
Research Paper

Spatial kernel models capturing field heterogeneity for accurate estimation of genetic potential

Motoyuki Ishimori¹⁾, Hideki Takanashi¹⁾, Masaru Fujimoto¹⁾, Hiromi Kajiya-Kanegae¹⁾, Junichi Yoneda²⁾, Tsuyoshi Tokunaga²⁾, Nobuhiro Tsutsumi¹⁾ and Hiroyoshi Iwata^{*1)}

¹⁾ Graduate School of Agricultural and Life Sciences, The University of Tokyo, 1-1-1 Yayoi, Bunkyo, Tokyo 113-8657, Japan

²⁾ EARTHNOTE Co. Ltd., 1388 Sokei, Ginoza, Okinawa 904-1303, Japan

According to Fisher's principles, an experimental field is typically divided into multiple blocks for local control. Although homogeneity is supposed within a block, this assumption may not be practical for large blocks, such as those including hundreds of plots. In line evaluation trials, which are essential in plant breeding, field heterogeneity must be carefully treated, because it can cause bias in the estimation of genetic potential. To more accurately estimate genotypic values in a large field trial, we developed spatial kernel models incorporating genome-wide markers, which consider continuous heterogeneity within a block and over the field. In the simulation study, the spatial kernel models were robust under various conditions. Although heritability, spatial autocorrelation range, replication number, and missing plots directly affected the estimation accuracy of genotypic values, the spatial kernel models always showed superior performance over the classical block model. We also employed these spatial kernel models for quantitative trait locus mapping. Finally, using field experimental data of bioenergy sorghum lines, we validated the performance of the spatial kernel models. The results suggested that a spatial kernel model is effective for evaluating the genetic potential of lines in a heterogeneous field.

Key Words: spatial kernel, field heterogeneity, variety evaluation trial, genetic potential, genomic selection, bioenergy sorghum.

Introduction

Field heterogeneity can often be a confounding factor in agronomical field experiments, such as variety evaluation trials. Therefore, Fisher's principles of experimental design, namely (1) replication, (2) randomization, and (3) blocking (local control), are essential for correctly interpreting the results of field experiments (Fisher 1935). According to Fisher's principles, multiple plots assigned to a variety are located in separate randomized blocks. In genotype evaluation trials, large blocks containing plots of all genotypes are often treated as a single replication unit (i.e., the complete block design), and the effects of the blocks are statistically estimated based on the assumption that the field is homogenous within a block.

However, field heterogeneity can exist even within a block (Gusmão 1986). Therefore, spatial analysis is necessary for capturing the heterogeneity within a block in addition to that among blocks. A study by Cullis and Gleeson

(1991) was the first to put forth a practical model for the two-dimensional spatial analysis of field experimental data. Subsequently, Gilmour *et al.* (1997) distinguished the components of field heterogeneity using an extended statistical model. Their study revealed that spatial variations over a field are mainly derived from (1) global heterogeneity, (2) highly local (natural) heterogeneity, and (3) artificial (extraneous) procedures. Although the contributions of these factors might vary across trials, spatial correlation patterns among neighboring plots can be captured using geostatistical modeling.

Geostatistical modeling has been applied to spatial variability analyses in agronomical fields since the 1980s (Vieira *et al.* 1983). Several studies showed that spatial analysis improved the interpretation of plant breeding trial data, something that is crucial for highly efficient selection (Duarte and Vencovsky 2005, Wu *et al.* 1998). To appropriately select superior genotypes, genotypic values must be accurately estimated based on massive phenotypic data obtained from breeding trials. In genomic selection (GS), genotypic values estimated based on phenotypic data are required for training a prediction model based on genome-wide markers (Meuwissen *et al.* 2001), and such models can then be used for predicting the genotypic values of non-phenotyped genotypes. Lado *et al.* (2013) reported the

Communicated by Takeshi Hayashi

Received May 21, 2020. Accepted May 19, 2021.

First Published Online in J-STAGE on August 27, 2021.

