Data Preservation Through Structured Research Data Management

Gylling Data Management, Inc.

"Data" is Available Everywhere

- Internet
- Popular publications
- Research journals
- Historical experiments that exist in nearly every research organization

Is this Useful Data?

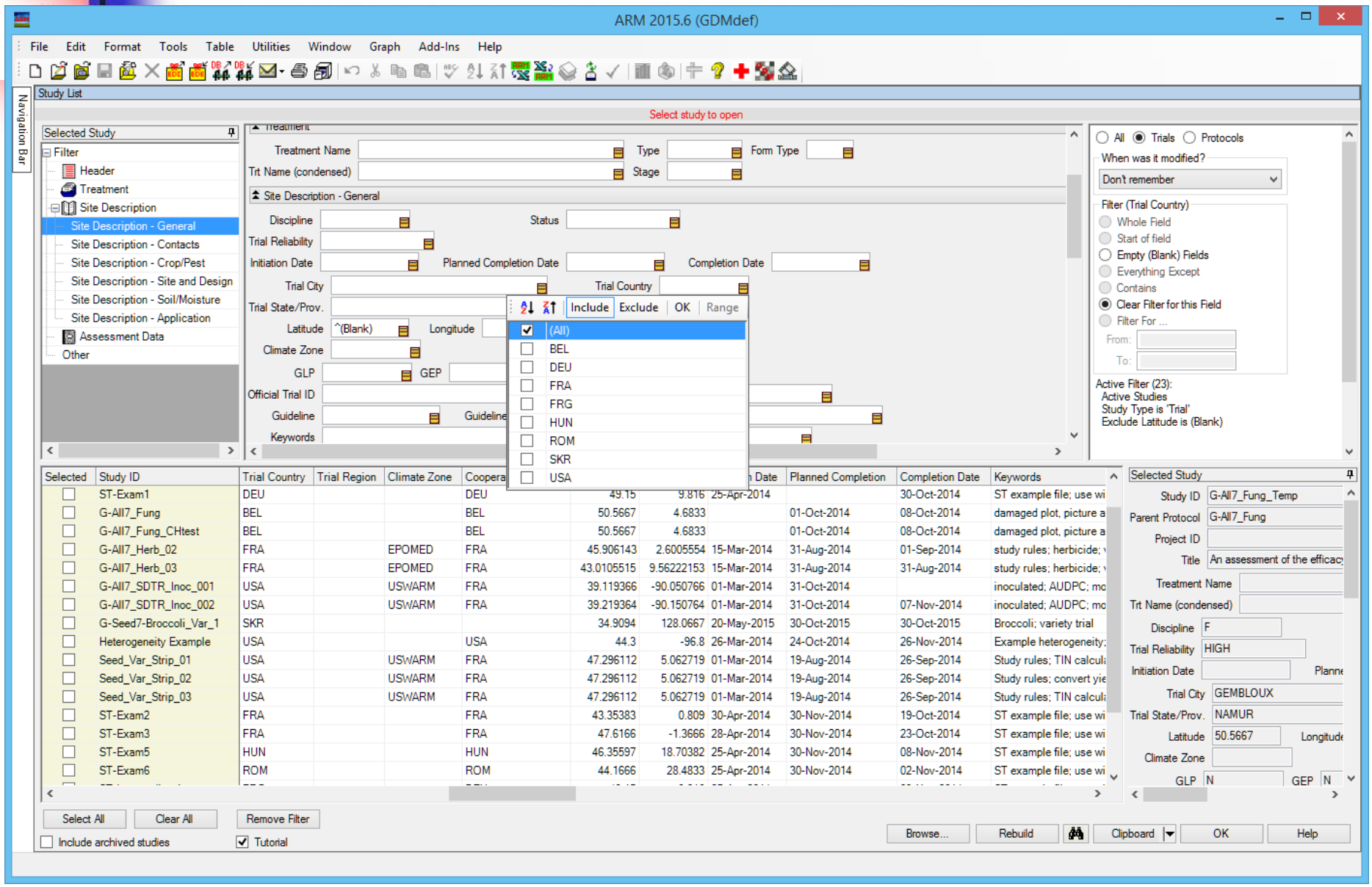
What About "Good Data"?

- Well documented data
- With a full description of assessment parameters
- Includes detailed crop information
- Complete trial site information:
  - Growing conditions (temperature, moisture)
  - Site, soil description and analyses
  - Trial maintenance activities

What About "Good Data"?

- Stored electronically in a readily useable standard format
- Indexed to quickly find relevant trials
- Using standardized terminology
- All values included defined units
- Can be quickly combined with similar trials

Study Management Tools



Study Management Tools

- Track progress of studies
- Search current and historical trials
- Extract information for mapping, etc.

Trial Management Steps

1. Prepare treatment/entry list
2. Plan required number of replicates by either: best guess, consult statistician, or perform power analysis
3. Create randomization
4. Define plot size
5. Calculate treatment quantity to apply

Trial Management Steps

6. Establish trial
   - Plant
   - Apply treatments
   - Record site location and other information
7. Make assessments
8. Review and analyze assessments
9. Prepare key graphs
10. Print final reports

Trial Management Steps

These steps are the same whether using:

- Paper
- Office/general purpose software
- Project management software
- ARM
  - Provides automations, and
  - Improves efficiency, quality, and consistency

Overview of Trial Management Software Requirements

- General Requirements:
  - Structure so trials are entered consistently
  - Dictionaries to standardize vocabulary
  - Enter information only once
- Resulting Benefits:
  - Portability across languages and platforms
  - Automation of routine tasks
  - Efficiency and accuracy

Protocol Components no. 1

- Clearly-defined treatments with formulation and rate details

Trt Line	Trt No.	Type	Treatment Name	Form Conc	Form Unit	Form Type	Rate	Rate Unit	Other Rate	Other Rate Unit	Appl Code	Appl Description
1	1	CHK	Untreated Check									
2	2	FUNG	TUB	250	G/L	EC	0.5	L/ha	125	g A/ha	A	pre-emergence
3	3	FUNG	TUB	250	G/L	EC	1	L/ha	250	g A/ha	A	pre-emergence
4	4	FUNG	TILT 250	250	G/L	EC	0.5	L/ha	125	g A/ha	B	early post
5	5	FUNG	MICO 60	600	G/L	EC	1.5	L/ha	900	g A/ha	B	early post
6	5	FUNG	FUNGOL	200	G/L	SC	1.25	L/ha	250	g A/ha	B	early post

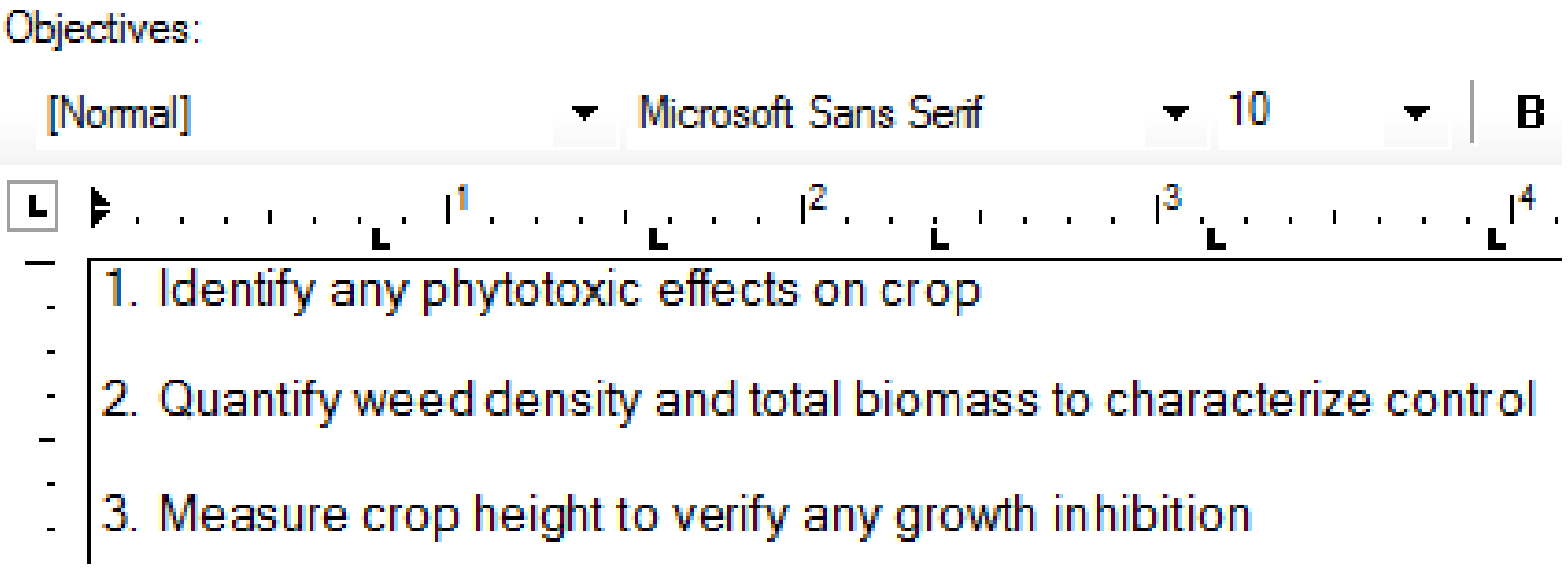
Protocol Components no. 2

- Description of required assessments

Assessment Data - Line 8								
Column Number	1		2		3		4	
Pest Type					W	*Weed*	W	*Weed*
Pest Name								
Crop Name								
Description	crop injury		crop height		weed density		weed biomass	
Rating Date								
Rating Type	PHYTO		HEIGHT		COUNT		BIOMAS	
Rating Unit	%		cm		PLANT		g	
Sample Size, Unit	1	PLOT	25	PLANT	1	m2	1	m2
Collection Basis, Unit								
Number of Subsamples	1		25		1		1	

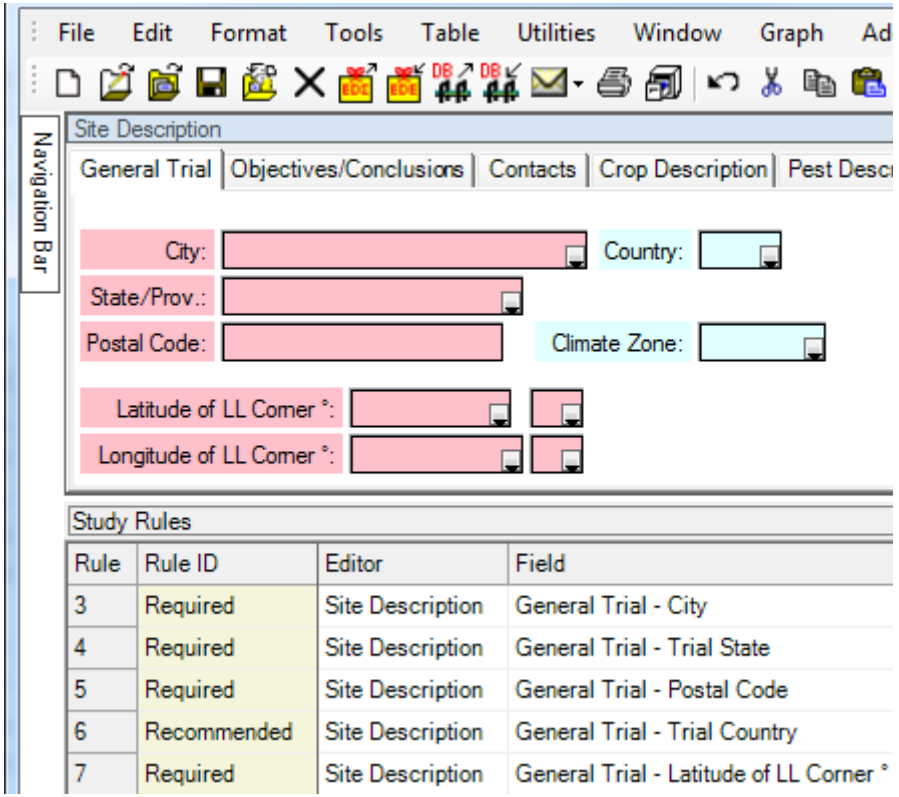
Protocol Components no. 3

- Define objectives



Protocol Components no. 4

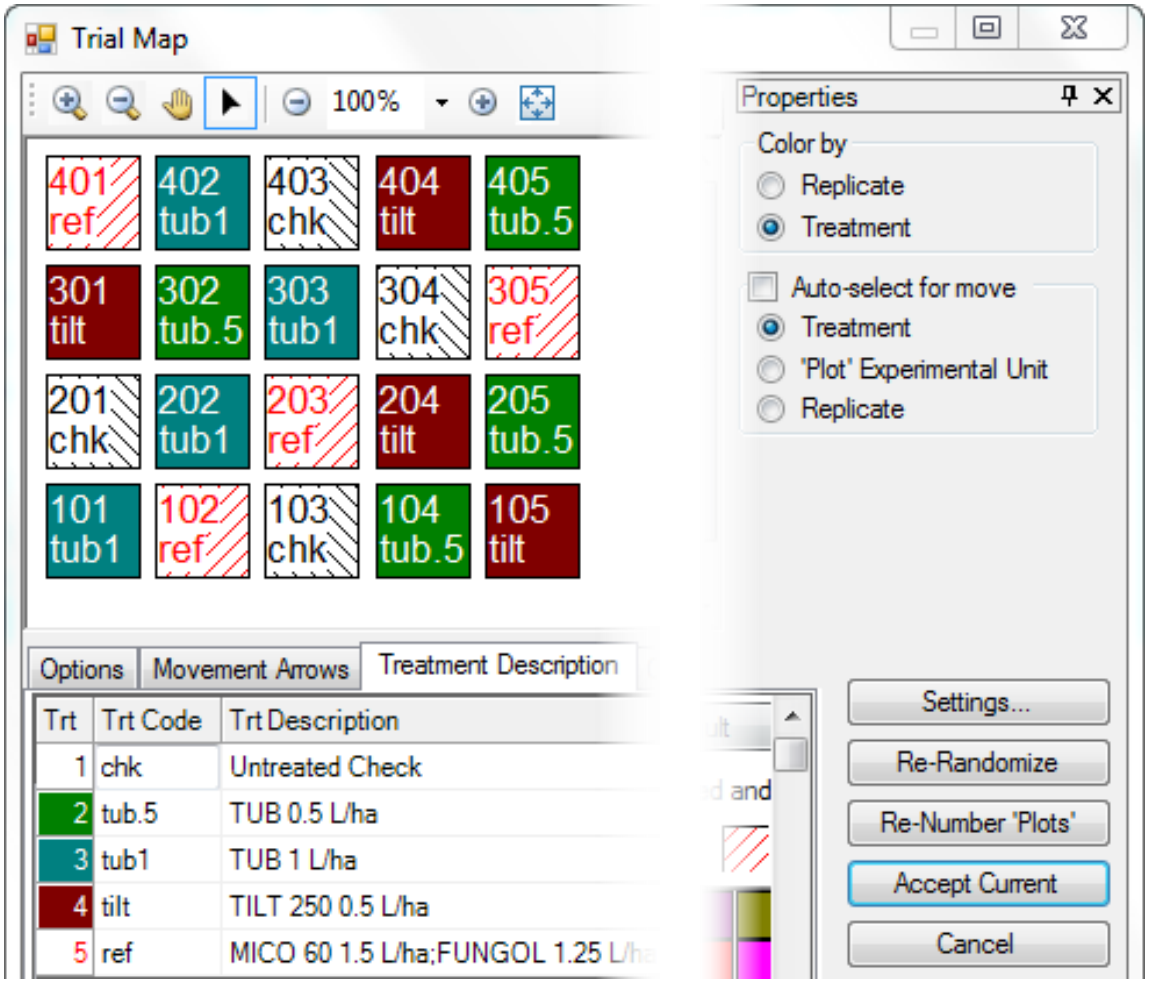
- Study rules that clearly identify key information to record in each trial created from the protocol



