




## Article

# Joint Modeling of Genetics and Field Variation in Plant Breeding Trials Using Relationship and Different Spatial Methods: A Simulation Study of Accuracy and Bias

Éder David Borges da Silva <sup>1,2,\*</sup> , Alencar Xavier <sup>3,4</sup>  and Marcos Ventura Faria <sup>2</sup> 

<sup>1</sup> Department of Plant Breeding, Corteva Agriscience, Rio Verde Research Station, Rio Verde 75909-796, Goiás, Brazil

<sup>2</sup> Departamento de Agronomia, Universidade Estadual do Centro Oeste, Guarapuava 85040-167, Paraná, Brazil; mfaria@unicentro.br

<sup>3</sup> Department of Biostatistics, Corteva Agriscience, Johnston, IA 50131, USA; alencar.xavier@corteva.com

<sup>4</sup> Department of Agronomy, Purdue University, West Lafayette, IN 47907, USA

\* Correspondence: eder.silva@corteva.com



**Citation:** Borges da Silva, É.D.; Xavier, A.; Faria, M.V. Joint Modeling of Genetics and Field Variation in Plant Breeding Trials Using Relationship and Different Spatial Methods: A Simulation Study of Accuracy and Bias. *Agronomy* **2021**, *11*, 1397. <https://doi.org/10.3390/agronomy11071397>

Academic Editor: Jacqueline Batley

Received: 21 May 2021

Accepted: 5 July 2021

Published: 12 July 2021

**Publisher's Note:** MDPI stays neutral with regard to jurisdictional claims in published maps and institutional affiliations.



**Copyright:** © 2021 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

**Abstract:** Modelling field spatial patterns is standard practice for the analysis of plant breeding. Jointly fitting the genetic relationship among individuals and spatial information enables better separability between the variance due to genetics and field variation. This study aims to quantify the accuracy and bias of estimative parameters using different approaches. We contrasted three settings for the genetic term: no relationship (I), pedigree relationship (A), and genomic relationship (G); and a set of approaches for the spatial variation: no-spatial (NS), moving average covariate (MA), row-column adjustment (RC), autoregressive  $AR1 \times AR1$  (AR), spatial stochastic partial differential equations, or SPDE (SD), nearest neighbor graph (NG), and Gaussian kernel (GK). Simulations were set to represent soybean field trials at  $F_{2.4}$  generation. Heritability was sampled from a uniform distribution  $U(0,1)$ . The simulated residual-to-spatial ratio between residual variance and spatial variance ( $V_e:V_s$ ) ranged from 9:1 to 1:9. Experimental settings were conducted under an augmented block design with the systematic distribution of checks accounting for 10% of the plots. Relationship information had a substantial impact on the accuracy of the genetic values ( $G > A > I$ ) and contributed to the accuracy of spatial effects (30.63–42.27% improvement). Spatial models were ranked based on an improvement to the accuracy of estimative of genetic effects as  $SD \geq GK \geq AR \geq NG \geq MA > RC \geq NS$ , and to the accuracy of estimative of spatial effects as  $GK \geq SD \geq NG > AR \geq MA > RC$ . Estimates of genetic and spatial variance were generally biased downwards, whereas residual variances were biased upwards. The advent of relationship information reduced the bias of all variance components. Spatial methods SD, AR, and GK provided the least biased estimates of spatial and residual variance.

**Keywords:** field plot variation; spatial adjustment; genomic selections; soybean breeding; spatial modeling

## 1. Introduction

Commercial breeding programs rely extensively on data-driven tools to efficiently improve crop genetics. The data analysis is based on modeling incorporate spatial and genetic relationship information.

In terms of modeling spatial patterns, plant breeding trials have historically utilized moving averages based on neighbor plots [1,2] as a simple method to account for within-field variation at higher resolutions than what is captured by the experimental blocks [3,4]. Spatial covariates have also been used in combination with other methods to increase accuracy and decrease bias [5,6], and it has shown favorable results when paired with genomic analysis [7]. Nevertheless, the spatial approaches have shifted from covariates to covariances with the advent of computing power, in which case the within-field variation

is reparametrized as structured random effects [8]. Under this framework, kernel methods, such as kriging and splines [9], are incorporated seamlessly into mixed models for plant breeding [10].

The current gold standard for the parameterization of spatial patterns in breeding trials is the two-dimensional auto-regressive structure referred to as  $AR1 \times AR1$  [11], where the dimensional correlations ( $\rho_x, \rho_y$ ) are estimated from data via restricted maximum likelihood [11–17].  $AR1 \times AR1$  has served as a complementary approach to common experimental design [18], however, convergence can be a major issue due to the multiplicity of parameters being estimated [17]. Two kriging-type Matérn processes [19,20] that can be utilized as stable alternatives to  $AR1 \times AR1$  are the stationary Gaussian kernels [21], and the stochastic partial differential equation (SPDE), which correspond to a spline-like generalization of the  $AR1$  process [22,23]. Yet, literature is scarce on the direct comparison among the multiple geostatistical methods when it comes to the analysis of plant breeding trials.

In terms of genetic information, modeling relationships are known to increase the accuracy of estimative genetic effects [24], with more impact in case of no replication or a few numbers of replications [25–29]. In addition, the joint model of genetic relationship has been reported to improve the signal separability between genetic and spatial variation [7,30,31].

The benefits in terms of accuracy attributed to relationship, spatial adjustment, and the combination of both have not been benchmarked, and their benefits remain unknown for early breeding stages. The goal of this study is to assess the accuracy and bias gains while comparing the combination of different parameterization approaches for the genetics and spatial patterns using simulation. Accuracy was measured on the coefficients as the correlation between estimated and true simulated values, and bias was measured on the variance components as the difference between estimated and true simulated values. Our simulation settings aim to reproduce the non-replicated trials from the early stages of varietal breeding programs, where the control of field plot variation is critical.

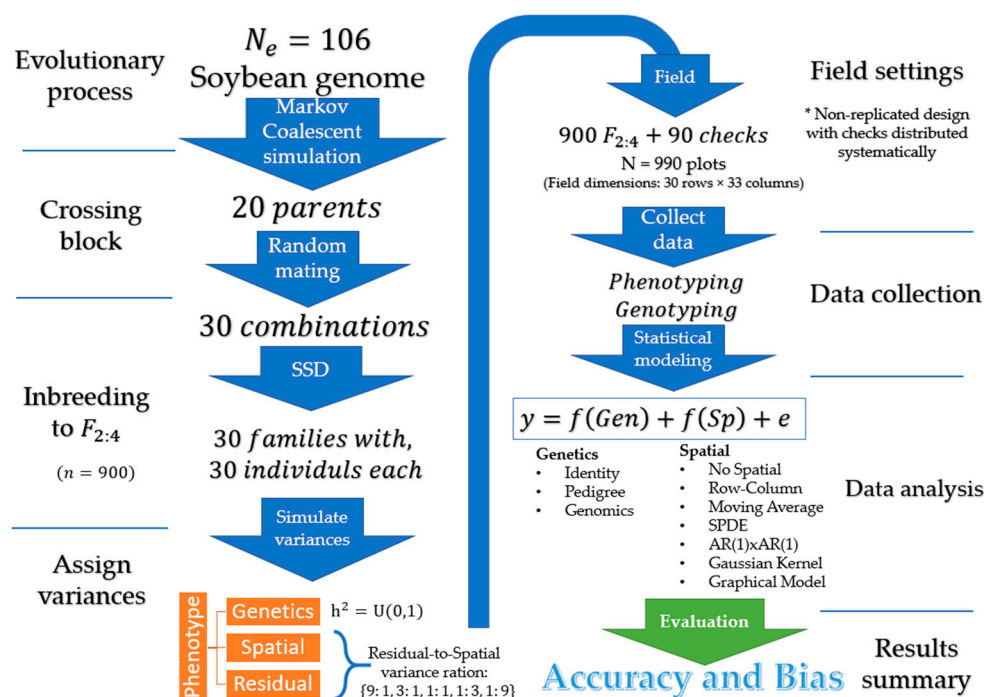
## 2. Material and Methods

### 2.1. Simulated Data

This section describes the general simulation settings, including the simulated genome, creation of the founder population and breeding parents, the experimental design, and simulated variance components. An overall representation of the simulation settings is provided in Figure 1. The next sections are reserved for the description of statistical models and the implementation details.

#### 2.1.1. Breeding Scenario

Simulation settings recreated early generation trials of a varietal breeding program, using soybeans (*Glycine max* L.) as template species to recreate the genomic settings. Soybean was chosen for its commercial importance [32]. Field settings were based on a single location containing balanced families, non-replicated entries, and performance checks. Entries were simulated to be at  $F_{2:4}$  generation, where selection would be performed at a single-observation level.



**Figure 1.** Illustration of the simulation steps: Create founder germplasm, form breeding population, assign variances, allocate plots into the field, data observation, statistical analysis, a summary of accuracy and bias.

