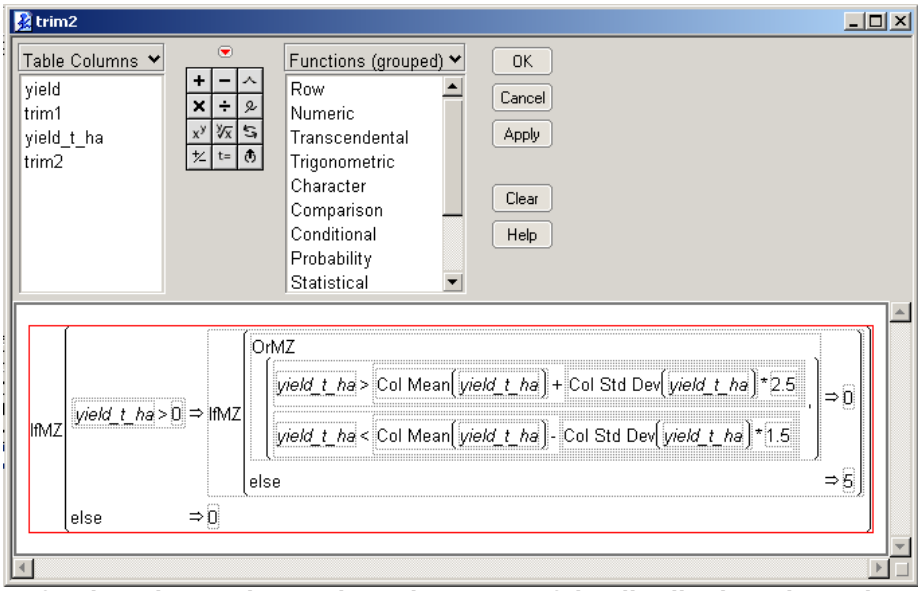
It is also possible to see where these points are in the paddock using a biplot. Choose Analyse/Fit Y by X. Then with a single click select 'Column 1' and cast it as an 'X,Factor' by pressing the 'X,Factor' box. Select 'Column2' and cast it as a 'Y,Factor' in the same way. Press OK.



The points to be trimmed are concentrated at the end of runs and where the comb may have been left down without harvesting.

sydney.edu.au/agriculture/acpa

The trimming procedure is the same for all 3 YieldCalc files. The different files just use different formulas for the yield calculation to account for different recorded units of measure. The JDYieldCalcB file also has a different formula for trimming the distribution (trim2). The JD yield monitor files usually have a longer tail at the low end of the distribution. If you find this is truncating too much data from your JD files, the formula in the trim2 column can be accessed by RIGHT CLICKING in the box containing the column name in the spreadsheet and then LEFT CLICKING on 'Formula'. This opens the box below and you can change the value 1.5 to 2 or 2.5 to gradually include more of the data at the lower end of the range. Press APPLY and OK .



The formula box for the 'trim2' column where the extent of the distribution trimmed can be changed

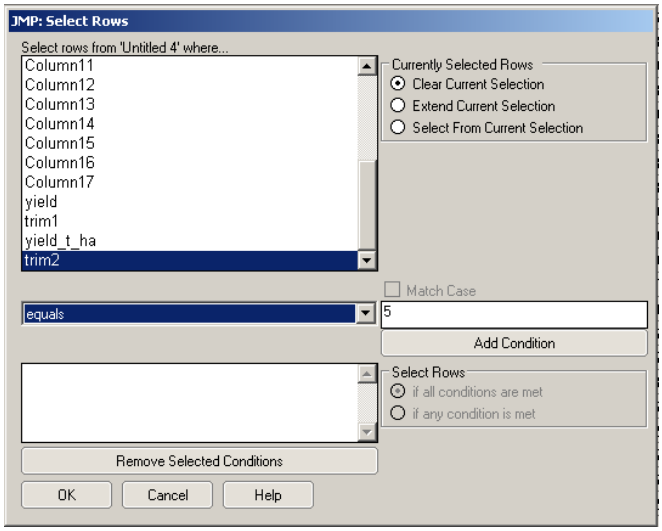
This will recalculate the trimming so it is necessary to then go back and clear the row colours before beginning Step 5 again. Clear the row colours in the spreadsheet by clicking on the top of the spreadsheet to Activate it, then Rows/Clear Row States.

