An augmented additive-dominance (AD) model for analysis of multi-parental spring wheat F$_2$ hybrids

Krishna D. Bondalapati[1], Jixiang Wu[2*], Karl D. Glover[2]

[1]Department of Mathematics & Statistics, South Dakota State University, Brookings, SD 57007, USA
[2]Plant Science Department, South Dakota State University, Brookings, SD 57007, USA
*Corresponding author: Jixiang Wu, Plant Science Department, South Dakota State University, Northern Plains Biostress 244D, Brookings, South Dakota, 57007, USA

*Corresponding author: Jixiang.wu@sdstate.edu

Abstract

Many plant breeding trials often include a large number of genotypes with possibly no repeated field plots. Ineffective control of field variation could result in an inflated residual variance and/or biased estimation/prediction of genetic effects. Three hundred thirty seven (337) multi-parental spring wheat (*Triticum aestivum* L.) F$_2$ hybrids were grown in two locations in South Dakota without repeated field plots in 2009. Two agronomic traits (grain yield and plant height) were measured. An augmented additive-dominance (AD) model including a sub-block component was proposed for analyzing the data. This augmented AD model was first investigated by simulated data followed by actual data analysis. Simulation results showed that the augmented AD models yielded unbiased estimates with and without presence of sub-block effect. Actual data analysis revealed that grain yield and plant height were significantly influenced by the systematic field variation. Additive effects were significant for grain yield and plant height, and dominance effects were significant for plant height. Results also showed that most spring wheat lines developed by South Dakota State University breeding program (SD lines) were good general combiners for increasing grain yield and reducing plant height. Thus, this augmented AD model can reduce the impact of field variation on estimation and/or prediction of genetic effects.

Keywords: additive-dominance model; experimental design; field variation; spring wheat; variance components.
Abbreviations: AD_additive-dominance; AED_augmented experimental design; NN_nearest neighbor; ML_maximum likelihood; REML_restricted maximum likelihood; MINQUE_minimum norm quadratic unbiased estimation.

Introduction

Genetic data analysis aims to capture useful genetic information from field trials that can be used for plant improvement. Early generation trials are often focused on evaluation of a large number of genotypes where seed and/or land availability is limited. Therefore, non-replicated trials are generally preferred because the objective of a breeder at this stage is limited for selection (Federer, 1956; Federer and Crossa, 2012; Lin and Poushinsky, 1983, 1985). Therefore, appropriate genetic analysis for early generation trials, especially without field replications, is crucial to obtain useful information for crop improvement. When a genetic experiment is conducted under multiple environments, the phenotypic performance of an individual is related to its genotypic effect (G), environmental effect (E), and genotype-by-environment interaction effect (GE). Cockerham (1980) proposed a general genetic model to partition the total genetic effect (G) that has led to the development of various useful genetic models based on the biological meaning associated with a trait (Zhu, 1989, 1994; Zhu and Weir, 1994). The origin for the development of different genetic models is the concept of extendibility. For example, the additive-dominance (AD) model (Jenkins et al., 2009; Jenkins et al., 2006; Wu et al., 2010a) is one of the most popularly used genetic models, where the G effect can be partitioned into additive (A) and dominance (D) effects. This conventional AD model can be further extended to different genetic models such as an AD model with cytoplasmic effect (ADC model) (Wu et al., 2010b), AD model with maternal effects (ADM model) (Zhu, 1994), AD model with additive-by-additive interaction effects (ADAA model) (Wu et al., 2006b) etc. Above mentioned genetic analyses were suitable for parents and their bi-parental populations. In some cases, some F$_1$ or F$_2$ populations are derived from three or four parents in order to increase the possibility of combining more desirable alleles into one genotype (Kadariya et al., 2011). An AD model works appropriately under a multi-parental mating design with repeated observations (Kadariya et al., 2011). However, the statistical conclusions for multi-environment experiments are not unknown. Therefore, numerical evaluation on data analysis for multi-parent crosses under multi-environments will help extend the use of such genetic data. On the other hand, these genetic models normally require data with replications or a randomized complete block (RCB) design, which is commonly used in field experiments to control field variation (Cochran and Cox, 1957). The assumption under an RCB design is that experimental units are homogeneous within each block. However, the assumption may be violated especially when a large number of genotypes are included in one replication. Further, partitioning of blocking component (B) as separation of field variation has been rarely reported in genetic data analysis (Wu et al., 2013; Wu et al., 1998). Federer (1956) introduced the class of augmented experimental designs (AEDs) to achieve control of variability within each block in a practical and efficient manner (Federer, 1956, 1961; Federer and Crossa, 2012). These designs are also considered