### 2.1.2. Genome

Simulated genomic parameters were based on a soybean genome [33]. The genome contained 20 chromosomes, with an average length of 128 cM, covering and 950 Mb in genome size. The simulated genome contained 46,000 segregating polymorphic sites.

### 2.1.3. Breeding Lines

A set of founder lines was set with an effective population size ( $N_e$ ) of 106 [34,35]. The founder lines would represent the entire diversity of the species, not the breeding program. From the 106 founders, we recreated an evolutionary process from which 20 breeding lines were produced. The evolutionary process was based on the Markovian Coalescent Simulator [36] that recreates multiple cycles of drift, mutation, and selection. The 20 breeding lines were utilized as parents for the populations under evaluation.

### 2.1.4. Breeding Trial

From the 20 breeding parents, 30 cross-combinations were performed at random. The 30 biparental families had 30 individuals each, totaling 900 breeding lines. The breeding trial included the 900 breeding lines in addition to 90 check plots, composed of 3 genotypes with 30 replications each. These were allocated into a rectangular field (30 rows × 33 columns) under a non-replicated augmented design, with checks systematically placed in the field [37]. Breeding lines, checks, and field layout were simulated separately in each iteration.

### 2.1.5. Genotyping and Breeding Values

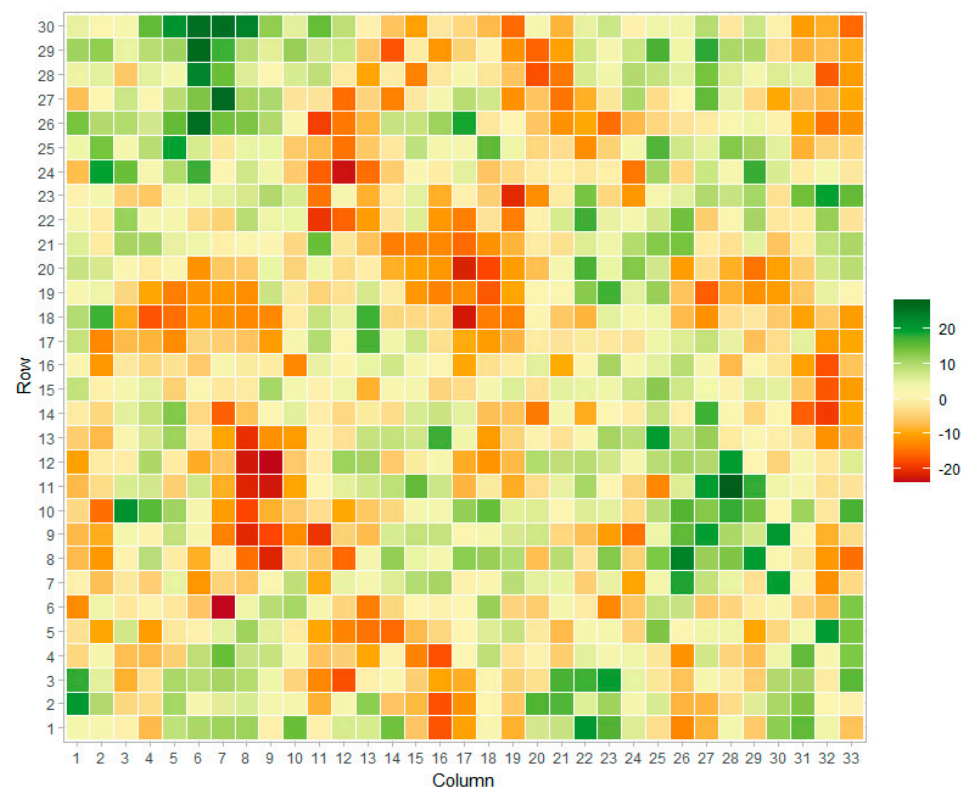
Breeding lines were simulated and genotyped with 20,000 markers to mimic commercial SNP arrays available for soybeans, hence capturing a subset of all simulated polymorphic sites (46,000). True breeding values were obtained from the true genomic information under an infinitesimal model, hence all polymorphic sites had a small value sampled from a normal distribution. The observed SNP data would not include all true segregating loci, reproducing previous methodological frameworks [38]. The true breeding values were subsequently rescaled to display the variance parameters described next.

### 2.1.6. Population Parameters

A different set of variances were sampled for each run. First, a heritability value was sampled from a uniform distribution as  $h^2 \sim U(0, 1)$ , from which the ratio between genetic variance ( $\sigma_g^2$ ) and residual variance ( $\sigma_e^2$ ) was estimated as follows:  $\sigma_e^2 = (1 - h^2) \div h^2$ , assuming  $\sigma_g^2 = 1$ . Spatial variance ( $\sigma_s^2$ ) was computed as a proportion of the residual variance, following the ratios:  $\delta(\sigma_s^2 : \sigma_e^2) = 1:9, 1:3, 1:1, 3:1$  and  $9:1$ , thus  $\sigma_s^2 = \sigma_e^2 \delta$ . In terms of intra-class correlation coefficients,  $ICC = \sigma_s^2 \div (\sigma_s^2 + \sigma_e^2)$ , the rates correspond to  $ICC = 0.10, 0.25, 0.50, 0.75$ , and  $0.90$ . The fixed ratios  $\delta$  enables contrasting scenarios where the pure error is prevalent over spatial signal, and vice-versa. Variance components were subsequently rescaled to add up to 1.

### 2.1.7. Spatial Variation

Field spatial patterns were simulated from a stationary Matérn covariance function [39]. This function is commonly utilized for the simulation of spatial gradients in two-dimensional projections [40–42]. The signal from Matérn processes can be adequately recovered from most spatial procedures such as Gaussian kernels, splines, and auto-regressive moving averages [43]. A simulated field is presented in Figure 2.



**Figure 2.** Example of simulated field signal using the Matérn function.

## 2.2. True and Statistical Model

### 2.2.1. True Model

The true model utilized to build the response ( $y$ ) was based on a random effect model set as the linear combination of an overall mean ( $\mu$ ), genetics ( $g$ ), spatial effect ( $s$ ), and i.i.d. residuals ( $e$ ). Thus,

$$y = \mu + g + s + e$$

Genetic, spatial, and residual terms were assumed to be normally distributed as  $g \sim N(0, G\sigma_a^2)$ ,  $s \sim N(0, S\sigma_a^2)$  and  $e \sim N(0, I\sigma_e^2)$ . The covariance matrices correspond to the genomic relationship build from genome-wide polymorphism ( $G$ ) and the Matérn covariance structure ( $S$ ) described above. For simplicity, the true genetic term did not include dominance or epistasis effects.

### 2.2.2. Statistical Model

The statistical model consists of the parametrizations that attempt to recover the true simulated values. The technical details of the model parametrizations are described in the next two sections. The statistical model under evaluation consisted of:

$$y = \hat{\mu} + \hat{g} + \hat{s} + \hat{e}$$

The notation “hat” here represents the estimated coefficients and matrices, whereas the non-hat regards the simulated (or true) values. In the statistical model,  $y$  is a vector containing the response variable,  $\hat{\mu}$  is the estimated intercept,  $\hat{g}$  is the vector of estimated genetic effects,  $\hat{s}$  is the vector of estimated spatial effects and  $\hat{e}$  is the vector of residuals. The genetic term was parameterized as  $\hat{g} \sim N(0, K\hat{\sigma}_g^2)$ , under the scenarios where the relationship among individuals is parametrized as unknown ( $K = I$ ), based on single-generation pedigree information ( $K = \hat{A}$ ), and based on the estimated genomic information ( $K = \hat{G}$ ) observed from the SNP array. The spatial term was set one of the following parametrizations: no spatial adjustment (NS), row-column adjustment (RC), moving average covariate (MA), AR1  $\times$  AR1 structure (AR), SPDE structure (SD), nearest neighbor graph (GM), or Gaussian kernel (GK).

### 2.2.3. Accuracy and Bias

The accuracy of coefficients is defined as the Pearson correlation between the estimate and true parameters. Hence, for the vector of genetic and spatial coefficients, the accuracy is defined as  $\text{cor}(g, \hat{g})$  and  $\text{cor}(s, \hat{s})$ . The accuracy of selection for the top 10% estimated as the Jaccard coefficient of coincidence, measuring the percentage overall between the top-ranked individuals from the estimate and true breeding values. Bias was measured as the difference between the estimate and true parameters (i.e.,  $\hat{\theta} - \theta$ ). Measures of bias were computed for the variance components of the genetic, spatial, and residual terms.

## 2.3. Genetic Parametrizations

The genetic term of our statistical model was set as a random effect, with coefficients normally distributed,  $\hat{g} \sim (0, K\hat{\sigma}_g^2)$ . Three parametrizations of the correlation matrix ( $K$ ) were:

### 2.3.1. No Relationship ( $K = I$ )

This parametrization is based on unstructured levels, which translated into the assumption that the relationship among genotypes is unknown. Under this setting, the genetic variance estimates largely rely on the replicated checks.