Step 5 is optional but is a good way to initially check that the trimming is working for your data sets. Before continuing, Rows/Clear Row States will set the spreadsheet back to the correct state.

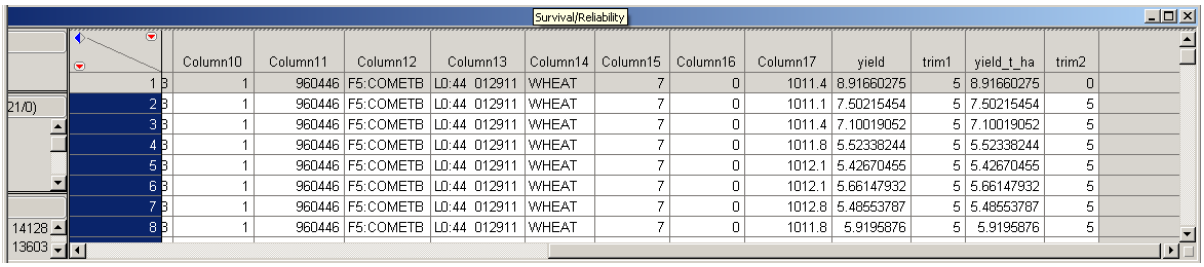
6. Trim the data set: Rows/Row Selection/Select Where

Highlight the 'trim2' column from the list offered, select 'equals' from the pull down menu and type 5 into the vacant box next to the pull-down menu arrow. Press OK.

This will highlight all the rows in the spreadsheet that we want to retain for mapping. For further processing it is not necessary to keep all the columns so we will select and save those that are important. This is achieved by holding the curser over the box with the column name in it and clicking once. The column will be highlighted and the top rectangle will be blue. Hold down the Control Key (Ctrl) and select the next column to save. Save columns 1,2,8, and yield_t_ha. If elevation has been recorded it is good to save that as well. It should be in Column 17.

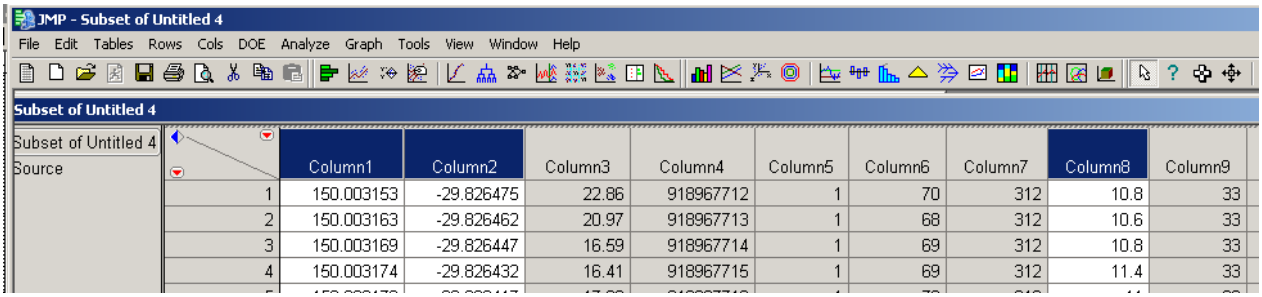


Select the points where 'trim2' equals '5'