Support for Typical Experimental Designs

- Randomize and appropriately analyze
  - Completely Random Design
  - Randomized Complete Block (RCB)
  - Latin Square
  - Lattice Designs (Incomplete Block)
  - Multi-Factor Designs
    - RCB with Factorial Arrangement of Treatments
    - Split-Plot
    - Strip-Block (Criss-Cross)

Randomize Treatments



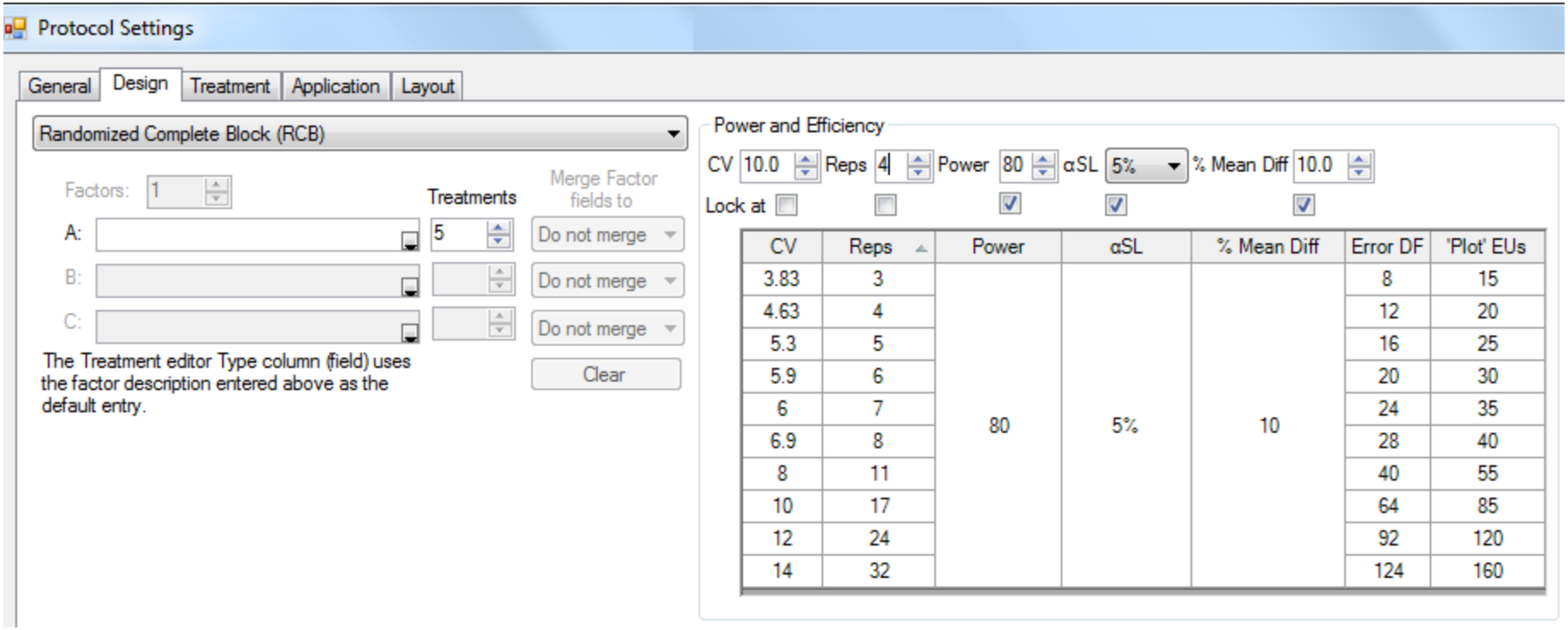
Power and Efficiency Planner, Plan Experiments to Have:

- A reasonable chance of distinguishing anticipated treatment differences
- The optimum number of replicates required to meet objectives
- An efficient experimental design and randomization for desired precision
- Cost-effective utilization of the available experimental area

Why is Planning Critical?

- Can reduce costs by selecting optimum number of replicates and samples
- Expected treatment differences are typically < 10%, and frequently < 5%, so small precision gains can help to:
  - Distinguish an actual treatment difference (reject null hypothesis H_0)
  - Strengthen evidence of no treatment diff.) (do not reject null hypothesis H_0)

Power and Efficiency Planner



CV	Reps	Power	αSL	% Mean Diff	Error DF	'Plot' EUs
3.83	3	80	5%	10	8	15
4.63	4				12	20
5.3	5				16	25
5.9	6				20	30
6	7				24	35
6.9	8				28	40
8	11				40	55
10	17				64	85
12	24				92	120
14	32				124	160

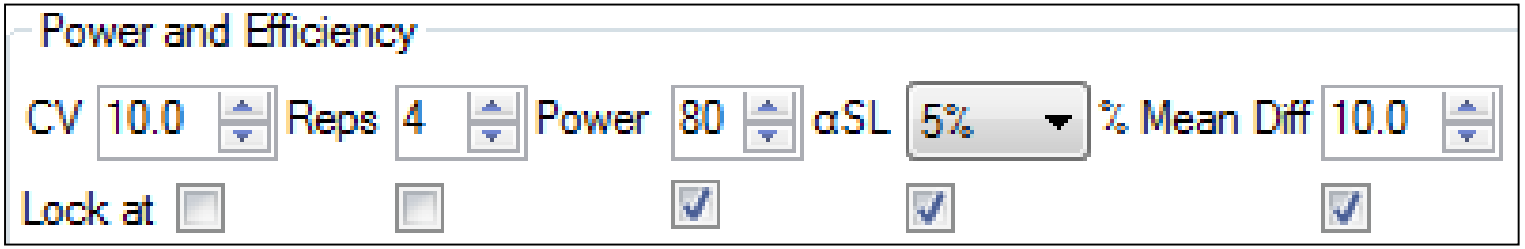
Power and Efficiency Planner

- Help plan experiments that successfully detect expected treatment differences
- Available in both protocols and trials so:
  - Protocol writers can more effectively plan experiments
  - Trialists can verify whether CV expectations are realistic based on local experience for specified crop(s)

Power and Efficiency Planner

- Calculates "power" based on:
  - Estimated CV of key assessment (e.g. yield)
  - Number of replicates
  - Power = Level of certainty to detect "real" treatment effects (80% or 90%)
  - Alpha Significance Level (e.g. 5%, 10%)
  - % Mean Diff = estimated treatment effect, expressed as percentage of overall (grand) mean across treatments of key assessment

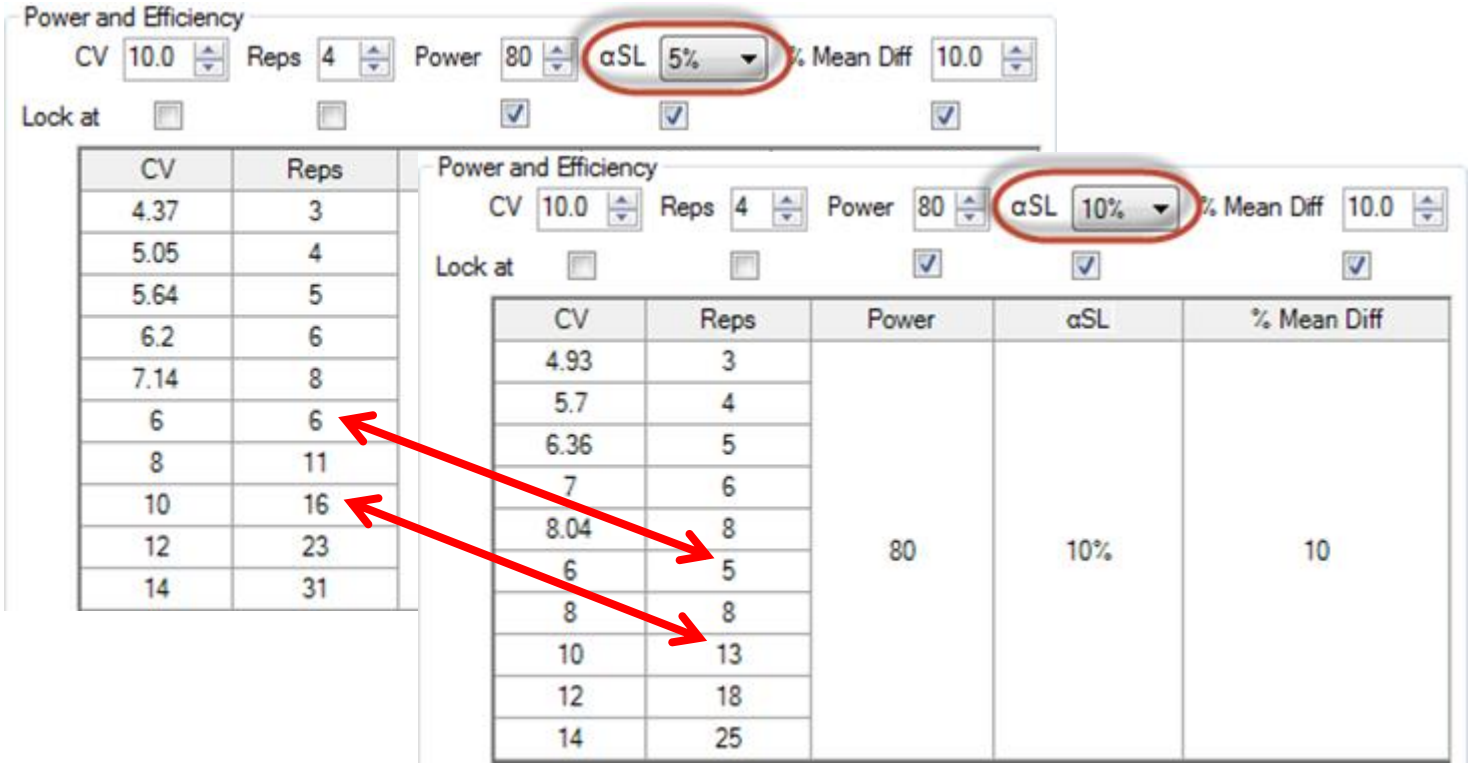
Power and Efficiency Planner



- "Lock at" to keep 3-4 columns constant
- Calculates table of possible values for "unlocked" columns (e.g. Rep or CV)
- Values entered by protocol writer are carried into trials created from protocol, conveying protocol expectations

