**Relationship between Mean Square Errors and Wheat Grain Yields in Two Long-term Nutrient Management Experiments**

Melissa Golden, Bruno Figueiredo, Mariana Ramos Del Corso, Nicole Remondet, Mame Diatite-Koumba, Shawntel Ervin, Daniel Alidekki, Jagmandeep Dhillon, Ethan Driver, Bill Jones, James Lasquites, Samuel Zoca, Bill Raun

**Abstract**

Rationale

Grain yield data collected from long-term experiments conducted on the same plots are often combined over years. The impact of the environment on treatment response from year to year is thus ignored.

Objective

The objectives of this work were to determine the relationship between mean square errors (MSE) and grain yield level, and to determine the frequency/rarity for combining year to year data.

Methods

Grain yield MSE’s were calculated by-year from two long-term winter wheat experiments (Experiment 222 and Experiment 502). Both trials employ a randomized complete block experimental design with 13 and 14 treatments, respectively. Experiment 222 (1969 to 2014) and Experiment 502 (1971-2014) have been continuously planted to winter wheat (Triticum aestivum L), and both annually receive an array of different nitrogen (N), phosphorus (P) and potassium (K) fertilizer rates. Mean square error terms were computed once sums of squares for replication and treatment (fixed effects) were removed from the analysis of variance model. Relationships between average wheat grain yields, coefficients of variation (CV) and MSE were evaluated over 46 and 44 years for experiments 222 and 502, respectively.

Results

As grain yields increased, MSE’s increased for both Experiments 222 and 502. More importantly was noting the extreme variability in MSE’s between years that would prevent combining data from consecutive years.

F-statistics (large MSE/small MSE, 40dfn, 40dfd, alpha=0.05) computed from two and three consecutive years were consistently more than the minimum F-value required to legitimately combine sites. In other words, sites were consistently different for any two or three year periods.

Finding no relationship between year and MSE suggests that homogeneity of treatment response with time, would not be expected.

Conclusion

Data analysis reported here, including 90 site years, shows that combining any two or three consecutive year-periods is not advisable, rather, individual years should be analyzed and reported on independently.

**Introduction**

In long-term experiments, where the same treatments are applied to specific plots, year after year, interpretation of yield response is often conducted over longer periods of time without considering year-to-year homogeneity/heterogeneity of experimental error.

Since the advent of modern statistics, experimental errors have long been considered and discussed. Early work by Mercer and Hall (1911) noted that the degree of confidence which may be attached to the results from field experiments depends on experimental error. Mercer and Hall (1911) further reported that experimental error could be reduced by repeating the experiment over a long period of time. This coming from one of the oldest long-term experiments in the world, at Rothamsted Experiment Station ([www.rothamseted.ac.uk](http://www.rothamseted.ac.uk), visited March 22, 2015).

Fisher (1926) reported that estimates of field errors coming from any specific experiment may or may not be a valid estimate of the actual errors affecting the averages, or differences of averages, and that are required to estimate experimental error. Fisher (1926) goes on to state that the estimate of error afforded by the “replicated trial” depends on differences between plots treated alike. The diligence of this early work was further reflected in his noting that a valid estimate of error will be obtained by arranging the plots in a deliberately random fashion. The value of replication (provides the opportunity to estimate experimental error) and randomization (provides the opportunity to have an unbiased estimate of experimental error) were thus cemented into the chronicles of biometry.

The value of this work was substantiated when Dr. Jose Crossa said “analyzing data over locations and/or years can be problematic as this can ignore treatment by environment interactions that are prevalent in agricultural studies” (personal communication March 23, 2015, Dr. Jose Crossa, Head Statistics, CIMMYT, Mexico, DF).

The objectives of this study were to determine the relationship between MSE and yield level, the year to year variability in MSE’s and the frequency for combining year to year data from long-term experiments.

**Materials and Methods**

Two long-term winter wheat experiments, Experiment 222 (Stillwater, OK) and Experiment 502 (Lahoma, OK) were used to determine the efficacy/feasibility of combining year to year grain yield data from long-term experiments. Experiments 222 and 502 were established in 1969 and 1970, respectively, under conventional tillage. Experiment 222 is located at the Agronomy Research Station in Stillwater, OK at an altitude of 272 masl, on a well-drained, deep and slowly permeable Kirkland silt loam (fine, mixed, thermic Udertic Paleustoll). Experiment 502 is located at the North Central Research Station in Lahoma, OK at an altitude of 396 masl on a well-drained, deep and moderately permeable Grant silt loam (fine-silty, mixed, thermic Udic Argiustoll) soil.