	Column10	Column11	Column12	Column13	Column14	Column15	Column16	Column17	yield	trim1	yield_t_ha	trim2
1	1	960446	F5:COMETB	L0:44 012911	WHEAT	7	0	1011.4	8.91660275	5	8.91660275	0
2	1	960446	F5:COMETB	L0:44 012911	WHEAT	7	0	1011.1	7.50215454	5	7.50215454	5
3	1	960446	F5:COMETB	L0:44 012911	WHEAT	7	0	1011.4	7.10019052	5	7.10019052	5
4	1	960446	F5:COMETB	L0:44 012911	WHEAT	7	0	1011.8	5.52338244	5	5.52338244	5
5	1	960446	F5:COMETB	L0:44 012911	WHEAT	7	0	1012.1	5.42670455	5	5.42670455	5
6	1	960446	F5:COMETB	L0:44 012911	WHEAT	7	0	1012.1	5.66147932	5	5.66147932	5
7	1	960446	F5:COMETB	L0:44 012911	WHEAT	7	0	1012.8	5.48553787	5	5.48553787	5
8	1	960446	F5:COMETB	L0:44 012911	WHEAT	7	0	1011.8	5.9195876	5	5.9195876	5

The points that are selected are highlighted in the spreadsheet



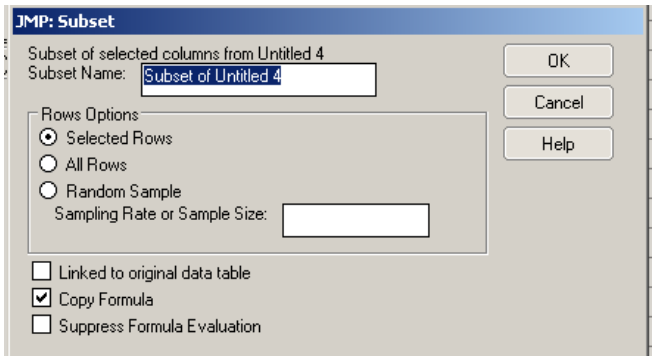
	Column1	Column2	Column3	Column4	Column5	Column6	Column7	Column8	Column9
1	150.003153	-29.826475	22.86	918967712	1	70	312	10.8	33
2	150.003163	-29.826462	20.97	918967713	1	68	312	10.6	33
3	150.003169	-29.826447	16.59	918967714	1	69	312	10.8	33
4	150.003174	-29.826432	16.41	918967715	1	69	312	11.4	33

Clicking once in the top of the box that contains the column name will select a column

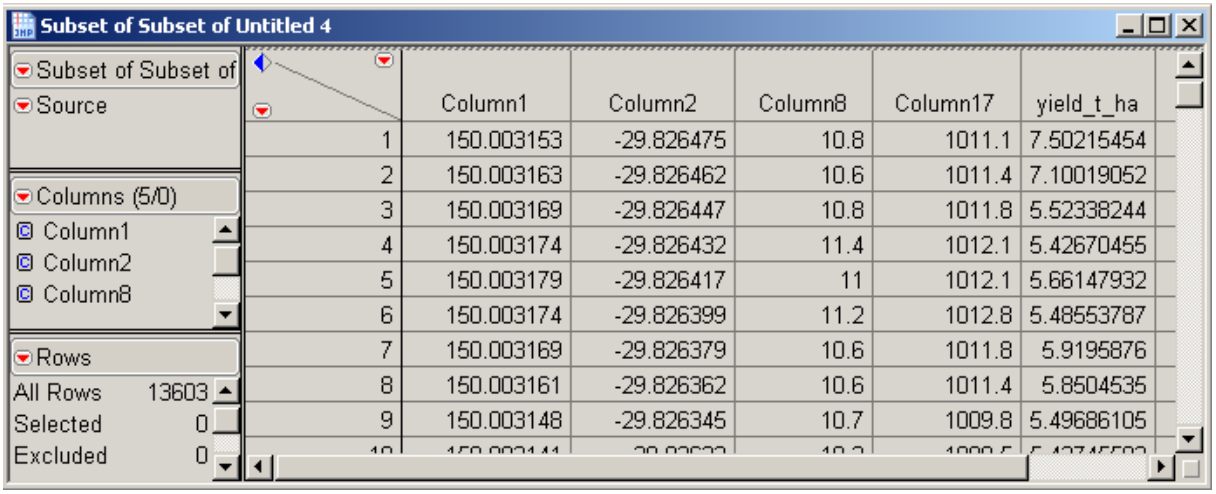
GRDC Grains Research & Development Corporation

7. Subset the Spreadsheet: Tables/Subset

Ensure that the 'Selected Rows' button is ON and the 'Copy Formula' box is ticked.Press OK.