Power and Efficiency Planner

Compare effect of significance level on minimum replicates for CV=6% vs. 10%



Power and Efficiency Planner

- Consider impact of Replicates on precision to detect treatment differences

Click on column heading to sort

Power and Efficiency

CV 5.0 Reps 4 Power 80 αSL 5% % Mean Diff 5.0

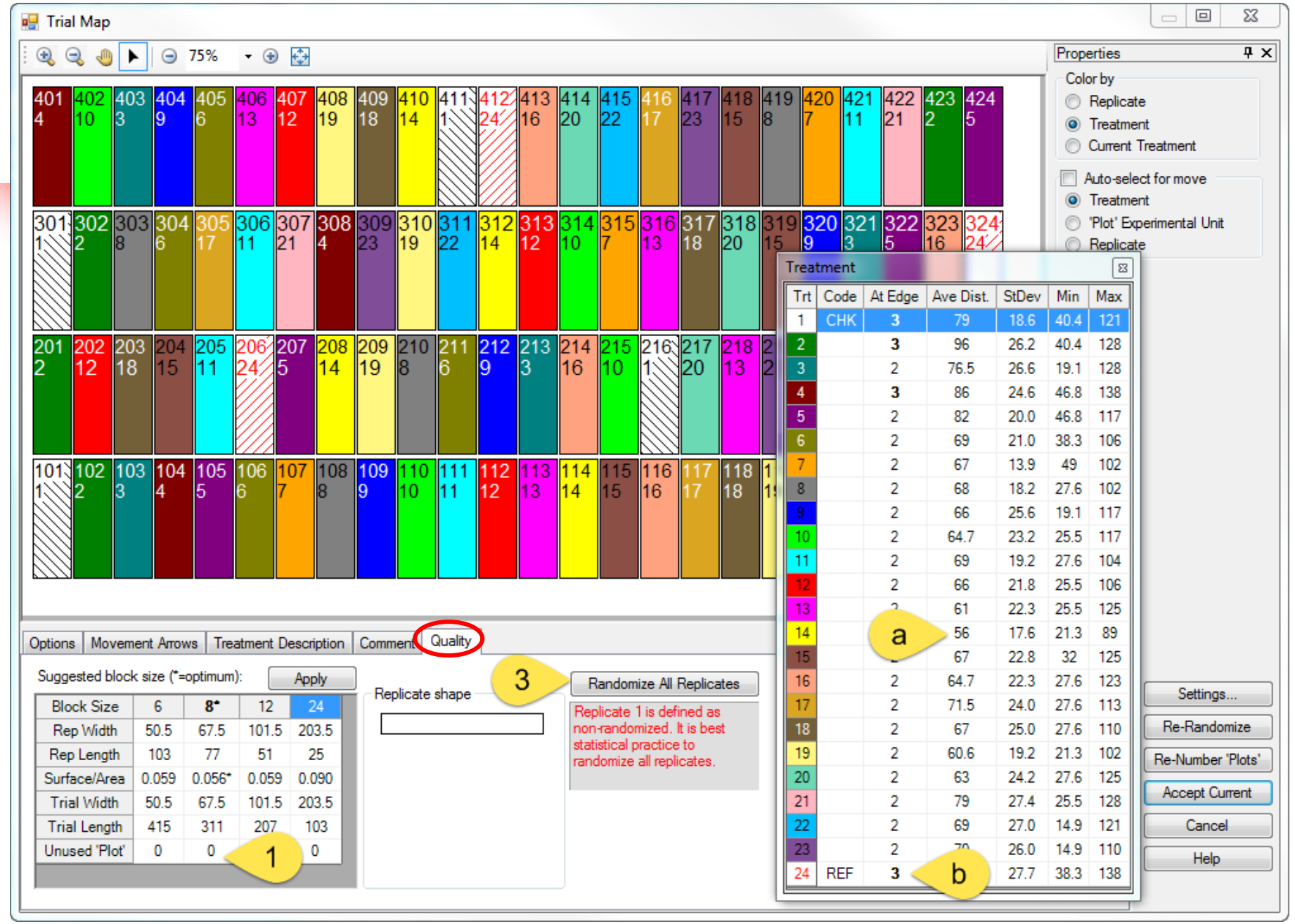
Lock at

CV	Reps	Power	αSL	% Mean Diff	Error DF	'Plot' EUs
5	3	80	5%	13	8	15
	4			10.8	12	20
	5			9.4	16	25
	6			8.5	20	30
	8			7.26	28	40
	9			7	32	45
	12			6	44	60
	17			5	64	85
	26			4	100	130
	45			3	176	225

Randomization Quality Review

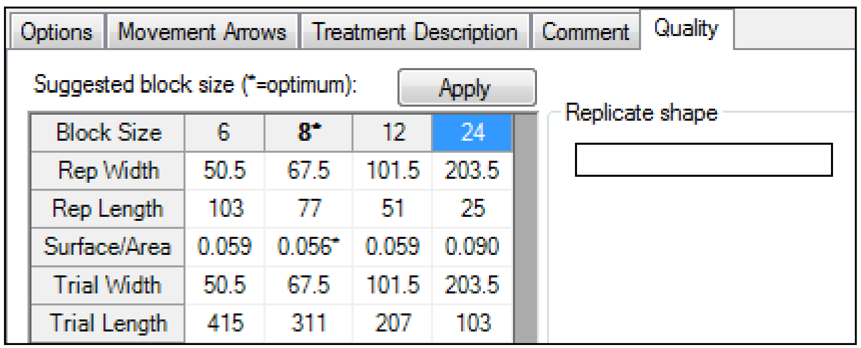
Goal is to improve experiment precision:

1. Arrange replicates as squares, not strips
2. Equalize treatment distribution
   a. Balance average distance from all other treatments
   b. Balance "Edge effect" across treatments
3. Randomize all replicates


Trial Map
75%
Properties
Color by
Replicate
Treatment
Current Treatment
Auto-select for move
Treatment
'Plot' Experimental Unit
Replicate
Treatment
Trt Code At Edge Ave Dist. StDev Min Max
1 CHK 3 79 18.6 40.4 121
2 3 96 26.2 40.4 128
3 2 76.5 26.6 19.1 128
4 3 86 24.6 46.8 138
5 2 82 20.0 46.8 117
6 2 69 21.0 38.3 106
7 2 67 13.9 49 102
8 2 68 18.2 27.6 102
9 2 66 25.6 19.1 117
10 2 64.7 23.2 25.5 117
11 2 69 19.2 27.6 104
12 2 66 21.8 25.5 106
13 61 22.3 25.5 125
14 56 17.6 21.3 89
15 67 22.8 32 125
16 2 64.7 22.3 27.6 123
17 2 71.5 24.0 27.6 113
18 2 67 25.0 27.6 110
19 2 60.6 19.2 21.3 102
20 2 63 24.2 27.6 125
21 2 79 27.4 25.5 128
22 2 69 27.0 14.9 121
23 2 26.0 14.9 110
24 REF 3 27.7 38.3 138
a
b
Options
Movement Arrows
Treatment Description
Comment
Quality
Suggested block size (*=optimum):
Apply
Block Size 6 8* 12 24
Rep Width 50.5 67.5 101.5 203.5
Rep Length 103 77 51 25
Surface/Area 0.059 0.056* 0.059 0.090
Trial Width 50.5 67.5 101.5 203.5
Trial Length 415 311 207 103
Unused 'Plot' 0 0 0
1
Replicate shape
3
Randomize All Replicates
Replicate 1 is defined as non-randomized. It is best statistical practice to randomize all replicates.
Settings...
Re-Randomize
Re-Number 'Plots'
Accept Current
Cancel
Help

Arrange Replicates as Squares not Strips

"Optimum" is smallest surface-to-area ratio

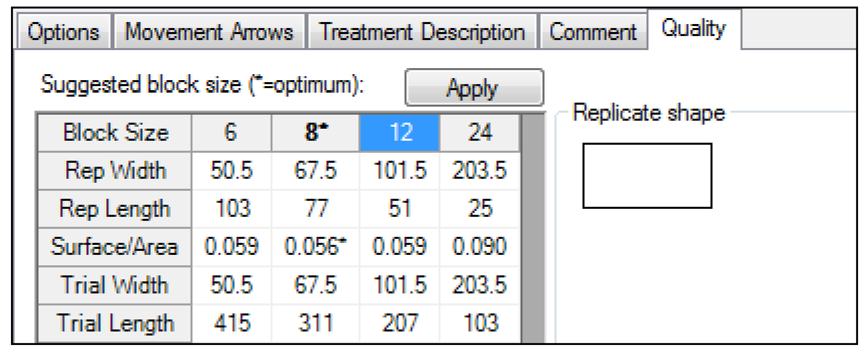


Options | Movement Arrows | Treatment Description | Comment | Quality

Suggested block size (*=optimum): Apply

Replicate shape

Block Size	6	8*	12	24
Rep Width	50.5	67.5	101.5	203.5
Rep Length	103	77	51	25
Surface/Area	0.059	0.056*	0.059	0.090
Trial Width	50.5	67.5	101.5	203.5
Trial Length	415	311	207	103

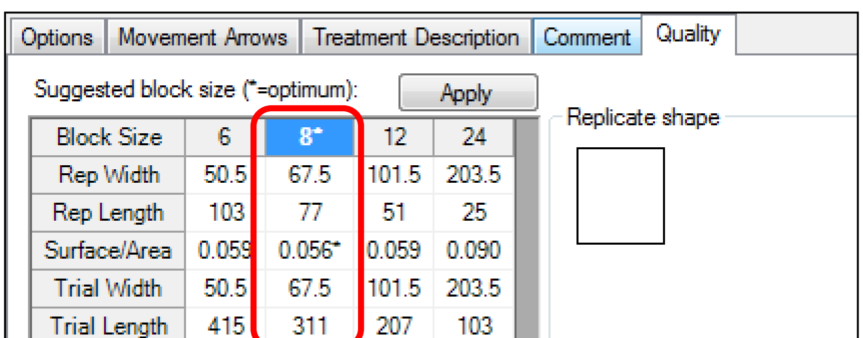


Options | Movement Arrows | Treatment Description | Comment | Quality

Suggested block size (*=optimum): Apply

Replicate shape

Block Size	6	8*	12	24
Rep Width	50.5	67.5	101.5	203.5
Rep Length	103	77	51	25
Surface/Area	0.059	0.056*	0.059	0.090
Trial Width	50.5	67.5	101.5	203.5
Trial Length	415	311	207	103

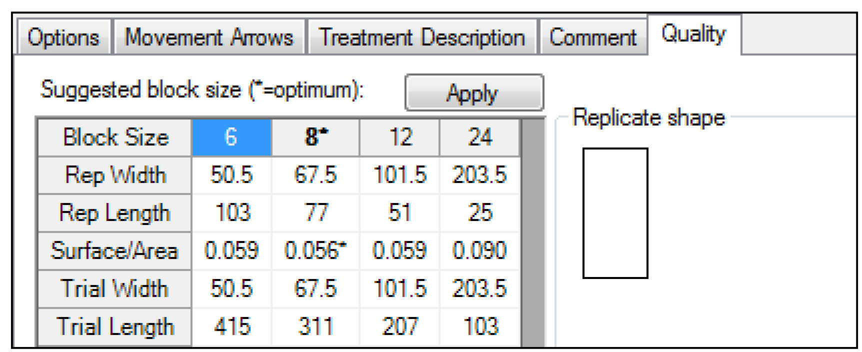


Options | Movement Arrows | Treatment Description | Comment | Quality

Suggested block size (*=optimum): Apply

Replicate shape

Block Size	6	8*	12	24
Rep Width	50.5	67.5	101.5	203.5
Rep Length	103	77	51	25
Surface/Area	0.059	0.056*	0.059	0.090
Trial Width	50.5	67.5	101.5	203.5
Trial Length	415	311	207	103



Options | Movement Arrows | Treatment Description | Comment | Quality

Suggested block size (*=optimum): Apply

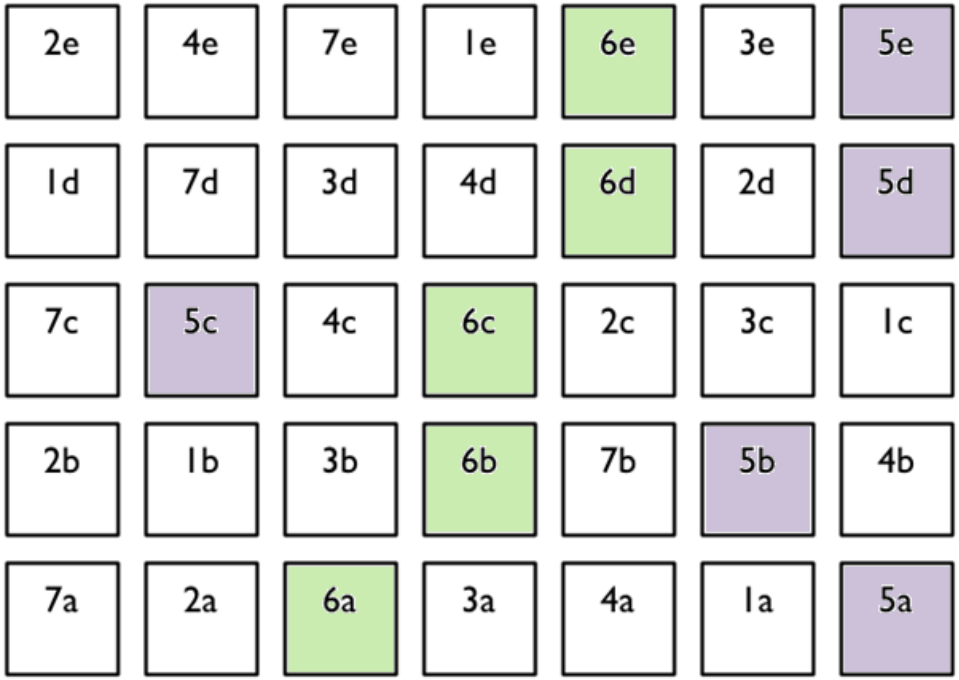
Replicate shape

Block Size	6	8*	12	24
Rep Width	50.5	67.5	101.5	203.5
Rep Length	103	77	51	25
Surface/Area	0.059	0.056*	0.059	0.090
Trial Width	50.5	67.5	101.5	203.5
Trial Length	415	311	207	103

Equalize Treatment Distribution

"Undesirable" layout of 7 treatments and 5 replicates in Randomized Complete Block:

- Trt. 6 in middle 3 columns of all reps
- Trt. 5 in right 2 cols for all but one plot

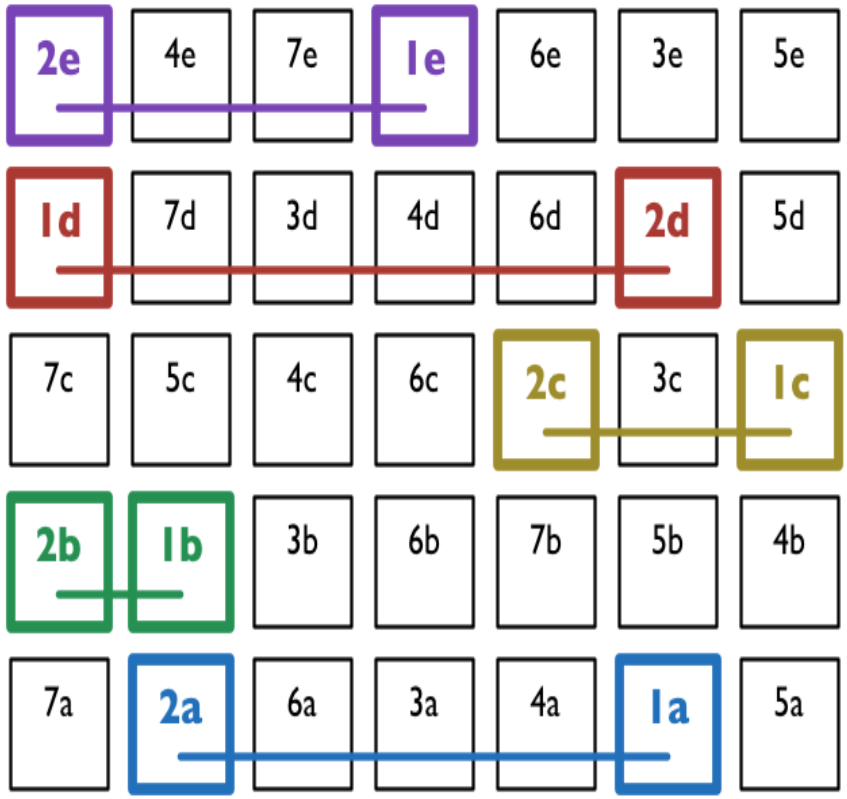


Uses "Average Distance of Treatment" Comparison (ADTC)

- van Es and van Es, "Spatial Nature of Randomization and Its Effect on the Outcome of Field Experiments", Agronomy Journal, 85:420-428 (1993).
- Comparison between treatments 1 and 2 is taken from 5 plots for each treatment.
- Measure the plot-to-plot distance for each plot containing treatment 1 to the paired plot within replicate containing treatment 2, for a total of 5 distances.
- ADTC for the treatment pair 1-2 is the average of the 5 distances.