### 2.3.2. Pedigree Relationship ( $K = \hat{A}$ )

In this study, the pedigree relationship matrix ( $\hat{A}$ ) was based on a single generation, hence capturing relationship at the full- and half-sibling level. Such parametrization is well known in animal breeding and it is computationally efficient, since the sparse precision  $\hat{A}^{-1}$  ( $\hat{A}$ ), the matrix can be directly computed [44].

### 2.3.3. Genomic Relationship ( $K = \hat{G}$ )

The genomic relationship matrix [28] was computed as  $G = \alpha MM'$ , where  $M$  is the genotypic matrix containing individuals as rows and markers as column, coded as {0,1,2} then centered column-wise, and  $\alpha$  is the normalizing factor computed as the sum of marker variance under Hardy–Weinberg equilibrium.

## 2.4. Spatial Parametrizations

The spatial term of our statistical model ( $\hat{s}$ ) was parametrized in multiple ways with the purpose of benchmarking different strategies. These included:

### 2.4.1. No Spatial Term

This baseline parametrization sets  $s_{NS} = 0$ . Not adding a spatial term is herein informative to contrast the effect of adding spatial parametrization on the accuracy of coefficients and bias of variance components.

### 2.4.2. Row-Column Effect

Two random effects, parametrizing each row and column as an independent level. This parametrization captures major linear patterns but no intra-field spatial gradients. The row-column term is described as  $s_{RC} = s_{row} + s_{col}$ , assuming normality as  $s_{row} \sim N(0, \mathbf{I}\sigma_{row}^2)$  and  $s_{col} \sim N(0, \mathbf{I}\sigma_{col}^2)$ .

### 2.4.3. Moving Average Covariate

The moving average is a fixed effect covariate with one degree of freedom,  $s_{MA} = x\beta$ , also referred to as Papadakis covariate or Nearest Neighbors covariate [7,45]. The covariate is computed for each observation by average the observed phenotypic values of surrounding plots. We averaged the values of phenotypes within the distance of two rows and two columns.

### 2.4.4. AR1 $\times$ AR1

The two-dimensional autoregressive model AR1  $\times$  AR1 term is here described as  $s_{AR} \sim N(0, \mathbf{R})$ , where the spatial structure is dictated by  $\mathbf{R} = (\Sigma_r \otimes \Sigma_c) \sigma_{AR}^2$ , and  $\Sigma$  represents the autoregressive matrices that specify the correlation ( $\rho$ ) among levels. Two key properties of this method are: (1) involves estimating three parameters ( $\sigma_{AR}^2$ ,  $\rho_r$ ,  $\rho_c$ ) via maximum likelihood and (2) the precision matrix ( $\mathbf{R}^{-1}$ ) is sparse and can be computed directly as without building  $\mathbf{R}$  [13,17,46].

### 2.4.5. Stochastic Partial Differential Equations (SPDE)

Under this framework, the spatial term is defined as  $s_{SD} \sim N(0, \mathbf{B})$ , where  $\mathbf{B}$  is a spatial covariance structure build via SPDE [23,47], defined by the function  $f(k) = (\kappa^2 - \Delta)^{0.5} \alpha(s)$ , where  $\kappa$  is scale parameter,  $\Delta$  is the Laplacian ( $\sum_i \partial^2 / \partial k_i^2$ ), and  $\alpha$  is a smother parameter ( $\alpha = \nu + 0.5\gamma$ ) with dimension domain  $\gamma$ .

### 2.4.6. Nearest Neighbor Graph

This parametrization generalizes the moving average covariate into a dynamic nearest neighbor function [48,49] using the adjacent matrix ( $\mathbf{H}$ ) as design matrix of random effects, thus  $s_{GM} = \mathbf{H}\gamma$  and  $\gamma \sim N(0, \mathbf{I}\sigma_\gamma^2)$ . The matrix  $\mathbf{H}$  was set to connect observations within two rows and two columns.

### 2.4.7. Gaussian Kernel

This uses a stationary Gaussian kernel to parametrize the covariance matrix of the spatial term [50]. In this method  $s_{GK} \sim N(0, \mathbf{U}\sigma_{GK}^2)$ , where  $\mathbf{U} = \exp(-\alpha \mathbf{D}^2)$ , as  $\mathbf{D}$  is the Euclidean distance among plots based on row and columns, and  $\alpha$  is a normalizing factor set ad hoc as 0.25.

## 2.5. Computation

The package AlphaSimR [38,51] was utilized to simulate the founder parents, to simulated the genome, to execute the Markovian coalescent simulator that generated the breeding parents, to simulate the crosses, to simulate the genotyping array, and to simulate the true breeding value.



Models were fit under the hierarchical Bayesian framework. Parameters were computed through Integrated Nested Laplace Approximation (INLA) to avoid computing Markov Chain Monte Carlo [22,52] with the R package INLA [47,52,53].

Covariance structures for SPDE and  $AR1 \times AR1$  are available in the INLA package [47]. The moving average covariate and the graphical model design matrix were based on the functions SPC and SPM, respectively, implemented in the bWGR package [54]. The genomic and pedigree relationship matrices, as well as the Gaussian kernel, were built using native R functions.

Subsequent statistical analyses of results were performed using the software R [55]. The code was run in parallel, distributed over 960 cores using the package doParallel [56]. Each scenario ( $\delta$ ) was simulated 2000 times, computing all combinations of genetic and spatial parametrization.

### 3. Results

Accuracies and biases are summarized in Table 1 and Figures 3 and 4. Results indicate that adding relationship information has a major impact on the accuracy of genetic coefficients, whereas spatial adjustment was beneficial when the residual-to-spatial ratio is 1:1 or above. The least biased variance components were provided by SPDE, and methods that provided poor spatial adjustment led to upper biased residual variance.

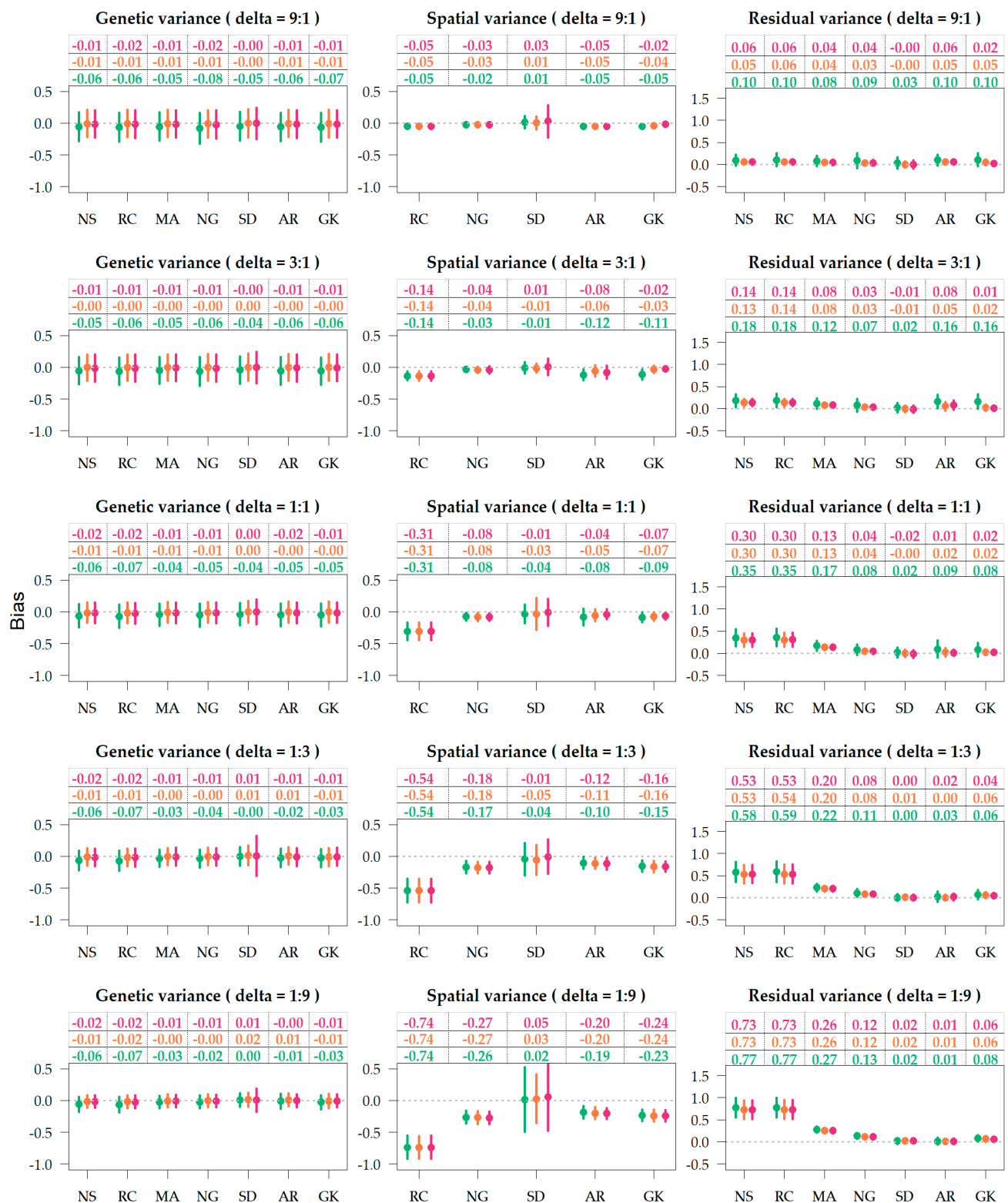
**Table 1.** Marginal treatment averages and standard error across residual-to-spatial variance ratios.

