

Revealing GE Interactions from Trial Data without Replications

Jixiang Wu^{1*}, Johnie Jenkins², Jack C. McCarty²

¹Department of Agronomy, Horticulture, and Plant Science, South Dakota State University, Brookings, SD, USA

²Crop Science Research Laboratory, USDA-ARS, Mississippi State, MS, USA

Email: *Jixiang.Wu@sdstate.edu

How to cite this paper: Wu, J.X., Jenkins, J. and McCarty, J.C. (2019) Revealing GE Interactions from Trial Data without Replications. *Open Journal of Statistics*, 9, 407-419.

<https://doi.org/10.4236/ojs.2019.93027>

Received: May 15, 2019

Accepted: June 27, 2019

Published: June 30, 2019

Copyright © 2019 by author(s) and Scientific Research Publishing Inc.

This work is licensed under the Creative Commons Attribution International License (CC BY 4.0).

<http://creativecommons.org/licenses/by/4.0/>



Open Access

Abstract

Detecting genotype-by-environment (GE) interaction effects or yield stability is one of the most important components for crop trial data analysis, especially in historical crop trial data. However, it is statistically challenging to discover the GE interaction effects because many published data were just entry means under each environment rather than repeated field plot data. In this study, we propose a new methodology, which can be used to impute replicated trial data sets to reveal GE interactions from the original data. As a demonstration, we used a data set, which includes 28 potato genotypes and six environments with three replications to numerically evaluate the properties of this new imputation method. We compared the phenotypic means and predicted random effects from the imputed data with the results from the original data. The results from the imputed data were highly consistent with those from the original data set, indicating that imputed data from the method we proposed in this study can be used to reveal information including GE interaction effects harbored in the original data. Therefore, this study could pave a way to detect the GE interactions and other related information from historical crop trial reports when replications were not available.

Keywords

GE Interaction, Historical Crop Trial Data, Imputation

1. Introduction

Replication is often required for valid data processing and statistical tests [1]. In addition, it provides a great prospect to dissect potential interaction effects among factors of interest like genotype-by-environment (GE) interaction [2]. It is very common for most multi-environment crop trial data like multi-location

and multi-year trial data sets to be reported by various institutions and made available for public users. However, most of the published crop trial data only include entry means under each environment rather than the original repeated field plot data. Without replications, it would be statistically challenging to detect GE interaction effects, which are highly related to yield stability, from the crop trial reports. Therefore, it would be a great addition to develop a new method that could be effectively used to detect GE interactions when original replicated field trial data are not available.

Because the entire original data can be treated as missing data, in order to successfully recover the potential genetic information, we will need an imputation method to generate “alternative replicated field data” that can be used to recover the information from the original data. There are two major categories of imputations: single imputation (SI) and multiple imputation (MI). With SI, missing values are filled by some type of predicted values like mean imputation, regression imputation, and/or matching methods [3] [4] [5]. Although SI has been widely used, one shortcoming is that it does not reflect the full uncertainty created by missing data and almost always underestimates the variance. For example, the regression imputation method is based on an estimated regression model to predict or impute the missing values. This could cause relationships to be over-identified and suggest greater precision in the imputed values than is warranted. In order to deal with the problem of increased noise due to data imputation, MI, which repeats multiple times resulting in multiple imputed data sets, is recommended, especially when data are missing at random (MAR) [4]. With MI, the imputation uncertainty is accounted for by creating these multiple data sets. The MI follows three basic steps: imputation, analysis, and pooling [6] [7]. With MI bias can be reduced and estimates are more precise. MI has several desirable features. The first feature is that introducing appropriate random error into the imputing process makes it possible to get approximately unbiased estimates of all parameters. The second feature is that repeated imputation allows researchers to get better estimates of the standard errors. The third feature is that MI can be used with any kind of data and analyses.

Unlike many other missing data being imputed, it is well-known that the entire original field measurements were unavailable except only entry means under each environment. Therefore, an important step is to propose probability density function for each entry/genotype based on the published results that can be used to impute entire “original data” so that the genetic information including GE interactions harbored in the original data can be detected, accordingly. In the present study, our objectives included 1) to propose a new procedure to generate a new data set with repeated measurements from given entry means and 2) to numerically validate the new method with a data set containing six locations, 28 potato genotypes, and three replications in each of six locations [8]. The purpose of this study is to provide an alternative method and computer tool to improve data analysis and statistical tests and thus to reveal more information harbored

in the historical crop trial data when replications were not available.

