

Article

# A Multi-Trait Gaussian Kernel Genomic Prediction Model under Three Tuning Strategies

Kismiantini <sup>1</sup>, Abelardo Montesinos-López <sup>2</sup>, Bernabe Cano-Páez <sup>3</sup>, J. Cricelio Montesinos-López <sup>4</sup>, Moisés Chavira-Flores <sup>5</sup>, Osva A. Montesinos-López <sup>6,\*</sup> and José Crossa <sup>7,8,\*</sup>

- <sup>1</sup> Statistics Study Program, Universitas Negeri Yogyakarta, Yogyakarta 55281, Indonesia
- <sup>2</sup> Centro Universitario de Ciencias Exactas e Ingenierías (CUCEI), Universidad de Guadalajara, Guadalajara 44430, Jalisco, Mexico
- <sup>3</sup> Facultad de Ciencias, Universidad Nacional Autónoma de México (UNAM), México City 04510, Mexico
- <sup>4</sup> Department of Public Health Sciences, University of California Davis, Davis, CA 95616, USA
- <sup>5</sup> Instituto de Investigaciones en Matemáticas Aplicadas y Sistemas (IIMAS), Universidad Nacional Autónoma de México (UNAM), México City 04510, Mexico
- <sup>6</sup> Facultad de Telemática, Universidad de Colima, Colima 28040, Colima, Mexico
- <sup>7</sup> International Maize and Wheat Improvement Center (CIMMYT), Km 45, Carretera Mexico, Veracruz 52640, Edo. de México, Mexico
- <sup>8</sup> Colegio de Postgraduados, Montecillos 56230, Edo. de México, Mexico
- \* Correspondence: oamontes1@ucol.mx or oamontes2@hotmail.com (O.A.M.-L.); j.crossa@cgiar.org (J.C.)

**Abstract:** While genomic selection (GS) began revolutionizing plant breeding when it was proposed around 20 years ago, its practical implementation is still challenging as many factors affect its accuracy. One such factor is the choice of the statistical machine learning method. For this reason, we explore the tuning process under a multi-trait framework using the Gaussian kernel with a multi-trait Bayesian Best Linear Unbiased Predictor (GBLUP) model. We explored three methods of tuning (manual, grid search and Bayesian optimization) using 5 real datasets of breeding programs. We found that using grid search and Bayesian optimization improve between 1.9 and 6.8% the prediction accuracy regarding of using manual tuning. While the improvement in prediction accuracy in some cases can be marginal, it is very important to carry out the tuning process carefully to improve the accuracy of the GS methodology, even though this entails greater computational resources.

**Keywords:** kernels; multi-trait; Bayesian optimization; grid search; genomic selection



**Citation:** Kismiantini; Montesinos-López, A.; Cano-Páez, B.; Montesinos-López, J.C.; Chavira-Flores, M.; Montesinos-López, O.A.; Crossa, J. A Multi-Trait Gaussian Kernel Genomic Prediction Model under Three Tuning Strategies. *Genes* **2022**, *13*, 2279. <https://doi.org/10.3390/genes13122279>

Academic Editor: Rodney A. Lea

Received: 2 November 2022

Accepted: 1 December 2022

Published: 3 December 2022

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



**Copyright:** © 2022 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/>).

## 1. Introduction

Genomic selection (GS) is frequently used for genetic improvement and has many advantages over phenotype-based selection [1]. Nevertheless, breeders face an adversity of challenges to improve the accuracy of the GS methodology, similar to multi-trait (MT) genomic prediction models, which take advantage of correlated traits to improve prediction accuracy [2] under multiple environments. Consequently, to accurately predict breeding values or phenotypic values is a challenge of primordial interest in GS, as the goal is to increase genetic gain. For this reason, when the traits of interest do not have a complex genetic architecture, this achievement is usually simple to accomplish. However, for complex heritable traits, traits with complex genetic architecture (such as grain yield) and with strong epistatic effects, this goal has limited success [3,4].

Reproducing Kernel Hilbert Spaces (RKHS) regression is a popular method in plant and animal breeding [5,6] for the prediction of complex traits and modeling complex interactions more efficiently. The central idea of an RKHS regression is to project the given original input data available in a finite dimensional space onto an infinite dimensional Hilbert space. Kernel methods can incorporate any statistical machine learning algorithm to the resulting transformed data, after using a kernel function. Empirical experience indicates that generally better results are accomplished with the transformed input. For

this reason, RKHS methods are becoming more popular for analyzing nonlinear patterns in datasets collected in plant and animal breeding.

RKHS methods are very attractive because, in addition to being efficient for capturing nonlinear patterns, they are also efficient for data compression, as the transformed input has less dimensionality than the original input; this is to say, when the input is a matrix of dimensions  $n \times p$ , with  $p \gg n$ , the transformed input has a dimension of order  $n \times n$ , which has less dimension and can reduce the computational complexity required during the training process. There are many transformations (kernel functions) used to capture nonlinear patterns in the original input, and each type of transformation is specialized for capturing some type of nonlinear pattern. However, it is impossible to capture all patterns with conventional linear statistical methods [5,6].

It should be noted that RKHS methods are not limited to regression as they are also powerful in the context of classification problems, where they are also efficient and popular. Support vector machine (SVM), which was proposed to the computer science community in the 1990s by Vapnik [7], is one of the most popular methods for classification based on kernels.

In the context of GS, RKHS methods are increasingly accepted as rising evidence aids in increasing the accuracy of predictions using linear methods. For example, in a study about body weight of broiler chickens, Long et al. [8] reported a better prediction accuracy of RKHS methods over linear models. Crossa et al. [9] and Cuevas et al. [10] in wheat and maize found that the RKHS methods outperformed the linear methods. However, some authors have also reported minimal differences between RKHS methods and linear models [11,12], which is expected when the nonlinear patterns in the data are minimal or non-existing.

Moreover, empirical evidence has shown that MT models are more efficient than single-trait (ST) models [13]. Some reasons why MT models are chosen over ST models [14] are that: (1) they capture complex relationships between correlated traits in a more efficient way, (2) they take advantage of the degree of correlation between lines and traits, (3) MT models offer better interpretability than ST models, (4) they are computationally more parsimonious to train than ST models, (5) more precise estimates of random effects of lines and genetic correlations between traits are obtained, which allows for improvement of the index selection, (6) they become more efficient for indirect selection as the precision of genetic correlation parameter estimates increases, and (7) they improve hypothesis testing because they reduce type I and II errors [2] due to a more precise estimates of parameters.

However, the prediction performance of RKHS methods over conventional linear models, is not improved when a proper tuning process is not achieved. For example, when the Gaussian kernel is implemented, the bandwidth hyperparameter is set to the median of the average distances or to 1, which in some cases is not optimal and can cause the resulting prediction performance to be worse than conventional linear models. This implies that when a nonlinear kernel is performed in a model, an additional process is required to select the optimal hyperparameters to increase the prediction accuracy. However, it is also true that certain default hyperparameters frequently do an acceptable prediction performance but are not optimal. For this reason, to acquire the full power of any statistical machine learning method, a careful fine-tuning process should always be carried out.

Based on the above-mentioned considerations, we do a benchmarking study in this paper to compare the prediction performance implementing a multi-trait Gaussian kernel in the context of genomic prediction using the multivariate Bayesian Genomic Best Linear Unbiased Predictor (GBLUP) model. Since the Gaussian kernel only depends on one hyperparameter, only this hyperparameter was tuned under the following three strategies: (1) no tuning setting to the bandwidth parameter, (2) tuning using the grid search method, and (3) tuning using the Bayesian optimization method. This benchmark was carried out using 5 real datasets collected in real plant breeding programs.

## 2. Material and Methods

### 2.1. Dataset 1. Japonica

This dataset contains information on the phenotypic performance of four traits (GY = Grain Yield, PHR = Percentage of Head Rice Recovery, GC = percentage of Chalky Grain, PH = Plant Height) of rice as reported by Monteverde et al. [15] and evaluated over the course of five years (2009, 2010, 2011, 2012 and 2013). The genotypes evaluated were 93, 292, 316, 316 and 134 lines for years 2009, 2010, 2011, 2012 and 2013, respectively. This dataset contains 54 environmental covariates but, in this application, these covariates were not included in the analysis. In this dataset, a total of 1051 assessments were evaluated over five years. In this dataset, the genotypes evaluated were 320 and for each 44,598 markers remained after quality control that were coded with 0, 1 and 2. For more details about the data, see Monteverde et al. [15].

### 2.2. Dataset 2. Indica

This dataset contains information on the same traits as the Japonica dataset [15], with only three environments (years 2010, 2011 and 2012). In each year (environment), 327 genotypes were evaluated. Although this dataset contained environmental covariates they were not used in this study. The total number of observations in this balanced dataset was 981 since each line was included once in each environment. The genotyping-by-sequencing (GBS) markers datasets were filtered to retain markers with 50% missing data after imputation and a minor allele frequency (MAF) > 0.05. The markers remaining after quality control were 92,430 SNPs for each line and were coded as 0, 1 and 2, where 0 was used if the SNP was homozygous for the major allele, 1 if the SNP was heterozygous and 2 if the SNP was homozygous for the other allele. For more details about the data, see Monteverde et al. [15].