		Accuracy			Bias		
		$\text{cor}(\hat{g}, \hat{g})$	$\text{cor}(\hat{s}, \hat{s})$	Selection	$\hat{\sigma}_g^2 - \sigma_g^2$	$\hat{\sigma}_s^2 - \sigma_s^2$	$\hat{\sigma}_e^2 - \sigma_e^2$
Relationship	I	0.569 (0.015)	0.662 (0.039)	0.395 (0.010)	−0.050 (0.004)	−0.146 (0.036)	0.185 (0.035)
	A	0.739 (0.009)	0.683 (0.038)	0.516 (0.009)	−0.005 (0.001)	−0.141 (0.036)	0.142 (0.035)
	G	0.782 (0.009)	0.692 (0.037)	0.562 (0.009)	−0.012 (0.001)	−0.134 (0.037)	0.141 (0.035)
	NS	0.661 (0.036)	−	0.463 (0.027)	−0.030 (0.006)	−	0.374 (0.069)
	RC	0.661 (0.036)	0.394 (0.025)	0.463 (0.027)	−0.034 (0.007)	−0.362 (0.070)	0.377 (0.069)
	MA	0.705 (0.028)	0.672 (0.034)	0.497 (0.022)	−0.020 (0.005)	−	0.155 (0.021)
Spatial	NG	0.708 (0.027)	0.765 (0.034)	0.499 (0.022)	−0.024 (0.007)	−0.120 (0.025)	0.076 (0.009)
	SD	0.717 (0.025)	0.778 (0.039)	0.509 (0.020)	−0.007 (0.005)	−0.004 (0.008)	0.007 (0.004)
	AR	0.711 (0.026)	0.678 (0.065)	0.502 (0.021)	−0.018 (0.006)	−0.101 (0.015)	0.046 (0.012)
	GK	0.714 (0.027)	0.786 (0.037)	0.505 (0.022)	−0.023 (0.006)	−0.114 (0.021)	0.057 (0.011)



**Figure 3.** Accuracy of genetic effect, spatial effect, and selection, under varying residual-to-spatial variance ratios ( $\delta = \sigma_e^2 : \sigma_s^2$ ). Colors correspond to the genetic relationship: Identity (green), pedigree (orange), genomics (red). Spatial adjustment methods on x-axis: no spatial adjustment (NS), row-column (RC), moving average covariate (MA), autoregressive AR1  $\times$  AR1 (AR), SPDE (SD), nearest neighbor graph (NG), and Gaussian kernel (GK).





**Figure 4.** Variance bias of genetic, spatial and residual terms, under varying residual-to-spatial variance ratios ( $\delta = \sigma_e^2 : \sigma_s^2$ ). Colors correspond to the genetic relationship: Identity (green), pedigree (orange), genomics (red). Spatial adjustment methods on x-axis: no spatial adjustment (NS), row-column (RC), moving average covariate (MA), autoregressive AR1  $\times$  AR1 (AR), SPDE (SD), nearest neighbor graph (NG), and Gaussian kernel (GK).

Across the various scenarios, the accuracy of genetic effects varied from 0.569 to 0.782. Based on the marginal averages (Table 1), the genetic accuracy increased from no relation-

ship (0.569) to a pedigree-based relationship (0.739) and increased further with genomic information (0.782). With regards to the contribution of the spatial term to genetic accuracy, the RC was equivalent to NS (0.661), whereas all other methods (0.705–0.717) provided an improvement on estimate genetic accuracy of approximately 0.05. The performance of spatial methods in the accuracy of genetic effects can be summarized as follows:  $SD \geq GK \geq AR \geq NG \geq MA > RC \geq NS$ .

Higher gains in genetic accuracy were observed when adding relationships to NS, as it goes from no relationship (0.525) to pedigree (0.709) and genomics (0.749), increasing 0.184 and 0.224 in genetic accuracy, respectively, (Supplementary Tables S1 and S2) when compared to models with sophisticated spatial adjustment, such as SD, which goes from no relationship (0.598) to pedigree (0.756) and genomics (0.799) where the average gains were 0.158 and 0.201 for pedigree and genomic relationship, respectively (Supplementary Table S2). The benefits of this relationship also increase as the residual-to-spatial ratio decreased, as the average improvement of pedigree over no relationship is 0.136 at  $\delta = 9 : 1$  and 0.170 when  $\delta = 1 : 9$ . Similarly, the improvements attributed to the genomic relationship were 0.176 at  $\delta = 9 : 1$  and 0.213 at  $\delta = 1 : 9$  (Supplementary Table S1).

Accuracies of spatial effects varied from 0.241 to 0.940 across simulated scenarios (Supplementary Table S3). The accuracy of spatial effects increased with spatial variance increase, such that little to no gain is observed when the residual-to-spatial variance ratio is 9:1 and 3:1, and negligible gains when the ratio is 1:1 (Figure 3). The contribution of adding a relationship to the spatial accuracy was small, (Supplementary Table S4), but beneficial. It provided a marginal increase from 0.021 and 0.030, going from no relationship (0.662) to pedigree (0.683) and genomic (0.692) relationship (Table 1).

Figure 3 shows that SD, AR, and GK provided the most accurate spatial predictions for scenarios where the spatial signal is greater than the random noise (1:3 and 1:9). In noisy scenarios (9:1 and 3:1),  $AR1 \times AR1$  did not perform as well, and the methods that provided the most accurate predictions were SD, GK, and the NG. Based on the marginal performance of accuracy in spatial effects (Table 1), the models can be ranked as follows:  $GK \geq SD \geq NG > AR \geq MA > RC$ .

The accuracy of selection of the top 10% ranked from 0.306 to 0.607 (Supplementary Table S5). This metric provided similar results to those observed on genetic accuracy since the accuracy to selection is an alternative measure of the accuracy of genetic effects with specific application to breeding purposes. The advent of relationships provided a great impact on selection (Table 1 and Supplementary Table S6), with an improvement from 0.395 without a relationship, to 0.516 and 0.562 when using pedigree and genomic relationship information, respectively, with the corresponding enhancement of 30.63% and 42.27%. The different spatial parameterizations also differed with regards to their contribution to selection accuracy, particularly on scenarios where the spatial signal was strongest (1:9 and 1:3). Figure 3 shows that in the extreme case (1:9), the enhancement from NS to SD is 0.120, both with genomic relationship, from 0.410 to 0.530 and without relationship 0.080, from 0.260 to 0.340.

The biases of the variance components are illustrated in Figure 4, and the marginal averages across delta scenarios are provided in Table 1. Supplementary Tables S3 and S4 present the marginal averages for the different combinations of relationship, spatial method, and delta. For all purposes, bias values closer to zero are desirable and an estimator is considered unbiased if the bias is exactly zero. Herein, bias indicates if certain parametrizations are under- or over-estimating the variances attributed to the different model terms. For the methods NS and MA, estimators of spatial variance ( $\hat{\sigma}_s^2$ ) are not computed, so the bias is not available. On the row-column parametrization (RC) the estimated spatial variance was the sum of the variances attributed to the row and column terms.

The dispersion around genetic variance estimates was nearly constant (Figure 4), as the simulations were based on sampling the heritability from a uniform distribution (0,1). The genetic and spatial variance components were generally biased downwards (Supplementary Tables S7–S10), and the residuals were biased upwards (Supplementary

Tables S11 and S12). Table 1 indicates that genetic variance is more biased downwards when the genetic term is fit without relationship ( $-0.050$ ) compared to with pedigree ( $-0.005$ ) and genomic ( $-0.012$ ) relationship, where pedigree was less biased than genomics. Genetic variances were least biased for spatial parametrization SD and most biased for RC and NS. In marginal terms, the best results were observed from the combination of SD and genomic information (Supplementary Table S8). Figure 4 shows MA and NG providing relatively unbiased genetic variances when associated to pedigree information.

Figure 4 provides three trends related to the bias of spatial variance: (1) in scenarios with deltas of 9:1 and 3:1, the bias of the spatial variance component is low as the absolute variance is low; (2) models unable to capture the field gradients (NS and RC) display upper biased residual variances, as the variance of spatial trends shifts towards the residuals; (3) SD was the least biased but displayed the highest dispersion.