2. Materials and Methods

2.1. Linear Model for GE Analysis

The linear model used for an observation y_{hij} , which represents the environment h , the genotype i , and the block j nested to the environment h , can be expressed as follows:

$$y_{hij} = \mu + E_h + G_i + GE_{hi} + B_{j(h)} + e_{hij} \quad (1)$$

In order to detect GE interaction effects, replication with each environment is required. Without replication, the GE interaction effects and random error are confounded and they cannot be separated and the GE interaction and block effects should be omitted from model (1).

2.2. Model Used for Data Imputation

The linear model for an observation under a single environment can be described as model (2) without including environmental and GE interaction effects:

$$y_{ij} = \mu + G_i + B_j + e_{ij} \quad (2)$$

In model (2), G_i may include GE interaction effect where it may exist. If we assume block effects and random error follow two independent normal distributions, then y_{ij} follows the following normal distribution in (3)

$$y_{ij} \sim N(\mu + G_i, \sigma_B^2 + \sigma^2) \quad (3)$$

Given the above distribution in (3), if we know $\mu + G_i$ and $\sigma_B^2 + \sigma^2$, we can generate y_{ij} under each single environment. Blocking is used for local control of field variation within each environment; however, block effects may not impact the results of variance components for genotypic effects and random error and prediction of genotypic effects if model (2) is applied. Therefore, to simplify, we can assume there are no block effects and they can be omitted during the data imputation process. If so, Equation (3) can be simplified as in the following normal distribution in (4) when there are not block effects:

$$y_{ij} \sim N(\mu + G_i, \sigma^2) \quad (4)$$

Actual values for μ , G_i , and σ^2 are unknown. If we can substitute $\mu + G_i$ and σ^2 with estimates $\hat{\mu} + \hat{G}_i$ and $\hat{\sigma}^2$, then we can impute each y_{ij} accordingly.

$$\hat{y}_{ij} \sim N(\hat{\mu} + \hat{G}_i, \hat{\sigma}^2) \quad (5)$$

where $\hat{\mu}$ is an estimated population mean; \hat{G}_i is an estimated/predicted genotypic effect for genotype i and $\hat{\sigma}^2$ is an estimated variance for random error. In many trial reports, individual genotypic means under each environment were available and thus can be used to substitute $\mu + G_i$ and mean square error

(MSE) can be used to substitute σ^2 . MSE value for each environment can be derived from the coefficient of variation or least significant difference (LSD).

2.3. Data Source

The data set (plrv) used for our imputation analysis, as a demonstration, is currently available in the R package *agricolae* [8]. The data set contains six environments, 28 potato genotypes, and three replications in each environment. There were three agronomic traits in the data while only yield was used for this study. The major reason for using this data set as a demonstration is that it is publicly available [8] and interested parties can generate repeatable results via the codes developed by the author of this study.

2.4. Data Imputation and Analysis

Data imputation: Phenotypic means for yield for 28 potato genotypes at each of six locations were calculated. The unit used for potato was not provided in the package. Interested readers may contact the package developer for more detailed information. With the data, MSE for each environment was calculated by the ANOVA method subject to model 2. Both phenotypic means and MSE under each environment were used to generate imputed data. Assuming that data were normally distributed, individual observations with no block effects for each environment were imputed following the normal distribution Equation (5) with the use of the `norm` R function [9]. The process is repeated for each of six environments. The number of replications for data imputation was the same one used in each original experiment. All imputed data sets are combined as a multi-environment data set for the following analysis. In order to compare the impact of replications, the imputed data were repeated for 10, 20, 50, 100, 200, and 500 times, respectively. Both results from individual imputed data sets and pooled results are reported and compared.