*Corresponding author (e-mail: hiroiwata@g.ecc.u-tokyo.ac.jp)

efficiency of spatial analysis in obtaining better estimates of genotypic values for GS. Similarly, Bernal-Vasquez *et al.* (2014) examined the efficiency of spatial analysis as the first stage of GS. Although these studies employed spatial analysis to estimate the prerequisite genotypic values for GS, they did not employ prediction models combined directly with spatial analysis.

To the best of our knowledge, a study by Elias *et al.* (2018) was the first to incorporate spatial analysis into GS models for considering field heterogeneity. In their model, a spatial kernel, transformed from physical distance on the field, was simultaneously fit with a genomic kernel based on genome-wide markers. GS models combining spatial analysis significantly increase the prediction accuracy based on simulated as well as real data. Because each field has specific heterogeneity, which can be expressed using approximate covariance structure, various spatial kernels based on different covariance structures should be examined for agronomical field experiments (Richter and Kroschewski 2012). For the practical utilization of GS models with spatial analysis, the effects of factors specific to agronomical fields, such as trait heritability, spatial autocorrelation range, replication number, and missing plots, should also be validated.

To this end, in the present study, we comprehensively assessed the efficiency of GS models with spatial analysis through a simulation study and validation with real data. As opposed to the classical block model in which field homogeneity is assumed within a block, we focused on the robustness of the spatial kernel models. Furthermore, we investigated whether GS models combined directly with spatial analysis were more efficient than the ones combined indirectly with spatial analysis, such as the models developed by Lado *et al.* (2013) and Bernal-Vasquez *et al.* (2014). Moreover, we proposed the application of the developed spatial GS models for quantitative trait locus (QTL) mapping. Finally, we validated the developed spatial GS models using a dataset from a field experiment with bioenergy sorghum [*Sorghum bicolor* (L.) Moench] lines.

Materials and Methods

Marker data for the simulation study

For the simulation study, we utilized the marker score data of diverse rice germplasm (Dilla-Ermita *et al.* 2017) available at <https://snp-seek.irri.org/> (accessed on June 25, 2021). The marker score data included 248 accessions and 40,840 markers with missing records. We imputed missing genotypes using Beagle v4.0 (Browning and Browning 2007). After imputation, we filtered out the markers with high heterozygosity (more than 20%) because the marker genotypes of the germplasm were expected to be homozygous. From the 248 accessions, we selected 200 accessions for a balanced design in the simulation study. Then, we filtered out markers with low minor allele frequency (below 5%) and those that exhibited a strong linkage ($r^2 > 0.8$) to

other markers on the same chromosome, retaining markers that appear earliest in ascending order of their physical positions. Finally, the marker score data chosen for the simulation study contained 200 accessions and 7,913 markers.

QTL settings for the simulation study

In the simulation study, QTLs were assumed to be located on markers. The number of QTLs was fixed at 20 through all simulations because it did not influence the estimation accuracy of genotypic values in a preliminary analysis. Although the detection power in QTL mapping might be influenced by the number of QTLs, we performed all simulations with 20 QTLs in this study to focus on the estimation of genotypic values. The markers on which QTLs were located were randomly selected in each simulation. Moreover, we did not consider linkage disequilibrium (LD) among markers for QTL mapping (i.e., multiple QTLs in LD could be included). The effect of each QTL was randomly generated to ensure a normal distribution (mean of 0 and standard deviation of 1). Only the additive effect was considered at each QTL (no dominance and epistasis). Genotypic values of the 200 rice accessions were calculated based on the marker scores and simulated marker effects and referred to as the “true” genotypic values. The following formula was used:

$$g_i = \sum_{j=1}^J Q_{ij}a_j,$$

where g_i is the true genotypic value of the i^{th} accession; J is the number of QTLs ($n = 20$); Q_{ij} is the score of i^{th} accession for the marker on which the j^{th} QTL is located, which is represented as $[-1, 0, 1]$; and a_j is the genetic effect of the j^{th} QTL.