Make a new spreadsheet with only the data selected

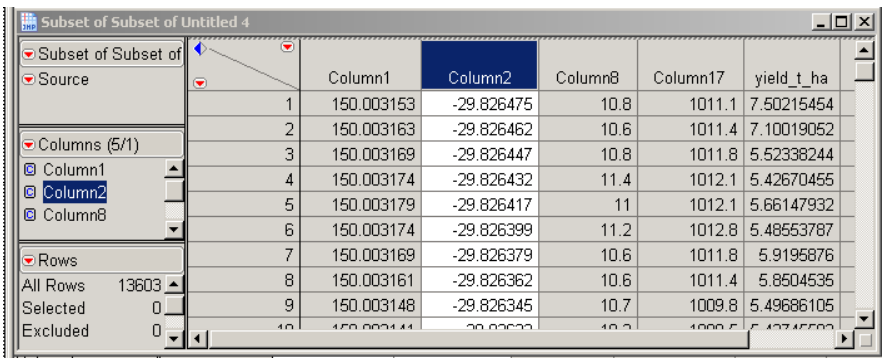


	Column1	Column2	Column8	Column17	yield_t_ha
1	150.003153	-29.826475	10.8	1011.1	7.50215454
2	150.003163	-29.826462	10.6	1011.4	7.10019052
3	150.003169	-29.826447	10.8	1011.8	5.52338244
4	150.003174	-29.826432	11.4	1012.1	5.42670455
5	150.003179	-29.826417	11	1012.1	5.66147932
6	150.003174	-29.826399	11.2	1012.8	5.48553787
7	150.003169	-29.826379	10.6	1011.8	5.9195876
8	150.003161	-29.826362	10.6	1011.4	5.8504535
9	150.003148	-29.826345	10.7	1009.8	5.49686105

The new spreadsheet contains only the columns and rows that were selected

Move 'Column 2' to the beginning of the Spreadsheet by first highlighting the column by placing the cursor in the box containing the column name and clicking once.

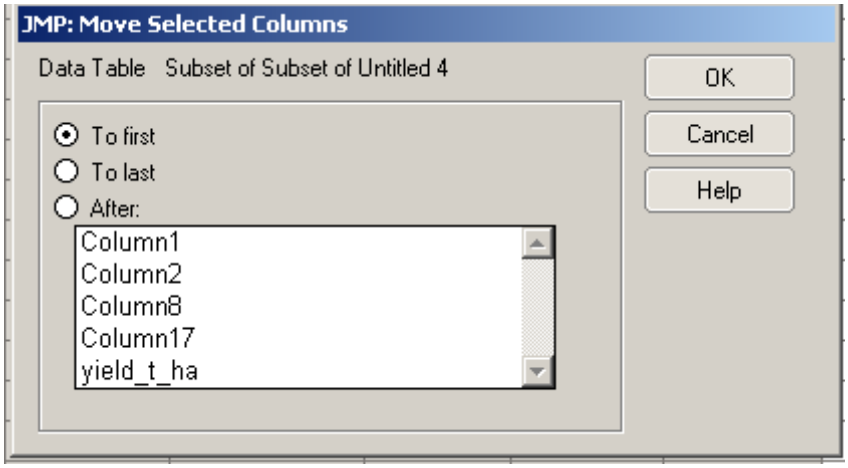
Then: Cols/Reorder Columns/Move Selected Columns...