The unbiasedness of residual variance relies on the accuracy of genetic and spatial terms, as more accurate models yield lower residual variances. Table 1 shows that the bias of residual variance was equally reduced when pedigree (0.142) and genomic relationships (0.141) were utilized in comparison to no relationship (0.185). The spatial parameterizations that provided the least residual bias were SD and AR (Table 1), closely followed by the Gaussian kernel and the Neared Neighbor graph. In one instance, SPDE was combined with genomic information, and the average residual biased was negative  $-0.003$  (Supplementary Table S12). The unbiased SD + G could be an artifact of the simulation settings, as SD is the spatial method closest related to the Matérn function, and both true and statistical models utilize genomic information, so it is expected that results would look considerably different if the true model was based on linear gradients.

#### 4. Discussion

Commercial plant breeding programs require making the most out of all information available for decision making, which often entails utilizing sophisticated statistical models that enable the usage of complex inputs, such as genomic relationships and spatial correlations [7,52,57]. Modeling additional sources of information leads to more accurate results and, consequently, better use of the available resources [58,59].

Spatial patterns found in field experiments often constitute a large amount of non-target signals. Ignoring or not properly accounting for exogenous sources of variation causes a reduction in the accuracy of coefficients due to contaminated signals [57,60] and, consequently, inaccurate selections [3,12]. Failing to control and accommodate nuisance parameters is critical when information is scarce [61], which is a common scenario in observational experiments of a non-replicated nature [37,62].

The noisy data problem is critical in the early stages of plant breeding, where decisions have large economic repercussions and are solely based on the information at hand [63]. Normal circumstances dictate that a large number of entries are placed into small plots set under non-replicated experimental design [64]. Plots may be grown primarily for seed increase in early trials; often, breeders perform visual selection relying on checks to discriminate genetic signals from field variation [65]. However, the infusion of information into single-loc evaluations, such as local spatial trends [66] and genomic [67], enables accurate data-driven selections of higher-yielding entries. Besides agronomic performance, models fitting both genomics and spatial information have been beneficial to improve the selection of quality traits, such as grain composition in soybeans [68].

Simulations from this study demonstrate that when there is enough spatial signal in the field, the adequate parameterization of field variation will improve the accuracy of genetic effects under un-replicated trials. In another study with wheat [69], it was reported that trials with replication across locations may also benefit from spatial adjustment. The main findings generally align with previous reports from simulation and real data from various crop and tree species, including studies performed on wheat, rye, barley, triticale, cotton, orange, lupin, tea, and soybeans [3,11,12,58,70–78].

Spatial modeling is a mature statistical approach and is becoming a standard pipeline procedure for analyzing field trials in plant breeding. Autoregressive procedures that account for field variation in agricultural trials were introduced in the 1940s [79,80], with multiple iterations of improvement [2]. Over the past 20 years,  $AR1 \times AR1$  became the benchmark model, where the success of this approach is largely due to the sparsity, speed, and availability of its implementation [81]. However, reports indicate some convergence problems with the  $AR1 \times AR1$  parameterization under certain conditions, such as its applicability to narrow field layouts [17,62]. This problem has served as motivation to the search for alternative frameworks that may provide equally satisfying results with more algorithmic stability. Previous studies suggested the use of Linear Variance (LV) models [82,83] to improve the stability of  $AR1 \times AR1$  models, particularly for REML-based implementations (e.g., ASREML, SAS). No convergence problems were observed in this study with the  $AR1 \times AR1$  implementation fit as a random term with Bayesian estimation correlations and variance components. Even without convergence problems,  $AR1 \times AR1$  parametrization displayed poor performance under the scenarios with a residual-to-spatial variance ratio lower than 1:1 (Figure 3), which we attributed to the small spatial variance. This issue was not observed in the scenario 1:3 and 1:9, where the spatial variance component is greater than the residual variance. Alternatives to  $AR1 \times AR1$  commonly involved splines and kernels within the linear model framework [46,78,84,85] as well as new methodologies derived from machine learning, including random forest [86,87] and deep learning [87]. This study utilized a set of methods besides  $AR1 \times AR1$ , with comparable performance. Similar results were reported [46,83]. We also showed evidence that adding relationship information provided additional benefits to the estimation of both genetics and spatial effects (Table 1).

We find it worth pointing out that, unlike the spatial parameterization, the use of genetic relationship information had not been utilized in plant breeding until recently [10,88], largely motivated by the success of animal breeding and the availability of genomic information. Nowadays, commercial breeding programs have genomic information from every entry grown in the field, and breeding wants to make the most out of this information. A general limitation to the use of genomic information is the cost of genotyping; whereas pedigree information is free, the genomics relationship provided higher accuracies than the pedigree relationship, as supported by numerous reports [26,29,89]. The advantages of genomics over pedigree are believed to be due to the Mendelian segregation and selection bias not captured by pedigree [25,90], and unreliable record-keeping of pedigree information in plant breeding [91,92]. Yet, the pedigree information is a suitable replacement for genomics when such information is not available (Table 1, Figures 3 and 4).

Our simulation settings endeavored to be as realistic as possible. The information utilized in this study is generally available for analysis of field trials in early breeding stages: (1) spatial information is obtained from the field layout map and (2) relationship information is available either through pedigree or genomics, as genotyping is a routine operation in most commercial plant breeding programs. Yet, the advantages of utilizing both sources of information jointly are not of common knowledge, and literature on the topic is scarce. Our simulations provided some evidence that such modeling is beneficial to both genetic effects and parameter estimation (Table 1), leading to more accurate selections in breeding stages where information in the entry basis is not abundant.

We acknowledge that  $AR1 \times AR1$ , SPDE, and the Gaussian kernel have shared statistical properties with the Matérn function utilized in the true model. As the field variation was simulated as gradients, these conditions are unfavored for row-column adjustments. Thus, terms of rows and columns were not coupled to other gradient-based models. However, row-column adjustment can be as good or better than more sophisticated parameterizations [58], in one specific study with real data.

Gradient-based spatial adjustments are complementary to standard experimental designs [18]. Increments in performance help unbalanced and un-replicated experiments. Among these, the p-rep designs have been emphasized in plant breeding trials in the

early stages [64,93,94]. In such designs, supplementing models with genomic relationship information tends to provide more robust coefficients and variance components without known tradeoffs observed on studies with real and simulated data [21,25,26,95–97]. Theoretical connections between the current work and generalizations to more complex designs and multi-environmental evaluations are provided in the Supplementary Text T1.

Our simulated study provides a new insight on spatial modeling: the residual-to-spatial variance ratio as a determinant factor to define when accuracy gains can be expected from the analysis. Based on the scenarios considered in this study, results indicate that more noise than spatial signal translates into no added benefit on fitting spatial terms.

We envision two directions for future studies: (1) assess and benchmark the benefits provided by fitting spatial and relationship information in multi-environmental trials, with a varying number of locations; (2) evaluate the benefits from adding other sources of information for single-environmental trials, such as genotype-by-environment interactions, dominance, and pleiotropy. The latter might include the simultaneous evaluation of multiple traits, optimization of experimental designs accounting for the number of plots, locations, and phenotyping costs, which become relevant with the advent of phenomics and high-throughput phenotyping.

## 5. Conclusions

The addition of a relationship matrix provided greater improvement to the accuracy of the genetic coefficients than accounting for the spatial variation. The study shows that genomic relationships provide more accurate results than pedigree, and both relationships are better than no relationship.

The modeling field plot variation is beneficial when the spatial variance is equal to or greater than the residual variance. Accounting for field variation using SPDE and Gaussian kernel provided the highest accuracy of genetic effects, closely followed by  $AR1 \times AR1$ , Nearest Neighbor Graph, and moving average covariate. In this study, the row-column parametrization was equivalent to no spatial adjustment. We reiterate that our results are conditional to the simulated scenario, as this study emphasized complex field patterns created through the Matérn function.

In terms of parameter estimation, variance components were least biased when relationship information was utilized. Concerning spatial variation, SPDE generally provided the least biased variance components, followed by  $AR1 \times AR1$ .

**Supplementary Materials:** The following are available online at <https://www.mdpi.com/article/10.3390/agronomy11071397/s1>, Table S1: Marginal genetic accuracy and standard error by delta ratios and treatments, Table S2: Marginal genetic accuracy and standard error by relationship and spatial methods, Table S3: Marginal spatial accuracy and standard error by delta ratios and treatments, Table S4: Marginal spatial and standard error by relationship and spatial methods, Table S5: Marginal selection accuracy (top 10%) and standard error by delta ratios and treatments, Table S6: Marginal selection accuracy (top 10%) and standard error by relationship and spatial methods, Table S7: Marginal bias of genetic variances and standard error by delta ratios and treatments, Table S8: Marginal bias of genetic variances and standard error by relationship and spatial methods, Table S9: Marginal bias of spatial variances and standard error by delta ratios and treatments, Table S10: Marginal bias of spatial variances and standard error by relationship and spatial methods, Table S11: Marginal bias of residual variances and standard error by delta ratios and treatments, Table S12: Marginal bias of residual variances and standard error by relationship and spatial methods, Text T1: Factors influencing accuracy within- and across-locations.