Simulated field layout

We simulated multiple blocks located in a field with each block as a replication unit. A single block comprised 200 plots, arranged in 10 rows (Y-axis) and 20 columns (X-axis). The rows and columns were assumed to have the same scale. Each plot included a single accession, that is, one set of 200 accessions was located in each block. In each simulation, the 200 accessions were randomly assigned to plots in each block using the randomized complete block design. To examine the effects of replication numbers, simulation studies were performed with two or four replication blocks. In the simulation study with two blocks, the field comprised 10 rows and 50 columns, and the two blocks were located along the X-axis (the space between the blocks was 10 columns). In the simulation study with four blocks, the field comprised 26 rows and 50 columns, and the four blocks were located along both X- and Y-axes (the space between the blocks was 6 rows and 10 columns, respectively).

To investigate the effects of missing data on the modeling of field heterogeneity, we performed simulations with 0%, 20%, and 50% missing plots. Missing plots were

randomly selected from across the field in each simulation. Simulations with missing data were only performed with four replication blocks.

Simulation of field heterogeneity

Field heterogeneity was simulated using two variables, namely heritability (h^2) and spatial autocorrelation range. In this study, low (20%) and intermediate (50%) heritability estimates were examined because preliminary analysis revealed that 80% heritability was rather high for spatial analysis. We assumed that the phenotypic values were affected by (1) random noise and (2) spatial autocorrelations in heterogeneity among neighboring plots, in addition to heritability. Random noise represents any factors specific to each plot that are not explicitly explained (e.g., the measurement error). Spatial autocorrelations in heterogeneity represent fluctuations in biotic (e.g., insect damage) and abiotic (e.g., fertilization and irrigation) conditions and are the main target of this study. We assumed that the effects of random noise explained 20% of the phenotypic variation under 50% h^2 , while field heterogeneity explained other non-genetic variations (30%). Under 20% h^2 , three conditions for random noise (20%, 40%, and 60%) were examined, which corresponded to the proportion of spatial variance (i.e., the degree of field heterogeneity) at 60%, 40%, and 20%, respectively.

The effect of field heterogeneity, which was designated as the spatial effect, was assessed for each plot. Phenotypic values were calculated using the following formula:

$$y_k = g_k + q_k + e_k,$$

where y_k is the phenotypic value of the k^{th} plot across blocks; g_k is the genotypic value of the accession assigned to the k^{th} plot; q_k is the spatial effect of the k^{th} plot; and e_k is the unexpected effect of the k^{th} plot. The vectors \mathbf{q} and \mathbf{e} correspond to field heterogeneity and random noise, respectively.

The spatial effect \mathbf{q} was assumed to be autocorrelated among neighboring plots. In all simulations, we generated spatial effects based on the variogram γ , which defines the dissimilarity between two plots in a field. In the random field (RF) approach, the variogram model γ can be derived from a corresponding covariance function \mathbf{C} :

$$\gamma(h) = C_0(1 - C_h),$$

where h is the Euclidean distance between two plots along the X- and Y-axes. Nugget effects, which are attributed to non-spatial variance, were not included in the model. Although we used $C_0 = 0.025$ as the sill (the maximum value of γ), the value did not reveal any importances in this study, because the simulated spatial effect was scaled to heritability. We examined variogram models based on three covariance functions: exponential, Gaussian, and spherical, which are described as follows:

$$C_h(\text{exponential}) = \exp\left(\frac{-h}{a}\right),$$

$$C_h(\text{Gaussian}) = \exp\left(\frac{-h^2}{a^2}\right),$$

$$C_h(\text{spherical}) = \left[1 - 1.5\left(\frac{h}{a}\right) + 0.5\left(\frac{h^3}{a^3}\right)\right] \text{ if } h \leq a; \text{ else } 0,$$

where a is the range parameter for adjusting the dissimilarity together with plot distance. We examined two range parameters: “narrow” ($a = 5$) and “wide” ($a = 30$). The larger the value of a is, the more correlated the spatial effects between distant plots are. Only the 20 nearest observations were used for the simulation instead of all plots (i.e., local kriging). We assumed that heterogeneity is stretched across the field with no tendencies toward rows or columns (i.e., isotropy). Based on the variogram γ , the spatial effect \mathbf{q} was generated, which was scaled to heritability based on the variance of \mathbf{q} . The variogram model was implemented with gstat (Gräler *et al.* 2016) in R (R Core Team 2019).