The average annual precipitation at Stillwater and Lahoma is 80 cm. The experimental design for both trials is a randomized complete block with four replications. There were 12 and 13 treatments that comprised application of different rates of N, P, and K fertilizers in both Experiment 222 and Experiment 502. Nitrogen, P, and K were applied as ammonium nitrate (34% N), triple super phosphate (20% P), and potassium chloride (53% K), respectively, in both experiments. Plots were permanent from year to year and received fixed rates of N, P, and K every year. Individual plots at Stillwater are 6.1 m wide and 18.3 long, and at Lahoma are 4.9 m wide and 18.3 m long. Experimental plots were conventionally tilled every year in the summer to a 15-cm depth using a disk plow. Plots were then harrowed before fertilization using a spike tooth harrow every year. Winter wheat was planted for 45 (Stillwater) and 44 (Lahoma) continuous years in 25.4-cm wide rows at seeding rates of 67 kg ha-1. In some years, the seeding rate was increased to 110 kg ha-1 in anticipation of poor germination and emergence due to unfavorable soil moisture conditions at seeding. Since 1992, winter wheat has been planted in 19.1-cm rows at Stillwater. At both sites, varieties were changed with time due to increased genetic yield potential and the need for rust resistance (Table 1). In both experiments, preplant fertilizer was broadcast and incorporated in late August to mid-September. Winter wheat was planted in late September to early October every year. Grain yield data was collected in each plot and year using a Massey-Ferguson 8XP experimental combine at both sites.

For both long-term data sets, analysis of variance was conducted by-year, accounting for replication and treatment in the model. Mean yields, and mean square errors (MSE) were recorded, and coefficients of variation (CV) determined accordingly. Linear relationships between mean grain yields, MSE, CV, year, and computed Least Significance Difference (LSD, alpha = 0.05) tests were also evaluated. Bartlett’s test for evaluating homogeneity of MSE terms from the by-year AOV’s was also performed, both year to year, and over three year sequences (F = large MSE/small MSE, 40 df numerator, 40 df denominator). These F statistics were then plotted as a function of year. Comprehensive discussion of both Experiment 502 and Experiment 222 can be found in the following publications (Raun et al., 1998, Raun et al., 2011 and, Westerman et al., 1994).

**Results**

Grain yield and MSE

For both Experiment 502 and Experiment 222, MSE’s tended to increase as yields increased (Figures 1 and 2). Numerically this makes sense as the actual numbers on both sides (yield and MSE) would be expected to be larger. Nonetheless, these relationships were weak (coefficient of determination, r2 , both less than 0.19).

MSE and Year

Mean square errors from by-year analysis of variance on wheat grain yields increased slightly in Experiment 502, and showed no measurable change in Experiment 222, with time (Figures 3 and 4).

Year and CV

At both locations, the coefficient of variation (CV = (square root (MSE/experimental mean) \*100)) from the analysis of variance model showed no change with time (Figures 5 and 6). This would suggest that the experimental errors associated with conducting these trials remained relatively constant, despite known influences of the environment on treatment response. This was encouraging because it says that the methods (planting, tillage, harvest, weed control, other) employed when conducting these trials were essentially repeatable and also because CV’s averaged less than 12 and 16 for experiments 222 and 502, respectively, and this over 46 and 44 years.

Combining Ensuing 2 and 3 Year Data

F-tests computed (large MSE/small MSE) for two and three ensuing year periods, in Experiment 502 and Experiment 222, are reported in Figures 7-10, respectively. For Experiment 502, 18 of 41 possible combinations of 2-year consecutive year periods had F values in excess of 1.7. In other words, ensuing two year periods should not have been combined 43% of the time (Figure 7). For consecutive 3-year periods (1971-2014), F values exceeded 1.7, 73% of the time, further restricting combining data over years (Figure 8).

This same data analysis for Experiment 222 showed that consecutive 2-year data should not be combined 55% of the time, and this value jumped to 90% when 3-year periods were considered (Figures 9 and 10, F value of 1.74 needed to declare significantly different).

**Conclusions**

Comprehensive data from the two long-term experiments analyzed show that combining any two, or three year consecutive data sets is not advisable based on the failure to meet homogeneity of error variance tests. This further suggests that combining year to year data essentially ignores the importance that environment has on treatment response. Furthermore, when considering that this data comes from long-term experiments conducted on the exact same location, the influence of environment from one year to the next was never the same.

**References**

Fisher, R.A. 1926. The arrangement of field experiments. J. of the Ministry of Agric. 33:503-514.

Mercer, W.B., and A.D. Hall. 1911. The experimental error of field trials. J. of Agricultural Sci. 4:107-132.

Raun, W.R., G.V. Johnson, S.B. Phillips and R.L. Westerman. 1998. Effect of long-term nitrogen fertilization on soil organic C and total N in continuous wheat under conventional tillage in Oklahoma. Soil & Tillage Res. 47:323-330.

Raun, William R., John B. Solie, and Marvin L. Stone. 2011. Independence of yield potential and crop nitrogen response. Precision Agric. 12:508-518.