**Author Contributions:** É.D.B.d.S. and A.X. conducted the research, wrote the article, designed the simulation and performed data analysis. M.V.F. contributed with the big picture feedback, revised the analysis and interpretation of results. All authors have read and agreed to the published version of the manuscript.

**Funding:** This research received no external funding.

**Institutional Review Board Statement:** Not applicable.



**Informed Consent Statement:** Not applicable.

**Data Availability Statement:** All data utilized in this study was generated through simulation. Software and all simulation parameters are described in the manuscript. The R script is available on GitHub (<https://github.com/Ederdb/SpatialCorrection> (accessed on 7 July 2021)).

**Acknowledgments:** Authors thank the Midwest Paraná State University (UNICENTRO) and acknowledge Corteva Agriscience for providing the computational resources to run simulations.

**Conflicts of Interest:** The authors declare that this research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

## References

1. Bartlett, M.S. Nearest Neighbour Models in the Analysis of Field Experiments. *J. R. Stat. Society. Ser. B* **1978**, *40*, 147–174. [[CrossRef](#)]
2. Wilkinson, G.N.; Eckert, S.R.; Hancock, T.W.; Mayo, O. Nearest Neighbour (Nn) Analysis of Field Experiments. *J. R. Stat. Soc. Ser. B* **1983**, *45*, 151–178. [[CrossRef](#)]
3. Stroup, W.W.; Baenziger, P.S.; Miltz, D.K. Removing Spatial Variation from Wheat Yield Trials: A Comparison of Methods. *Crop Sci.* **1994**, *34*, 62–66. [[CrossRef](#)]
4. Wu, T.; Mather, D.E.; Dutilleul, P. Application of Geostatistical and Neighbor Analyses to Data from Plant Breeding Trials. *Crop Sci.* **1998**, *38*, 1545–1553. [[CrossRef](#)]
5. Magnussen, S. Bias in Genetic Variance Estimates Due to Spatial Autocorrelation. *Theoret. Appl. Genet.* **1993**, *86*, 349–355. [[CrossRef](#)] [[PubMed](#)]
6. Taye, G.; Njuho, P.M. Smoothing Fertility Trends in Agricultural Field Experiments. *Statistics* **2008**, *42*, 275–289. [[CrossRef](#)]
7. Lado, B.; Matus, I.; Rodríguez, A.; Inostroza, L.; Poland, J.; Belzile, F.; del Pozo, A.; Quincke, M.; Castro, M.; Zitzewitz, J. von Increased Genomic Prediction Accuracy in Wheat Breeding Through Spatial Adjustment of Field Trial Data. *G3 Genesgenomesgenet.* **2013**, *3*, 2105–2114. [[CrossRef](#)]
8. Robinson, G.K. That BLUP Is a Good Thing: The Estimation of Random Effects. *Statist. Sci.* **1991**, *6*, 15–32. [[CrossRef](#)]
9. Cressie, N. The Origins of Kriging. *Math. Geol.* **1990**, *22*, 239–252. [[CrossRef](#)]
10. Piepho, H.P.; Möhring, J.; Melchinger, A.E.; Büchse, A. BLUP for Phenotypic Selection in Plant Breeding and Variety Testing. *Euphytica* **2008**, *161*, 209–228. [[CrossRef](#)]
11. Cullis, B.R.; Gleeson, A.C. Spatial Analysis of Field Experiments—An Extension to Two Dimensions. *Biometrics* **1991**, *47*, 1449–1460. [[CrossRef](#)]
12. Gilmour, A.R.; Cullis, B.R.; Verbyla, A.P.; Verbyla, A.P. Accounting for Natural and Extraneous Variation in the Analysis of Field Experiments. *J. Agric. Biol. Environ. Stat.* **1997**, *2*, 269–293. [[CrossRef](#)]
13. Dutkowski, G.W.; Silva, J.C.E.; Gilmour, A.R.; Lopez, G.A. Spatial Analysis Methods for Forest Genetic Trials. *Can. J. For. Res.* **2002**, *32*, 2201–2214. [[CrossRef](#)]
14. Möhring, J.; Piepho, H.-P. Comparison of Weighting in Two-Stage Analysis of Plant Breeding Trials. *Crop Sci.* **2009**, *49*, 1977–1988. [[CrossRef](#)]
15. Williams, E.R.; Piepho, H.-P. An Evaluation of Error Variance Bias in Spatial Designs. *J. Agric. Biol. Environ. Stat.* **2018**, *23*, 83–91. [[CrossRef](#)]
16. Williams, E.R.; Piepho, H.P. A Comparison of Spatial Designs for Field Variety Trials. *Aust. New Zealand J. Stat.* **2013**, *55*, 253–258. [[CrossRef](#)]
17. Piepho, H.-P.; Möhring, J.; Pflugfelder, M.; Hermann, W.; Williams, E.R. Problems in Parameter Estimation for Power and AR(1) Models of Spatial Correlation in Designed Field Experiments. *Int. J. Fac. Agric. Biol.* **2015**, *10*, 3–16.
18. Borges, A.; González-Reymúndez, A.; Ernst, O.; Cadenazzi, M.; Terra, J.; Gutiérrez, L. Can Spatial Modeling Substitute for Experimental Design in Agricultural Experiments? *Crop Sci.* **2019**, *59*, 44–53. [[CrossRef](#)]
19. Stein, M.L. *Interpolation of Spatial Data: Some Theory for Kriging*; Springer: New York, NY, USA, 2012; ISBN 978-1-4612-7166-6.
20. Perdikaris, P.; Venturi, D.; Royset, J.O.; Karniadakis, G.E. Multi-Fidelity Modelling via Recursive Co-Kriging and Gaussian-Markov Random Fields. *Proc. R. Soc. A Math. Phys. Eng. Sci.* **2015**, *471*, 1–23. [[CrossRef](#)]
21. Elias, A.A.; Rabbi, I.; Kulakow, P.; Jannink, J.-L. Improving Genomic Prediction in Cassava Field Experiments Using Spatial Analysis. *G3 Genesgenomesgenet.* **2018**, *8*, 53–62. [[CrossRef](#)]
22. Rue, H.; Held, L. *Gaussian Markov Random Fields: Theory and Applications*; Monographs on Statistics and Applied Probability; Chapman & Hall/CRC: Boca Raton, FL, USA, 2005; ISBN 978-1-58488-432-3.
23. Lindgren, F.; Rue, H.; Lindström, J. An Explicit Link between Gaussian Fields and Gaussian Markov Random Fields: The Stochastic Partial Differential Equation Approach: Link between Gaussian Fields and Gaussian Markov Random Fields. *J. R. Stat. Soc. Ser. B* **2011**, *73*, 423–498. [[CrossRef](#)]
24. Muir, W.M. Comparison of Genomic and Traditional BLUP-Estimated Breeding Value Accuracy and Selection Response under Alternative Trait and Genomic Parameters: Comparison of BLUP and GEBV Selection. *J. Anim. Breed. Genet.* **2007**, *124*, 342–355. [[CrossRef](#)]