The random noise \mathbf{e} was assumed to follow a normal distribution with a mean of 0. As mentioned above, the standard deviation of \mathbf{e} was adjusted to correspond to the heritability setting of each simulation.

Simulations of field heterogeneity were run 50 times for each different condition (i.e., variogram models, blocks, the levels of heritability, spatial autocorrelation ranges, and random noises).

Statistical models for simultaneous estimation of marker and spatial effects (one-step method)

To simultaneously estimate spatial and marker effects, we employed the following formula:

$$y_k = \mu + \sum_{z=1}^Z M_{kz}a_z + q_k + e_k, \quad (1)$$

where y_k is the phenotypic value of the k^{th} plot across blocks; μ is the mean value across all plots; Z is the number of markers ($n = 7913$); M_{kz} is the scaled and centered marker score of accession assigned to the k^{th} plot at the z^{th} marker; a_z is the genetic effect of the z^{th} marker; q_k is the spatial effect of the k^{th} plot; and e_k is the residual [where $\mathbf{e} \sim N(0, \mathbf{I}\sigma_e^2)$].

For estimation of the marker effect \mathbf{a} , BayesB was used as the prior (Meuwissen *et al.* 2001). In BayesB, the distribution of marker effects is assumed to follow two-component mixtures, with a point of mass at zero and a scaled t -slab. The proportion of markers with a non-zero effect is specified by π , which follows the beta distribution in the range $[0, 1]$.

The spatial effect \mathbf{q} , was assumed to follow the distribution $\mathbf{q} \sim N(0, \mathbf{K}\sigma_q^2)$. \mathbf{K} is an $n \times n$ spatial kernel matrix, where n is the total number of plots in the field. Elias *et al.* (2018) examined three kernel functions [Gaussian (*gaus*), power, and spherical (*sph*)] for \mathbf{K} . The three models reflect a specific spatial correlation between plots, which were

tested together in another study on field heterogeneity (Richter and Kroschewski 2012). In addition to these kernels, we examined the exponential (*expo*) kernel. The *expo* and *power* models are different parameterizations of essentially the same model (Piepho *et al.* 2015). The formulas for these models are as follows:

$$\text{expo} = \exp\left(\frac{-S_{kr}}{h}\right),$$

$$\text{gaus} = \exp\left(\frac{-S_{kr}^2}{h^2}\right),$$

$$\text{power} = f^{S_{kr}},$$

$$\text{sph} = \begin{cases} 1 - 1.5 \left(\frac{S_{kr}}{h}\right) + 0.5 \left(\frac{S_{kr}^3}{h^3}\right) & \text{if } (S_{kr} \leq h); \\ \text{else } 0, \end{cases}$$

where S_{kr} is the component of an $n \times n$ spatial distance matrix (\mathbf{S}) and h or f is a hyperparameter for standardization. S_{kr} was calculated as the Euclidean distance between the k^{th} and r^{th} plots using the following formula:

$$S_{kr} = \sqrt{(\text{row}_k - \text{row}_r)^2 + (\text{column}_k - \text{column}_r)^2},$$

where row_k and row_r are the row numbers and column_k and column_r are the column numbers of the k^{th} and r^{th} plots.