Data analysis: First, phenotypic means for different genotypes in each environment were calculated for the original data set and each multi-environment data set. Second, linear mixed model (LMM) approaches such as restricted maximum likelihood (REML) and minimum norm quadratic unbiased estimation (MINQUE) [10] [11] can be used to analyze each imputed data set subject to model (1) mentioned above. Variance components for genotypic effects and GE interaction effects were estimated by MINQUE approach [12]. Genotypic effects and GE interaction effects were predicted using the adjusted unbiased prediction (AUP) method [13]. Mean and its confidence interval (CI) of 95% for each parameter were calculated. All data analyses were conducted under the R environment [9]. The MINQUE package [14] with `minque` approach for variance component estimation and AUP approach [13] for random effect prediction was used for our imputed data analysis. The R scripts for data imputation and other related data analyses were developed by the first author of this study and will be available upon request.

3. Results

3.1. Original and Imputed Phenotypic Means for 28 Potato Genotypes

The phenotypic means for 28 potato genotypes under six environments calculated from the original data set are provided in **Table 1**. Generally, wider ranges among six environments were observed compared to the ranges among genotypes within each environment (**Table 1**), indicating that environmental effects played a more important role on yield than genotypes. Some genotypes were observed to be more adapted to specific environments. For example, Canchan was more adapted to Hyo02 (47.78) but Desiree was less adapted to the same environment (8.89). On the other hand, genotype Desiree was more adapted to the environment SR03 (11.42) than Canchan to the same environment (2.42), indicating that genotype-by-environment (GE) interactions also played an important role on potato yield. Therefore, it will be interesting to investigate GE interaction effects in the yield trial analysis.

Phenotypic means and their 95% confidence intervals (ranges of 2.5% and 97.5% percentiles) for 28 entries under each environment over 50 imputed data sets are provided in **Table 2**. Comparing the results in **Table 2** and **Table 1**, the imputed means and original means were close to each other with a maximum difference of 1.35 and a mean difference of 0.28. The correlation coefficient between the original phenotypic means and imputed phenotypic means was almost 1.0.

Table 1. Individual phenotypic yield means for 28 genotypes in each of six environments+.

Geno	Ayac	Hyo02	LM02	LM03	SR02	SR03
102.18	24.925	28.8889	32.037	46.778	13.5185	11.7696
104.22	21.451	53.5185	39.198	50.418	16.0494	7.0988
121.31	23.460	41.2963	38.395	63.704	2.5000	11.2551
141.28	31.844	60.4630	33.951	77.568	19.2346	15.4774
157.26	19.670	41.3889	45.160	76.986	23.9506	14.5556
163.9	17.538	29.5370	28.889	32.029	12.7160	7.7954
221.19	15.414	32.0370	31.025	43.453	8.5432	7.4376
233.11	24.283	50.5556	29.198	47.333	13.1481	7.4815
235.6	29.914	73.5185	40.370	56.136	14.8025	17.0679
241.2	20.444	36.0185	35.741	46.247	11.0864	8.5053
255.7	26.067	47.0370	32.679	40.792	19.1852	17.7778
314.12	17.325	49.4444	34.506	57.246	8.5309	1.9876
317.6	26.614	53.4259	42.346	64.012	14.8148	10.7425
319.20	25.775	56.6667	32.963	86.808	21.2099	9.1235
320.16	30.329	31.1111	35.284	43.290	13.6296	4.4444
342.15	19.897	40.7407	27.531	38.800	17.5926	11.5185

Continued

346.2	21.575	32.6852	25.556	32.037	18.0247	13.1733
351.26	31.749	50.1852	29.259	72.024	20.3704	13.1070
364.21	26.639	52.4074	37.901	57.066	13.5185	16.8261
402.7	19.297	42.5000	31.235	49.765	12.8395	9.2284
405.2	28.667	35.7407	32.346	43.259	16.7901	17.1166
406.12	19.587	59.8148	37.778	53.588	13.8272	11.5046
427.7	26.089	55.6482	44.444	58.336	21.2346	11.3889
450.3	28.724	50.1852	36.889	72.242	15.4321	13.7037
506.2	25.000	46.7593	45.556	53.250	18.1482	10.8848
Canchan	21.327	47.7778	21.605	59.247	9.6296	2.4211
Desiree	18.765	8.8889	20.370	27.427	10.0617	11.4202
Unica	21.301	72.2222	47.840	57.535	18.2469	17.4787

+: The original data set is available in R package agricolae [8]; however, the unit for yield is not provided.