as important alternatives to traditional experimental designs when genetic material is limited for replication and/or in the presence of many genotypes for testing. In AEDs, local error control can be achieved through replication of one or more checks or standard genotypes following a traditional experimental design. Augmented block designs are one of the widely used designs that control systematic field variation in one direction (Federer, 1961; Federer and Crossa, 2012). Few studies exist in the literature pertaining to the use of augmented block designs for screening large numbers of genotypes (Elouafi and Nachit, 2004; Kehel et al., 2010). Another commonly used approach to control field variation is the use of nearest neighbor (NN) based methods (Besag and Kempton, 1986; Cullis and Gleeson, 1991; Gleeson and Cullis, 1987; Kempton et al., 1994; Wilkinson et al., 1983; Williams, 1986). These methods are based on the assumption that errors between neighboring plots exhibit a higher correlation compared to the remote plots and have proven to be more effective than conventional incomplete blocking (Wilkinson et al., 1983). However, neither AEDs nor NN methods have been commonly addressed with respect to partitioning of the total genetic effect. A major limitation of the above mentioned studies is the detection of genetic effect components. Therefore, direct application of these methods in genetic studies may be limited, especially for complex genetic studies having GE interactions (Cockerham, 1980; Wu et al., 2010a). Thus, it will be of great help if AEDs or NN-based methods can be more fully integrated with complex genetic models. In this study, our objective was to propose an augmented AD genetic model from a multi-parental mating system using a two-location spring wheat (Triticum aestivum L.) data set. Our first component was to evaluate this extended genetic model with simulated data. Our second component was to apply this model to analyze two agronomic traits (grain yield and plant height). The ultimate goal of this study was to provide a general genetic model framework to incorporate systematic field variation to improve capturing desirable genetic effects for crop improvement.

Results

Because the Briggs and Oxen check genotypes were planted in the border of each row at both locations, it was appropriate to treat each row as a sub-block (incomplete blocks). Our simulations and actual data analysis were based on such a structure. The results are reported as follows.

Model evaluations

The data set included F_2 unreplicated populations within each environment. In addition, parents were not included among the field plots. Therefore, appropriateness of this extended AD model for the given data structure needed to be evaluated through simulations. Even though simulations were carried out for various cases, only three representative cases are reported in this study (Table 1). The first case assumed zero for all the variance components except residuals. This was designed to assess Type I error for all variance components. In the second set, all variance components except for the sub-block effect were set at 20 and the variance component for sub-block effect was set at zero. This was designed to assess the Type I error for the sub-block effect and testing power for all the other components. In the third set, pre-defined values for all variance components were set at 20 to assess the testing power. Two hundred simulations were carried out to evaluate the Type-I error and testing power at a nominal

probability of 0.05. Results showed that each variance component estimate was unbiased or slightly biased for both the conventional AD model (without sub-block effect in the AD model) and augmented AD model when there was no sub-block effect (sets 1 and 2 in Table 1). However, the conventional AD model yielded more biased estimates compared to the augmented AD model when there was sub-block effect (set 3 in Table 1). Mean square error (MSE) for dominance and dominance-by-environment effects was higher (set 1 in Table 1) due to no-replication within each location. Testing powers for environment, dominance and dominance-by-environment effects were low (set 2 and 3 in Table 1) due to no repeated measurements for these F^2 populations. Testing power for these components can be improved by increasing the number of environments and with repeated plots for genotypes within each environment. Overall, both models yielded similar results when the sub-block variance component was zero and augmented AD model yielded more unbiased results when there was sub-block effect. Therefore, the augmented AD model can be considered as a better choice for this given data structure.

Actual data analysis

Both full and reduced models were applied to actual spring wheat data analyses. Variance components and proportional variance components are listed in Tables 2 and 3. Since additive effects, equivalent to general combining ability, are more important for plant breeders, they were predicted using the adjusted unbiased prediction (AUP) method (Zhu, 1993) and are summarized in Table 4.