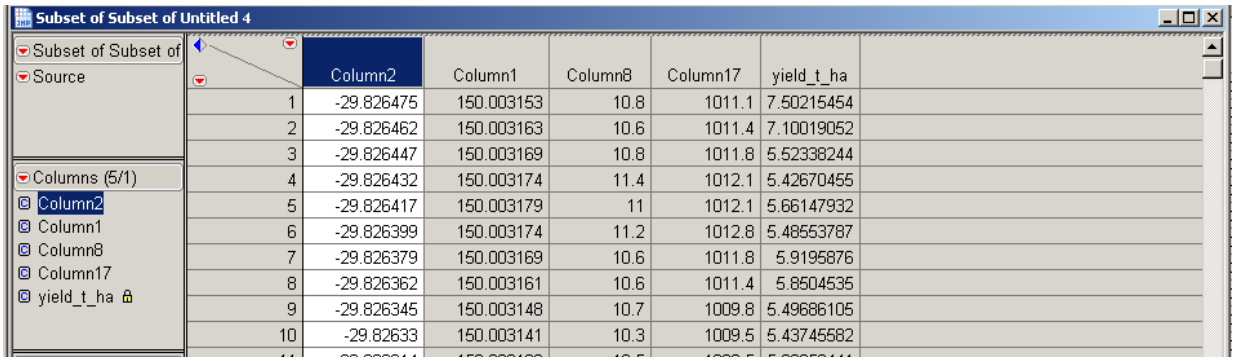
	Column1	Column2	Column8	Column17	yield_t_ha
1	150.003153	-29.826475	10.8	1011.1	7.50215454
2	150.003163	-29.826462	10.6	1011.4	7.10019052
3	150.003169	-29.826447	10.8	1011.8	5.52338244
4	150.003174	-29.826432	11.4	1012.1	5.42670455
5	150.003179	-29.826417	11	1012.1	5.66147932
6	150.003174	-29.826399	11.2	1012.8	5.48553787
7	150.003169	-29.826379	10.6	1011.8	5.9195876
8	150.003161	-29.826362	10.6	1011.4	5.8504535
9	150.003148	-29.826345	10.7	1009.8	5.49686105

Column 2 is selected

Ensure that the 'To First' button is ON and press OK.



Highlight column 2 and move it to the beginning of the spreadsheet



	Column2	Column1	Column8	Column17	yield_t_ha
1	-29.826475	150.003153	10.8	1011.1	7.50215454
2	-29.826462	150.003163	10.6	1011.4	7.10019052
3	-29.826447	150.003169	10.8	1011.8	5.52338244
4	-29.826432	150.003174	11.4	1012.1	5.42670455
5	-29.826417	150.003179	11	1012.1	5.66147932
6	-29.826399	150.003174	11.2	1012.8	5.48553787
7	-29.826379	150.003169	10.6	1011.8	5.9195876
8	-29.826362	150.003161	10.6	1011.4	5.8504535
9	-29.826345	150.003148	10.7	1009.8	5.49686105
10	-29.82633	150.003141	10.3	1009.5	5.43745582

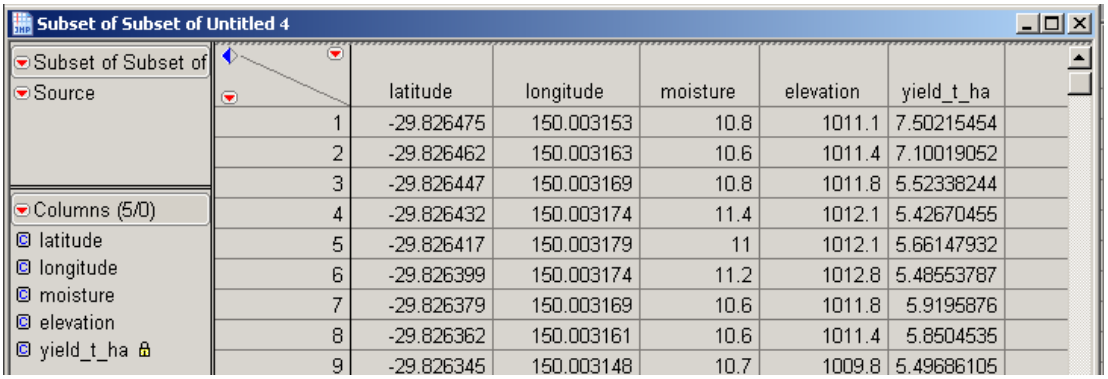
Column 2 has been moved to the beginning of the spreadsheet