The extent of correlation in kernel structures changes with the hyperparameters (Elias *et al.* 2018). Generally, correlations are closer to zero under small hyperparameters in any model and vice versa. The extent of dependence on hyperparameters is relatively strong for the *gaus* and *power* kernels but mild for the *expo* and *sph* kernels. The value of h ranges from 0 to the maximum value of the factors of \mathbf{S} ($\max_{k,r} S_{kr}$), while the value of f ranges from 0 to 1. The optimal values of h and f depend on datasets. Based on preliminary analysis, we selected three values covering the ranges of h and f , which were sufficient for determining the appropriate hyperparameter. For h , we examined three values, H1–H3, which were calculated using the following formula

$$h = \frac{\max(\mathbf{S})}{3^2} x^2,$$

where $x = 1, 2$, or 3 . Similarly, for f , we examined three values, F1–F3, which were calculated using the following formula

$$f = 1.001 - \left(\frac{x}{3}\right)^3,$$

where $x = 3, 2$, or 1 . When h or $f \rightarrow 0$, the correlation is close to zero even between neighboring plots. On the contrary, when $h \rightarrow \max_{k,r} S_{kr}$ or $f \rightarrow 1$, even farther plots are closely correlated with one another.

For comparison with the field heterogeneity models based on the spatial kernel \mathbf{K} , we also examined the perfor-

mance of a classical block model that treats each block as a single replication unit. The model is described using the following formula:

$$y_k = \mu + \sum_{z=1}^Z M_{kz} a_z + b_e + e_k, \quad (2)$$

where b_e is the random effect of the e^{th} block when the k^{th} plot is located in the e^{th} block. The block effects \mathbf{b} follow the multivariate Gaussian distribution $N(0, \mathbf{I}\sigma_b^2)$, where σ_b^2 follows a scaled inverse chi-square distribution.

The genotypic value of the i^{th} accession, \hat{g}_i , was calculated as $\hat{g}_i = \hat{\mu} + \sum_{z=1}^Z M_{iz} \hat{a}_z$. The estimation accuracy of genotypic values is the average value of correlation coefficients between \mathbf{g} and $\hat{\mathbf{g}}$ with 50 simulations. In the simulation study, the optimum value of the hyperparameter (h or f) can be determined based on estimation accuracy because \mathbf{g} is obvious.

To infer the QTL detection power, we used the estimated marker variances, which were calculated as the variance of the genotype score \times estimated effects of markers, \mathbf{v} [$v_z = \text{var}(M_z a_z)$]. The total genetic variance was calculated as the variance of estimated genotypic values. Only markers over the 99.5 percentile of the empirical distribution of the estimated variances of marker effects, which corresponded to the top 40 markers with high variances, were selected to examine associations with QTLs. Considering LD around QTLs, we assumed that QTLs were detected when the top 40 markers were located within 500 kbp from the true QTL. If several QTLs were located within 500 kbp from the markers, we judged that all QTLs in the region were detected. Other markers located away from any QTL were classified as false positives. We used the average values of QTL detection power (% , detected QTLs/20 simulated QTLs) and the false-positive rate (% , markers not associated with any QTL/the top 40 markers) over 50 simulations.

Statistical models for the estimation of spatial effects followed by marker effects (two-step method)

The one-step method described above can be divided into two steps, i.e., (1) the estimation of block effects and (2) the estimation of marker effects, in that order. To evaluate the efficiency of the one-step method, we examined the performance of the two-step method. The formulas for the first step are as follows:

$$y_k = \mu + d_k + q_k + e_k \text{ (spatial kernel models)}, \quad (3)$$

$$y_k = \mu + d_k + b_e + e_k \text{ (block model)}, \quad (4)$$

where d_k is the genotype effect assigned to the k^{th} plot. The genotype effect \mathbf{d} was assumed to follow the multivariate normal distribution $N(0, \mathbf{W}\sigma_d^2)$. \mathbf{W} is an $n \times n$ (where n is the number of plots) binary matrix, in which the factor \mathbf{W}_{kr} is 1 when the same genotype is assigned to the k^{th} and r^{th} plots, while \mathbf{W}_{kr} is 0 in other cases. The estimated line

value \hat{p} is calculated as $\hat{p}_i = \hat{\mu} + \hat{d}_k$ (the i^{th} genotype is assigned to the k^{th} plot).