Variance components

Estimated variance components associated with full and reduced models are listed in Table 2. Results showed that sub-block effects significantly influenced grain yield and plant height (model I in Table 2). Also, residual variance (Ve) decreased by 29% for grain yield and 22% for plant height by using the full model (Model I vs Model II in Table 2). This suggests that field variations affected grain yield and plant height and could cause biased results if field variation was not controlled. Additive effects were significant for both grain yield and plant height. Dominance effects and additive-by-environment interaction effects were significant only for plant height. Proportional additive variance is equivalent to narrow-sense heritability. The estimated narrow-sense heritability for grain yield was about 28.9% based on the conventional AD model while 42.1% on the augmented AD model. The estimated narrow-sense heritability for plant height was similar for both genetic models (37.9% vs 38.5%). Therefore, the augmented AD model could improve the accuracy of estimation for genetic variance components and heritability. However, based on the estimated narrow-sense heritability for these two traits, selection for these traits in advanced generations are recommended.

Additive effects

Additive effects, equivalent to general combining ability effects, are important genetic effects for inbred line development. Predicted additive effects for grain yield and plant height are summarized in Table 4. Thirty six out of 57 parents had significant additive effects associated with grain yield. Twelve parents (00S0262-4W, 2006-10129, 2006-10130, 'Danby', HW010, HW316, MN03119-4, N-33, NDSW0601, NDSW0612, S/W COMP.W, and SD06W117)

Table 1. Estimated Type-I error, testing power, and mean square error (MSE) for variance components using both conventional and extended AD models from 200 simulated data sets based on the spring wheat data structure.

	Pre-set value[a]	Model I[b]			Model II		
		Bias[c]	MSE[d]	Power[e]	Bias	MSE	Power
Set 1							
V_E	0	0.07	0.01	0.01	0.08	0.01	0.01
V_A	0	0.44	0.19	0.04	0.45	0.20	0.08
V_D	0	3.38	11.56	0.08	3.34	11.23	0.11
V_{AE}	0	0.48	0.23	0.07	0.46	0.22	0.08
V_{DE}	0	4.77	22.96	0.09	4.39	19.51	0.08
V_{SB}	0	0.16	0.03	0.04	-	-	-
V_e	20	0.01	0.01	1.00	0.02	0.01	1.00
Set 2							
V_E	20	0.18	3.46	0.70	0.19	3.42	0.72
V_A	20	0.20	0.35	0.99	0.20	0.34	0.99
V_D	20	-1.53	2.89	0.52	-1.33	2.30	0.51
V_{AE}	20	-0.24	0.22	1.00	-0.15	0.18	1.00
V_{DE}	20	0.78	1.58	0.43	0.57	1.26	0.50
V_{SB}	0	0.18	0.03	0.06	-	-	-
V_e	20	0.01	0.01	1.00	0.02	0.01	1.00
Set 3							
V_E	20	-0.56	4.99	0.60	0.45	4.97	0.60
V_A	20	0.64	0.78	0.99	1.32	2.23	0.97
V_D	20	1.81	3.94	0.60	4.55	23.03	0.46
V_{AE}	20	-0.02	0.15	1.00	3.97	16.02	1.00
V_{DE}	20	-0.64	1.28	0.44	3.68	16.87	0.31
V_{SB}	20	0.28	0.19	1.00	-	-	-
V_e	20	-0.08	0.02	1.00	14.99	224.69	1.00

[a] Pre-set variance component is set at zero to evaluate Type-I error and is set at a non-zero value to assess testing power at 0.05 probability level.
[b] Model I refers to extended AD model after including sub-block effect and model II refers to the conventional AD model without sub-block effect.
[c] Deviation of mean estimate from the pre-set value
[d] Represents Type-I error for pre-set variance component zero and testing power for non-zero pre-set value.
[e] Mean square error related to bias and variation.

Table 2. Estimated variance components for grain yield and plant height from both conventional AD model and extended AD models.