### 2.3. Dataset 3. Groundnut

This dataset was reported by Pandey et al. [16] with genotypic and phenotypic information for 318 genotypes and four environments. The traits measured were seed yield per plant (SYPP), pods per plant (NPP), pod yield per plant (PYPP) and yield per hectare (YPH). The environments were identified as: Aliyarnagar\_Rainy 2015 (ENV1), Jalgoan\_Rainy 2015 (ENV2), ICRISAT\_Rainy 2015 (ENV3), and ICRISAT Post-Rainy 2015 (ENV4).

This dataset contained a total of 1272 observations and is balanced, since each genotype was included once in each environment. For each genotype, 8268 single nucleotide polymorphism (SNP, or SNPs in plural) markers (coded with 0, 1 and 2) were available after quality control. For more details about the data, see Pandey et al. [16].

### 2.4. Dataset 4. Cotton

This dataset was proposed by Gapare et al. [17] with genotypic and phenotypic information for 859 genotypes and seven environments [Myall Vale (MV), Collarenebri (CO), Bourke (BK), Emerald (EM), St. George (SG), Breeza (BR), Darling Downs (DD)]. The traits analyzed for the study were fiber length and strength.

This dataset contains a total of 859 observations and is not balanced, since each genotype was not included in each environment. For each genotype, 5000 single nucleotide polymorphism (SNP, or SNPs in plural) markers (coded with 0, 1 and 2) were available after quality control. For more details about the data, see Gapare et al. [17].

### 2.5. Dataset 5. Disease

This dataset contains 438 wheat genotypes (lines), three traits. PTR that denotes *Pyrenophora tritici-repentis* (PTR), SN denotes *Parastagonospora nodorum*, a major fungal pathogen of wheat fungal taxon, and SB that denotes *Bipolaris sorokiniana* (SB), that causes seedling diseases, common root rot and spot blotch of several crops such as barley and wheat. These 438 lines were evaluated in the greenhouse for six replicates during a period

of time. The replicates were considered as environments (Env1, Env2, Env3, Env4, Env5, and Env6).

For the three traits evaluated, the total number of observations was  $438 \times 6 = 2628$ .

DNA samples were genotyped using 67,436 SNPs. For each marker, the genotype for each line was coded as the number of copies of a designated marker-specific allele carried by the line (absence = zero and presence = one). SNP markers with unexpected heterozygous genotypes were recoded as either AA or BB. Those markers that had more than 15% missing values or with  $MAF < 0.05$  were removed. A total of 11,617 SNPs were still available for analysis after quality control and imputation.

## 2.6. Multi-Trait Kernel Model

This model is given in Equation (1) as:

$$Y = 1_n \mu^T + X_E \beta_E + Z_L g + Z_{EL} g_E + \epsilon \quad (1)$$

where  $Y$  is the matrix of phenotypic response variables of order  $n \times n_T$  and ordered first by environments and then by lines,  $n_T$  denotes the number of traits,  $1_n$  is a vector of ones of length  $n$ ,  $\mu^T$  is a vector of intercepts for each trait of length  $n_T$ ,  $T$  denotes the transpose of a vector or matrix, that is,  $\mu = [\mu_1, \dots, \mu_{n_T}]^T$ ,  $X_E$  is the design matrix of environments of order  $n \times I$ ,  $I$  denotes the number of environments,  $\beta_E$  is the matrix of coefficients for environments with a dimension of  $I \times n_T$ ,  $Z_L$  is the design matrix of lines of order  $n \times J$ ,  $J$  denotes the number of lines,  $g$  is the matrix of random effects of lines of order  $J \times n_T$  distributed as  $g \sim MN_{J \times n_T}(0, K, \Sigma_T)$ , that is, with a matrix-variate normal distribution with parameters  $M = 0$ ,  $U = G$  and  $V = \Sigma_T$ ,  $K$  is the Gaussian kernel (GK) that mimics a covariance matrix to capture the degree of similarity between lines, such as the genomic relationship matrix (Linear kernel) proposed by [18] that was built with marker data of order  $J \times J$  and  $\Sigma_T$  is the variance-covariance matrix of traits of order  $n_T \times n_T$ .  $Z_{EL}$  is the design matrix of the genotype  $\times$  environment interaction of order  $n \times JI$ ,  $g_E$  is the matrix of genotype  $\times$  environment interaction random effects distributed as  $g_E \sim MN_{JI \times n_T}(0, Z_E Z_E^T \odot Z_g K Z_g^T, \Sigma_T)$ , where  $\Sigma_E$  is a diagonal variance-covariance matrix of environments of order  $I \times I$  and  $\odot$  denotes the Hadamard product.  $\epsilon$  is the residual matrix of dimension  $n \times n_T$  distributed as  $\epsilon \sim MN_{n \times n_T}(0, I_{IJ}, R)$ , where  $R$  is the residual variance-covariance matrix of order  $n_T \times n_T$ .

The GK was computed using the GK function:

$$K(x_i, x_j) = e^{-\gamma \|x_i - x_j\|^2}, \text{ with } \gamma > 0 \quad (2)$$

where  $x_i$  and  $x_j$  are the marker vectors for the  $i$ th and  $j$ th individuals (genotypes), respectively [19,20]. It is necessary to point out that the GK function was reparametrized (Caamal-Pat, et al. [1]) as:

$$K(x_i, x_j) = e^{\log \rho \|x_i - x_j\|^2} \text{ with } \rho \in (0, 1) \quad (3)$$

using the variable change ( $\rho = e^{-\gamma}$ ). Subsequently, the three strategies for tuning the bandwidth ( $\gamma$ ) hyperparameter used in this implementation are listed as follows:

(1) Manual tuning (no tuning, denoted as NT) setting the value of  $\gamma = 1$ , which is equivalent to setting  $\rho = e^{-1}$ .

(2) Tuning using a grid search (GrS) strategy with 26 values in the grid for the values of  $\rho$  between 0.01 and 0.999 with increments of 0.04, this means that 26 values of  $\rho$  were evaluated. The average of the normalized root mean square error of each predicted trait ( $NRMSE = \frac{1}{T} \sum_{t=1}^n NRMSE_t$ ,  $t = 1, \dots, T$ ) was used as metric for choosing the optimal  $\rho$  value in the inner testing set.

(3) Tuning  $\rho$  using the Bayesian optimization (BO) method. The average NRMSE was also used as metrics to select the optimal  $\rho$  value in the inner testing set.

The implementation of this model with the three strategies of tuning the  $\rho$  hyperparameter of the GK was carried out in the R statistical software [21,22].

### 2.7. Evaluation of Prediction Performance

In each of the five datasets, the seven outer fold cross validation was implemented (Montesinos-López et al. [19]). For this reason, 7 – 1 folds were assigned to the outer-training set, while the remaining were assigned to the outer-testing set until each of the 7 folds were tested once. For tuning the bandwidth hyperparameter of the Gaussian kernel five nested cross-validations was used; that is to say, the outer-training was divided into five groups where four were used for the inner training set (80% of the training) and one for the validation (inner-testing) set (20% of the outer training). Next, the average of the five validation folds was reported as the metric of prediction performance to select the optimal hyperparameter (bandwidth of the Gaussian kernel). Using this optimal hyperparameter (band width), the multi-trait kernel model (1) was then refitted with the whole outer-training set (the 7 – 1 folds), and finally, the prediction of each outer-testing set was obtained.

The prediction accuracy was reported in terms of the average normalized root mean square error for each trait ( $NRMSE_t = \frac{1}{7} \sum_{k=1}^7 NRMSE_{t,k} = \frac{1}{7} \sum_{k=1}^7 \frac{RMSE_{t,k}}{\bar{y}_{t,k}}$ , for  $t = 1, \dots, n_T$ , where  $n_T$  is the number of predicted traits;  $RMSE_{t,k}$  and  $\bar{y}_{t,k}$  denote the NRMSE and the mean of the  $t$ -th trait for the  $k$ th fold, respectively), where  $RMSE_{t,k} = \sqrt{\frac{1}{n_k} \left( \sum_{i=1}^{n_k} (y_{it} - \hat{f}_t(x_i))^2 \right)}$  denoting the root mean square error of the  $t$ -th trait for the  $k$ th fold. In addition to the NRMSE for each trait, we also reported the average NRMSE of all traits as follows:  $NRMSE = \frac{1}{T} \sum_{t=1}^T NRMSE_t$ . These metrics were computed under the three strategies for tuning the bandwidth ( $\gamma$ ) hyperparameter used in the implementation of the model, so that  $NRMSE_{NT}$ ,  $NRMSE_{GrS}$  and  $NRMSE_{BO}$  denote the NRMSE of no tuning (NT), grid search (GrS) and Bayesian optimization (BO) tuning strategies, respectively. The relative efficiencies were also reported and that were computed as:

$$RE_{GrS} = \frac{NRMSE_{NT}}{NRMSE_{GrS}}$$

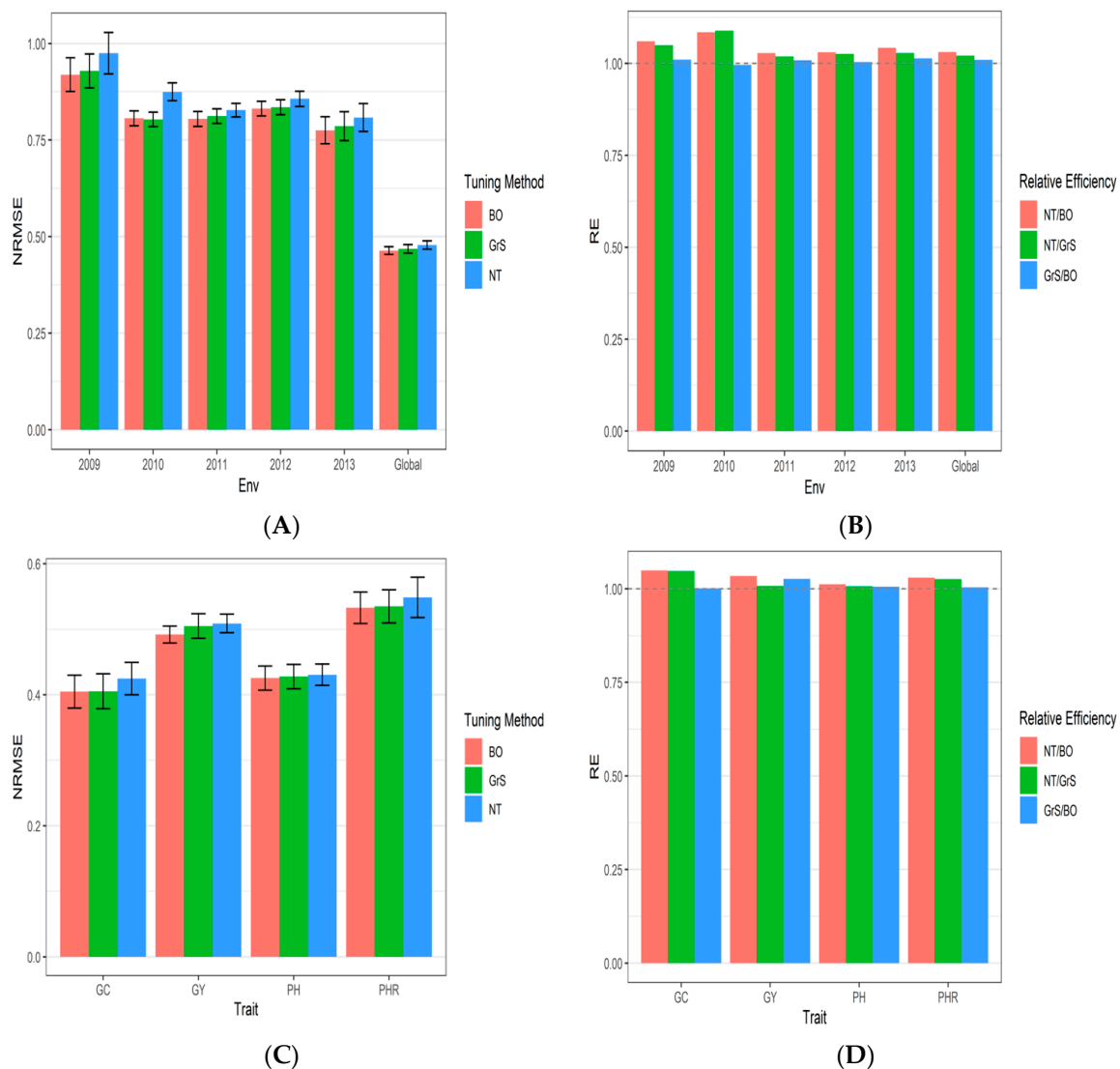
$$RE_{BO} = \frac{NRMSE_{NT}}{NRMSE_{BO}}$$

When  $RE_{GrS} > 1$  ( $RE_{BO} > 1$ ), the best performance prediction in terms of NRMSE was obtained using the GrS (BO) strategy, when  $RE_{GrS} < 1$  ( $RE_{BO} < 1$ ), the NT strategy was superior in terms of prediction accuracy and when  $RE_{GrS} = 1$  ( $RE_{BO} = 1$ ), both strategies of hyperparameter tuning were equally efficient. We also computed the relative efficiency in terms of NRMSE between the grid search strategy (GrS) and Bayesian optimization (BO) strategy ( $RE_{GrS/BO} = NRMSE_{GrS}/NRMSE_{BO}$ ) and the interpretation is the same as the previous example.

## 3. Results

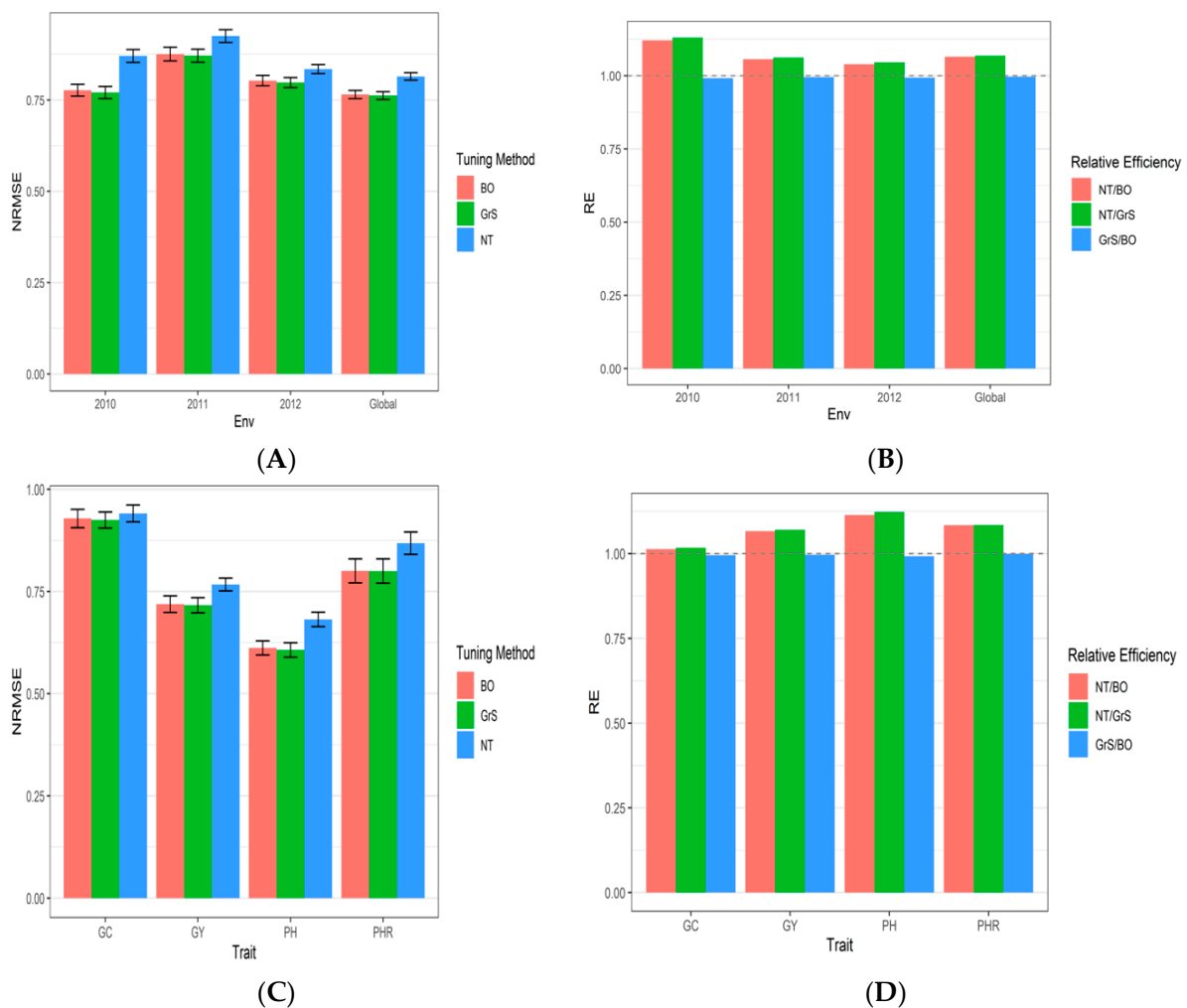
The results are provided in three sections for Japonica, Indica and Groundnut datasets 1–3, respectively. The results from dataset 1 (Japonica) are given in Table A1, Figure 1A–D, and Appendix B Table A4. The results from dataset 2 (Indica) are in Table A2, Figure 2A–D, and Appendix B Table A5. The results from dataset 3 (Groundnut) are shown in Table A3, Figure 3A–D, and Appendix B Table A6.

The results from dataset 4 (Cotton) can be found in Supplementary Materials Tables S1 and S2 and Figures S1A–S1D, whereas results from dataset 5 (Diseases) are found in Supplementary Materials Tables S3 and S4 and Figure S2A–D.

