College of Food, Agricultural and Natural Resource Sciences
UNIVERSITY OF MINNESOTA

# Alternative Statistical Analysis Options for Unreplicated Paired Design Experiments

Paulo H. Pagliari and Jeffrey S. Strock
Department of Soil, Water, and Climate, University of Minnesota

## SITE DESCRIPTION

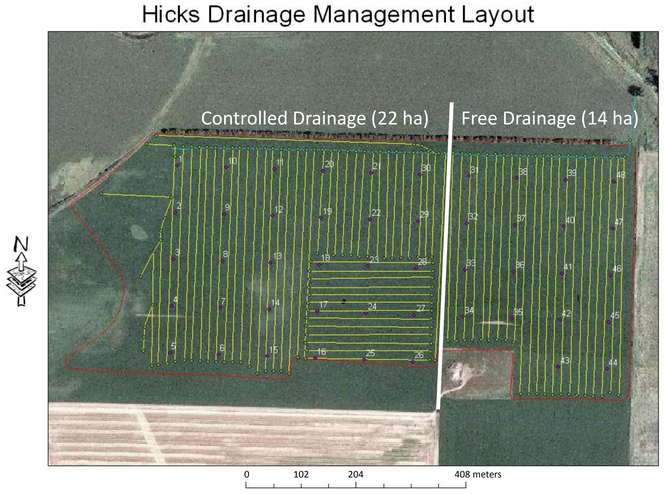
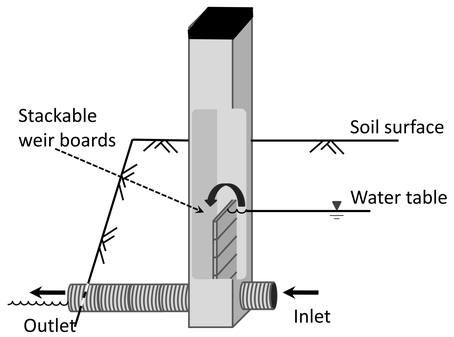



Figure 1. Areal picture of the farmers field were the experiment was conducted.

- This experiment was conducted at one farmers field totaling 36 hectares (ha) without replications.
- The field was split into two sites:
  - one site was 22 ha and was designated to be the controlled drainage plot.
  - the second site was 14 ha and was designated to be the free drainage plot.
- The tile drainage system was installed in 2006.
- Calibration Period:
  - In 2006 both sites were managed as free drainage.
  - In 2007 both sites were managed as controlled drainage.
- Treatment Period:
  - In 2008 and 2009 the controlled drainage site was managed as controlled drainage and the free drainage site was managed as free drainage.




- In controlled drainage the height of the water table can be adjusted by adding or removing weir boards to the structure.
- Therefore, the amount of water in the field and available for plant use can also be adjusted.

## RESEARCH QUESTIONS

- Does controlled drainage practice reduces the total amount of water flowing out of the field?
- Does controlled drainage practice reduces the total load of nutrients flowing out of the field?

## NATURE OF THE DATA

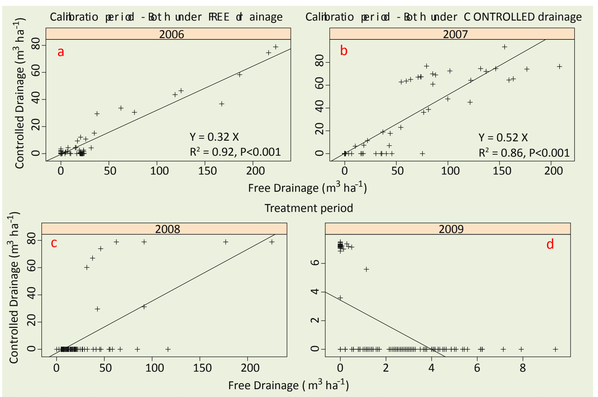



Figure 2. Water flow in the controlled drainage site as a function of water flow in the free drainage site (a-d). Note different Y and X scales in panel d.

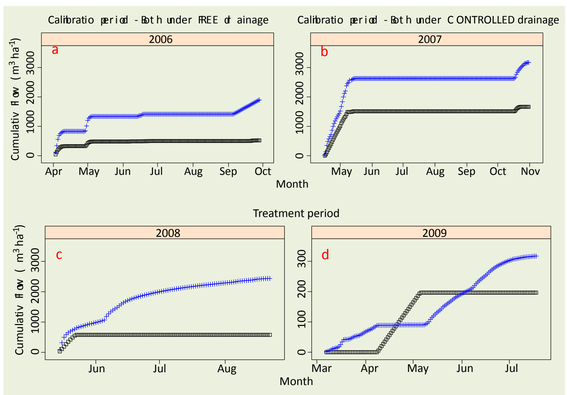



Figure 3. Cumulative water flow in the controlled and free drainage sites (a-d). Note different Y and X scales in panel d.