Variance component[a]	Model I[b]		Model II	
	Grain yield	Plant height	Grain yield	Plant height
V_E	57905.54 **	1.67 *	62669.78 **	2.55 **
V_A	17755.65 **	24.04 **	16375.25 **	27.93 **
V_D	4645.55	15.57 **	16981.24 *	18.69 **
V_{AE}	2313.76	6.94 **	3571.75	6.80 **
V_{DE}	3668.75	0.00	350.54	0.00
$V_{SB(E)}$	7132.56 **	6.64 **	-	-
V_e	13757.80 **	15.83 **	19442.55 **	20.35 **

[a] Variance components calculated based on F_1, $V_E=\sigma_E^2$; $V_A=\sigma_A^2$; $V_D=\sigma_D^2$; $V_{AE}=\sigma_{AE}^2$; $V_{DE}=\sigma_{DE}^2$; $V_{SB}=\sigma_{SB}^2$; and $V_e=\sigma_e^2$.
[b] Model I refers to extended AD model after including sub-block effect and model II refers to the conventional AD model without sub-block effect.
*, ** Significant at 0.05 and 0.01, respectively.

exhibited negative additive effects, indicating that these parents are more likely to reduce grain yield if used as parents for crossing. Parents 00S0219-10W, 2006-10126, 'Glenn', ND04/3-20 and 20 SD lines (29 SD lines in total) exhibited positive additive effects for grain yield. Therefore, these parents can be used as general combiners to improve grain yield. Thirty seven of 57 parents exhibited significant additive effect for plant height. Twenty parents (00S0120-3W, 00S0262-4W, 98S0113-20-23, HW316, 'Kelby', 'RB07' and 14 SD lines) exhibited significant negative additive effects for plant height, indicating that they can be used as general combiners to reduce the plant height. On the other hand, 17 parents (00S0129-10W, 2006-10129, 'Alice', BC98334-10W-8W, Danby, Glenn, 'Howard', HW010, N-31, N-33, ND04/3-20, NDSW0612, SD06W117, SD3997, SD4073, SD98W175-1-14, and 'Wendy') exhibited significant positive additive effects for plant height, indicating that these parents are likely to increase plant height if used as parents. Overall, grain yield and plant height were

significantly influenced by field variations and this extended AD model improved the data analysis by including it in the AD model. Genetic effects revealed that some parents, especially SD lines, could be used as general combiners for improving yield and reducing plant height.

Discussion

Experiments in many plant breeding trials at early stages often involve a large number of test entries, which can increase the possibility of selecting the desirable genotypes or genetic effects. One challenging issue is that some test populations could be derived from more than two parents such as some F_2 populations used in this study. The second critical issue commonly associated with these experiments is no replication when availability of seed or land is limited. This study aimed to target these two issues with the applications of appropriate statistical model and methodologies. Based on the data structure used in this study,

Table 3. Estimated variance components expressed as proportion to the phenotypic variance for two spring wheat traits from the extended AD model (model I).

Variance component[a]	Grain yield	Plant height
V_A/V_P	0.42 **	0.38 **
V_D/V_P	0.11	0.25 **
V_{AE}/V_P	0.06	0.11 **
V_{DE}/V_P	0.08	0.00
V_e/V_P	0.33 **	0.26 **

[a] The total variance V_P is the sum of V_A, V_D, V_{AE}, V_{DE}, and V_e. *, ** Significant at 0.05 and 0.01, respectively

Table 4. Predicted additive effects for two spring wheat traits from the extended AD model (model I).