Westerman, R.L, R.K. Boman, W.R. Raun and G.V. Johnson. 1994. Ammonium and nitrate nitrogen in soil profiles of long-term winter wheat fertilization experiments. Agron. J. 86:94-99.

**List of Figures**

Figure 1. Relationship between mean square errors (MSE) from by-year analysis of variance and mean grain yield level, Experiment 502, 1971 to 2014.

Figure 2. Relationship between mean square errors (MSE) from by-year analysis of variance and mean grain yield level, Experiment 222, 1969 to 2014.

Figure 3. Relationship between mean square errors (MSE) from by-year analysis of variance and year, Experiment 502, 1971 to 2014.

Figure 4. Relationship between mean square errors (MSE) from by-year analysis of variance and year, Experiment 222, 1969 to 2014.

Figure 5. Relationship between the coefficient of variation (CV) from by-year analysis of variance and year, Experiment 502, 1971-2014.

Figure 6. Relationship between the coefficient of variation (CV) from by-year analysis of variance and year, Experiment 222, 1969-2014.

Figure 7. F-statistic (large MSE/small MSE) plotted over time for 2 ensuing years, Experiment 502, 1971-2014. F-statistics above the red line (39 dfn, 39 dfd, 0.05 = 1.70) note 2 year periods where data should not have been combined.

Figure 8. F-statistic (large MSE/small MSE) plotted over time for 3 ensuing years, Experiment 502, 1971-2014. F-statistics above the red line (39 dfn, 39 dfd, 0.05 = 1.70) note 3 year periods where data should not have been combined.

Figure 9. F-statistic (large MSE/small MSE) plotted over time for 2 ensuing years, Experiment 222, 1971-2014. F-statistics above the red line (36 dfn, 36 dfd, 0.05 = 1.74) note 2 year periods where data should not have been combined.

Figure 10. F-statistic (large MSE/small MSE) plotted over time for 3 ensuing years, Experiment 222, 1971-2014. F-statistics above the red line (36 dfn, 36 dfd, 0.05 = 1.74) note 3 year periods where data should not have been combined.

Fig. 1

Fig. 2

**Fig. 3**

**Fig. 4**

**Fig. 5**

**Fig. 6**

**Fig. 7**

**Fig. 8**

**Fig. 9**

**Fig. 10**

**Table 1. Winter wheat varieties planted in Experiments 502 and 222, 1969-present.**

|  |  |  |
| --- | --- | --- |
| Year | Experiment 502 | Experiment 222 |
| 1969 | - | Scout 66 |
| 1970 | - | Scout 66 |
| 1971 | Scout 66 | Scout 66 |
| 1972 | Scout 66 | Scout 66 |
| 1973 | Scout 66 | Scout 66 |
| 1974 | Scout 66 | Scout 66 |
| 1975 | Triumph 64 | Triumph 64 |
| 1976 | Triumph 64 | Triumph 64 |
| 1977 | Osage | Triumph 64 |
| 1978 | Osage | Osage |
| 1979 | TAM 101 | Osage |
| 1980 | TAM 101 | Osage |
| 1981 | TAM 101 | TAM 101 |
| 1982 | TAM 101 | TAM 101 |
| 1983 | TAM 101 | TAM 101 |
| 1984 | TAM 101 | TAM 101 |
| 1985 | TAM 101 | TAM 101 |
| 1986 | TAM 101 | TAM 101 |
| 1987 | TAM 101 | TAM 101 |
| 1988 | TAM 101 | TAM 101 |
| 1989 | TAM 101 | TAM 101 |
| 1990 | TAM 101 | TAM 101 |
| 1991 | TAM 101 | TAM 101 |
| 1992 | TAM 101 | TAM 101 |
| 1993 | Karl | Karl |
| 1994 | Karl | Karl |
| 1995 | Tonkawa | Tonkawa |
| 1996 | Tonkawa | Tonkawa |
| 1997 | Tonkawa | Tonkawa |
| 1998 | Tonkawa | Tonkawa |
| 1999 | Tonkawa | Tonkawa |
| 2000 | Custer | Custer |
| 2001 | Custer | Custer |
| 2002 | Custer | Custer |
| 2003 | Custer | Custer |
| 2004 | Custer | Custer |
| 2005 | Overley | Endurance |
| 2006 | Overley | Endurance |
| 2007 | Overley | Endurance |
| 2008 | Overley | OK Field |
| 2009 | OK Field | Endurance |
| 2010 | GoLead | Bullet |
| 2011 | Centerfield | Bullet |
| 2012 | Endurance | Billings |
| 2013 | OK9935C | Billings |
| 2014 | Doublestop-CL | Ruby Lee |

**Added Items (not part of the publication)**

**Results**

For the LSD’s that were computed by-year and plotted over time, there was a slight trend for these to increase (Figure X). But more important than the trend was that these values ranged from 0.22 to 0.86 Mg/ha (3.2 to 10.8 bu/ac).