Table 2. Phenotypic yield means for 28 genotypes (IMean) over 50 imputed data sets and their 95% confidence intervals (LL = low limit and UL = up limit) in each of six environments.

Gen	Ayac			Hyo02			LM02		
	IMean	LL	UL	IMean	LL	UL	IMean	LL	UL
102.18	24.47	18.85	30.09	29.09	22.93	34.39	32.47	26.52	38.24
104.22	22.00	13.82	28.95	53.60	48.06	59.64	39.66	35.18	44.78
121.31	23.04	15.72	29.98	41.49	34.66	47.19	38.29	31.18	43.35
141.28	31.98	24.26	41.22	60.93	53.87	67.61	34.02	26.18	40.85
157.26	20.32	12.31	26.91	41.53	35.09	46.82	44.98	39.70	50.44
163.9	17.74	10.57	25.27	29.37	22.24	35.28	29.05	23.40	34.71
221.19	15.90	8.75	21.69	32.22	26.82	37.47	31.91	27.30	37.91
233.11	25.14	17.51	33.23	51.06	45.00	56.81	29.66	24.40	36.44
235.6	31.23	22.88	39.75	73.73	66.93	80.15	40.38	33.63	45.94
241.2	20.20	12.34	30.88	35.60	28.76	42.78	36.10	30.37	41.68
255.7	26.40	20.50	33.72	47.54	40.20	53.09	32.90	26.52	38.41
314.12	17.55	9.54	24.42	49.81	43.79	55.57	34.67	29.46	40.45
317.6	26.58	20.56	33.66	53.58	47.98	59.87	42.37	36.92	47.28
319.2	25.94	18.23	32.33	55.52	48.78	62.70	32.76	28.03	37.06
320.16	30.53	22.94	39.13	30.61	25.44	36.50	35.04	29.11	41.20
342.15	19.52	11.79	27.26	40.08	34.42	46.37	27.91	22.04	33.05
346.2	21.92	15.25	27.30	32.87	26.81	40.79	25.49	18.85	32.24
351.26	30.57	24.79	36.35	50.19	45.26	56.09	29.39	22.07	34.91
364.21	27.33	23.19	32.62	52.30	44.99	59.36	38.65	33.07	42.85
402.7	19.29	12.64	26.46	41.01	35.50	47.04	31.70	25.34	37.72
405.2	29.18	22.79	36.11	36.38	28.64	43.13	31.99	28.15	35.67