ID	Parents	Grain yield		Plant height	
		Estimate	Pvalue	Estimate	Pvalue
1	00S0120-3W	-33.29	0.29	-4.66	0.00
2	00S0219-10W	104.29	0.04	2.75	0.01
3	00S0262-4W	-86.74	0.00	-2.83	0.00
4	02V03*N1545	-26.67	0.47	1.87	0.09
5	2006-10126	68.97	0.01	0.12	0.98
6	2006-10129	-180.54	0.00	3.75	0.00
7	2006-10130	-107.95	0.04	1.69	0.10
8	98S0113-20-23	-28.62	0.58	-14.46	0.00
9	ALICE	-76.55	0.07	4.62	0.00
10	BC98334-10W-8W	-55.32	0.09	2.73	0.03
11	Danby	-147.29	0.01	4.81	0.00
12	GLENN	39.58	0.00	1.88	0.00
13	HOWARD	-2.05	0.98	2.57	0.00
14	HW010	-165.47	0.00	1.79	0.00
15	HW316	-128.21	0.00	-1.52	0.03
16	KELBY	35.48	0.07	-2.59	0.01
17	LOLO	59.46	0.10	0.76	0.77
18	MN03119-4	-40.53	0.00	-0.74	0.07
19	MTHW0471	-11.41	0.60	-0.54	0.28
20	N-31	-43.50	0.18	5.86	0.00
21	N-33	-76.08	0.00	1.96	0.02
22	ND04/3-20	37.16	0.01	2.99	0.00
23	NDSW0601	-136.19	0.03	-1.36	0.72
24	NDSW0612	-113.75	0.00	1.10	0.00
25	RB07	-38.32	0.07	-0.87	0.04
26	S/W COMP.W	-49.67	0.00	0.47	0.73
27	SD02W129	246.84	0.00	0.74	0.86
28	SD06W117	-230.89	0.00	6.20	0.00
29	SD3851	36.72	0.01	-0.56	0.23
30	SD3868	-1.70	0.99	0.47	0.96
31	SD3900	14.38	0.22	0.43	0.95
32	SD3942	81.81	0.00	-8.62	0.02
33	SD3943-21	150.77	0.00	-2.03	0.00
34	SD3944	147.85	0.00	-1.10	0.00
35	SD3948	38.17	0.02	-0.40	0.27
36	SD3956	25.21	0.11	-0.63	0.25
37	SD3997	-10.46	0.90	3.12	0.00
38	SD4011	72.00	0.00	-3.19	0.00
39	SD4018	-17.87	0.45	-2.24	0.00
40	SD4027	89.50	0.00	-1.26	0.01
41	SD4032	89.57	0.00	-1.34	0.00
42	SD4036	52.97	0.00	-0.50	0.19
43	SD4056	76.23	0.01	2.28	0.10
44	SD4070	39.24	0.00	-0.61	0.05
45	SD4073	89.37	0.00	1.72	0.00
46	SD4076	87.39	0.00	-2.04	0.00
47	SD4078	139.91	0.00	-4.05	0.00
48	SD4081	74.34	0.00	-1.45	0.00
49	SD4085	35.26	0.01	-2.24	0.00
50	SD4101	26.31	0.02	0.02	0.99
51	SD4102	14.50	0.30	-2.06	0.00
52	SD4105	39.90	0.14	-2.15	0.00
53	SD4106	42.87	0.04	-0.23	0.84
54	SD98W175-1	68.97	0.01	0.12	0.98
55	SD98W175-1-14	-50.56	0.43	6.80	0.00
56	TRAVERSE	12.40	0.83	1.40	0.07
57	WENDY	-12.38	0.85	5.34	0.00

*, ** Significant at 0.05 and 0.01, respectively.

we proposed to use each row as a sub-block because each row included one or two common genotypes. Simulation showed that when the sub-block component is found to be insignificant, it can safely be removed from consideration because when there is no significant variation due to sub-blocks, both the extended AD model (Model I in Table 1) and the conventional AD model (Model II in Table 1) perform equally well. However, when there was a significant variation due to the sub-block component, the extended AD model in this study appeared to more desirable because it could yield less biased estimation for variance components compared to the AD model without including sub-block effects. The results were consistent with our previous study where an AD model was based on bi-parental crosses under a generalized lattice experimental design (Wu et al., 2013). This study was generalized from our previous study with extension to multi-parental cross data. In our previous study, a generalized lattice experimental design was applied because the field plots were arranged rectangularly; however, the field layout in this study did not allow the use of a generalized lattice design. With the applications of linear mixed model approaches, the simulated results for the AD model derived from bi-parental and/or multi-parental data structures were consistent (Wu et al., 2013). The actual data analysis showed that the augmented AD model was more preferred over the conventional AD model. Therefore, the augmented AD model can be recommended for a general use. Though augmented AD models such as used in previous this study (Wu et al., 2013) and this study could provide better results compared to the conventional AD models, it must be pointed out that testing power for dominance effects and dominance-by-environment interaction effects is relatively low compared to additive effects without field replications. To improve testing power for dominance effects and their interactions with environments, repeated field plots are highly recommended.