Distances, Treatments 1-2

- Average distance = 3 plots = 24 feet for 8 foot wide plots



Unequal Treatment Distribution

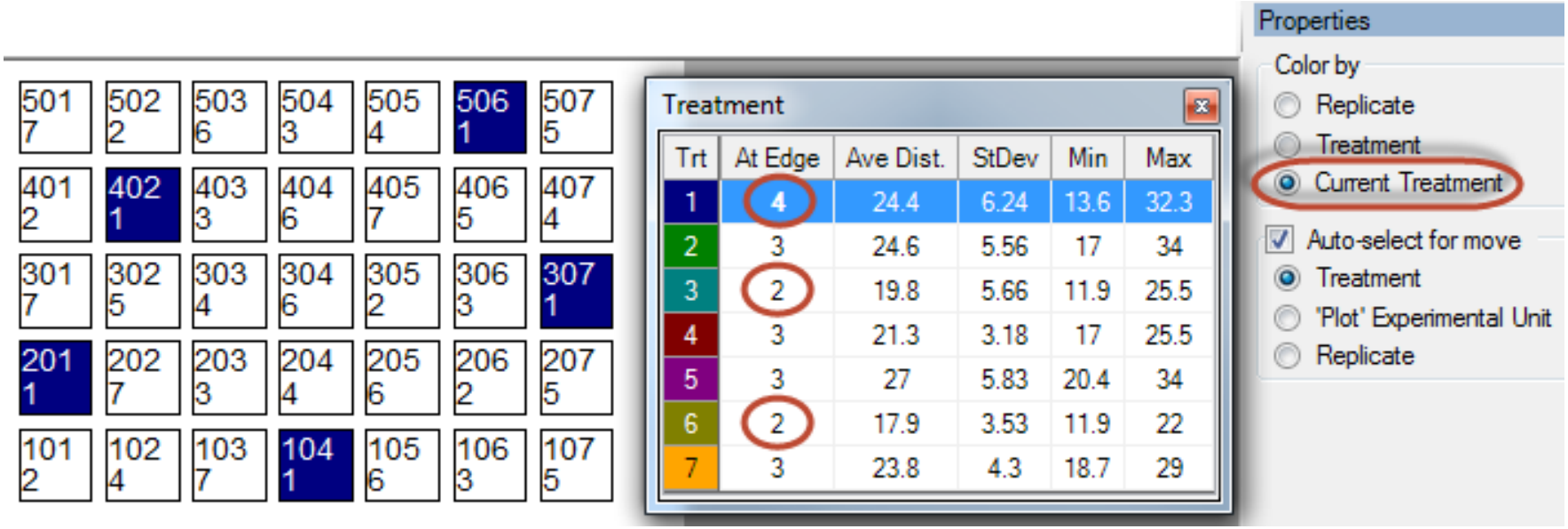
- Average distance from 17.9 to 24.6
- Ranges from 11.9(T3,T6) to 34(T2,T5)
- Error variances for treatments may not be homogeneous

Treatment

Trt	At Edge	Ave Dist.	StDev	Min	Max
1	4	24.4	6.24	13.6	32.3
2	3	24.6	5.56	17	34
3	2	19.8	5.66	11.9	25.5
4	3	21.3	3.18	17	25.5
5	3	27	5.83	20.4	34
6	2	17.9	3.53	11.9	22
7	3	23.8	4.3	18.7	29

Unbalanced "Edge effect"

- Treatment 1 occurs at edge 4 times, T2 and T3 at edge only 2 times



Balanced Treatment Distribution and Edge Effect

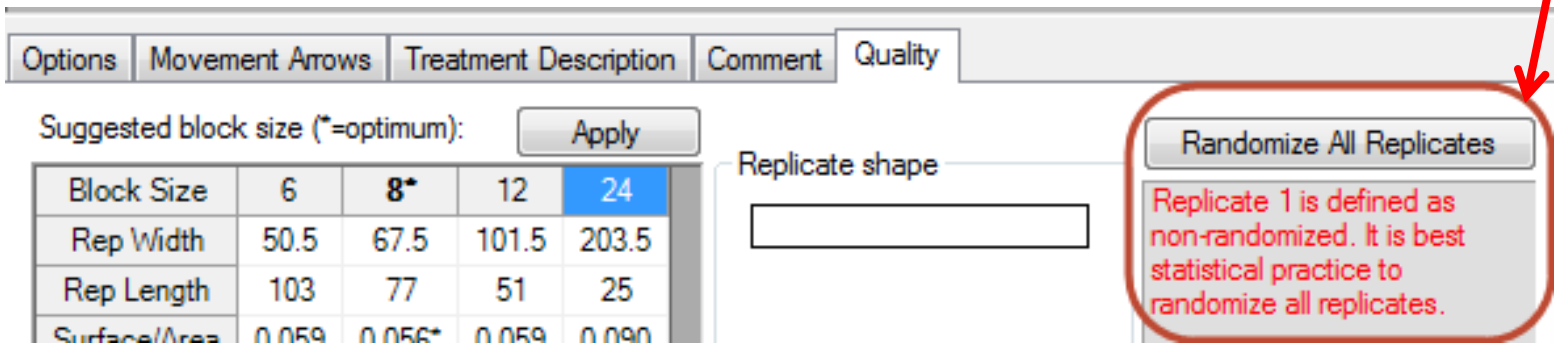
- Average distance from 21.3 to 24.4
- Distances range from 18.7 to 27.2
- "Edge effect" is balanced

Treatment

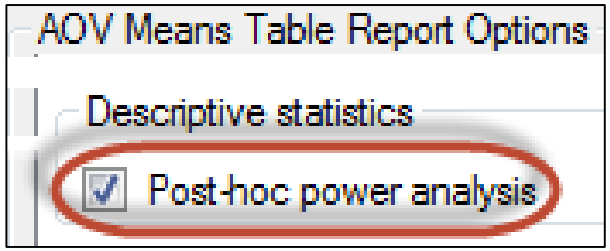
Trt	At Edge	Ave Dist.	StDev	Min	Max
1	2	22	2.15	20.4	25.5
2	3	23.8	3.57	18.7	27.2
3	3	24.4	1.76	22	27.2
4	3	22.4	3.47	18.7	25.5
5	3	22	3.4	18.7	27.2
6	3	21.3	2.58	18.7	25.5
7	3	22.7	2.56	18.7	25.5

Randomize All Replicates

- This frame displays when a non-randomized replicate is defined in Settings
- Select "Randomize All Replicates" to follow recommended statistical practice



Post-hoc Power Analysis



- Optional descriptive statistic on AOV Means Table report
- Lists, for each assessment column, the minimum number of replicates required to statistically separate treatment means based on Treatment P(F) and current significance level
- Use for planning future trials

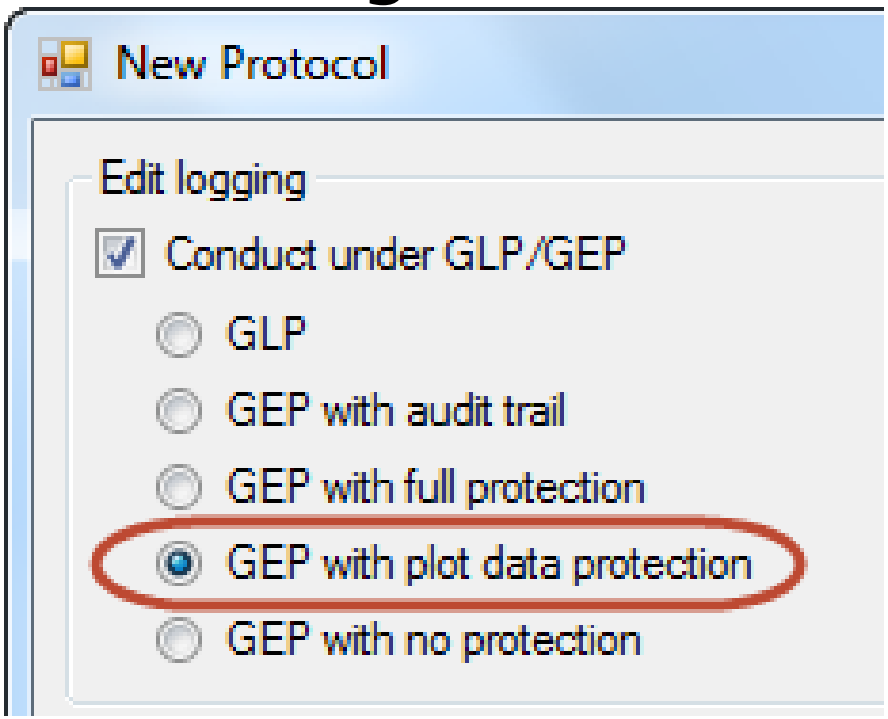
Post-hoc Power Analysis

- In example, LSD can distinguish 25% mean difference (largest existing difference is 18%)
- Current AOV Trt P(F) is 0.2979, so use 0.30+ significance level to separate treatment means
- Need 8+ replicates to reject null hypothesis at 0.05 significance

Crop Variety	CEZANNE
Trt No.	24
2	85.33 a
3	81.67 a
4	98.00 a
5	95.33 a
LSD P=.05 (% mean diff)	21.808 (25%)
Standard Deviation	10.915
CV	12.12
Grand Mean	90.083
Minimum Replicates (power = 80)	8
Largest Mean Difference (% mean diff)	16.333 (18%)
Treatment F	1.541
Treatment Prob(F)	0.2979

Data Protection

- Offer a protection that allows only trial owners to change assessment data

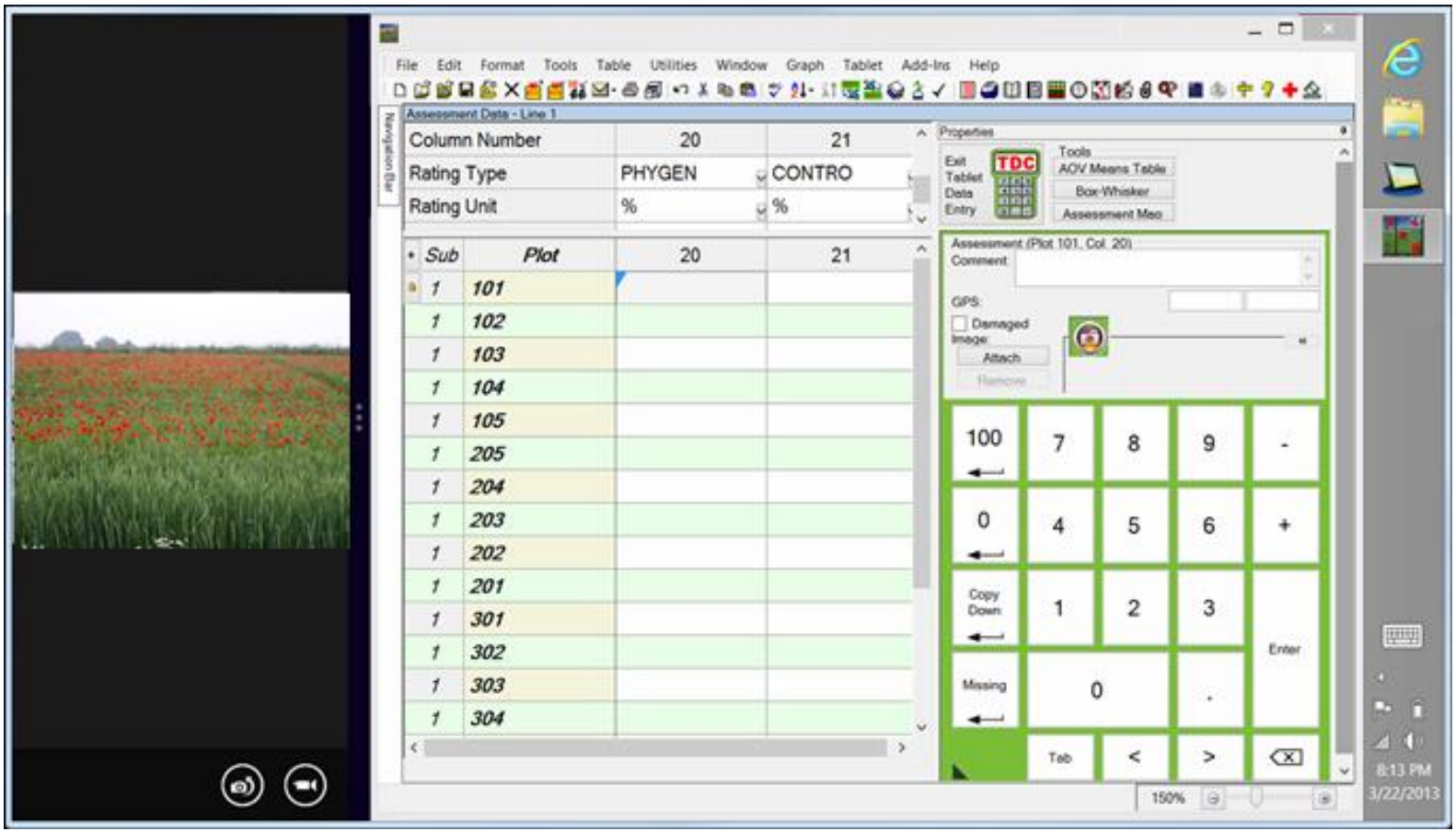


Robust Data Collection Tools

- Enter data only once to avoid transcription errors
- Employ appropriate range checking for assessed values
- Perform data quality checks before leaving trial site (analyze, graph)
- Include photographs that illustrate or support measurements & observations