- There was a strong linear relationship between the sites under free or controlled drainage during the calibration period (Fig. 2 a and b).
- In contrast, during the treatment period now clear relationship can be established between the two sites (Fig. 2 c and d).

- The site under free drainage had greater cumulative water flow per hectare than the controlled drainage site in all four years (Fig. 3 a-d).
- The greater water flow per hectare in the free drainage site is mainly due to landscape positioning.

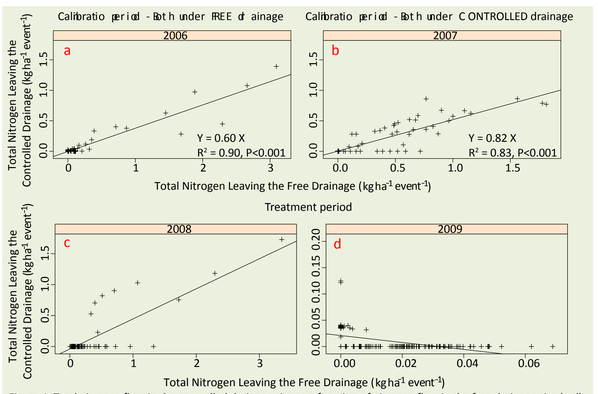



Figure 4. Total nitrogen flow in the controlled drainage site as a function of nitrogen flow in the free drainage site (a-d). Note different Y and X scales in panel d.

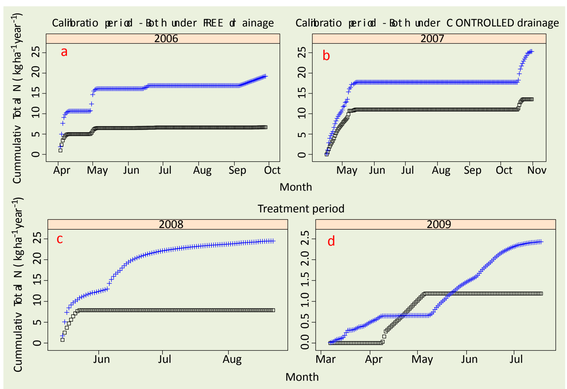



Figure 5. Cumulative total nitrogen flow in the controlled and free drainage sites (a-d). Note different Y and X scales in panel d.

- There was also a strong linear relationship between total nitrogen flow in the two sites during the calibration years (Fig 4 a and b).
- No relationship was observed between the two sites during the treatment period (Fig. 4 c and d).

- Cumulative total nitrogen flow followed the pattern observed for the water flow in the controlled and free drainage sites (Fig. 5 a-d).

## WHERE IS THE STATISTICAL ANALYSIS?

- Because this is a unreplicated experiment, only one samples is collected in each experimental unit each year.
- Current statistical approaches do not help in answering the research questions.
- What can be done?
- How can we assess the effects of controlled drainage on water drainage and nutrient movement out of the field?

## PROPOSED STATISTICAL PROCEDURES

### Method 1 – Calculation of Different Variance Components:

- First, estimate the variance component ($\sigma_{ti}^2$) for each site (i) using the appropriate flow values:

| Controlled Drainage | Free Drainage |
|---|---|
| ~~2006~~ 2007 2008 2009 | 2006 ~~2007~~ 2008 2009 |

$$s_{tc}^2 = \frac{1}{n-1}\sum(x_{ic} - x_{.c})^2 = 571{,}327 \qquad s_{tf}^2 = \frac{1}{n-1}\sum(x_{if} - x_{.f})^2 = 1{,}183{,}557$$

- Second, estimate the pooled within-treatments sample variance component ($\sigma_\varepsilon^2$):

| Controlled Drainage | Free Drainage |
|---|---|
| 2006 2007 2008 2009 | 2006 2007 2008 2009 |

$$s_e^2 = \frac{1}{n-1}\sum(x_{i.} - x_{..})^2 = 1{,}224{,}076$$

- Compare values within each site using t-tests based on the variance component for each site ($\sigma_{ti}^2$):

$$T = \frac{(X_1 - X_2)}{\sqrt{\frac{s^2}{n}}} \qquad T = \frac{(1662 - 577)}{\sqrt{\frac{571{,}327}{3}}} = 2.49,\ 2.49 > 1.96 \quad \text{(t-value at 0.05)}$$

- Compare values between each site using t-test based on the pooled within treatment sample variance ($\sigma_\varepsilon^2$), as indicated above.

Table 1. Summary of statistical analysis for the effect of controlled drainage on water flow.

| Year | Controlled Drainage | | | Free Drainage | | |
|---|---|---|---|---|---|---|
| | Water Flow ($m^3$ $ha^{-1}$ $year^{-1}$) | | | | | |
| 2006 | 520 | | B | 1896 | a | A |
| 2007 | 1662 | a | B | 3170 | | A |
| 2008 | 577 | b | B | 2436 | a | A |
| 2009 | 197 | b | A | 317 | b | A |

Means followed by the same lower case letter in the column, or upper case letter in the row are not significantly different at the p-value 0.05.