Materials and Methods

Plant materials and experiment

The data set used in this study included 334 F_2 spring wheat populations generated from 57 parents. Among the 334 F_2 populations, 269 were obtained from bi-parental crossing, 49 were obtained from tri-parental crossing and 16 were obtained from tetra-parental crossing. All 334 F_2 populations were grown in 2009 at Aurora, SD, and 323 F_2 populations were grown in Watertown, SD. In addition to the F_2 populations, 'Briggs' and 'Oxen' were grown as comparative checks and were repeated 21 and 20 times in Aurora and Watertown, respectively. In addition, Briggs was used as a plot filler on eight occasions in each location and 'Chris' was used 3 times in Aurora and 16 times in Watertown. None of the genotypes Briggs, Oxen, or Chris were used as parents for F_2 populations. All F_2 populations were planted at each location without replications. Field layout in Aurora was a rectangular arrangement of plots with 20 rows and 21 columns. Individual plots were planted to measure 1.5 by 4.6 m consisting of 7 rows (21.4 cm row spacing). Prior to heading, however, all plots were trimmed to a final length of 4.1 m. Briggs was planted twice within each row in columns 1 and 21, while Oxen was planted in each row at 11th column. All the F_2 populations and fillers (Briggs and Chris) were randomly planted in the remaining plots. Layout in Watertown was similar to Aurora except the last two rows (19 & 20) were planted parallel to rows 1 and 2 due to space constraint.

Statistical model

The linear model presented in eq. (1) is a general genetic model. Based on the spring wheat data structure and field layouts used in this study, complex genetic models can be derived from eq. (1).

$$y = \mu + E + G + GE + SB(E) + e \qquad (1)$$

Where, μ represents the population mean, E represents the environmental effect, G represents the total genotypic effect, *GE* represents genotype-by-environment interaction effect, and SB(E) represents the sub-block effect within each environment. A random error is represented by *e*. A sub-block can be either a single or multiple rows or columns or a combination of rows and columns dependent on field layout and/soil conditions. It is recommended that at least one control line be embedded in each sub-block and the plots within each sub-block are more similar than between sub-blocks. If a sub-block includes a complete/incomplete replication of check genotypes, the general model in eq. (1) refers to augmented block designs. Moreover, if a RCB design is employed, the block effect can be integrated as a sub-block effect including a complete replication of genotypes. Therefore, the model in eq. (1) applies to a more generalized experiment including the experimental design used in this study. In an AD genetic model used in this study, the genetic effect (G) is partitioned into additive (A) and dominance (D) effects (Cockerham, 1980; Jenkins et al., 2009; Jenkins et al., 2006). No other genetic effects are assumed. In the same manner, the *GE* effect can be partitioned into additive-by-environment (*AE*) and dominant-by-environment (*DE*) effects. The total G effect in eq. (1) is different for parents and F_1 or F_2 generations derived from a cross of two, three, and four homozygous lines. For generalization, assume that female is a F_1 cross of parents *j* and *k*; and male is a F_1 cross of parents *l* and *m*. Generalized additive-dominance (AD) model for the cross of $F_{1(jk)}$ and $F_{1(lm)}$ in the ith sub-block nested with in h[th] environment can be expressed as:

$$
\begin{aligned}
F_1: y_{hijklm(F_1)} = {} & \mu + E_h + SB_{i(h)} + \frac{1}{2}\left(A_j + A_k + A_l + A_m\right) \\
& + \frac{1}{4}\left(D_{jl} + D_{jm} + D_{kl} + D_{km}\right) \\
& + \frac{1}{2}\left(AE_{hj} + AE_{hk} + AE_{hl} + AE_{hm}\right) \\
& + \frac{1}{4}\left(DE_{hjl} + DE_{hjm} + DE_{hkl} + DE_{hkm}\right) \\
& + e_{hijkl(F_1)} \qquad (2)
\end{aligned}
$$

$$
\begin{aligned}
F_2: y_{hijklm(F_2)} = {} & \mu + E_h + SB_{i(h)} \\
& + \frac{1}{2}\left(A_j + A_k + A_l + A_m\right) \\
& + \frac{1}{8}\big(D_{jj} + D_{kk} + D_{ll} + D_{mm} + D_{jl} + D_{kl} \\
& \quad + D_{jm} + D_{km}\big) \\
& + \frac{1}{2}\left(AE_{hj} + AE_{hk} + AE_{hl} + AE_{hm}\right) \\
& + \frac{1}{8}\big(DE_{hjj} + DE_{hkk} \\
& \quad + DE_{hll} + DE_{hmm} + DE_{hjl} + DE_{hkl} \\
& \quad + DE_{hjm} + DE_{hkm}\big) \\
& + e_{hijkl(F_2)} \qquad (3)
\end{aligned}
$$

where $\mu, E_h,$ and $SB_{i(h)}$ were defined as above.