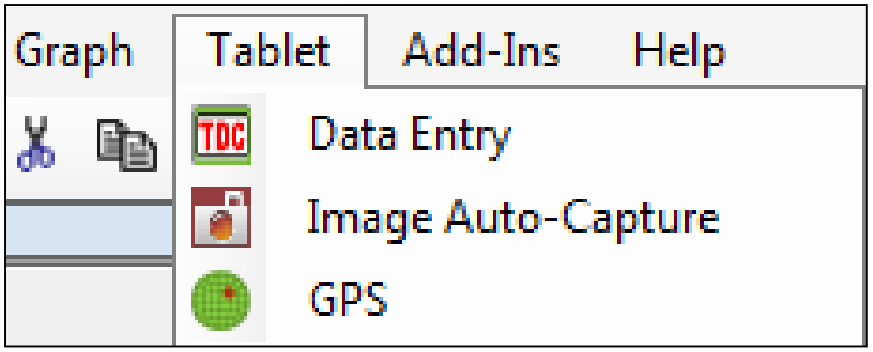
ARM Tablet Data Collector

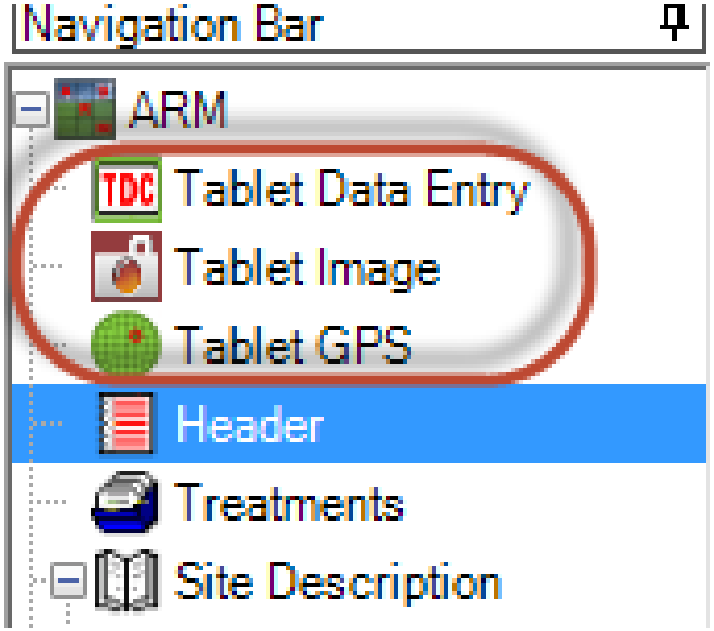
Win 8.1 Pro tablet + ARM + special ARM tablet features



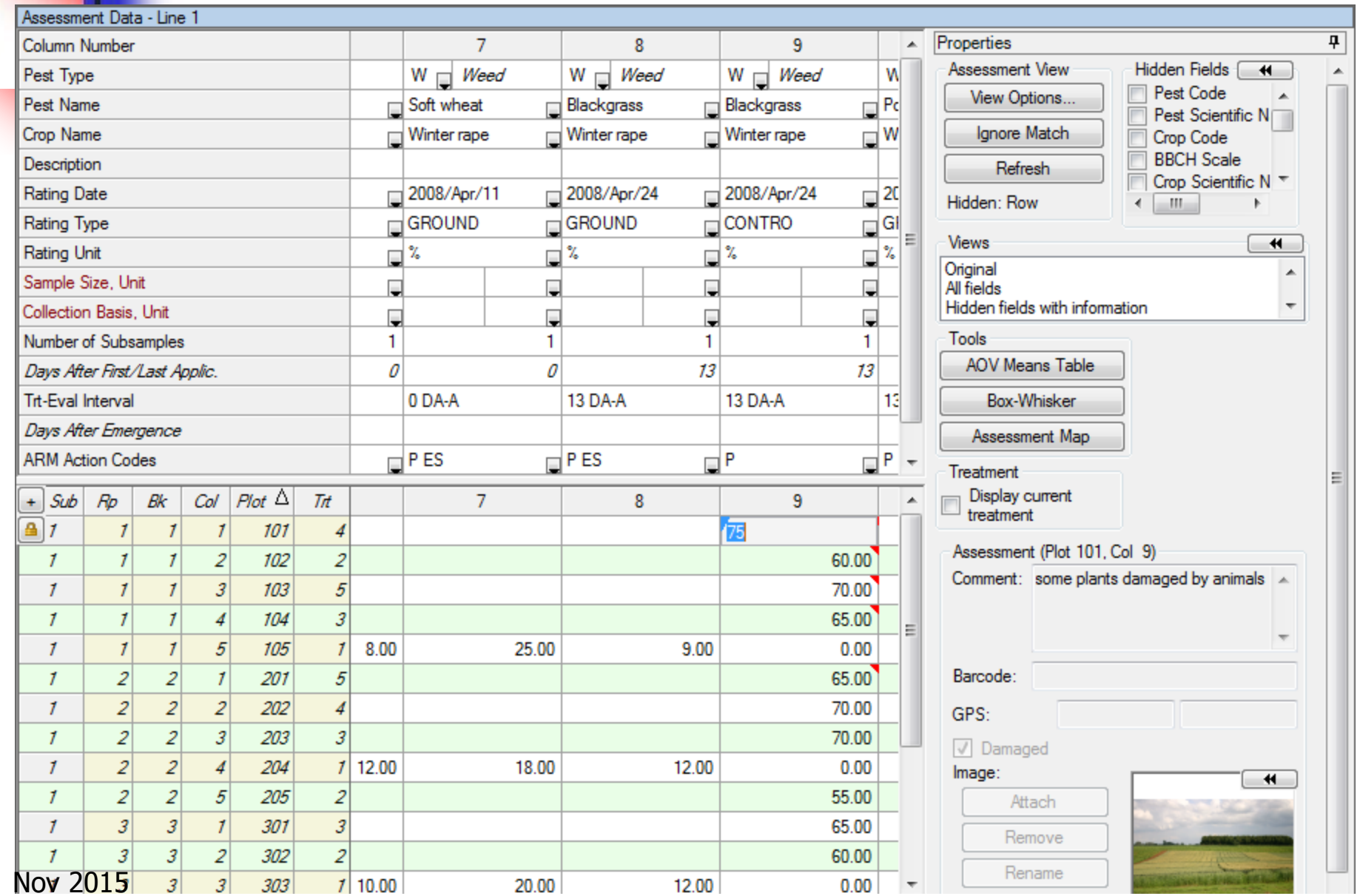
Tablet Data Collector Features

- Tablet Data Entry
- Tablet Image Capture
- Tablet GPS





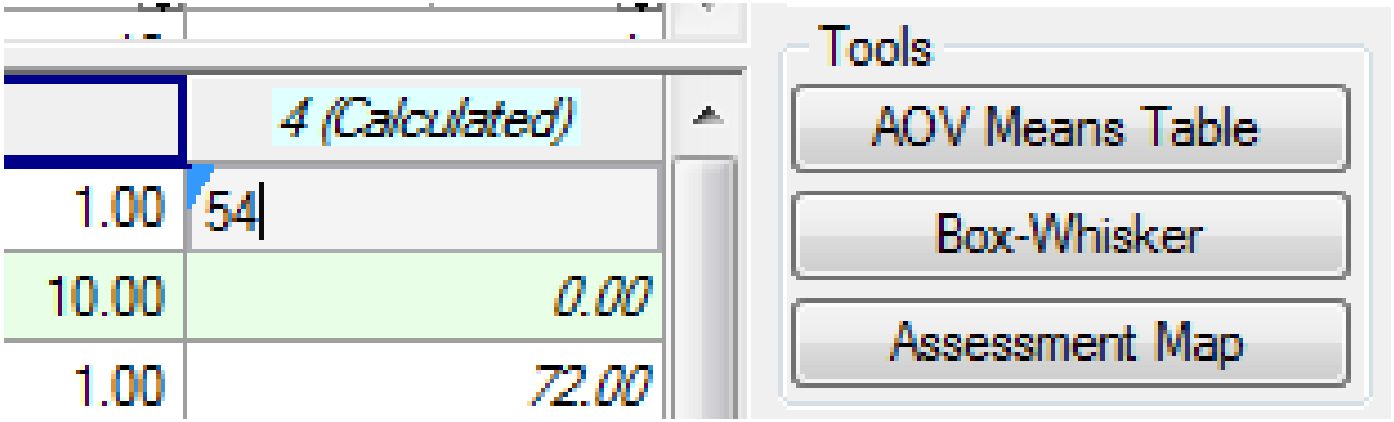
Assessment - Linked Image



Nov 2015

Assessment Review Tools

- Analysis of Data
- Graph of Variability
- Assessment Map (look for site effect)