The second step can be solved using the elements of \hat{p} as the values of a response variable, as follows:

$$p_i = \mu + \sum_{z=1}^Z M_{iz}a_z + e_i. \quad (5)$$

The estimated genotypic value \hat{g} can be calculated in the same manner as in the one-step method. All one-step and two-step methods were implemented using the BGLR package in R (Pérez and de los Campos 2014). The posterior density was calculated based on 15,000 sampling iterations, in which the first 5,000 samples were discarded.

Field experimental data of bioenergy sorghum lines

To validate the performance of the developed spatial kernel models, we applied the one-step method to field experimental data of bioenergy sorghum lines. The experiment comprised a breeding line evaluation trial on marginal lands located in Mexico (Ishimori *et al.* 2020). Although the experiment was performed for 3 years (2013 to 2015), field data obtained in 2014 alone were used in this study. Sorghum plants were grown under standard fertilization conditions [N:P:K = 17:60:50 (kg·ha⁻¹)] for bioenergy sorghum.

The experiment included 598 genotypes (inbred lines and F₁ hybrids). We employed the randomized complete block design without check varieties in the experiment; however, the design could be treated as an incomplete block design because there were some missing plots. We set two blocks (A and B) in the experimental field, each including 598 plots (totaling 1,196 plots), with one variety assigned to each plot. The block comprised 23 rows (Y-axis) and 27 columns (X-axis). The two blocks were aligned along the X-axis, with three additional columns arranged between the blocks. Each row comprised a ridge across the blocks for drip irrigation. To reduce border effects, we planted one genotype between rows and around the borders of the blocks.

Each plot included five plants. The phenotypic value was calculated as the mean value of two healthy plants from each plot. Although the spatial distance between the plots, S , was calculated in the same manner as that in the simulation study, we used a real scale (m) along both X- and Y-axes on a two-dimensional field as follows:

$$S_{kr} = \sqrt{(X_k - X_r)^2 + (Y_k - Y_r)^2}.$$

The distance between two neighboring plots was 0.9 × 1.5 m, measured from the center of the plots. The minimum and maximum distances between two plots over the field were 0.9 and 60.2 m, respectively.

Because the true genotypic values are unknown for the real dataset, the optimal hyperparameter (h or f) is also unknown for the estimation of genotypic values. Therefore, we used the kernel averaging approach proposed by de los

Campos *et al.* (2010). The formula for a one-step method is as follows:

$$y_k = \mu + \sum_{z=1}^Z M_{kz}a_z + \sum_{t=1}^T q_{kt} + e_k, \quad (6)$$

where q_{kt} is the spatial effect of the k^{th} plot evaluated by the spatial kernel \mathbf{K}_t with the t^{th} value of the hyperparameter [$q_{kt} \sim N(0, \mathbf{K}_t \sigma_{qt}^2)$]. The kernel averaging approach is equivalent to a model (1) with a single kernel for the spatial effect if variance components are known, suggested by de los Campos *et al.* (2010). We used the kernel averaging approach with the three values of the hyperparameter (H1–H3 or F1–F3). The performance of the kernel averaging approach was also validated in the simulation study.

To validate the optimal spatial kernel model for field heterogeneity in the real dataset, we performed 20 simulations of 5-fold cross-validation (CV). All genotypes, but not plots, were randomly divided into five subsets. Each subset included two plots to which a genotype was allocated when there were no missing data. In one CV round, four subsets were used for training the model and one subset was used for validation. A set of predicted phenotypic values across all plots was obtained after the completion of a CV round. The predicted phenotypic values of plots (\hat{y}) were calculated as follows:

$$\hat{y}_k = \hat{\mu} + \sum_{z=1}^Z M_{kz}\hat{a}_z + \hat{q}_k \text{ (the spatial kernel models)}$$

$$\hat{y}_k = \hat{\mu} + \sum_{z=1}^Z M_{kz}\hat{a}_z + \hat{b}_e \text{ (the block model).}$$

The correlation coefficients between phenotypic values from the real dataset and the predicted phenotypic values from each CV round were calculated, and then mean values across 20 simulations were used to decide the optimal spatial model. Using the optimal model, we estimated the spatial and marker effects on data from all plots. Other procedures were following the simulation study. In this study, two traits, culm juice Brix and culm length, were used for model validation.