A_j (A_k, A_l, A_m) is the additive effect;

D_{jj} $(D_{kk}, D_{ll}, D_{jk}, D_{jl}, D_{kl}, D_{jm}, D_{kl})$ is the dominance effect;

AE_{hj} $(AE_{hk}, AE_{hl}, AE_{hm})$ is the additive-by-environment interaction (*AE*) effect;

DE_{hjj} ($DE_{hkk}, DE_{hll}, DE_{hjk}, DE_{hjl}, DE_{hkl}, DE_{hjm}, DE_{hkl}$) is the dominance-by-environment interaction (DE) effect, respectively.

The eq.s (2) and (3) can be converted to different mating designs. For example, if $j=k=l=m$, then the AD model in eq.s (2) and (3) becomes an AD model for parents. If $j=k$ and $l=m$, then the above equations are for a bi-parental cross. If $j=k$ and $m \neq l$ or $j \neq k$ and $m=l$, then these two equations are for a tri-parental cross. If $j \neq k$ and $m \neq l$ then the above equations represent an AD model for tetra-parental cross.

Methodology

There are three commonly used linear mixed model approaches: maximum likelihood (ML), restricted maximum likelihood (REML), and minimum norm quadratic unbiased estimation (MINQUE) (Hartley and Rao, 1967; Rao, 1971; Searle et al,. 1992). Since MINQUE approach does not require iterations (Rao, 1971), it has less computational load and it was used in this study to perform simulations and the actual data analyses. MINQUE requires prior values to estimate variance components. Since different prior values yield similar variance component estimations (Zhu, 1989), we set $\alpha_u = 1$ ($u = 1,2,\dots 7$) for both simulation study and actual data analysis. In addition to the variance components, genotypic effects were predicted using the adjusted unbiased prediction (AUP) approach (Zhu, 1993). A *g*-fold non-pseudo jackknife technique described by Wu et al. (2013) was applied to calculate the standard deviations and to conduct statistical tests for these parameters of interest. Our various simulations showed that $g=10$ is an ideal choice for this study to control Type-I error and to obtain desirable testing power. All results presented in this study were based on using five repetitions of this 10-fold jackknife process and produced very consistent results among replications. Thus, this non-pseudo jackknife technique was employed to detect significance for each parameter. Based on the data structure, this study included two important components. The first was a model evaluation for the spring wheat data set using three parameter configurations (Table 1). Bias, empirical Type I error, testing power, and mean square error (MSE) (Wu et al., 2006a; Wu et al., 2006b) were calculated for each variance component based on 200 simulated data sets. The second component was the actual data analysis for two agronomic traits (grain yield and plant height). All data analyses were completed using R-software and GenMod (an R-package) (Wu, 2012) which can be used for various genetic data analyses.

Conclusions

In this study, we proposed an augmented AD genetic model for multi-parental spring wheat F_2 hybrids with control of field variation when repeated field plots for F_2 hybrids were not available in two locations in South Dakota. Simulation results showed that the augmented AD models yielded unbiased estimates with and without presence of field row effect and can be recommended for a general use. Augmented AD model analysis revealed that grain yield and plant height were significantly influenced by the systematic field variations and could improve estimation of heritability for these traits. The results also showed that most spring wheat lines developed by South Dakota State University breeding program were good general combiners for increasing grain yield and reducing plant height. Thus, this augmented AD model can provide more accurate genetic information for crop improvement.

References

Besag J, Kempton R (1986) Statistical analysis of field experiments using neighbouring plots. Biometrics. 42:231-251.

Cochran WG, Cox GM (1957) Experimental Design, 2nd edn. Wiley, New York.

Cockerham CC (1980) Random and fixed effects in plant genetics. Theor Appl Genet. 56:119-131.

Cullis BR, Gleeson AC (1991) Spatial analysis of field experiments- An extension to two dimensions. Biometrics. 47:1449-1460.

Elouafi I, Nachit MM (2004) A genetic linkage map of the Durum x Triticum dicoccoides backcross population based on SSRs and AFLP markers, and QTL analysis for milling traits. Theor Appl Genet. 108:401-413.

Federer WT (1956) Augmented (hoonuiaku) designs. Hawaii Plant Rec. 55:191-208.

Federer WT (1961) Augmented designs with one-way elimination of heterogeneity. Biometrics. 17:447-473.