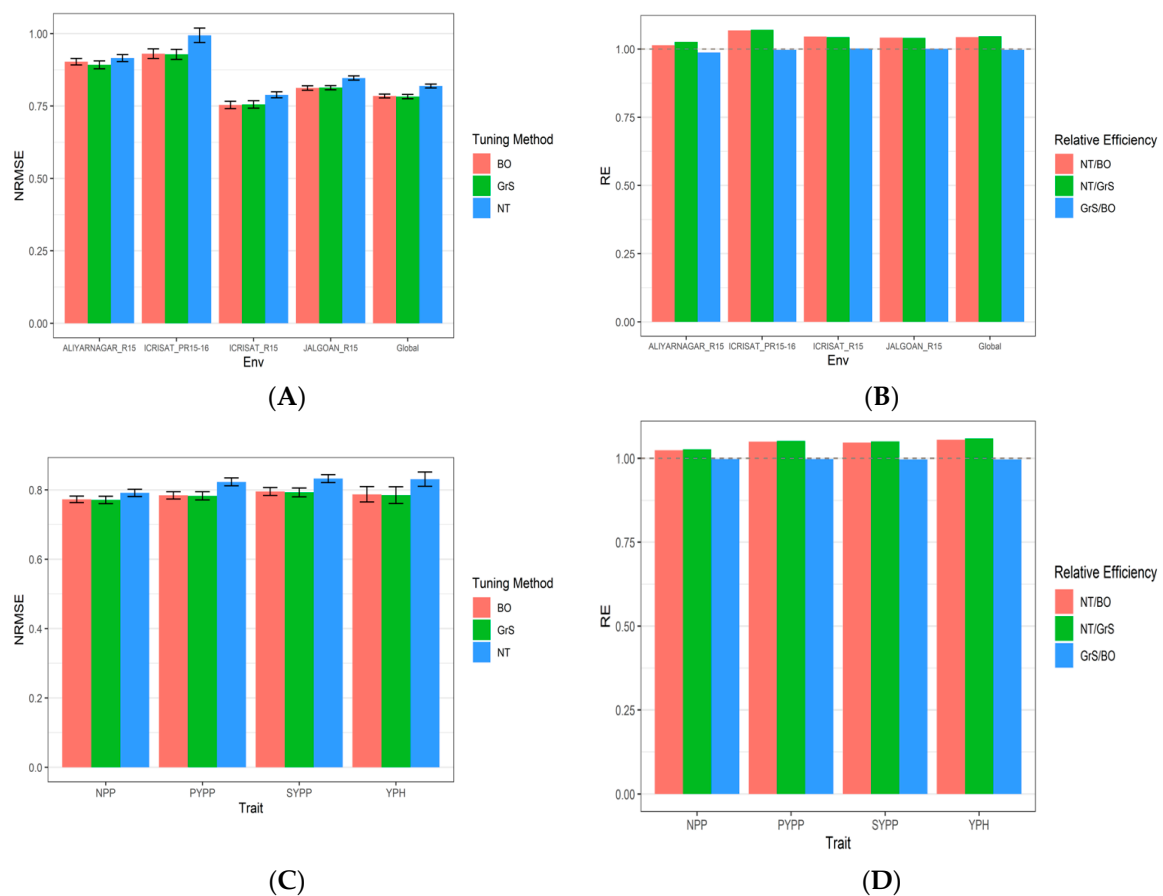


**Figure 1.** (A) The prediction performance for dataset 1, Japonica dataset in terms of normalized root mean squared error (NRMSE) for each year (2009–2013), across years (Global), and across traits with three strategies of tuning (BO, GrS and NT) under 7 Fold Cross-Validation (7FCV). (B) The relative efficiency for each environment (2009–2013) and across environments (Global) and across traits (CG, GY, PH and PHR) with three strategies of tuning (BO, GrS and NT) under 7 Fold Cross-Validation (7FCV). (C) The prediction performance in terms of normalized root mean squared error (NRMSE) for each trait (CG, GY, PH and PHR) across years with three strategies of tuning (BO, GrS and NT) under 7 Fold Cross-Validation (7FCV). (D) The relative efficiency for each trait (CG, GY, PH and PHR) across years with three strategies of tuning (BO, GrS and NT) under 7 Fold Cross-Validation (7FCV). When  $RE > 1$  the denominator method outperforms the numerator in terms of prediction performance.





**Figure 2.** (A) Prediction performance for dataset 2, Indica dataset in terms of normalized root mean squared error (NRMSE) for each year (2010–2012) across traits (CG, GY, PH and PHR), across years and across traits (Global) with three strategies of tuning (BO, GrS and NT) under 7 Fold Cross-Validation (7FCV). (B) The relative efficiency for each environment (2010–2012) across traits and across environments and across traits (Global) with three strategies of tuning (BO, GrS and NT) under 7 Fold Cross-Validation (7FCV). (C) Prediction performance in terms of normalized root mean squared error (NRMSE) for each trait (CG, GY, PH and PHR) across years with three strategies of tuning (BO, GrS and NT) under 7 Fold Cross-Validation (7FCV). (D) The relative efficiency for each trait (CG, GY, PH and PHR) across environments with three strategies of tuning (BO, GrS and NT) under 7 Fold Cross-Validation (7FCV). When  $RE > 1$  the denominator method outperforms the numerator in terms of prediction performance.



**Figure 3.** (A) The prediction performance for dataset 3, Groundnut dataset in terms of normalized root mean squared error (NRMSE) for each environment across traits (ALIYARNAGAR\_R15, ICRISAT\_PR15-16 ICRISAT\_R15 and JALGOAN\_R15), across environments and traits (Global) with three tuning strategies (BO, GrS and NT) under 7 Fold Cross-Validation (7FCV). (B) The relative efficiency for each environment across traits (ALIYARNAGAR\_R15, ICRISAT\_PR15-16 ICRISAT\_R15 and JALGOAN\_R15), across environments and traits (Global) with three tuning strategies (BO, GrS and NT) under 7 Fold Cross-Validation (7FCV). (C) The prediction performance in terms of normalized root mean squared error (NRMSE) for each trait (NPP, PYPP, SYPP, YPH) across environments with three tuning strategies (BO, GrS and NT) under 7FCV. (D) The relative efficiency for each trait (NPP, PYPP, SYPP, YPH) across environments with three tuning strategies (BO, GrS and NT) under 7 Fold Cross-Validation (7FCV). When  $RE > 1$ , the denominator method outperforms the numerator in terms of prediction performance.

### 3.1. Dataset 1 Japonica

The prediction performance for each environment and across environments (Global) at Japonica's dataset in terms of normalized root mean squared error (NRMSE) and relative efficiency (RE) comparing the three methods of hyperparameter tuning (no tuning, grid search (GrS), and Bayesian optimization), under the 7FCV strategy are provided. NRMSE\_GC, NRMSE\_GY, NRMSE\_PH and NRMSE\_PHR denote the NRMSE of traits CC, GY, PH and PHR.

As can be seen in Figure 1 and Table A1, in terms of NRMSE for the GC trait the best performance under the GrS strategy was in environments 2009 (1.124), 2010 (0.913), 2012 (0.887), and 2013 (0.730), while the environments with best RNMSE under the BO strategy were 2011 (0.774), 2012 (0.887), and Global (0.404). For trait GY, the best RMSE values were observed under the BO strategy in environments 2010 (0.818), 2011(0.875), 2012 (0.858), 2013(0.865) and global (0.491). The exception was in the 2009 environment where the best NRMSE value was 0.984 under the NT strategy.



For the PH variable, the best predictions (lower NRMSE) were observed under the BO strategy [2010 (0.639), 2011 (0.757), 2013 (0.664) and Global (0.425)]. In the 2009 environment, the lowest NRMSE was 0.695 under the NT tuning strategy, while in the 2012 environment, the lowest NRMSE=0.653 was observed under the GrS strategy. For trait PHR, the best performance in terms of NRMSE of most environments was observed under the BO strategy [2009 (0.838), 2011 (0.811), 2012 (0.925), 2013 (0.837) and global (0.532)]. The year 2010 was an exception, as the best NRMSE was 0.827 and was found under the GrS strategy (see Figure 1 and Table A1). The standard error of prediction performance for every environment and across environments (Global) is provided in Appendix B Table A4.

Across traits, the prediction performance can be observed in Figure 1A (Table A1), where the best predictions (lower NRMS) were observed under the BO and GrS strategies and the worst under the NT strategy. In addition, across traits we can observe that the RE of comparing the NT strategy versus BO strategy for each environment and across environments were 1.060 (2009), 1.0849 (2010), 1.028 (2011), 1.03 (2012), 1.0427 (2013) and 1.031 (Global) (Figure 1B; Table A1). This indicates that the BO method outperformed NT strategy in terms of NRMSE in all the environments mentioned by 6.03% (2009), 8.49% (2010), 2.83% (2011), 3% (2012), 4.27% (2013), and 3.1% (Global) (Figure 1B; Table A1), respectively. While the RE of comparing the NT strategy versus GrS strategy for each environment and across environments were 1.0494 (2009), 1.0889 (2010), 1.0193 (2011), 1.0261 (2012), 1.0284 (2013) and 1.0212 (Global) (Figure 1B; Table A1). This result indicates that the GrS method outperformed NT strategy in terms of NRMSE in all the environments mentioned by 4.94% (2009), 8.89% (2010), 1.93% (2011), 2.61% (2012), 2.84% (2013), and 2.12% (Global). Finally, the RE of comparing the GrS method versus the BO method were 1.0104 (2009), 0.9963 (2010), 1.0008 (2011), 1.004 (2012), 1.013 (2013) and 1.009 (Global) (Figure 1B; Table A1). This means that the BO strategy is slightly better in terms of prediction performance than the GrS method since the RE were slightly superior to one.