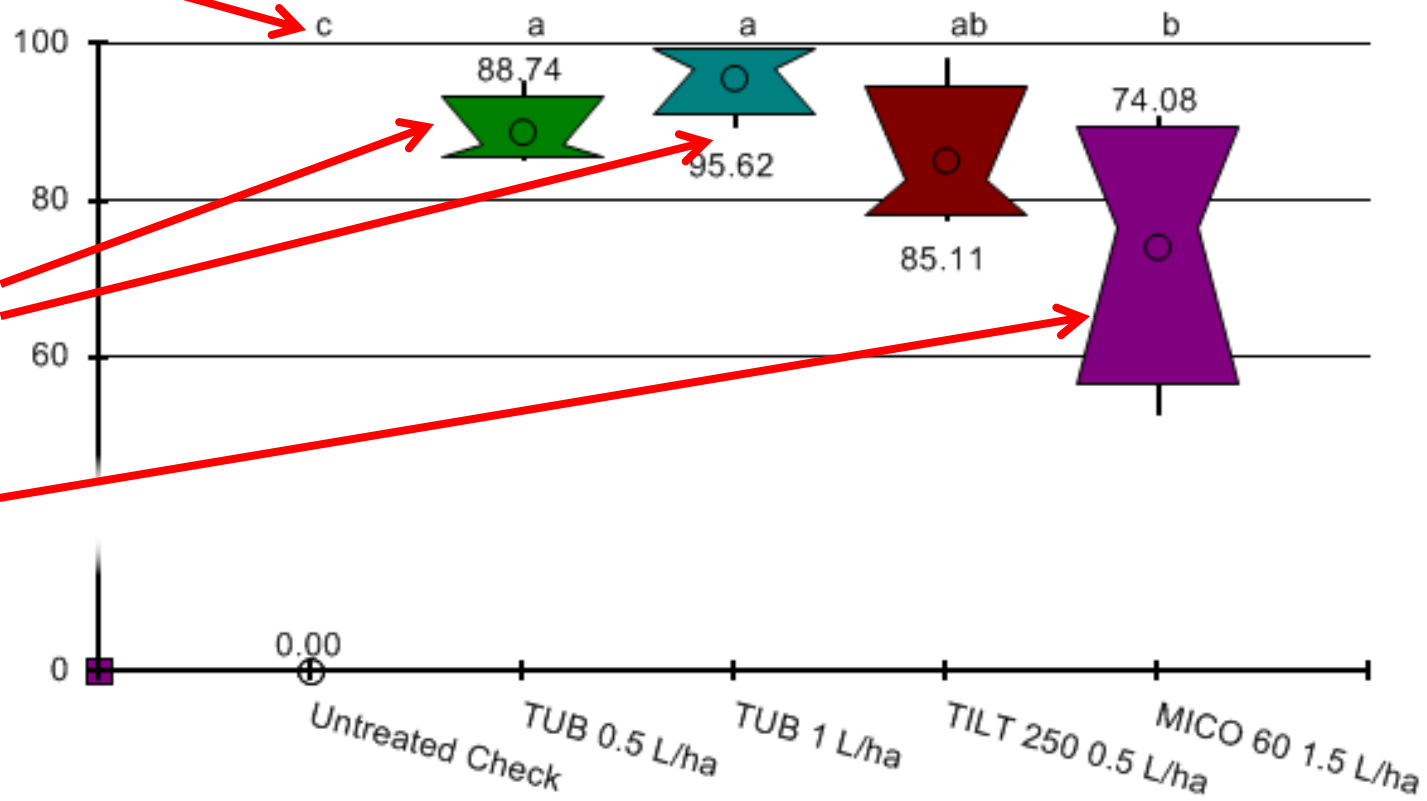
Analysis of Data

- Duncan's Test at 5%
- Coefficient of Variation

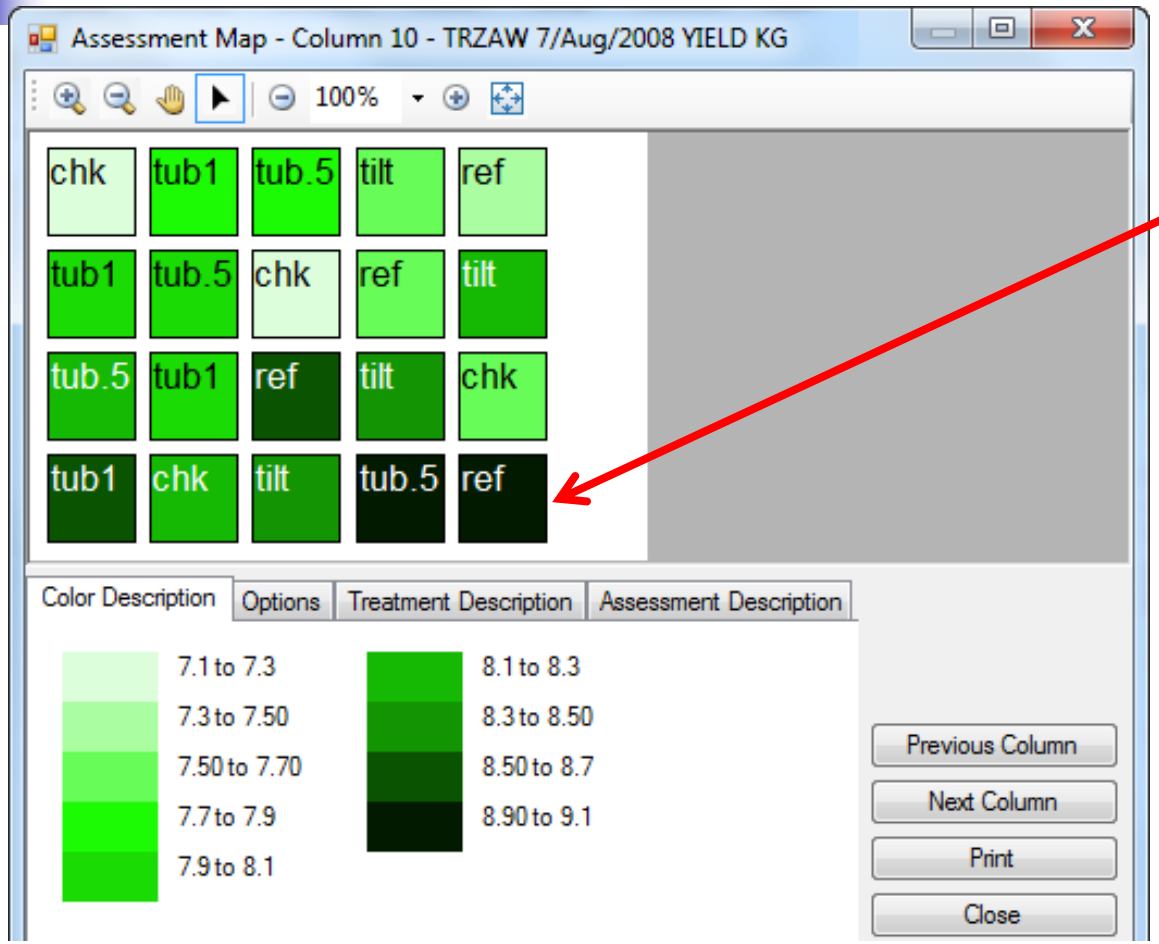
Pest Code				SEPTTR	SEPTTR
Description				severity	control
Rating Date				2/Jul/2008	2/Jul/2008
Rating Type				PESSEV	PESSEV
Rating Unit				%	%UNCK
Sample Size, Unit				10 LEAF	10 LEAF
Trt No.	Treatment Name	Rate	Rate Unit	Appl Code	7	8
1	Untreated Check			ABC	15.51 a	0.00 c
2	TUB	0.5	l/ha	ABC	1.74 b	88.74 a
3	TUB	1	l/ha	ABC	0.83 b	95.62 a
4	TILT 250	0.5	l/ha	ABC	2.35 b	85.11 ab
5	MICO 60	1.5	l/ha	AB	3.88 b	74.09 b
	FUNGOL	1.25	l/ha	C		
LSD (P=.05)					3.146	12.750
Standard Deviation					2.042	8.275
CV					42.01	12.04
Bartlett's X2					10.194	6.963
P(Bartlett's X2)					0.037*	0.073
Skewness					1.7361*	-1.3261*
Kurtosis					2.3213*	0.1148
Replicate F					4.360	2.117
Replicate Prob(F)					0.0270	0.1514
Treatment F					35.175	89.729
Treatment Prob(F)					0.0001	0.0001

Variability Graph (Box-Whisker)

- Duncan's Test at 5%
- Stable across replicates
- More variable across replicates