Name the columns with more appropriate descriptors by placing the cursor directly on the name and clicking TWICE. The name will be highlighted and can be edited. A shortcut here is that once you have edited the first column name, pressing the TAB key will move the cursor directly to the next column name and it will be ready to edit.

The names will be:

Rename Column2 with:*latitude*
Rename Column8 with: *moisture*

Rename Column1 with: *longitude*
Rename Column17 with: *elevation*

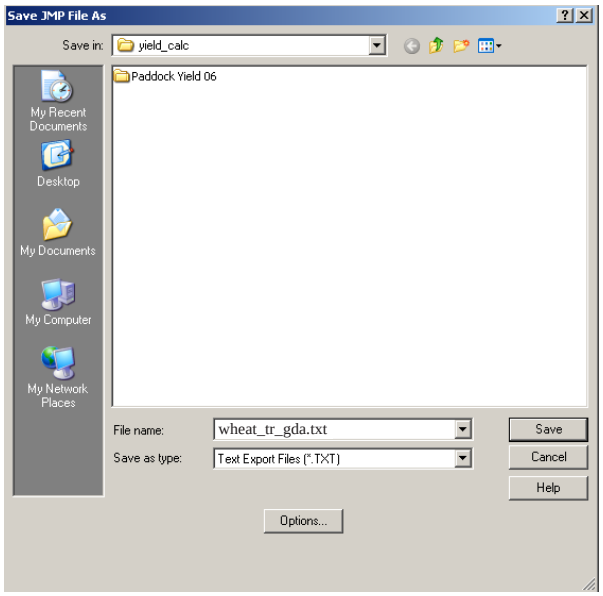


	latitude	longitude	moisture	elevation	yield_t_ha
1	-29.826475	150.003153	10.8	1011.1	7.50215454
2	-29.826462	150.003163	10.6	1011.4	7.10019052
3	-29.826447	150.003169	10.8	1011.8	5.52338244
4	-29.826432	150.003174	11.4	1012.1	5.42670455
5	-29.826417	150.003179	11	1012.1	5.66147932
6	-29.826399	150.003174	11.2	1012.8	5.48553787
7	-29.826379	150.003169	10.6	1011.8	5.9195876
8	-29.826362	150.003161	10.6	1011.4	5.8504535
9	-29.826345	150.003148	10.7	1009.8	5.49686105

The columns have been labelled with suitable names

GRDC
Grains Research & Development Corporation

8. Save the data as a .txt file: File/Save As



Save the data as a .txt file with a meaningful name

Locate the appropriate storage location and give the file a suitable name. The name should reflect the fact that the data has been trimmed or cleaned and that it contains GPS location in latititude and longitude. An example is: wheat_tr_gda.txt.

For the 'Save as Type' window, select 'Text Export Files (*.TXT) from the drop-down menu and then press SAVE. A trimmed yeld file in .txt format has been created and can be used in further mapping and analysis.

NOTE: If the original yield data file has an already calculated yield value that is to be used but the file still has erroneous points, the DisTrim.jmp file can be used from Step 2. The column with the yield values from the original file must be labelled 'yield' prior to joining in Step 3. All other steps remain the same.

Brett Whelan & James Taylor
Australian Centre for Precision Agriculture