In Figure 1C (Table A1), the prediction performance is provided for each trait across environments, while in Figure 1D the relative efficiencies of comparing NT versus BO, NT versus GrS, and GrS and BO, are provided and show that in all traits, the best strategies for tuning in terms of NRMSE are the BO and GrS method without relevant differences between the GrS and BO method.

### 3.2. Dataset 2 Indica

NRMSE\_GC, NRMSE\_GY, NRMSE\_PH and NRMSE\_PHR denote the NRMSE of traits CC, GY, PH, and PHR, respectively. Figure 2 and Table A2 shows that in terms of RMSE for the GC trait, the best performance under the GrS strategy was in environments 2010 (0.918), 2011 (0.92), 2012 (0.943), and Global (0.924). For trait GY the best RMSE values were observed under the GrS strategy in environments 2010 (0.915), 2011(0.825), and Global (0.716). However, in 2012, the best NRMSE value was 0.984 under the NT strategy.

For PH trait, the best predictions (lower NRMSE) were observed under the GrS strategy [2010 (0.422), 2012 (0.692), and Global (0.607)], with the exception in 2011 environment where the lowest NRMSE was 0.87 under the BO tuning strategy. For trait PHR the best performance in terms of NRMSE of most environments was also observed under the GrS strategy [2011 (0.866), 2012 (0.8), and global (0.8)], except in year 2010, where the best NRMSE was 0.819 using the BO strategy. Further details are given in Figure 2 and Table A2. The standard error of prediction performance for every environment (Global) is provided in Appendix B Table A5.

As we summarize across traits for each environment, we can see in Figure 2A and Table A2, that the best predictions (lower NRMS) were observed under the BO and GrS strategies and the worst under the NT strategy. We can also observe that the RE of comparing the NT strategy versus BO strategy for each environment and across environments were 1.12 (2010), 1.056 (2011), 1.039 (2012), and 1.064 (Global) (Figure 2B and Table A2). This indicates that the BO method outperformed NT strategy in terms of NRMSE in all environments by 12% (2010), 5.6% (2011), 3.9% (2012), and 6.4% (Global). The RE of com-

paring the NT strategy versus GrS strategy for each environment and across environments were 1.129 (2010), 1.061 (2011), 1.046 (2012), and 1.068 (Global) (Figure 2B and Table A2). This indicates that the GrS method outperformed the NT strategy in terms of NRMSE in all the environments by 12.9% (2010), 6.1% (2011), 4.6% (2012), and 6.8% (Global). Finally, the RE of comparing the GrS method versus the BO method were 0.991 (2010), 0.995 (2011), 0.993 (2012), and 0.996 (Global) (Figure 2B and Table A2). These results indicate that the BO strategy is slightly worse in terms of prediction performance than the GrS method since in most cases, the RE is less than one.

The prediction performance in terms of NRMSE of each trait across environments are given in Figure 2C and Table A2 and the relative efficiencies of comparing NT versus BO, NT versus GrS and GrS and BO are given in Figure 2D and Table A2, where in three out of four traits the best strategies for tuning are the BO and GrS, while the worst was the NT strategy. It should be noted that there are no relevant differences between the BO and GrS methods.

### 3.3. Dataset 3 Groundnut

Here, NRMSE\_NPP, NRMSE\_PYPP, NRMSE\_SYPP and NRMSE\_YPH denote the NRMSE of traits NPP, PYPP, SYPP and YPH. As shown in Table A3, in terms of NRMSE for the NPP trait, the best performance under the GrS strategy was in environments ALIYARNAGAR\_R15 (0.808), ICRISAT\_R15 (0.786) and Global (0.77), while the environments with the best RNMSE under the BO strategy were ICRISAT\_PR15-16 (0.902) and JALGOAN\_R15 (0.808). For trait PYPP, the best NRMSE values were observed under the BO strategy in ICRISAT\_PR15-16 (0.954), ICRISAT\_R15 (0.772) and JALGOAN\_R15 (0.836). ALIYARNAGAR\_R15 and Global were the exception where the best NRMSE values were 0.934 and 0.782 under the GrS strategy (Figure 3)

In Figure 3 and Table A3 we can also see that in terms of NRMSE for the SYPP trait the best performance under the GrS strategy was in environments ALIYARNAGAR\_R15 (0.933), ICRISAT\_PR15-16 (0.944), ICRISAT\_R15 (0.787) and Global (0.792), while the environment with the best RNMSE under BO strategy was JALGOAN\_R15 (0.838). For trait YPH the best NRMSE values were observed under the GrS strategy in environments ALIYARNAGAR\_R15 (0.811), ICRISAT\_PR15-16 (0.915), JALGOAN\_R15 (0.767) and Global (0.784). In the case of environment ICRISAT\_R15, the best NRMSE value was 0.67 under the BO strategy. More details are provided in Table A3. The standard error of prediction performance for every environment and across environments (Global) are provided in the Appendix B Table A6.

Summarizing across traits for each environment, the best predictions (lower NRMS) were observed under the BO and GrS strategies in most cases, while the worst were under the NT strategy (Figure 3A; Table A3). Across traits we can also observe that the RE of comparing the NT strategy versus the BO strategy from each environment and across environments were 1.013 (ALIYARNAGAR\_R15), 1.045 (ICRISAT\_R15), 1.068 (ICRISAT\_PR15-16), 1.042 (JALGOAN\_R15) and 1.044 (Global) (Figure 3B; Table A3). This indicates that the BO method outperformed the NT strategy in terms of NRMSE in all environments by 1.3% (ALIYARNAGAR\_R15), 4.5% (ICRISAT\_R15), 6.8% (ICRISAT\_PR15-16), 4.2% (JALGOAN\_R15) and 4.4% (Global). Meanwhile, the RE of comparing the NT strategy versus GrS strategy for each environment and across environments were 1.026 (ALIYARNAGAR\_R15), 1.043 (ICRISAT\_R15), 1.07 (ICRISAT\_PR15-16), 1.04 (JALGOAN\_R15) and 1.046 (Global) (Figure 3B; Table A3). This indicates that the GrS method outperformed the NT strategy in terms of NRMSE in all environments by 2.6% (ALIYARNAGAR\_R15), 4.3% (ICRISAT\_R15), 7% (ICRISAT\_PR15-16), 4% (JALGOAN\_R15) and 4.6% (Global). Finally, the RE of comparing the GrS method versus the BO method were 0.987 (ALIYARNAGAR\_R15), 0.997 (ICRISAT\_R15), 1.002 (ICRISAT\_PR15-16), 1.001 (JALGOAN\_R15) and 0.997 (Global) (Figure 3B). This means that the BO strategy is slightly worse in terms of prediction performance than the GrS method since the RE in most of the cases was less than one. For more details, see Table A3.

The prediction performance in terms of NRMSE of each trait across environments is given in Figure 3C and Table A3, while the relative efficiencies of comparing NT versus BO, NT versus GrS and GrS and BO are given in Figure 3D and Table A3, where we can appreciate that in the four's traits evaluated, the best strategies for tuning were the BO and GrS, while the worst was the NT strategy. No relevant differences were observed between the BO and GrS methods.

#### 4. Discussion

As a predictive methodology, GS can help increase genetic gain by saving significant resources since candidate phenotypes do not need to be measured in the field, as they are predicted [23]. However, a number of factors still need to be improved for prediction performance to be optimized. One of these factors is the choice of the statistical machine learning that will be used. In this regard, there are statistical machine learning methods that can only capture linear patterns, while others are able to capture non-linear patterns [19].

Kernel methods are very attractive since they are able to capture non-linear patterns and are very versatile, as they can be used with many statistical machine learning methods. For example, kernel methods can be applied in conventional mixed models, in support vector machines and even in many others machine learning methods such as random forest, and gradient boosting machine providing a modified input. In conventional mixed models, the use of kernels is quite straightforward since the genomic relationship matrix that is provided in this case is replaced by a particular kernel, enhancing the power of mixed models to capture nonlinear patterns in the data [19]. However, many kernels such as the Gaussian kernel implemented in this study have hyper-parameters that must be appropriately tuned to guarantee successful implementation.

For this reason, in this study we evaluated the influence in terms of prediction performance of three tuning strategies (manual tuning, grid Search and Bayesian optimization) under a multi-trait Bayesian GBLUP model. We found that using the grid search and Bayesian optimization outperform the prediction accuracy of the manual tuning by 2.1% and 3.1%, respectively, in the japonica dataset, by 6.4% and 6.8%, respectively, in the Indica dataset, by 4.4% and 4.6%, respectively, in the groundnut dataset, by 1.9% and 2.1%, respectively, in the Cotton dataset, and by 2.3% and 2.7%, respectively, in the disease dataset.

About the time for implementing the tuning methods we found that the grid Search method required around 15% more time for its implementation than the Bayesian optimization method, and around 20 times more expensive in computation resources than the manual tuning. The tuning process is more expensive in terms of computational resources. The grid search approach was slightly more costly than the Bayesian optimization since the size of the grid contain 26 values, however the larger the number of values in the grid search the larger the computational resources required by the grid search.