Federer WT, Crossa J (2012) I.4 Screening Experimental Designs for Quantitative Trait Loci, Association Mapping, Genotype-by Environment Interaction, and Other Investigations. Frontiers Physiol. 3:156

Gleeson AC, Cullis BR (1987) Residual maximum likelihood (REML) estimation of neighbour model for field experiments. Biometrics. 43:277-287.

Hartley HO, Rao JNK (1967) Maximum-likelihood estimation for the mixed analysis of variance model. Biometrika. 54:93-108.

Jenkins J, McCarty J, Jr., Wu J, Gutierrez O (2009) Genetic variance components and genetic effects among eleven diverse upland cotton lines and their F2 hybrids. Euphytica. 167:397-408.

Jenkins JN, Wu J, McCarty JC, Saha S, Gutierrez O, Hayes R, Stelly DM (2006) Genetic effects of thirteen Gossypium barbadense L. chromosome substitution lines in topcrosses with upland cotton cultivars: I. Yield and yield components. Crop Sci. 46:1169-1178.

Kadariya M, Glover KD, Wu J, Gonzalez JL (2011) Multi-parental mating design analysis: Model evaluation and application in spring wheat. In: Song W (ed) 23rd Applied Statistics in Agriculture, Manhattan, KS, pp 111-127.

Kehel Z, Habash DZ, Gezan SA, Welham SJ, Nachit MM (2010) Estimation of Spatial Trend and Automatic Model Selection in Augmented Designs. Agron J. 102:1542.

Kempton RA, Seraphin JC, Sword AM (1994) Statistical analysis of two-dimensional variation in variety yield trials. J Agr Sci. 122:335-342.

Lin CS, Poushinsky G (1983) A modified augmented design for an early stage of plant selection involving a large number of test lines without replication. Biometrics. 39:553-561.

Lin CS, Poushinsky G (1985) A modified augmented design (Type 2) for rectangular plots. Can J Plant Sci 65:743-749

Rao CR (1971) Estimation of variance and covariance components—MINQUE theory. J Multivariate Anal. 1:257-275.

Searle SR, Casella G, McCulloch CE (1992) Variance components. Wiley, New York, NY

Wilkinson GN, Eckert SR, Hancock TW, Mayo O (1983) Nearest neighbour (NN) analysis of field experiments. J Roy Stat Soc B. 45:151-211.

Williams ER (1986) A neighbour model for field experiments. Biometrika. 73:279-287.

Wu J (2012) GenMod: An R package for various agricultural data analyses. ASA, CSSA, and SSSA 2012 International Annual Meetings, Cincinnati, OH, p 127.

Wu J, Bondalapati K, Glover K, Berzonsky W, Jenkins JN, McCarty JC (2013) Genetic analysis without replications: model evaluation and application in spring wheat. Euphytica. 190:447-458.

Wu J, Jenkins JN, McCarty JC (2010a) A generalized approach and computer tool for quantitative genetics study. In: Gadbury G (ed) Applied Statistics in Agriculture, Manhattan, KS, pp 85-106.

Wu J, Jenkins JN, McCarty JC, Saha S, Stelly DM (2006a) An additive-dominance model to determine chromosomal effects in chromosome substitution lines and other gemplasms. Theor Appl Genet. 112:391-399.

Wu J, Jenkins JN, McCarty JC, Wu D (2006b) Variance component estimation using the additive, dominance, and additive × additive model when genotypes vary across environments. Crop Sci. 46:174-179.

Wu J, McCarty JC, Jenkins JN (2010b) Cotton chromosome substitution lines crossed with cultivars: genetic model evaluation and seed trait analyses. Theor Appl Genet. 120:1473-1483.

Wu J, Xu H, Zhu J (1998) An approach to eliminating systematic errors in genetic analysis. In: Chen LS, Ruan SG, Zhu J (eds) International conference on mathematical biology. World Scientific Publishings Co., Singapore, pp 265-270.

Zhu J (1989) Estimation of genetic variance components in the general mixed model. North Carolina State University, Raleigh, NC.

Zhu J (1993) Methods of predicting genotype value and heterosis for offspring of hybrids (Chinese). J Biomath. 8:32-44.

Zhu J (1994) General genetic model and new analysis methods for quantitative traits (Chinese). J Zhejiang Agric Univ. 20:551-559.

Zhu J, Weir BS (1994) Analysis of cytoplasmic and maternal effects I. A genetic model for diploid plant seeds and animals. Theor Appl Genet. 89:153-159.