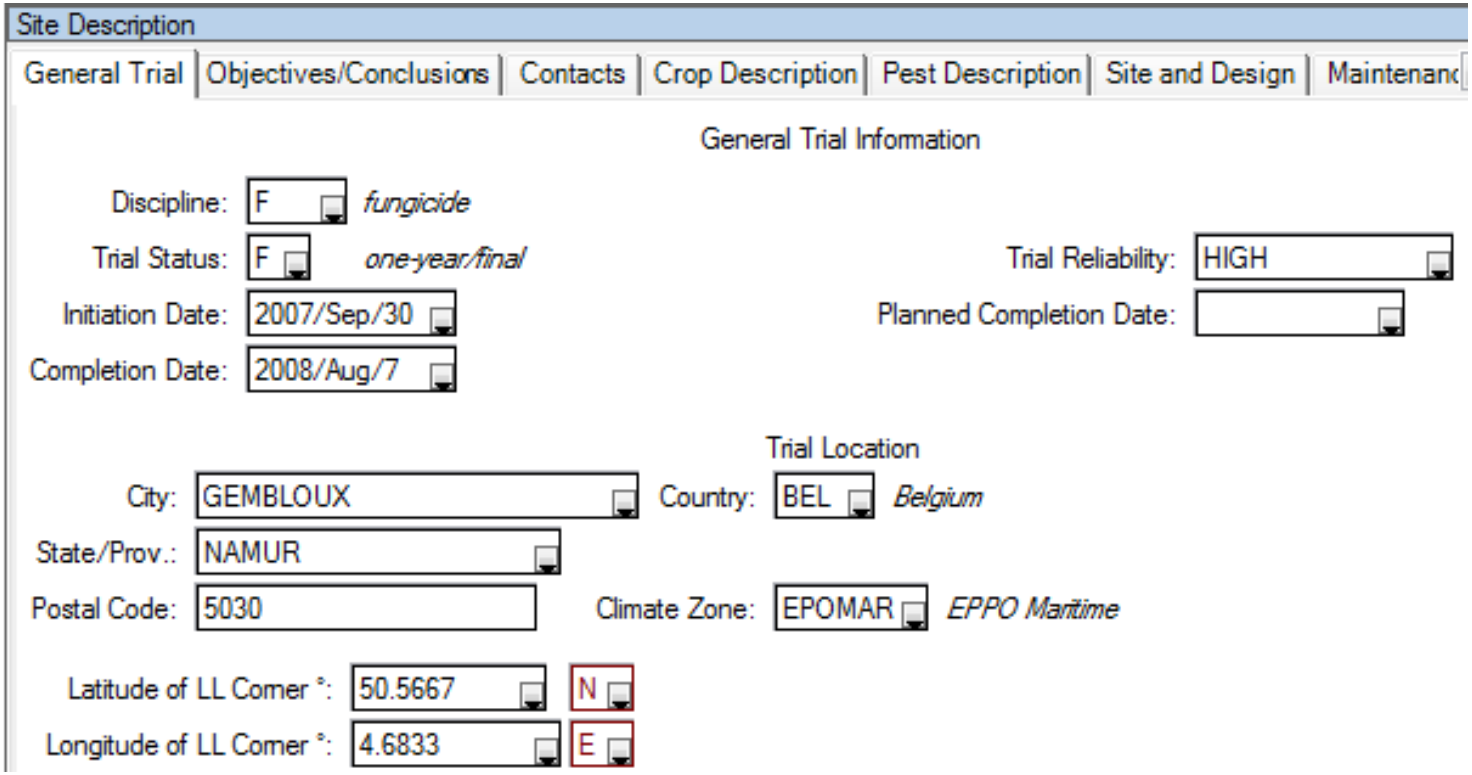
Assessment Map



Replicate 1 values darker (larger) than other replicates

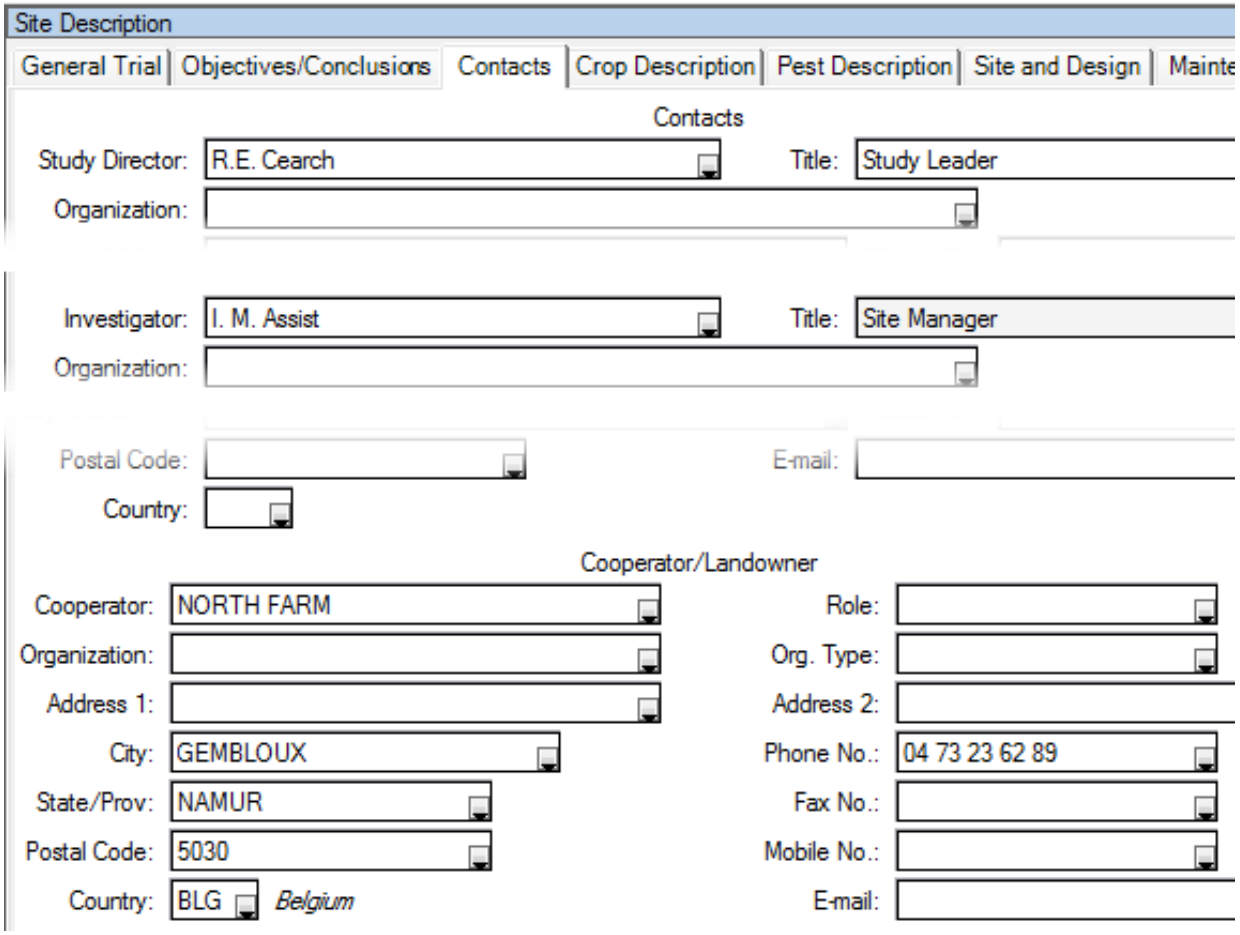
Site Information no. 1

- Trial Location



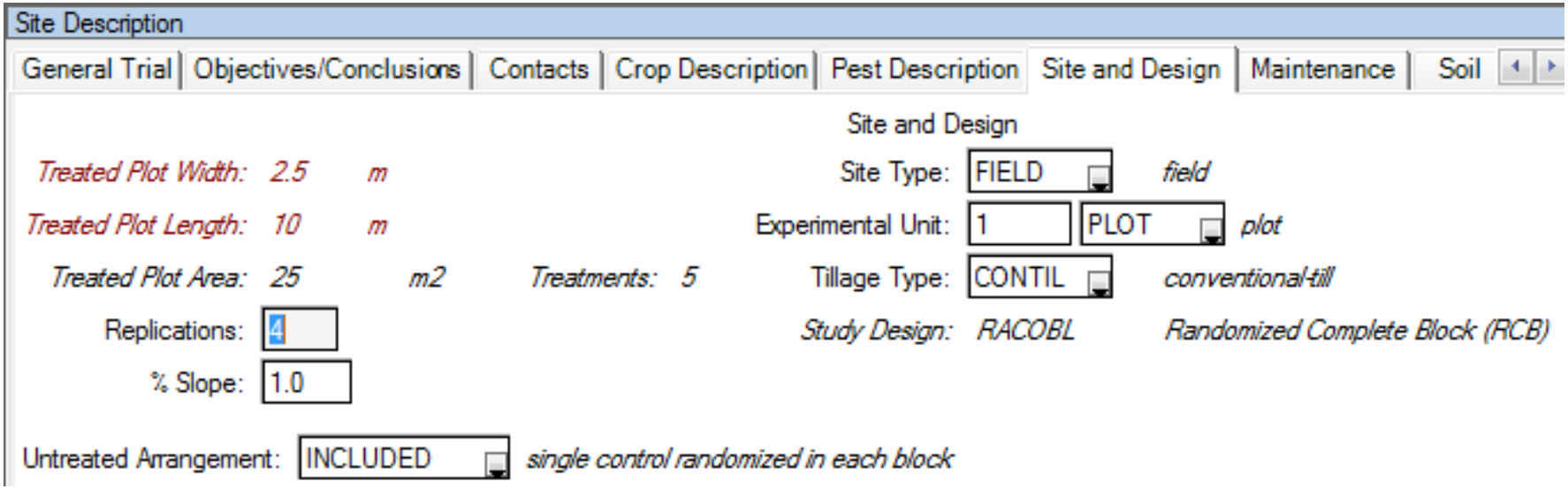
Site Information no. 2

- People



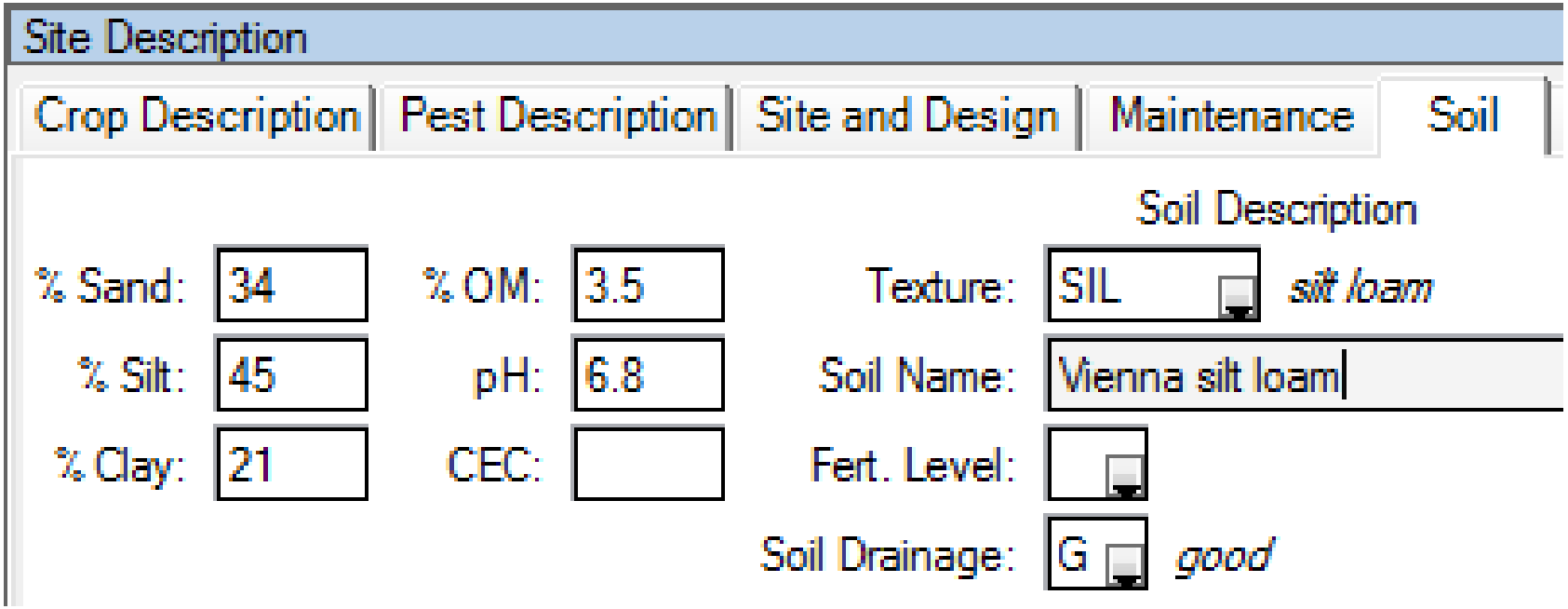
Site Information no. 3

- Site and Design



Site Information no. 4

- Soil



Site Information no. 5

- Application

Site Description

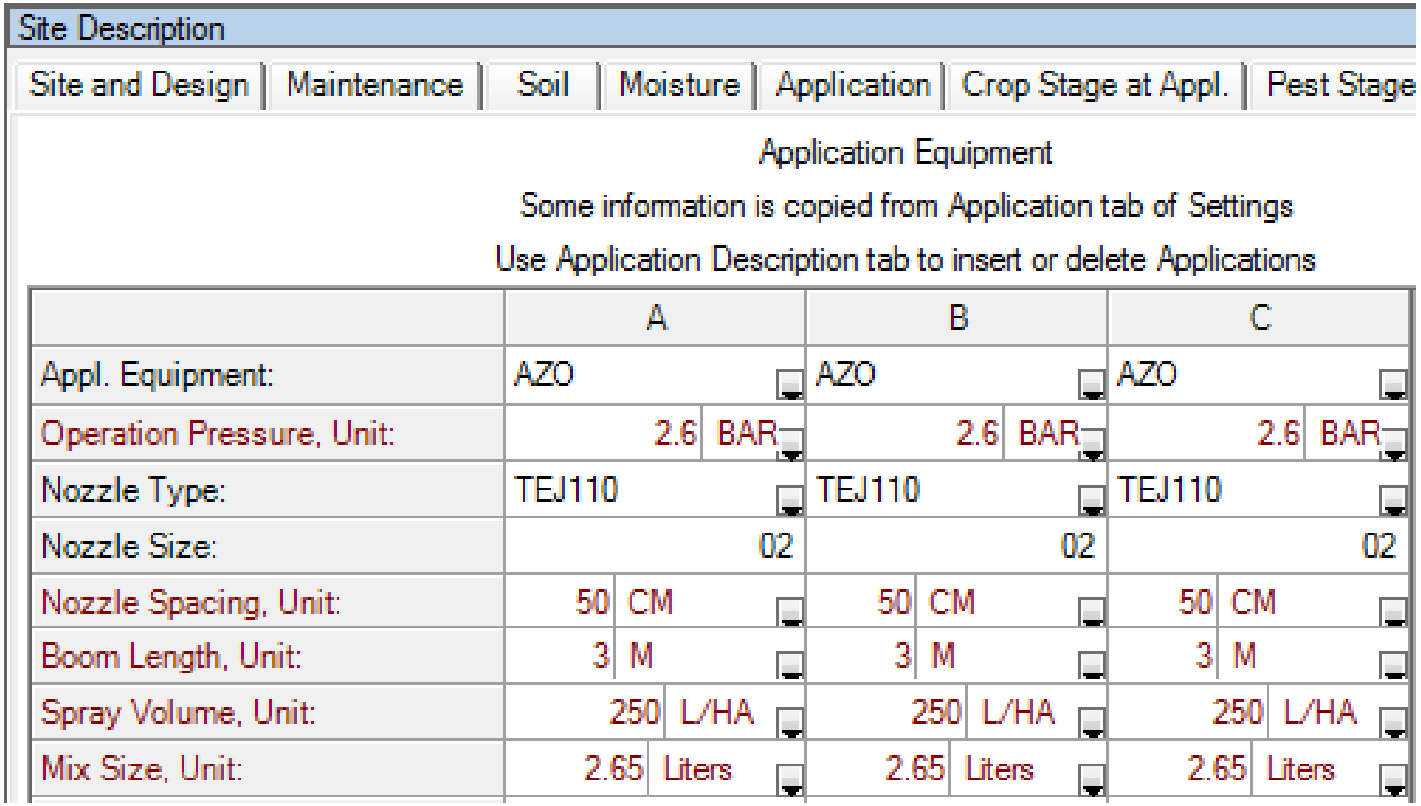
Contacts | Crop Description | Pest Description | Site and Design | Maintenan

Application Description

	A	B	C
Application Date:	2008/Apr/15	2008/Jun/3	2008/Jul/8
Appl. Start Time:	14:30	10:00	11:15
Appl. Stop Time:			
Application Method:	SPRAY	SPRAY	SPRAY
Application Timing:	POSPOS	POSPOS	POSPOS
Application Placement:	BROFOL	BROFOL	BROFOL
Applied By:			
Air Temperature, Unit:	17 C	17 C	19.5 C
% Relative Humidity:			
Wind Velocity, Unit:			
Wind Direction:			
Dew Presence (Y/N):			
Soil Temperature, Unit:	10 C	13 C	16 C
Soil Moisture:	MOIST	DRY	MOIST
% Cloud Cover:	50	20	10

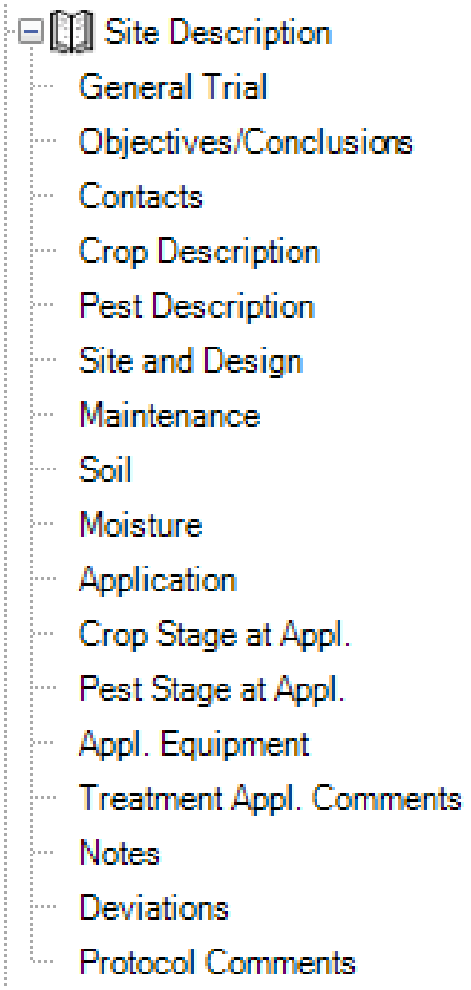
Site Information no. 6

- Application Equipment



Site Information no. 7

- Other site information as appropriate
  - Trial objectives and conclusions
  - Crop and pest details
  - Rainfall and irrigation
  - Notes and deviations from protocol



Management Reports

- Trial Map
- Applications: spray or seeding plan
- Plot Signs
- Site Information
- Labels: container, plot, seed, harvest
- Field Tour Sheet
- List of Treatments
- Statistical analysis of assessments

Label Reports

- Labels for:
  - Pre-measured quantity of products for each application, placed in small containers

--------- Container ----------
Container 1/Trt. Line
Brief Container 1/Trt. Line
Brief Cont. Spray Volume, 1/trt line
Brief Cont.,Material ID,SpVol, 1/trt line
Experimental Cont. 1/Line
Container 1/Trt. Line + Title
Container 1/Trt. Line + Appl
Container 1/Treatment
Container 1/Treatment + Title
Cont. Multi-Row Trt. 4"x2"
Cont. Multi-Row Lg. Plot #
Cont. Multi-Row Lg.#, File Name
Cont. Multi-Row Lg.#, Mix Size
Cont. Multi-Row Lg.#, Mix Details 99x68mm
Cont. Multi-Row Lg.#, Mix, Rate, Stage 4"x2"
Cont. Weight Audit (1 wide line)
Container Export (1 wide line)

Label Reports

- Labels for:
  - Identifying each plot

--------- Plot ----------
Plot # Spray Randomization (1/Trt.)
Large Plot # Spray Rand. (1/Trt.)
Plot 1/plot
Brief Plot 1/plot
Plot Soil Core Tube
AgCan Tyvec Plot 6"x3"
AgCan Tear-off Plot 6"x3"
AgCan Tear-off Plot (harvest order)
AgCan Brief Plot
Large Plot # Stake Label
Large Plot # Stake/no Trial ID
Large Bold Plot # Stake Label
Large Plot # Sample Label
Large Plot # Sample, bar code
Large Plot # Sample+Product, bar code
Plot Product Quantity
Plot Seed Tray (in trt. order)
Large Plot # Trt, Mix (Ridgetown)

Label Reports

- Labels for:
  - Identifying packets of seed to plant
  - Identifying small sacks of material harvested from each plot

------ Seed Packet ----------
Seed Packet (in trt. order)
Seed Packet (detailed, 1/Plot * Subs)
Seed Packet (detailed, 1/Plot, trt. order)
Seed Packet (detailed, 'n' blank pages)
Seed Packet (brief, 1/Treatment)
Seed Packet (brief, 'n' blank pages)
------ Harvest ----------
Plot Harvest (in harvest order)
Plot Harvest+Moisture,Weight fill-in
Plot Harvest+Range/Row, bar code
Harvest Bag (in harvest order)
Harvest Bag, bar code Plot (harvest order)
Harvest Bag (pooled, 1/Trt.)
Brief Harvest Bag (harvest order)
Brief Harvest Bag (pooled, 1/Trt.)
Brief Harvest Bag, bar code Trial,Trt,Plot
Brief Harvest Bag, bar code Trt,Plot
Brief Harvest Bag, bar code Plot
Brief Tear-off Harvest Bag 6.75cm x 5cm

Applications Report

- Spray/Seeding Plan

Product quantity to measure for each application rate

Trial ID: G-All7 Fung Location: Gembloux Trial Year:

Appl. no. 1

Reps: 4 Appl Code: A Plots: 2.5 by 10 meters
Spray vol: 200 L/ha Mix size: 2.15 liters (min 2.15)

Trt No.	Treatment Name	Form Conc	Form Unit	Form Type	Rate	Rate Unit	Appl Code	Spray Volume	Volume Unit	Mix Size	Mix Unit	Amt Product to Measure	Rep 1	2	3	4
3	TUB	250	G/L	EC	1	l/ha	ABC					10.75 ml/mx	101	202	301	402
1	Untreated Check						ABC						102	205	303	401
4	TILT 250	250	G/L	EC	0.5	l/ha	ABC					5.375 ml/mx	103	204	305	404
2	TUB	250	G/L	EC	0.5	l/ha	ABC					5.375 ml/mx	104	201	302	403
5	MICO 60	600	G/L	EC	1.5	l/ha	AB	250	L/HA	2.65	Liters	15.9 ml/mx	105	203	304	405