We found differences in prediction performance in each environment, and larger differences were observed when the environments represented years, since many times the year-to-year variability is significant. For example, in dataset 1 (Japonica) we observed higher prediction error in year 2009, compared to the other years and this can be attributed mostly to the effect of years and less to the unbalance in the number of genotypes evaluated in each environment. In addition, relevant differences in terms of prediction performance between environments were found in dataset 3 (Groundnut) where environments ALI-YARNAGAR\_R15 and ICRISAT\_PR15-16 resulted in the worst prediction performance and environments ICRISAT\_R15 and JALGOAN\_R15 with the best predictions. These results point out that even in the same year the location-to-location variability is considerably high.

In addition, for some datasets we found significant differences in terms of prediction performance between traits. For example, in dataset 1 (Japonica) the best predictions were observed in traits GC and PH and the worst in traits GY and PHR. While in dataset 2 (Indica) the best predictions were observed in traits GY and PH, while the worst in traits GC and PHR. In dataset 3 (Groundnut) the four traits showed a similar prediction performance.

The improvement in prediction accuracy of the grid search and Bayesian optimization is not very significant compared with the manual tuning; however, this is data dependent. By data dependent we mean that, if the dataset contain complex non-linear patterns, then using kernels with the appropriate implementation and tuned methods will result in important improvement in prediction accuracies with respect to not using kernels. However, if the data only contain linear patterns, we cannot expect an improvement in prediction accuracy using non-linear kernels. We also need to be aware that we are under a multi-trait framework, which lends to greater difficulty in selecting the bandwidth hyperparameter that can work simultaneously for all traits under study. In the context of tuning the bandwidth for Gaussian kernels, a greater gain in prediction accuracy was observed using grid search and Bayesian optimization as was shown by Montesinos-López et al. [20].

## 5. Conclusions

In this study we showed the importance of carefully tuning the Gaussian kernel to improve the prediction accuracy. We found that we can increase the prediction accuracy between 1.9% and 6.8% by tuning the Gaussian kernel using Bayesian optimization or the grid search method. We did not find any relevant differences between tuning with Bayesian optimization and the grid search method. In general, the results indicated that modest gain in prediction accuracy were obtained for some datasets, while in others major improvements were achieved. We encourage researchers dedicating sufficient time for the tuning process. It is also important to point out that the degree of improvement in prediction accuracy can be influenced by the metric used for evaluating the prediction performance in the validation set and for this reason our results are not conclusive. We encourage future benchmark studies to be able to see the influence of the metric used.

**Supplementary Materials:** The following supporting information can be downloaded at: <https://www.mdpi.com/article/10.3390/genes13122279/s1> with results from dataset 4 (cotton) and dataset 5 (disease). The results from dataset 4 (Cotton) are shown at the following Figures and Tables: (1) Figure S1. dataset 4 Cotton. A) Prediction performance in terms of normalized root mean squared error (NRMSE) for each environment (BK, BR, CO, DD, EM, MV and SG), across environments (Global) and across traits (LEN and STR) with three tuning strategies (BO, GrS and NT) under 7 Fold Cross-Validation (7FCV). B) Relative efficiency for each environment (BK, BR, CO, DD, EM, MV and SG), across environments (Global) and across traits with three tuning strategies (BO, GrS and NT) under 7-Fold Cross-Validation (7FCV). C) Prediction performance in terms of normalized root mean squared error (NRMSE) for each trait (LEN and STR) across environments with three tuning strategies (BO, GrS and NT) under 7 Fold Cross-Validation (7FCV). D) Relative efficiency for each trait (LEN and STR) across environments with three tuning strategies (BO, GrS and NT) under 7 Fold Cross-Validation (7FCV). When  $RE > 1$  the denominator method outperforms the numerator in terms of prediction performance; (2) Table S1. Prediction performance for every environment and across environments (Global) of the dataset 4 (Cotton) in terms of normalized root mean square error (NRMSE) under three tuning methods (BO, GrS and NT) for two cotton traits (LEN and STR) dataset. RE denotes relative efficiency. The RE in rows corresponding to BO were computed dividing the NRMSE under NT by the NRMSE under BO. While the RE in rows corresponding to GrS were computed dividing the NRMSE under NT by the NRMSE under GrS. While those RE in the rows corresponding to NT strategy were computed dividing the NRMSE under GrS by the NRMSE under BO, and (3) Table S2. Standard error of prediction performance for every environment and across environments (Global) of two traits (LEN and STR) for the dataset 4 (Cotton) in terms of normalized root mean square error (NRMSE) under three methods of tuning (BO, GrS and NT). RE denotes relative efficiency. Results from dataset 5 (Disease) are given at (1) Figure S2. dataset 5 (Disease). A) The prediction performance for dataset 5 (Disease) in terms of normalized root mean squared error (NRMSE) for each environment (Env1, Env2, Env3, Env4, Env5, Env6), across environments (Global) and across traits with three tuning strategies (BO, GrS and NT) under 7-fold cross-validation (7FCV). B) Relative efficiency for each environment (Env1, Env2, Env3, Env4, Env5, Env6), across environments (Global) and across traits (PTR, SB and SN) with three tuning strategies (BO, GrS and NT) under 7FCV. C) Prediction performance in terms of normalized root mean squared error (NRMSE) for each trait (PRT, SB and SN) across environments with three tuning strategies (BO, GrS



and NT) under 7-fold cross-validation (7FCV). D) Relative efficiency for each trait (PRT, SB and SN) across environments with three tuning strategies (PRT, SB and SN) under 7-fold cross-validation (7FCV). When  $RE > 1$  the denominator method outperforms the numerator in terms of prediction performance; (2) Table S3. Prediction performance for every environment and across environments (Global) of the dataset 5 (Disease) in terms of normalized root mean square error (NRMSE) under three tuning methods (BO, GrS and NT) and for 3 traits (SB, SN, SE). RE denotes relative efficiency. The RE in rows corresponding to BO were computed dividing the NRMSE under NT by the NRMSE under BO. While the RE in rows corresponding to GrS were computed dividing the NRMSE under NT by the NRMSE under GrS. While those RE in the rows corresponding to NT strategy were computed dividing the NRMSE under GrS by the NRMSE under BO, and (3) Table S4. Standard error of prediction performance for every environment and across environments (Global) of the dataset 5 (Disease) in terms of normalized root mean square error (NRMSE) under three methods of tuning (BO, GrS and NT) for three traits (PTR, SB, SN). RE denotes relative efficiency.

**Author Contributions:** Conceptualization, K., O.A.M.-L. and A.M.-L.; methodology, O.A.M.-L., A.M.-L., J.C.M.-L., K. and J.C.; software, B.C.-P. and M.C.-F.; formal analysis, O.A.M.-L., K. and A.M.-L.; investigation, J.C.M.-L., O.A.M.-L. and J.C.; writing—original draft preparation, O.A.M.-L., A.M.-L., K. and J.C.M.-L., writing—review and editing, K., O.A.M.-L., A.M.-L. and J.C.M.-L. All authors have read and agreed to the published version of the manuscript.

**Funding:** We are thankful for the financial support provided by the Bill and Melinda Gates Foundation [INV-003439, BMGF/FCDO, Accelerating Genetic Gains in Maize and Wheat for Improved Livelihoods (AG2MW)], the USAID projects [USAID Amend. No. 9 MTO 069033, USAID-CIMMYT Wheat/AGGMW, AGG-Maize Supplementary Project, AGG (Stress Tolerant Maize for Africa), and the CIMMYT CRP (maize and wheat). We are also thankful for the financial support provided by the Foundation for Research Levy on Agricultural Products (FFL) and the Agricultural Agreement Research Fund (JA) through the Research Council of Norway for grants 301835 (Sustainable Management of Rust Diseases in Wheat) and 320090 (Phenotyping for Healthier and more Productive Wheat Crops).

**Data Availability Statement:** Phenotypic and genomic data can be downloaded from the link: [https://github.com/osval78/Multivariate\\_Tuning\\_Kernel\\_Method](https://github.com/osval78/Multivariate_Tuning_Kernel_Method).

**Acknowledgments:** The authors are thankful to administrative, technical field support and Lab assistances that established the different experiments in the field as well as in the Laboratory at the different institutions that generated the data used in this study.

**Conflicts of Interest:** The authors declare no conflict of interest.

## Appendix A

**Table A1.** The prediction performance for every environment and across environments (Global) of the dataset 1 (Japonica) in terms of normalized root mean square error (NRMSE) under three methods of tuning (BO, GrS and NT) for the Japonica dataset. RE denotes relative efficiency. The RE in the rows corresponding to BO were computed dividing the NRMSE under NT by the NRMSE under BO. The RE in rows corresponding to GrS were computed dividing the NRMSE under NT by the NRMSE under GrS. RE in the rows corresponding to the NT strategy were computed dividing the NRMSE under GrS by the NRMSE under BO.