Continued

406.12	20.07	13.05	27.29	60.16	53.04	67.08	37.66	33.42	42.35
427.7	25.77	18.44	32.38	55.40	49.12	61.77	44.01	39.21	50.36
450.3	28.57	21.51	35.58	50.04	43.23	56.10	37.25	32.36	42.00
506.2	24.83	18.20	32.41	46.97	40.56	53.82	45.57	38.51	53.65
Canchan	21.92	14.10	28.43	48.28	40.48	56.56	21.42	16.69	27.67
Desiree	18.28	11.12	24.61	10.22	3.63	16.61	20.69	16.46	24.98
Unica	21.43	14.14	28.93	72.18	65.41	80.30	47.67	42.32	53.71
	LM03			SR02			SR03		
Gen	IMean	LL	UL	IMean	LL	UL	IMean	LL	UL
102.18	47.00	36.53	56.60	13.84	10.63	17.12	11.86	8.48	15.00
104.22	51.63	43.26	58.41	16.08	12.86	18.61	7.21	3.50	10.21
121.31	63.46	55.31	72.45	2.97	1.02	5.71	10.47	7.35	13.70
141.28	77.17	67.40	87.48	19.25	15.80	23.17	15.42	12.33	18.43
157.26	77.20	67.65	87.22	23.93	20.45	26.93	14.63	11.99	18.13
163.9	33.23	22.55	42.02	13.01	9.49	15.72	7.92	3.62	12.18
221.19	44.46	36.72	52.94	9.10	5.26	13.42	7.42	4.58	9.49
233.11	46.83	36.72	57.52	13.15	8.88	16.68	7.25	4.19	11.01
235.6	56.24	46.53	66.10	14.68	11.63	18.18	16.92	14.49	19.66
241.2	46.70	35.90	57.77	11.07	7.97	13.83	8.53	5.25	10.82
255.7	42.26	31.27	49.10	19.17	15.93	22.41	17.75	14.71	21.26
314.12	57.11	48.55	64.42	8.52	4.88	13.27	2.69	0.20	6.33
317.6	64.23	52.97	74.57	14.57	11.08	19.13	10.39	7.30	13.96
319.2	87.69	78.07	98.44	21.08	17.90	24.63	9.14	6.31	12.98
320.16	42.85	33.27	51.79	13.26	8.34	18.07	4.10	1.30	7.63
342.15	39.43	30.52	50.67	17.37	11.90	20.86	11.34	6.70	15.04
346.2	32.28	23.58	42.92	17.11	12.80	21.64	13.65	11.01	17.30
351.26	71.25	61.39	80.55	20.08	16.31	23.39	12.46	8.93	16.09
364.21	58.47	47.06	68.19	13.59	10.47	17.63	16.79	13.37	20.02
402.7	50.14	41.20	58.47	12.79	9.63	15.60	9.17	5.97	12.59
405.2	43.55	33.55	50.95	16.92	12.46	20.49	16.96	13.70	20.88
406.12	52.28	40.17	61.89	14.33	11.09	17.32	11.50	8.13	14.65
427.7	58.52	47.51	69.08	20.55	17.21	23.67	11.66	8.38	15.19
450.3	71.99	59.88	82.25	15.66	12.61	19.29	13.36	10.92	17.11
506.2	52.75	43.12	62.13	18.13	14.74	21.69	11.00	7.97	14.08
Canchan	59.32	50.79	70.21	9.33	5.33	14.15	2.69	0.05	5.54
Desiree	28.01	16.57	37.87	10.13	6.16	13.94	10.99	7.44	14.41
Unica	57.72	48.24	68.64	17.73	13.35	21.47	17.82	14.98	21.00

The results implied that the imputed phenotypic data represented the original data. In addition, the simulated 95% confidence intervals were highly related to the mean square error (MSE) for each of six environments. The largest confidence intervals were observed in environment LM-03 with the largest MSE of 87.08 and while the smallest confidence in SR-03 due to the small MSE in this environment (Table 2).

3.2. Imputed Entry Means for 28 Potato Genotypes

Correlation coefficients between phenotypic means from the original data and five sets of imputed means were obtained and are presented in Figure 1. The correlation coefficients between original phenotypic means and five sets of imputed phenotypic means were around 0.98 while coefficients among five sets of imputed phenotypic means were around 0.96. The results showed that phenotypic means obtained from each individual imputed data set were also highly consistent, implying that the imputed phenotypic means well represented the original phenotypic mean data.

3.3. Pooled Results

Due to some degree of uncertainty of imputed data, multiple imputed data sets were applied to reduce the bias potentially caused by single imputed data. The question is how many imputed data would be sufficient to adjust the bias. As mentioned in this study, we generated 10, 20, 50, 100, 200, and 500 imputed data sets, which were used to obtain the pooled phenotypic means for 28 genotypes under six environments, mean variance components for environment effects, genotypic effects, GE interaction effects, and random errors, and mean predicted environment effects, genotypic effects, and GE interaction effects. However, due to large amount of results, only summarized results were provided.

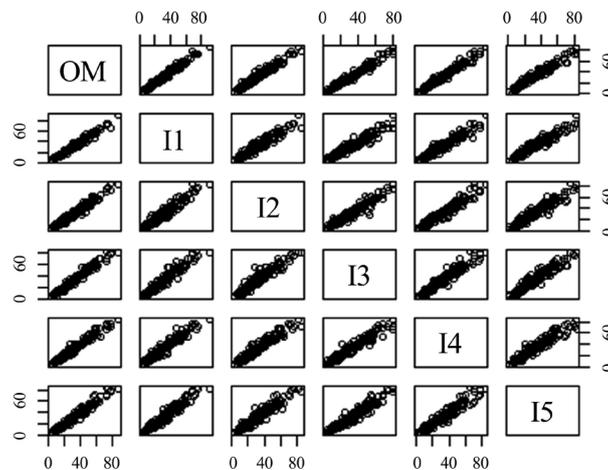


Figure 1. Correlations among individual phenotypic means from the original data set and five imputed data sets. OM = individual means from the original data set. I1 to I5 = individual means from the 1st five imputed data sets.