25. Habier, D.; Fernando, R.L.; Dekkers, J.C.M. The Impact of Genetic Relationship Information on Genome-Assisted Breeding Values. *Genetics* **2007**, *177*, 2389–2397. [CrossRef]
26. Habier, D.; Tetens, J.; Seefried, F.-R.; Lichtner, P.; Thaller, G. The Impact of Genetic Relationship Information on Genomic Breeding Values in German Holstein Cattle. *Genet. Sel. Evol.* **2010**, *42*, 1–12. [CrossRef]
27. Zhang, Z.; Todhunter, R.J.; Buckler, E.S.; Van Vleck, L.D. Technical Note: Use of Marker-Based Relationships with Multiple-Trait Derivative-Free Restricted Maximal Likelihood. *J. Anim. Sci.* **2007**, *85*, 881–885. [CrossRef]
28. VanRaden, P.M. Efficient Methods to Compute Genomic Predictions. *J. Dairy Sci.* **2008**, *91*, 4414–4423. [CrossRef]
29. Goddard, M.E.; Hayes, B.J.; Meuwissen, T.H.E. Using the Genomic Relationship Matrix to Predict the Accuracy of Genomic Selection: Predict the Accuracy of Genomic Selection. *J. Anim. Breed. Genet.* **2011**, *128*, 409–421. [CrossRef]
30. Damesa, T.M.; Möhring, J.; Worku, M.; Piepho, H.-P. One Step at a Time: Stage-Wise Analysis of a Series of Experiments. *Agron. J.* **2017**, *109*, 845–857. [CrossRef]
31. Xavier, A.; Hall, B.; Hearst, A.A.; Cherkauer, K.A.; Rainey, K.M. Genetic Architecture of Phenomic-Enabled Canopy Coverage in Glycine Max. *Genetics* **2017**, *206*, 1081–1089. [CrossRef]
32. FAO. FAO Global Statistical Yearbook, FAO Regional Statistical Yearbooks. Available online: <http://www.fao.org/faostat/en/#data/QC> (accessed on 6 January 2021).
33. Schmutz, J.; Cannon, S.B.; Schlueter, J.; Ma, J.; Mitros, T.; Nelson, W.; Hyten, D.L.; Song, Q.; Thelen, J.J.; Cheng, J.; et al. Genome Sequence of the Palaeopolyploid Soybean. *Nature* **2010**, *463*, 178–183. [CrossRef]
34. Tsuda, M.; Kaga, A.; Anai, T.; Shimizu, T.; Sayama, T.; Takagi, K.; Machita, K.; Watanabe, S.; Nishimura, M.; Yamada, N.; et al. Construction of a High-Density Mutant Library in Soybean and Development of a Mutant Retrieval Method Using Amplicon Sequencing. *BMC Genom.* **2015**, *16*, 1–18. [CrossRef] [PubMed]
35. Xavier, A.; Thapa, R.; Muir, W.M.; Rainey, K.M. Population and Quantitative Genomic Properties of the USDA Soybean Germplasm Collection. *Plant Genet. Resour.* **2018**, *16*, 513–523. [CrossRef]
36. Chen, G.K.; Marjoram, P.; Wall, J.D. Fast and Flexible Simulation of DNA Sequence Data. *Genome Res* **2009**, *19*, 136–142. [CrossRef] [PubMed]
37. Clarke, G.P.Y.; Stefanova, K.T. Optimal Design for Early-Generation Plant-Breeding Trials with Unreplicated or Partially Replicated Test Lines. *Aust. New Zealand J. Stat.* **2011**, *53*, 461–480. [CrossRef]
38. Faux, A.-M.; Gorjanc, G.; Gaynor, R.C.; Battagin, M.; Edwards, S.M.; Wilson, D.L.; Hearne, S.J.; Gonen, S.; Hickey, J.M. AlphaSim: Software for Breeding Program Simulation. *Plant Genome* **2016**, *9*, 1–14. [CrossRef]
39. Krainski, E.T.; Gómez-Rubio, V.; Bakka, H.; Lenzi, A.; Castro-Camilo, D.; Simpson, D.; Lindgren, F.; Rue, H. *Advanced Spatial Modeling with Stochastic Partial Differential Equations Using R and INLA*; CRC Press/Taylor and Francis Group: Boca Raton, FL, USA, 2019; ISBN 978-0-429-03189-2.
40. Storvik, G.; Frigessi, A.; Hirst, D. Stationary Space-Time Gaussian Fields and Their Time Autoregressive Representation. *Stat. Model.* **2002**, *2*, 139–161. [CrossRef]
41. Adin, A.; Martínez-Beneito, M.A.; Botella-Rocamora, P.; Goicoa, T.; Ugarte, M.D. Smoothing and High Risk Areas Detection in Space-Time Disease Mapping: A Comparison of P-Splines, Autoregressive, and Moving Average Models. *Stoch. Env. Res. Risk Assess* **2017**, *31*, 403–415. [CrossRef]
42. Laga, I.; Kleiber, W. The Modified Matérn Process. *STAT* **2017**, *6*, 241–247. [CrossRef]
43. Wiens, A.; Nychka, D.; Kleiber, W. Modeling Spatial Data Using Local Likelihood Estimation and a Matérn to Spatial Autoregressive Translation. *Environmetrics* **2020**, *31*, 1–20. [CrossRef]
44. Henderson, C.R. A Simple Method for Computing the Inverse of a Numerator Relationship Matrix Used in Prediction of Breeding Values. *Biometrics* **1976**, *32*, 69–83. [CrossRef]
45. Gezan, S.A.; White, T.L.; Huber, D.A. Accounting for Spatial Variability in Breeding Trials: A Simulation Study. *Agron. J.* **2010**, *102*, 1562–1571. [CrossRef]
46. Selle, M.L.; Steinsland, I.; Hickey, J.M.; Gorjanc, G. Flexible Modelling of Spatial Variation in Agricultural Field Trials with the R Package INLA. *Appl. Genet.* **2019**, *132*, 3277–3293. [CrossRef]
47. Lindgren, F.; Rue, H. Bayesian Spatial Modelling with R - INLA. *J. Stat. Softw.* **2015**, *63*, 1–25. [CrossRef]
48. Eppstein, D.; Paterson, M.S.; Yao, F.F. On Nearest-Neighbor Graphs. *Discret. Comput. Geom.* **1997**, *17*, 263–282. [CrossRef]
49. Tanner, H.G. On the Controllability of Nearest Neighbor Interconnections. In Proceedings of the 2004 43rd IEEE Conference on Decision and Control (CDC) (IEEE Cat. No.04CH37601), Nassau, Bahamas, 14–17 December 2004.
50. Xavier, A.; Muir, W.M.; Craig, B.; Rainey, K.M. Walking through the Statistical Black Boxes of Plant Breeding. *Appl. Genet.* **2016**, *129*, 1933–1949. [CrossRef]
51. Gaynor, R.C.; Gorjanc, G.; Hickey, J.M. AlphaSimR: An R Package for Breeding Program Simulations. *G3 Genes Genomes Genet.* **2021**, *11*, jkaa017. [CrossRef]
52. Rue, H.; Martino, S.; Chopin, N. Approximate Bayesian Inference for Latent Gaussian Models by Using Integrated Nested Laplace Approximations. *J. R. Stat. Soc. Ser. B* **2009**, *71*, 319–392. [CrossRef]
53. Holand, A.M.; Martino, S. *AnimalINLA: Bayesian Animal Models*; Norwegian University of Science and Technology: Trondheim, Norway, 2016.
54. Xavier, A.; Muir, W.M.; Rainey, K.M. BWGR: Bayesian Whole-Genome Regression. *Bioinformatics* **2019**, 1957–1959. [CrossRef] [PubMed]