Appl. no. 2

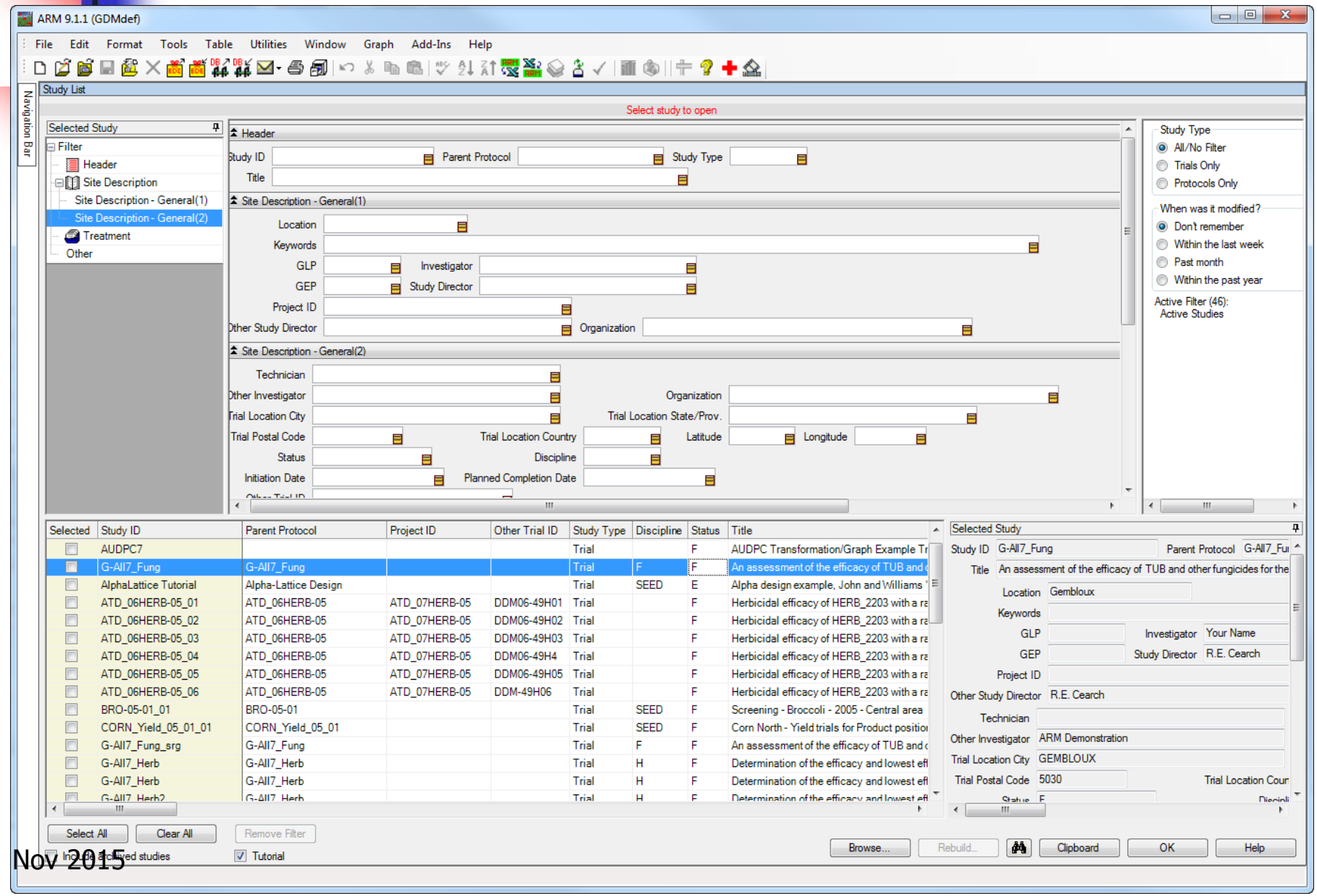
Reps: 4 Appl Code: B Plots: 2.5 by 10 meters
Spray vol: 200 L/ha Mix size: 2.15 liters (min 2.15)

Trt No.	Treatment Name	Form Conc	Form Unit	Form Type	Rate	Rate Unit	Appl Code	Spray Volume	Volume Unit	Mix Size	Mix Unit	Amt Product to Measure	Rep 1	2	3	4
3	TUB	250	G/L	EC	1	l/ha	ABC					10.75 ml/mx	101	202	301	402
1	Untreated Check						ABC						102	205	303	401
4	TILT 250	250	G/L	EC	0.5	l/ha	ABC					5.375 ml/mx	103	204	305	404
2	TUB	250	G/L	EC	0.5	l/ha	ABC					5.375 ml/mx	104	201	302	403
5	MICO 60	600	G/L	EC	1.5	l/ha	AB	250	L/HA	2.65	Liters	15.9 ml/mx	105	203	304	405

Data Analysis Reports

- Choices of different statistical methods:
  - Assessment Data Summary
  - AOV Means Table
  - Factorial AOV
  - Correlations
  - Dose-Response

Study Management Tools



Study Management Tools

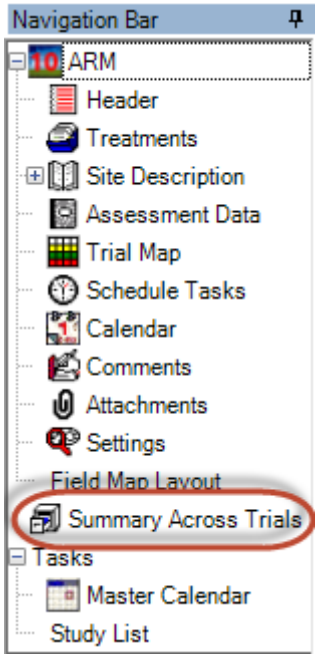
- Track progress of studies
- Search current and historical trials
- Extract information for mapping, etc.

Multi-Trial Summary

- Tools to analyze experiments over locations and years
- Easy selection of trials, treatments, and assessments to include
- Automated statistical analysis

ARM Summary Across Trials

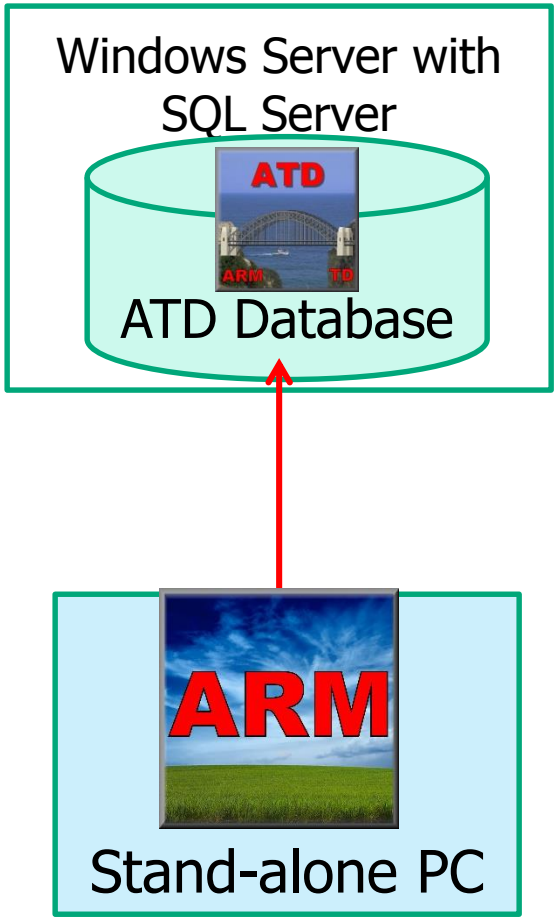
- Optional ARM add-in to summarize a trial series over locations and years
- Summarize selected treatments/entries across a wide range of trials
- View and arrange summary on a grid
- Export the report to Word, Excel, PDF
- Data graphs of across-trial means, trial clusters
- Export raw data to other statistics software



ATD Trial Database

- ARM clients connect to ATD Backend database either directly over a local network, or remotely over VPN

- Authorized ARM clients export trials to SQL Server ATD Backend database that resides on the shared server

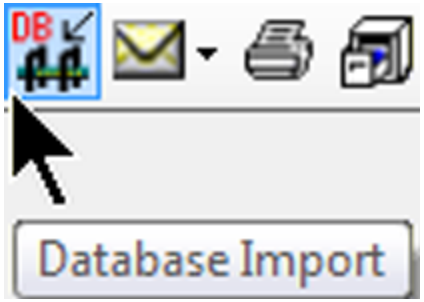


Using ATD with ARM

- Authorized ARM users export trials to ATD database using "Database Export" button on ARM toolbar
- All ARM users who install the ATD connection can import trials from ATD using "Database Import" to select 1 or more trials to import from Backend database (interface is similar to ARM study list)

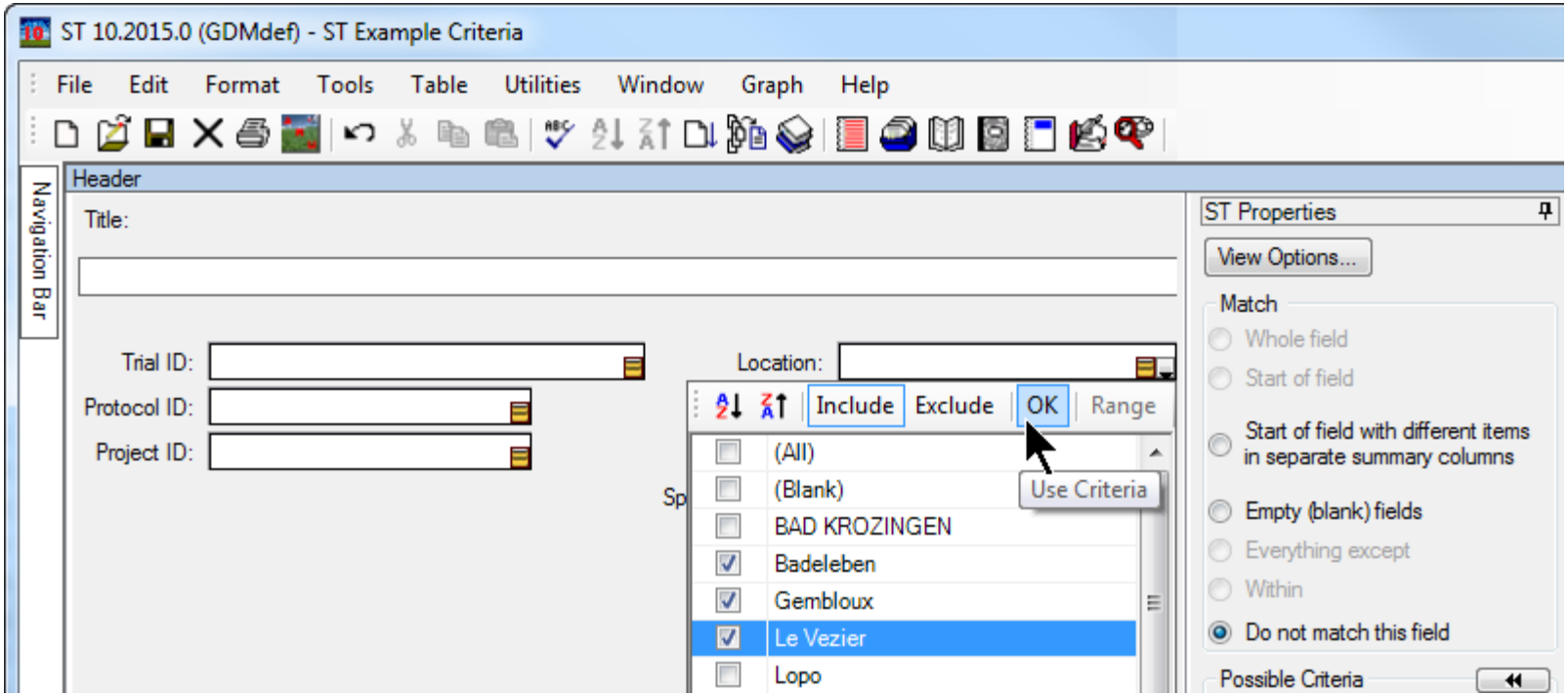
Using ATD with ARM

- ARM trials imported from ATD Backend database can be used in ARM like any standard trial: reviewing, graphing, and analyzing assessment data, or printing reports



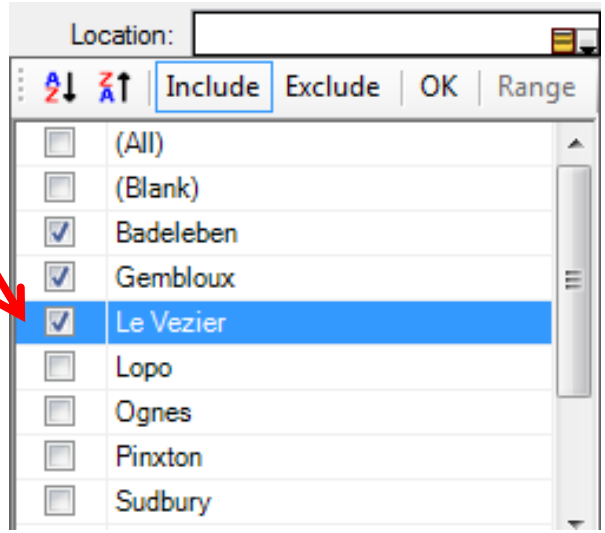
Using ATD with ST

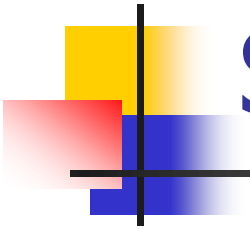
ST criteria/query selection screen connects directly to ATD



Using ATD with ST

- ST is the query and multi-trial summary interface for ATD.
- Select one or more field entries from drop-down lists showing unique field entries in ATD for the current ARM entry field.





Software Must Always "Grow"

- As research methods and objectives change and improve, software must also adapt to support those new research objectives and methods.
- "Unchanging" software:
  - Becomes less useful each year.
  - Can be costly by "losing" (not supporting) information gathered with new technology.