Figure 2 showed that phenotypic means from the original data set were highly correlated and consistent with pooled phenotypic means from multiple imputed data sets (correlation coefficients were almost close to 1). Figure 3 showed that predicted environmental effects from the original data were highly consistent with the pooled predicted environmental effects from different imputed data sets (correlation coefficients among these predicted environmental effects were close to 1). The same conclusions can be made for predicted genotypic effects (Figure 4) and predicted GE interaction effects (Figure 5). These results suggested that 10 repeated imputed data sets were sufficient to obtain unbiased phenotypic means and predicted environmental effects, genotypic effects, and GE interaction effects.

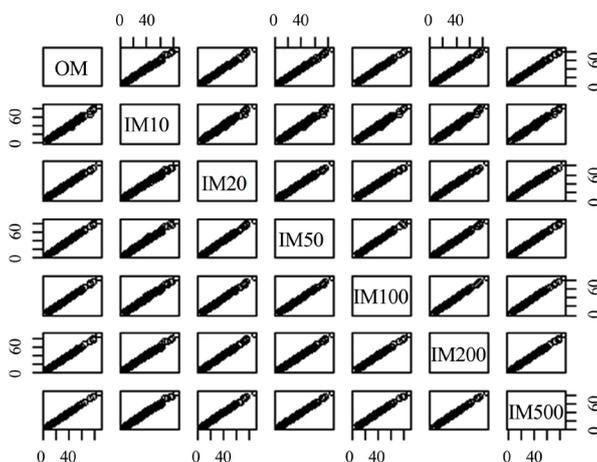


Figure 2. Correlations among individual phenotypic means from the original data set and individual phenotypic means from different multi-imputed data sets. OM = individual means from the original data set. IM10, IM20, IM50, IM100, IM200, and IM500 = pooled phenotypic means from 10, 20, 50, 100, 200, and 500 imputed data sets.

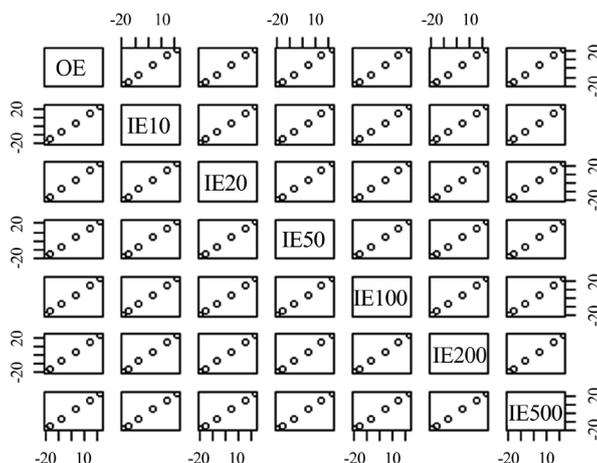


Figure 3. Correlations among predicted environmental effects from the original data set and mean environmental effects from different multi-imputed data sets. OE = environmental effects from the original data set. IE10, IE20, IE50, IE100, IE200, and IE500 = pooled environmental effects from 10, 20, 50, 100, 200, and 500 imputed data sets, respectively.