55. R Core Team. *R: A Language and Environment for Statistical Computing*; R Foundation for Statistical Computing: Vienna, Austria, 2020.
56. Ooi, H.; Corporation, M.; Weston, S.; Tenenbaum, D. DoParallel: Foreach Parallel Adaptor for the “parallel” Package (Version 1.0.16). Available online: <https://CRAN.R-project.org/package=doParallel> (accessed on 2 July 2020).
57. Oakey, H.; Verbyla, A.; Pitchford, W.; Cullis, B.; Kuchel, H. Joint Modeling of Additive and Non-Additive Genetic Line Effects in Single Field Trials. *Theor. Appl. Genet.* **2006**, *113*, 809–819. [[CrossRef](#)]
58. Bernal-Vasquez, A.-M.; Möhring, J.; Schmidt, M.; Schönleben, M.; Schön, C.-C.; Piepho, H.-P. The Importance of Phenotypic Data Analysis for Genomic Prediction—A Case Study Comparing Different Spatial Models in Rye. *BMC Genom.* **2014**, *15*, 1–17. [[CrossRef](#)]
59. Lorenz, A.; Nice, L. Training Population Design and Resource Allocation for Genomic Selection in Plant Breeding. In *Genomic Selection for Crop Improvement: New Molecular Breeding Strategies for Crop Improvement*; Varshney, R.K., Roorkiwal, M., Sorrells, M.E., Eds.; Springer International Publishing: Cham, Switzerland, 2017; pp. 7–22. ISBN 978-3-319-63170-7.
60. Hunt, C.H.; Smith, A.B.; Jordan, D.R.; Cullis, B.R. Predicting Additive and Non-Additive Genetic Effects from Trials Where Traits Are Affected by Interplot Competition. *J. Agric. Biol. Environ. Stat.* **2013**, *18*, 53–63. [[CrossRef](#)]
61. Basu, D. On the Elimination of Nuisance Parameters. In *Selected Works of Debabrata Basu*; DasGupta, A., Ed.; Selected Works in Probability and Statistics; Springer: New York, NY, USA, 2011; pp. 279–290. ISBN 978-1-4419-5825-9.
62. Robbins, K.R.; Backlund, J.E.; Schnelle, K.D. Spatial Corrections of Unreplicated Trials Using a Two-Dimensional Spline. *Crop Sci.* **2012**, *52*, 1138–1144. [[CrossRef](#)]
63. Martin, S.K.S.; Futi, X. Genetic Gain in Early Stages of a Soybean Breeding Program. *Crop Sci.* **2000**, *40*, 1559–1564. [[CrossRef](#)]
64. Williams, E.R.; John, J.A. A Note on the Design of Unreplicated Trials. *Biom. J.* **2003**, *45*, 751–757. [[CrossRef](#)]
65. Chandra, S. Efficiency of Check-Plot Designs in Unreplicated Field Trials. *Appl. Genet.* **1994**, *88*, 618–620. [[CrossRef](#)]
66. Rodríguez-Álvarez, M.X.; Boer, M.P.; van Eeuwijk, F.A.; Eilers, P.H.C. Spatial Models for Field Trials. *arXiv* **2016**, arXiv:1607.08255.
67. Michel, S.; Ametz, C.; Gungor, H.; Akgöl, B.; Epure, D.; Grausgruber, H.; Löschenberger, F.; Buerstmayr, H. Genomic Assisted Selection for Enhancing Line Breeding: Merging Genomic and Phenotypic Selection in Winter Wheat Breeding Programs with Preliminary Yield Trials. *Appl. Genet.* **2017**, *130*, 363–376. [[CrossRef](#)]
68. Bernardeli, A.; de Rocha, J.R.A.S.C.; Borem, A.; Lorenzoni, R.; Aguiar, R.; Silva, J.N.B.; Bueno, R.D.; Alves, R.S.; Jarquin, D.; Ribeiro, C.; et al. Modeling Spatial Trends and Enhancing Genetic Selection: An Approach to Soybean Seed Composition Breeding. *Crop Sci.* **2021**, *61*, 976–988. [[CrossRef](#)]
69. Qiao, C.G.; Basford, K.E.; DeLacy, I.H.; Cooper, M. Evaluation of Experimental Designs and Spatial Analyses in Wheat Breeding Trials. *Appl. Genet.* **2000**, *100*, 9–16. [[CrossRef](#)]
70. Besag, J.; Kempton, R. Statistical Analysis of Field Experiments Using Neighbouring Plots. *Biometrics* **1986**, *42*, 231–251. [[CrossRef](#)]
71. Zimmerman, D.L.; Harville, D.A. A Random Field Approach to the Analysis of Field-Plot Experiments and Other Spatial Experiments. *Biometrics* **1991**, *47*, 223–229. [[CrossRef](#)]
72. Kempton, R.A.; Seraphin, J.C.; Sword, A.M. Statistical Analysis of Two-Dimensional Variation in Variety Yield Trials. *J. Agric. Sci.* **1994**, *122*, 335–342. [[CrossRef](#)]
73. Grondona, M.O.; Crossa, J.; Fox, P.N.; Pfeiffer, W.H. Analysis of Variety Yield Trials Using Two-Dimensional Separable ARIMA Processes. *Biometrics* **1996**, *52*, 763–770. [[CrossRef](#)]
74. Brownie, C.; Gumpertz, M.L. Validity of Spatial Analyses for Large Field Trials. *J. Agric. Biol. Environ. Stat.* **1997**, *2*, 1–23. [[CrossRef](#)]
75. Duarte, J.B.; Vencovsky, R. Spatial Statistical Analysis and Selection of Genotypes in Plant Breeding. *Pesqui. Agropecuária Bras.* **2005**, *40*, 107–114. [[CrossRef](#)]
76. Stefanova, K.T.; Smith, A.B.; Cullis, B.R. Enhanced Diagnostics for the Spatial Analysis of Field Trials. *J. Agric. Biol. Environ. Stat.* **2009**, *14*, 392–410. [[CrossRef](#)]
77. Müller, B.U.; Schützenmeister, A.; Piepho, H.-P. Arrangement of Check Plots in Augmented Block Designs When Spatial Analysis Is Used. *Plant Breed.* **2010**, *129*, 581–589. [[CrossRef](#)]
78. Sun, M.; Goggi, S.A.; Matson, K.; Palmer, R.G.; Moore, K.; Ciazio, S.R. Thin Plate Spline Regression Model Used at Early Stages of Soybean Breeding to Control Field Spatial Variation. *J. Crop Improv.* **2015**, *29*, 333–352. [[CrossRef](#)]
79. Papadakis, J.S. Methode Statistique Pour Des Experiences Sur Champ. *Bull. Inst. Amel. Plantes A Salonique* **1937**, *23*.
80. Bartlett, M.S. The Approximate Recovery of Information from Replicated Field Experiments with Large Blocks. *J. Agric. Sci.* **1938**, *28*, 418–427. [[CrossRef](#)]
81. Gilmour, A.R.; Thompson, R.; Cullis, B.R. Average Information REML: An Efficient Algorithm for Variance Parameter Estimation in Linear Mixed Models. *Biometrics* **1995**, *51*, 1440–1450. [[CrossRef](#)]
82. Piepho, H.P.; Williams, E.R. Linear Variance Models for Plant Breeding Trials. *Plant Breed.* **2010**, *129*, 1–8. [[CrossRef](#)]
83. Boer, M.P.; Piepho, H.-P.; Williams, E.R. Linear Variance, P-Splines and Neighbour Differences for Spatial Adjustment in Field Trials: How Are They Related? *JABES* **2020**, *25*, 676–698. [[CrossRef](#)]
84. Rodriguez-Alvarez, M.X.; Boer, M.; Eilers, P.; van Eeuwijk, F. SpATS: Spatial Analysis of Field Trials with Splines (Version 1.0-15). Available online: <https://CRAN.R-project.org/package=SpATS> (accessed on 25 June 2019).
85. Selle, M.L.; Steinsland, I.; Powell, O.; Hickey, J.M.; Gorjanc, G. Spatial Modelling Improves Genetic Evaluation in Smallholder Breeding Programs. *Genet. Sel. Evol.* **2020**, *52*, 1–17. [[CrossRef](#)] [[PubMed](#)]

86. Hengl, T.; Nussbaum, M.; Wright, M.N.; Heuvelink, G.B.M.; Gräler, B. Random Forest as a Generic Framework for Predictive Modeling of Spatial and Spatio-Temporal Variables. *PeerJ* **2018**, *6*, 1–49. [[CrossRef](#)] [[PubMed](#)]
87. Georganos, S.; Grippa, T.; Niang Gadiaga, A.; Linard, C.; Lennert, M.; Vanhuyse, S.; Mboga, N.; Wolff, E.; Kalogirou, S. Geographical Random Forests: A Spatial Extension of the Random Forest Algorithm to Address Spatial Heterogeneity in Remote Sensing and Population Modelling. *Geocarto Int.* **2019**, 1–16. [[CrossRef](#)]
88. Crossa, J.; Campos, G.D.L.; Perez, P.; Gianola, D.; Burgueno, J.; Araus, J.L.; Makumbi, D.; Singh, R.P.; Dreisigacker, S.; Yan, J.; et al. Prediction of Genetic Values of Quantitative Traits in Plant Breeding Using Pedigree and Molecular Markers. *Genetics* **2010**, *186*, 713–724. [[CrossRef](#)] [[PubMed](#)]
89. Beukelaer, H.D.; Badke, Y.; Fack, V.; Meyer, G.D. Moving Beyond Managing Realized Genomic Relationship in Long-Term Genomic Selection. *Genetics* **2017**, *206*, 1127–1138. [[CrossRef](#)] [[PubMed](#)]
90. Sun, J.; Rutkoski, J.E.; Poland, J.A.; Crossa, J.; Jannink, J.-L.; Sorrells, M.E. Multitrait, Random Regression, or Simple Repeatability Model in High-Throughput Phenotyping Data Improve Genomic Prediction for Wheat Grain Yield. *Plant Genome* **2017**, *10*, 1–12. [[CrossRef](#)]
91. Becelaere, G.V.; Lubbers, E.L.; Paterson, A.H.; Chee, P.W. Pedigree- vs. DNA Marker-Based Genetic Similarity Estimates in Cotton. *Crop Sci.* **2005**, *45*, 2281–2287. [[CrossRef](#)]
92. Maccaferri, M.M.; Sanguineti, M.C.S.C.; Xie, C.X.; Smith, J.S.C.S.S.C.; Tuberosa, R.T. Relationships among Durum Wheat Accessions. II. A Comparison of Molecular and Pedigree Information. *Genome* **2007**, *50*, 385–399. [[CrossRef](#)] [[PubMed](#)]
93. Moehring, J.; Williams, E.R.; Piepho, H.-P. Efficiency of Augmented P-Rep Designs in Multi-Environmental Trials. *Appl. Genet.* **2014**, *127*, 1049–1060. [[CrossRef](#)] [[PubMed](#)]
94. Williams, E.; Piepho, H.-P.; Whitaker, D. Augmented P-Rep Designs. *Biom J* **2011**, *53*, 19–27. [[CrossRef](#)] [[PubMed](#)]
95. Schulz-Streeck, T.; Ogutu, J.O.; Gordillo, A.; Karaman, Z.; Knaak, C.; Piepho, H.-P. Genomic Selection Allowing for Marker-by-Environment Interaction. *Plant Breed.* **2013**, *132*, 532–538. [[CrossRef](#)]
96. Schulz-Streeck, T.; Ogutu, J.O.; Piepho, H.-P. Comparisons of Single-Stage and Two-Stage Approaches to Genomic Selection. *Appl. Genet.* **2013**, *126*, 69–82. [[CrossRef](#)] [[PubMed](#)]
97. Mao, X.; Dutta, S.; Wong, R.K.W.; Nettleton, D. Adjusting for Spatial Effects in Genomic Prediction. *JABES* **2020**. [[CrossRef](#)]