Tuning Type	Year	NRMSE_GC	NRMSE_GY	NRMSE_PH	NRMSE_PHR	NRMSE	RE
Bayesian Optimization	2009	1.1256	0.9846	0.7289	0.8383	0.919350	1.0603959
Bayesian Optimization	2010	0.9377	0.8181	0.6398	0.8298	0.806350	1.0849197
Bayesian Optimization	2011	0.7741	0.8752	0.7574	0.8119	0.804650	1.0283353
Bayesian Optimization	2012	0.8877	0.8585	0.6541	0.9253	0.831400	1.0304306
Bayesian Optimization	2013	0.7327	0.8658	0.6643	0.8374	0.775050	1.0427714
Bayesian Optimization	Global	0.4048	0.4919	0.4255	0.5328	0.463750	1.0310512

**Table A1.** *Cont.*

Tuning Type	Year	NRMSE_GC	NRMSE_GY	NRMSE_PH	NRMSE_PHR	NRMSE	RE
Grid Search	2009	1.1246	1.0185	0.7283	0.8445	0.928975	1.0494093
Grid Search	2010	0.9130	0.8299	0.6436	0.8270	0.803375	1.0889373
Grid Search	2011	0.7787	0.8943	0.7585	0.8156	0.811775	1.0193095
Grid Search	2012	0.8877	0.8689	0.6530	0.9298	0.834850	1.0261724
Grid Search	2013	0.7305	0.9151	0.6578	0.8401	0.785875	1.0284078
Grid Search	Global	0.4053	0.5049	0.4278	0.5348	0.468200	1.0212516
No Tuning	2009	1.3539	0.9841	0.6954	0.8661	0.974875	1.0104694
No Tuning	2010	1.1494	0.8527	0.6445	0.8527	0.874825	0.9963105
No Tuning	2011	0.8013	0.9069	0.7661	0.8355	0.827450	1.0088548
No Tuning	2012	0.9123	0.8961	0.6673	0.9511	0.856700	1.0041496
No Tuning	2013	0.7988	0.9049	0.6741	0.8550	0.808200	1.0139668
No Tuning	Global	0.4246	0.5088	0.4306	0.5486	0.478150	1.0095957

**Table A2.** The prediction performance for every environment and across environments (Global) of the dataset 2 (Indica) in terms of normalized root mean square error (NRMSE) under three methods (BO, GrS and NT) for the Indica dataset. RE denotes relative efficiency. The RE in rows corresponding to BO were computed dividing the NRMSE under NT by the NRMSE under BO. The RE in rows corresponding to GrS were computed dividing the NRMSE under NT by the NRMSE under GrS. Those RE in the rows corresponding to NT strategy were computed dividing the NRMSE under GrS by the NRMSE under BO.

Tuning Type	Year	NRMSE_GC	NRMSE_GY	NRMSE_PH	NRMSE_PHR	NRMSE	RE
Bayesian Optimization	2010	0.9201	0.9234	0.4439	0.8195	0.776725	1.1207956
Bayesian Optimization	2011	0.9305	0.8324	0.8707	0.8687	0.875575	1.0563344
Bayesian Optimization	2012	0.9492	0.7529	0.6981	0.8116	0.802950	1.0391058
Bayesian Optimization	Global	0.9287	0.7190	0.6117	0.8006	0.765000	1.0645425
Grid Search	2010	0.9181	0.9154	0.4229	0.8253	0.770425	1.1299607
Grid Search	2011	0.9206	0.8250	0.8728	0.8668	0.871300	1.0615173
Grid Search	2012	0.9433	0.7542	0.6926	0.8004	0.797625	1.0460429
Grid Search	Global	0.9248	0.7165	0.6070	0.8002	0.762125	1.0685583
No Tuning	2010	0.9470	1.0435	0.5777	0.9140	0.870550	0.9918890
No Tuning	2011	0.9390	0.8394	0.9642	0.9570	0.924900	0.9951175
No Tuning	2012	0.9461	0.8031	0.7495	0.8387	0.834350	0.9933682
No Tuning	Global	0.9409	0.7669	0.6816	0.8681	0.814375	0.9962418



**Table A3.** The prediction performance for every environment and across environments (Global) of the dataset 3 (Groundnut) in terms of normalized root mean square error (NRMSE) under three tuning methods (BO, GrS and NT) for Groundnut dataset. RE denotes relative efficiency. The RE in rows corresponding to BO were computed dividing the NRMSE under NT by the NRMSE under BO. While the RE in rows corresponding to GrS were computed dividing the NRMSE under NT by the NRMSE under GrS. While those RE in the rows corresponding to NT strategy were computed dividing the NRMSE under GrS by the NRMSE under BO.

Tuning Type	Environment	NRMSE_NPP	NRMSE_PYPP	NRMSE_SYPP	NRMSE_YPH	NRMSE	RE
Bayesian Optimization	ALIYARNAGAR_R15	0.8992	0.9442	0.9452	0.8242	0.9032	1.0139781
Bayesian Optimization	ICRISAT_R15	0.7872	0.7729	0.786	0.6701	0.75405	1.0458524
No Tuning	ALIYARNAGAR_R15	0.9152	0.9554	0.9597	0.833	0.915825	0.9879318
Bayesian Optimization	ICRISAT_PR15-16	0.9025	0.9547	0.9469	0.9191	0.9308	1.068194
Bayesian Optimization	JALGOAN_R15	0.8081	0.8361	0.8383	0.7674	0.812475	1.0423705
No Tuning	ICRISAT_PR15-16	0.9331	1.0378	1.0184	0.9878	0.994275	0.9975827
Grid Search	ALIYARNAGAR_R15	0.8902	0.9342	0.9337	0.8111	0.89230	1.0263645
Grid Search	ICRISAT_R15	0.7862	0.7755	0.7873	0.6737	0.755675	1.0436034
No Tuning	ICRISAT_R15	0.7866	0.823	0.8324	0.7125	0.788625	1.002155
Grid Search	ICRISAT_PR15-16	0.9026	0.9517	0.9441	0.9158	0.92855	1.0707824
Grid Search	JALGOAN_R15	0.8091	0.8377	0.8404	0.7671	0.813575	1.0409612
No Tuning	JALGOAN_R15	0.827	0.8753	0.8758	0.8095	0.8469	1.0013539
Bayesian Optimization	Global	0.7726	0.7841	0.7952	0.7871	0.78475	1.0440268
Grid Search	Global	0.7707	0.7825	0.7925	0.7845	0.78255	1.0469619
No Tuning	Global	0.7912	0.8229	0.8324	0.8307	0.8193	0.9971966

## Appendix B

**Table A4.** The standard error of prediction performance for each year and across years (Global) of the dataset 1 (Japonica) in terms of normalized root mean square error (NRMSE) under three methods of tuning (BO, GrS and NT). RE denotes relative efficiency.

Tuning Type	Year	NRMSE_SE_GC	NRMSE_SE_GY	NRMSE_SE_PH	NRMSE_SE_PHR	NRMSE_SE
Bayesian Optimization	2009	0.1457	0.0531	0.1201	0.0303	0.0436500
Bayesian Optimization	2010	0.0475	0.0239	0.0580	0.0244	0.0192250
Bayesian Optimization	2011	0.0344	0.0405	0.0606	0.0173	0.0191000
Bayesian Optimization	2012	0.0176	0.0408	0.0630	0.0317	0.0191375
Bayesian Optimization	2013	0.0594	0.0850	0.1139	0.0211	0.0349250
Bayesian Optimization	Global	0.0251	0.0129	0.0184	0.0240	0.0100500
Grid Search	2009	0.1326	0.0701	0.1150	0.0328	0.0438125
Grid Search	2010	0.0378	0.0296	0.0580	0.0227	0.0185125
Grid Search	2011	0.0353	0.0399	0.0619	0.0152	0.0190375
Grid Search	2012	0.0182	0.0410	0.0623	0.0310	0.0190625
Grid Search	2013	0.0599	0.1058	0.1102	0.0246	0.0375625
Grid Search	Global	0.0266	0.0187	0.0186	0.0253	0.0111500
No Tuning	2009	0.2366	0.0583	0.0931	0.0406	0.0535750
No Tuning	2010	0.0825	0.0189	0.0522	0.0315	0.0231375
No Tuning	2011	0.0325	0.0328	0.0587	0.0164	0.0175500
No Tuning	2012	0.0122	0.0419	0.0702	0.0332	0.0196875
No Tuning	2013	0.0614	0.0880	0.1179	0.0201	0.0359250
No Tuning	Global	0.0246	0.0141	0.0162	0.0307	0.0107000

**Table A5.** The standard error of prediction performance for each year and across years (Global) of the dataset 2 (Indica) in terms of normalized root mean square error (NRMSE) under three methods of tuning (BO, GrS and NT). RE denotes relative efficiency.

Tuning Type	Year	NRMSE_SE_GC	NRMSE_SE_GY	NRMSE_SE_PH	NRMSE_SE_PHR	NRMSE_SE
Bayesian Optimization	2010	0.0272	0.0275	0.0314	0.0428	0.0161125
Bayesian Optimization	2011	0.0300	0.0300	0.0333	0.0574	0.0188375
Bayesian Optimization	2012	0.0261	0.0125	0.0297	0.0457	0.0142500
Bayesian Optimization	Global	0.0225	0.0201	0.0172	0.0292	0.0111250
Grid Search	2010	0.0280	0.0264	0.0299	0.0481	0.0165500
Grid Search	2011	0.0245	0.0270	0.0356	0.0557	0.0178500
Grid Search	2012	0.0253	0.0133	0.0317	0.0398	0.0137625
Grid Search	Global	0.0197	0.0184	0.0177	0.0295	0.0106625
No Tuning	2010	0.0328	0.0168	0.0382	0.0528	0.0175750
No Tuning	2011	0.0235	0.0271	0.0366	0.0521	0.0174125
No Tuning	2012	0.0295	0.0077	0.0228	0.0387	0.0123375
No Tuning	Global	0.0207	0.0157	0.0176	0.0273	0.0101625

**Table A6.** The standard error of prediction performance for each environment and across environments (Global) of the dataset 3 (Groundnut) in terms of normalized root mean square error (NRMSE) under three methods of tuning (BO, GrS and NT). RE denotes relative efficiency.

