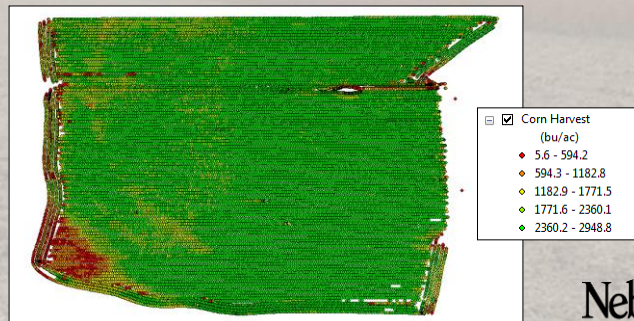


Yield Data Cleaning: Why it's Important

Joe D. Luck, Precision Agriculture Engineer
University of Nebraska-Lincoln Extension
Department of Biological Systems Engineering



UNIVERSITY OF
Nebraska
Lincoln

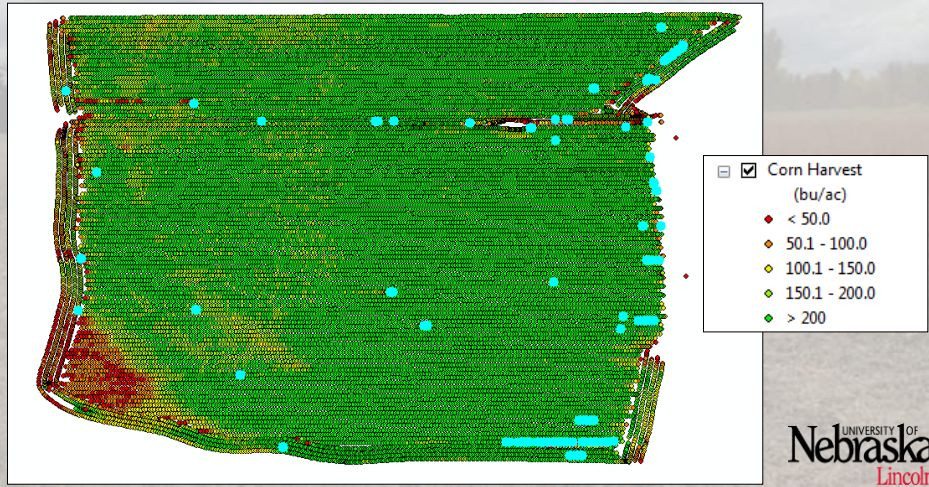
Discussion Topics

- Why do we need to waste time by cleaning our yield data after we've collected it?
- What tools do we have to clean the data?
 - Yield Editor
 - Microsoft Excel
- Examples of errors we can remove
- Farm Management Information Systems (FMIS)
- Suggests of what not to do when cleaning

UNIVERSITY OF
Nebraska
Lincoln

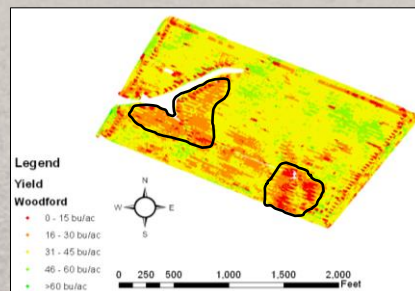
Errors in the data

Sensor or data collection errors can skew the yield



Why spend time removing errors?

- Let's discuss what things used to be like
- Yield data collection since the early 1990s
- FMIS tools were "primitive" at that time
 - Not user friendly
 - Very little "batch commands" were available
- Management zones were drawn by hand in many cases



Why spend time removing errors?