Table 2. Summary of statistical analysis for the effect of controlled drainage on total nitrogen flow.

| Year | Controlled Drainage | | | Free Drainage | | |
|---|---|---|---|---|---|---|
| | Total Nitrogen Flow (kg $ha^{-1}$ $year^{-1}$) | | | | | |
| 2006 | 6.68 | | B | 12.23 | a | A |
| 2007 | 13.56 | a | A | 16.10 | | A |
| 2008 | 7.88 | ab | B | 15.58 | a | A |
| 2009 | 1.79 | b | A | 1.55 | b | A |

Means followed by the same lower case letter in the column, or upper case letter in the row are not significantly different at the p-value 0.05.

### Method 2 – Calculation of Daily Ratios Between Two Treatments:

- First, calculate cumulative daily flow for the variable of interest (e.g. flow volume, nutrient flow volume, etc) in the two sites.
- Second, calculate the ratio between the cumulative flow in the controlled drainage over the free drainage for each day.
- Use a beta distribution and SAS to analyze the data set.

Proc Glimmix; Class year;
Model ratiovolume = year / dist=beta link=logit ddfm=kr;
Lsmeans year / ilink diff lines adjust=bon; Lsmeans year / ilink diff lines adjust=tukey; Lsmeans year / ilink diff lines adjust=simulate; Run;

Table 3. Summary of power analysis to detect a 10% difference in the ratio of cumulative water flow volume using a beta distribution.

| Label | Num Df | Den Df | FValue | ProbF | alpha | cp | nc | Power |
|---|---|---|---|---|---|---|---|---|
| 2006 vs 2007 | 1 | 374 | 5159 | 0.001 | 0.05 | 3.87 | 5159 | **1.00** |
| 2006 vs 2008 | 1 | 162 | 0 | 0.956 | 0.05 | 3.89 | 0 | **0.05** |
| 2006 vs 2009 | 1 | 159 | 207 | 0.001 | 0.05 | 3.90 | 207 | **1.00** |
| 2007 vs 2008 | 1 | 161 | 166 | 0.001 | 0.05 | 3.89 | 166 | **1.00** |
| 2007 vs 2009 | 1 | 159 | 24 | 0.001 | 0.05 | 3.90 | 24 | **0.99** |
| 2008 vs 2009 | 1 | 156 | 136 | 0.001 | 0.05 | 3.90 | 136 | **1.00** |

Table 4. Summary of statistical analysis for the effect of controlled drainage on water flow.

| Year | (Controlled Drainage) / (Free Drainage) | | | | | | | |
|---|---|---|---|---|---|---|---|---|
| | Water Flow Ratio ($m^3$ $ha^{-1}$ $year^{-1}$) | | | | | | | |
| | Fisher | | Bonferroni | | Tukey | | Simulated | |
| 2006 | 0.35 | c | 0.35 | c | 0.35 | c | 0.35 | c |
| 2007 | 0.58 | b | 0.58 | b | 0.58 | b | 0.58 | b |
| 2008 | 0.35 | c | 0.35 | c | 0.35 | c | 0.35 | c |
| 2009 | 0.68 | a | 0.68 | a | 0.68 | a | 0.68 | a |

Means followed by the same lower case letter in the column are not significantly different at the p-value 0.05.

Table 5. Summary of statistical analysis for the effect of controlled drainage on total nitrogen flow.

| Year | (Controlled Drainage) / (Free Drainage) | | | | | | | |
|---|---|---|---|---|---|---|---|---|
| | Total Nitrogen Flow Ratio ($m^3$ $ha^{-1}$ $year^{-1}$) | | | | | | | |
| | Fisher | | Bonferroni | | Tukey | | Simulated | |
| 2006 | 0.41 | c | 0.41 | c | 0.41 | c | 0.41 | c |
| 2007 | 0.61 | b | 0.61 | b | 0.61 | b | 0.61 | b |
| 2008 | 0.42 | c | 0.42 | c | 0.42 | c | 0.42 | c |
| 2009 | 0.65 | a | 0.65 | a | 0.65 | a | 0.65 | a |

Means followed by the same lower case letter in the column not significantly different at the p-value 0.05.

## CONCLUSIONS

- The lack of a linear relationship between the controlled drainage site and free drainage site after the treatment period started limits the number of statistical procedures that can be applied to make inferences on treatment effects.
- The use of t-test provide a simple way to compare unreplicated studies where only one observation per subject is possible during the coarse of one growing season.
- The use of a beta distribution to compare the cumulative flow ratio seems to be the best approach to compare data from unreplicated studies.