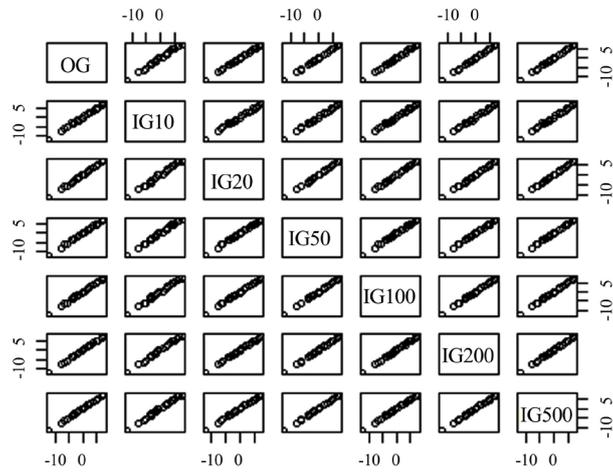


Figure 4. Correlations among predicted genotypic effects from the original data set and mean genotypic effects from different multi-imputed data sets. OG = genotypic effects from the original data set. IG10, IG20, IG50, IG100, IG200, and IG500 = pooled genotypic effects from 10, 20, 50, 100, 200, and 500 imputed data sets.

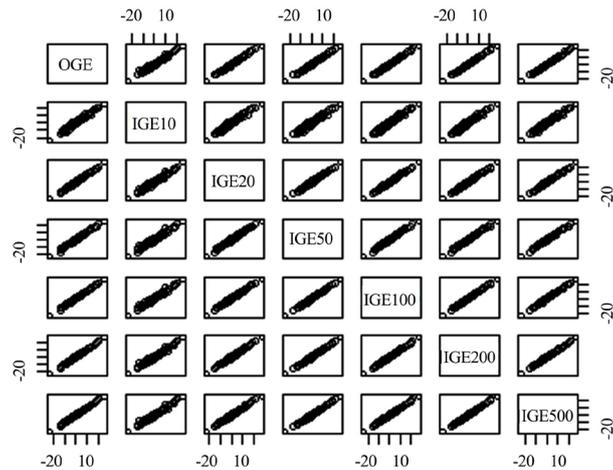


Figure 5. Correlations among predicted GE interaction effects from the original data set and mean GE interaction effects from different multi-imputed data sets. OG = GE interaction effects from the original data set. IG10, IG20, IG50, IG100, IG200, and IG500 = mean GE interaction effects from 10, 20, 50, 100, 200, and 500 imputed data sets.

In summary, the results from imputed data were highly consistent with those results from the original data set, which includes replication. The results used for our comparisons included phenotypic means, environmental effects, genotypic effects, and GE interaction effects. In addition, it appears that pooled results from 10 repeated imputed data sets were almost identical to the results from the original data set with replications.

4. Discussion

Crop trial data can provide important information to researchers. Revisiting the historical data and discovering more information will help researchers reveal more genetic information in different respects. As mentioned above, however,

many published trial data are summarized and the capability of using summarized data rather than the original repeated field plot data can be limited due to the lack of repeated field data. Therefore, it is crucial to generate new data sets that can be used to reveal genetic information comparable to the results from the original data with replications. This was our motivation to propose a new methodology in this study.

The key component in data imputation is to determine appropriate probability models, which can be used to generate simulated data to substitute multiple missing data points. Therefore, data imputation in this study can be considered as a simulation technique given particular probability models. Though the original field data from multi-environment crop trials were not available, the results such as entry means, numbers of replications, and mean square error provide information to determine a probability density function for each genotype/entry under each environment. With such a probability model for each genotype/entry, the entire data with replications can be imputed. Once data are imputed, various statistical data analyses for the imputed data can be followed like a linear-mixed model analysis in this study.

Due to the uncertainty of single imputed data set, multiple imputed data sets have been applied in this study to reduce the potential bias for each parameter. The question is how many independent imputed data sets are sufficient to represent the results from the original data set. Based on a demonstration data set, which is available in the R package *agricolae*, the correlation coefficient was around 0.98 between the phenotypic means from each of five individual imputed data sets and the phenotypic means from the original data while correlation coefficients was around 0.96 for the phenotypic means among five individual imputed data sets (**Figure 1**), showing that each imputed data set could be used to substitute the original data with replications. Our results also showed that 10 imputed data sets could sufficiently adjust the bias for this demonstration data set. However, it is likely that more imputed data sets would be required for a large MSE. It is possible sometimes that MSE values are not available on trial reports, one possible solution is that using a wide range of MSE values to impute multiple data sets. Such finding is important when individual MSEs in different environments are not available.