Since then, developers have spent a good deal of time working on automating this process

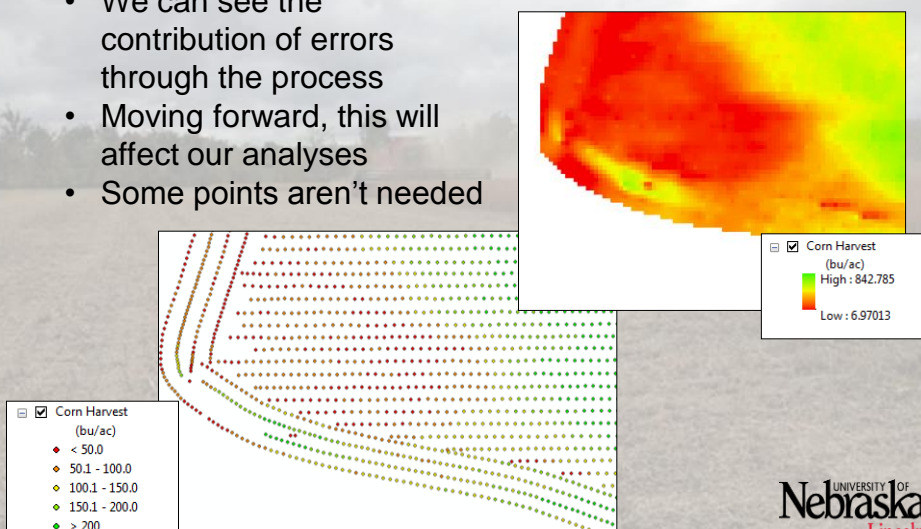
- Today, generating grid or contour maps is as easy as clicking a button
- The problem is that the errors we discussed can have a significant impact on grid or contour values

But isn't removing data a bad thing?

- It depends...
 - Raw yield data exists as individual points
 - To generate a grid or contour map, we interpolate (kriging)
 - In many cases, the yield data density is more than adequate to accurately estimate

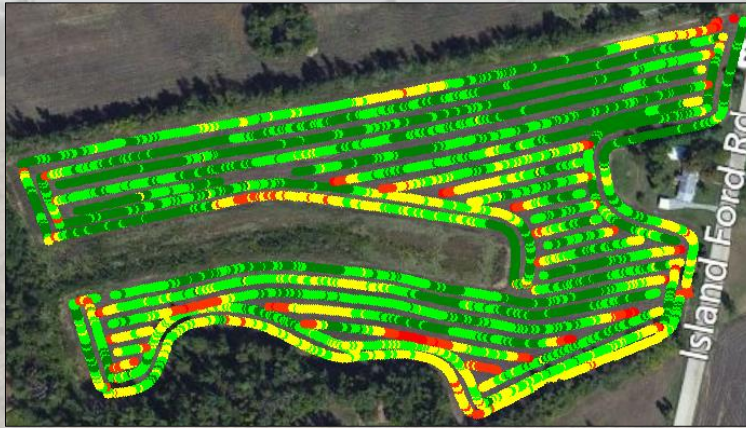
Point to Grid Interpolation

- We can see the contribution of errors through the process
- Moving forward, this will affect our analyses
- Some points aren't needed



Example of Cleaning Yield Data

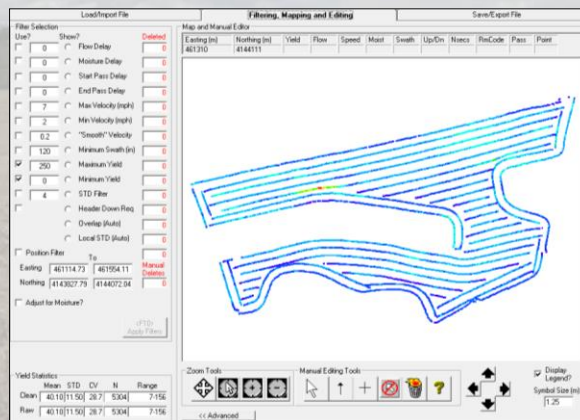
- Below is a raw yield data file
- What errors are noticed here?
- We can process these automatically



UNIVERSITY OF
Nebraska
Lincoln

Example of Cleaning Yield Data

- Yield Editor Software user interface
- Import SMS or Apex files
- Export .txt files to import back into our FMIS