Tuning Type	Environment	NRMSE_SE_NPP	NRMSE_SE_PYPP	NRMSE_SE_SYPP	NRMSE_SE_YPH	NRMSE_SE
Bayesian Optimization	ALIYARNAGAR_R15	0.0269	0.0164	0.0185	0.0287	0.0113125
Bayesian Optimization	ICRISAT_PR15-16	0.0299	0.0342	0.0386	0.0283	0.0163750
Bayesian Optimization	ICRISAT_R15	0.0228	0.0282	0.0255	0.0255	0.0127500
Bayesian Optimization	JALGOAN_R15	0.0249	0.0089	0.0110	0.0169	0.0077125
Bayesian Optimization	Global	0.0094	0.0105	0.0114	0.0222	0.0066875
Grid Search	ALIYARNAGAR_R15	0.0312	0.0228	0.0253	0.0310	0.0137875
Grid Search	ICRISAT_PR15-16	0.0322	0.0366	0.0403	0.0288	0.0172375
Grid Search	ICRISAT_R15	0.0223	0.0288	0.0268	0.0241	0.0127500
Grid Search	JALGOAN_R15	0.0246	0.0077	0.0086	0.0166	0.0071875
Grid Search	Global	0.0106	0.0119	0.0126	0.0241	0.0074000
No Tuning	ALIYARNAGAR_R15	0.0294	0.0205	0.0202	0.0262	0.0120375
No Tuning	ICRISAT_PR15-16	0.0429	0.0555	0.0581	0.0452	0.0252125
No Tuning	ICRISAT_R15	0.0239	0.0198	0.0216	0.0176	0.0103625
No Tuning	JALGOAN_R15	0.0224	0.0086	0.0078	0.0195	0.0072875
No Tuning	Global	0.0104	0.0113	0.0113	0.0205	0.0066875

## References

1. Caamal-Pat, D.; Pérez-Rodríguez, P.; Crossa, J.; Velasco-Cruz, C.; Pérez-Elizalde, S.; Vázquez-Peña, M. lme4GS: An R-Package for Genomic Selection. *Front. Genet.* **2021**, *12*, 680569. [[CrossRef](#)] [[PubMed](#)]
2. Montesinos-López, O.A.; Montesinos-López, A.; Cano-Paez, B.; Hernández-Suárez, C.M.; Santana-Mancilla, P.C.; Crossa, J. A Comparison of Three Machine Learning Methods for Multivariate Genomic Prediction Using the Sparse Kernels Method (SKM) Library. *Genes* **2022**, *13*, 1494. [[CrossRef](#)] [[PubMed](#)]
3. Cordell, H.J. Epistasis: What it means, what it doesn't mean, and statistical methods to detect it in humans. *Hum. Mol. Genet.* **2002**, *11*, 2463–2468. [[CrossRef](#)] [[PubMed](#)]

4. Golan, D.; Rosset, S. Effective genetic-risk prediction using mixed models. *Am. J. Hum. Genet.* **2014**, *95*, 383–393. [[CrossRef](#)] [[PubMed](#)]
5. Gianola, D.; Fernando, R.L.; Stella, A. Genomic-assisted prediction of genetic value with semi parametric procedures. *Genetics* **2006**, *173*, 1761–1776. [[CrossRef](#)] [[PubMed](#)]
6. Gianola, D.; van Kaam, J.B.C.H.M. Reproducing kernel Hilbert spaces regression methods for genomic assisted prediction of quantitative traits. *Genetics* **2008**, *178*, 2289–2303. [[CrossRef](#)] [[PubMed](#)]
7. Vapnik, V. *The Nature of Statistical Learning Theory*; Springer: New York, NY, USA, 1995.
8. Long, N.; Gianola, D.; Rosa, G.J.; Weigel, K.A.; Kranis, A.; González-Recio, O. Radial basis function regression methods for predicting quantitative traits using SNP markers. *Genet. Res.* **2010**, *92*, 209–225. [[CrossRef](#)] [[PubMed](#)]
9. Crossa, J.; de los Campos, G.; Pérez, P.; Gianola, D.; Burgueño, J.; Araus, J.L. Prediction of genetic values of quantitative traits in plant breeding using pedigree and molecular markers. *Genetics* **2010**, *186*, 713–724. [[CrossRef](#)] [[PubMed](#)]
10. Cuevas, J.; Montesinos-López, O.A.; Juliana, P.; Guzmán, C.; Pérez-Rodríguez, P.; González-Bucio, J.; Burgueño, J.; Montesinos-López, A.; Crossa, J. Deep Kernel for Genomic and Near Infrared Predictions in Multi-environment Breeding Trials. *G3-Genes Genomes Genet.* **2019**, *9*, 2913–2924. [[CrossRef](#)] [[PubMed](#)]
11. Tusell, L.; Pérez-Rodríguez, P.; Wu, S.F.X.-L.; Gianola, D. Genome-enabled methods for predicting litter size in pigs: A comparison. *Animal* **2013**, *7*, 1739–1749. [[CrossRef](#)] [[PubMed](#)]
12. Morota, G.; Koyama, M.; Rosa, G.J.M.; Weigel, K.A.; Gianola, D. Predicting complex traits using a diffusion kernel on genetic markers with an application to dairy cattle and wheat data. *Genet. Sel. Evol.* **2013**, *45*, 17. [[CrossRef](#)] [[PubMed](#)]
13. Aroju, S.K.; Cao, M.; Trolove, M.; Barrett, B.A.; Inch, C.; Eady, C.; Stewart, A.; Faville, M.J. Multi-Trait Genomic Prediction Improves Predictive Ability for Dry Matter Yield and Water-Soluble Carbohydrates in Perennial Ryegrass. *Front. Plant Sci.* **2020**, *11*, 1197. [[CrossRef](#)] [[PubMed](#)]
14. Montesinos-López, O.A.; Montesinos-López, A.; Crossa, J.; Cuevas, J.; Montesinos-López, J.C.; Salas-Gutiérrez, Z.; Lillemo, M.; Philomin, J.; Singh, R. A Bayesian Genomic Multi-output Regressor Stacking Model for Predicting Multi-trait Multi-environment Plant Breeding Data. *G3-Genes Genomes Genet.* **2019**, *9*, 3381–3393. [[CrossRef](#)] [[PubMed](#)]
15. Monte Verde, E.; Gutierrez, L.; Blanco, P.; Pérez de Vida, F.; Rosas, J.E.; Bonnet Carrère, V.; Quero, G.; McCouch, S. Integrating Molecular Markers and Environmental Covariates To Interpret Genotype by Environment Interaction in Rice (*Oryza sativa* L.) Grown in Subtropical Areas. *G3 Genes Genomes Genet.* **2019**, *9*, 1519–1531. [[CrossRef](#)]
16. Pandey, M.K.; Chaudhari, S.; Jarquin, D.; Janila, P.; Crossa, J.; Patil, S.C.; Sundravada, S.; Khare, D.; Bhat, R.S.; Radhakrishnan, T.; et al. Genome-based trait prediction in multi-environment breeding trials in groundnut. *Theor. Appl. Genet.* **2020**, *133*, 3101–3117. [[CrossRef](#)] [[PubMed](#)]
17. Gapare, W.; Liu, S.; Conaty, W.; Zhu, Q.H.; Gillespie, V.; Llewellyn, D.; Stiller, W.; Wilson, I. Historical Datasets Support Genomic Selection Models for the Prediction of Cotton Fiber Quality Phenotypes Across Multiple Environments. *G3 Genes Genomes Genet.* **2018**, *8*, 1721–1732. [[CrossRef](#)] [[PubMed](#)]
18. VanRaden, P.M. Efficient Methods to Compute Genomic Predictions. *J. Dairy Sci.* **2008**, *91*, 4414–4423. [[CrossRef](#)] [[PubMed](#)]
19. Montesinos-López, O.A.; Montesinos-López, A.; Crossa, J. (Eds.) *Multivariate Statistical Machine Learning Methods for Genomic Prediction*; Springer International Publishing: Cham, Switzerland, 2022; ISBN 978-3-030-89010-0.
20. Montesinos-López, O.A.; Carter, A.H.; Bernal-Sandoval, D.A.; Cano-Paez, B.; Montesinos-López, A.; Crossa, J. A Comparison Between Three Tuning Strategies for Gaussian kernels in the Context of Univariate Genomic Prediction. *Genes* **2022**. submitted for publication.
21. R Core Team. *R: A Language and Environment for Statistical Computing*; R Foundation for Statistical Computing: Vienna, Austria, 2022.
22. Pérez, P.; de los Campos, G. Genome-Wide Regression and Prediction with the BGLR Statistical Package. *Genetics* **2014**, *198*, 483–495. [[CrossRef](#)] [[PubMed](#)]
23. Meuwissen, T.H.E.; Hayes, B.J.; Goddard, M.E. Prediction of total genetic value using genome-wide dense marker maps. *Genetics* **2001**, *157*, 1819–1829. [[CrossRef](#)] [[PubMed](#)]