Though the method proposed in this study could help determine GE interaction from imputed data sets and increase the likelihood for statistical test and result validation, the imputation methods are based on the assumption of normal distribution for the original data, on which ANOVA analysis and mean comparisons are based. Many original field trial data included repeated field block; however, this information is often unavailable in the report. Our study showed that results from the imputed data without block effects were highly consistent with the results from the original data including blocks. Therefore, for simplicity, block effects could be omitted during the process of imputing trial data. In addition, data were imputed based on the MSE under each environment;

however, it appears that imputed data based on individual MSE under each environment and MSE over environments yielded almost identical results, suggesting either individual MSE or pooled MSE over environments could be used to impute trial data.

Statistical tests for each parameter of interest could follow several approaches. The first possible approach is jackknife based technique [15]. The second possible approach is to use a confidence interval test. With a large number of imputed data sets, we could construct a confidence interval (CI) for 95% or 99% and a CI statistical test can be employed. With the second one, a large number of imputed data sets will be required to provide more reliable CI tests for parameters of interest. Thus, the second approach could be computationally intensive if the original data set was large. However, with high-power servers and/or parallel algorithms, the time used to generate and analyze a large number of imputed data sets could be trivial.

Acknowledgements

This study was partially supported by USDA-NIFA hatch project (SD00H525-14) and South Dakota Soybean Research and Promotion Council (SD1900233).

Conflicts of Interest

The authors declare no conflicts of interest regarding the publication of this paper.

References

- [1] Kuehl, R.O. (1999) Design of Experiments: Statistical Principles of Research Design and Analysis. 2nd Edition, Duxbury Press, Pacific Grove.
- [2] Gray, E. (1982) Genotype \times Environment Interactions and Stability Analysis for Forage Yield of Orchardgrass Clones. *Crop Science*, **22**, 19-23.
<https://doi.org/10.2135/cropsci1982.0011183X002200010005x>
- [3] Enders, C.K. (2010) Applied Missing Data Analysis. The Guilford Press, New York.
- [4] Eekhout, I., *et al.* (2012) Missing Data: A Systematic Review of How They Are Reported and Handled. *Epidemiology*, **23**, 729-732.
<https://doi.org/10.1097/EDE.0b013e3182576cdb>
- [5] Roth, P.L. (1994) Missing Data: A Conceptual Review for Applied Psychologists. *Personnel Psychology*, **47**, 537-560.
<https://doi.org/10.1111/j.1744-6570.1994.tb01736.x>
- [6] Rubin, D.B. (1987) Multiple Imputation for Nonresponse in Surveys. Wiley, New York. <https://doi.org/10.1002/9780470316696>
- [7] van Buuren, S. (2012) Flexible Imputation of Missing Data. Chapman & Hall/CRC, Boca Raton. <https://doi.org/10.1201/b11826>
- [8] De Mendiburu, F. (2017) *Agricolae: Statistical Procedures for Agricultural Research*.
- [9] R Core Team (2012) R: A Language and Environment for Statistical Computing. R Foundation for Statistical Computing, Vienna.
- [10] Rao, C.R. (1971) Estimation of Variance and Covariance Components MINQUE

Theory. *Journal of Multivariate Analysis*, **1**, 257-275.

[https://doi.org/10.1016/0047-259X\(71\)90001-7](https://doi.org/10.1016/0047-259X(71)90001-7)

- [11] Patterson, H.D. and Thompson, R. (1971) Recovery of Inter-Block Information when Block Size Are Unequal. *Biometrika*, **58**, 545-554.
<https://doi.org/10.1093/biomet/58.3.545>
- [12] Zhu, J. (1989) Estimation of Genetic Variance Components in the General Mixed Model. North Carolina State University, Raleigh.
- [13] Zhu, J. (1993) Methods of Predicting Genotype Value and Heterosis for Offspring of Hybrids. *Journal of Biomathematics*, **8**, 32-40.
- [14] Wu, J. (2014) Minque: An R Package for Linear Mixed Model Analyses.
<http://cran.r-project.org>
- [15] Bondalapati, K.D., Wu, J. and Glover, K.D. (2014) An Augmented Additive-Dominance (AD) Model for Analysis of Multi-Parental Spring Wheat F2 Hybrids. *Australian Journal of Crop Science*, **8**, 1441-1447.