UNIVERSITY OF
Nebraska
Lincoln

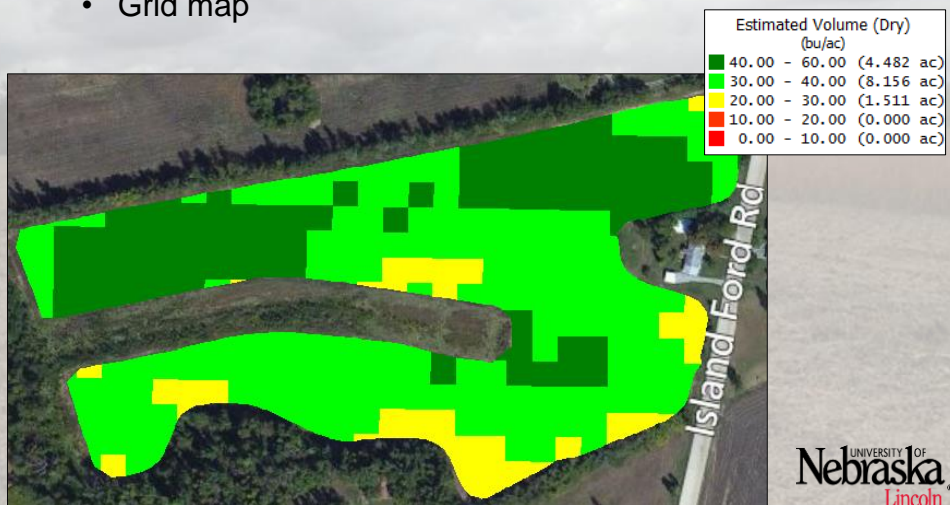
Example of Cleaning Yield Data

- After data are cleaned, we can bring that back into the GIS Software
- Notice that several data points have been removed



Example of Cleaning Yield Data

- Raw yield data
- Grid map

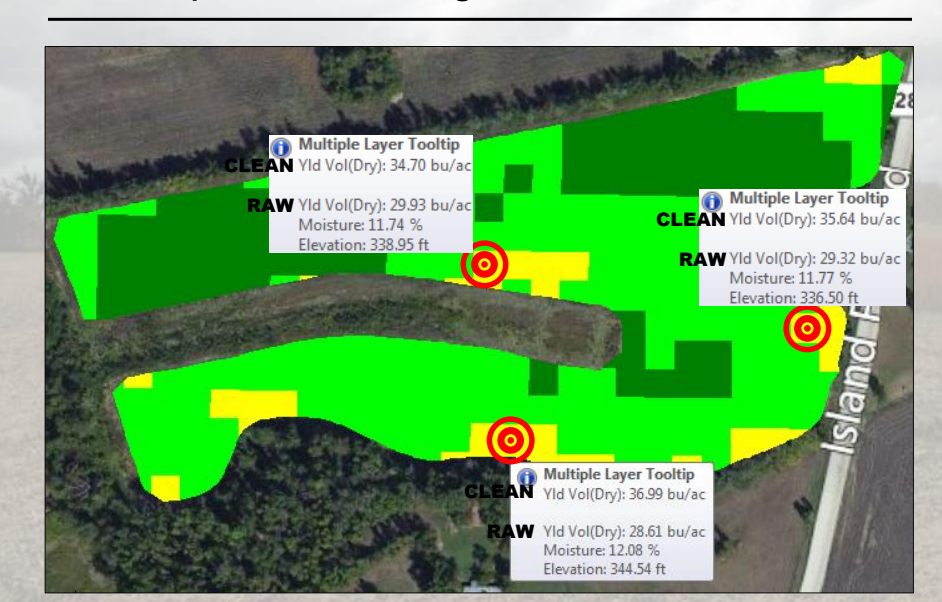


Example of Cleaning Yield Data

- Clean yield data
- Compare to previous map



Example of Cleaning Yield Data



Example of Cleaning Yield Data



UNIVERSITY OF
Nebraska
Lincoln

Using Excel to Remove Errors

- The .txt files can be imported into Excel
- Sort the data and delete records (low travel distance, swath width, moisture)
- Save the resulting data in a .txt file
- Import back into our GIS software

	E1	G	K	M	N	O	P	Q	R
	Distance_f	Duration_s	Swth_Width	Crop_Flw_M	Moisture_	Yld_Mass_W	Yld_Vol_We	Yld_Mass_D	Yld_Vol_Dr
1	0.0757	1.0000	15.0000	0.3464	16.2200	13283.3473	237.2136	13092.6922	233.8089
2	0.0772	1.0000	15.0000	2.6578	16.6100	100000.4000	1785.8039	98106.2748	1751.9786
3	0.0838	1.0000	15.0000	1.6976	19.4000	58800.6429	1050.0600	55756.8449	995.7039
4	0.0864	1.0000	15.0000	5.1783	19.4000	174137.9241	3109.7494	165123.7256	2948.7741
5	0.0868	1.0000	15.0000	2.9903	19.0800	100000.4000	1785.8039	95200.3808	1700.0853
6	0.0868	1.0000	15.0000	0.0115	14.2100	383.1031	6.8414	383.1031	6.8414
7	0.1045	1.0000	15.0000	3.5969	14.7800	100000.4000	1785.8039	100000.4000	1785.8039
8	0.1085	1.0000	15.0000	1.6535	16.2200	44235.8463	789.9623	43600.9318	778.6240
9	0.1197	1.0000	15.0000	0.3043	15.0700	7382.8166	131.8421	7376.7366	131.7335
10	0.1207	1.0000	15.0000	2.0751	18.4900	49944.4825	891.9069	47893.8208	855.2863

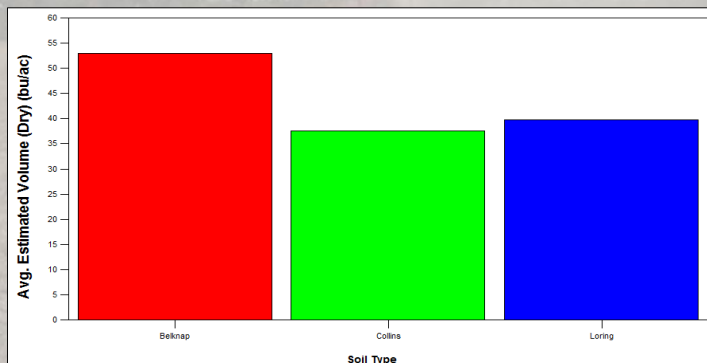
UNIVERSITY OF
Nebraska
Lincoln

Recommendations on how to Clean

- We've been talking about "physical" parameters to base our cleaning process on
- Other techniques include filtering based on statistics
- Others still use maximum and minimum yield thresholds
- We don't recommend using those as legitimate data may be deleted!

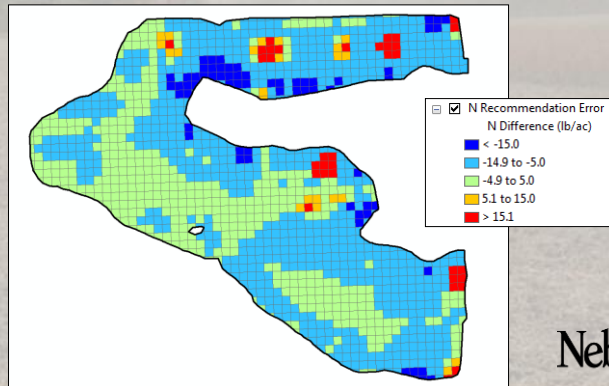
Why Cleaning Yield Data is Important

- In the end, we want to quantify our data
- Or, we may want to automate our Rx map process
- If we put bad data into these systems, we'll be rewarded with more bad data



Why Cleaning Yield Data is Important

- Automated Rx development can take in multiple data layers to estimate application rates
- Poor yield data input can create errors
- Nitrogen recommendation based on yield data:



UNIVERSITY OF
Nebraska
Lincoln

Summary

- Always store your raw data in a backup beforehand
- Post processing may take time, it will be worth it
- Automation of this process is likely to improve



UNIVERSITY OF
Nebraska
Lincoln