Detailed descriptions of marker score data are available in Ishimori *et al.* (2020). For this study, we modified the conditions to filter out markers under high LD; the threshold of $r^2 > 0.95$ between two markers (the original paper) was changed to $r^2 > 0.8$. The number of markers was 4,642 in this study.

Results

Effects of heritability, spatial autocorrelation range, and replication number

The simulation of field heterogeneity based on the exponential variogram model, with four replication blocks, is illustrated in Fig. 1. We also simulated field heterogeneity based on the Gaussian and spherical variogram models (Supplemental Fig. 1).

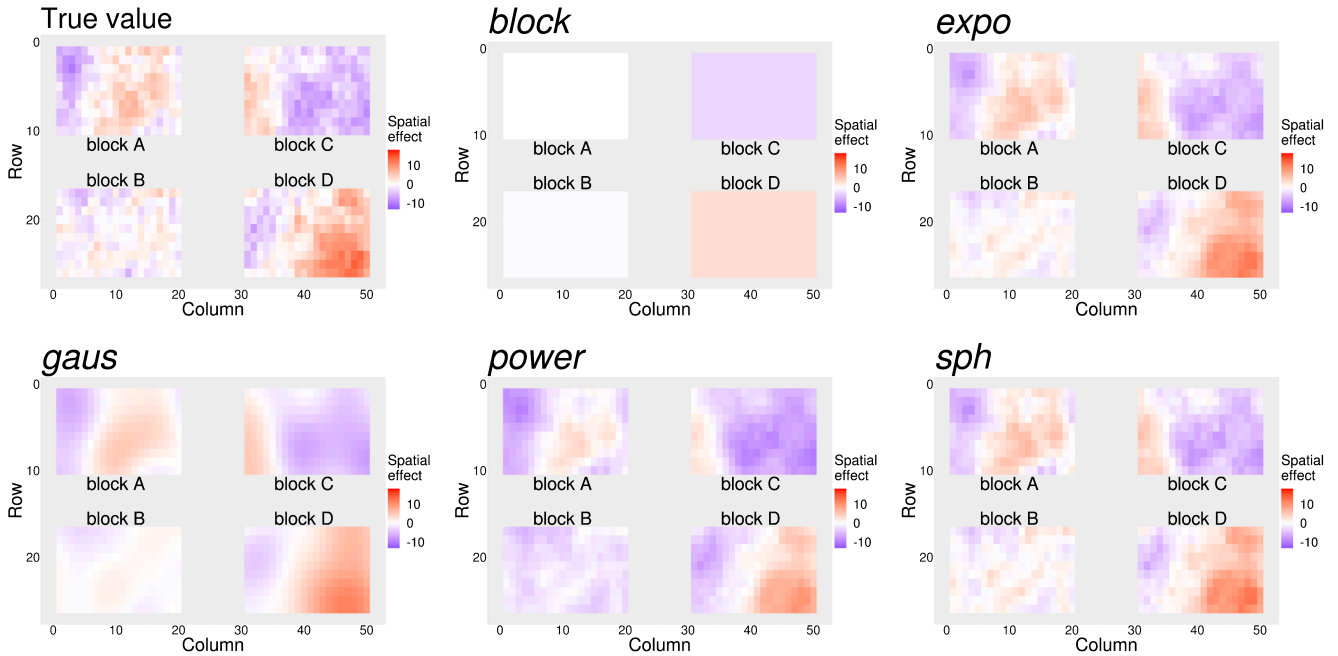


Fig. 1. Spatial analysis of field heterogeneity in a simulation based on an exponential variogram model. The example field has four experimental blocks (A–D) for replication. The true value, which was generated in each simulation, is shown in the upper left panel. In the *block* model (the upper central panel), the spatial effect of plots was estimated to be identical within a block. In spatial kernel models [exponential (*expo*), Gaussian (*gaus*), *power*, and spherical (*sph*)], the spatial effect is estimated for each plot.

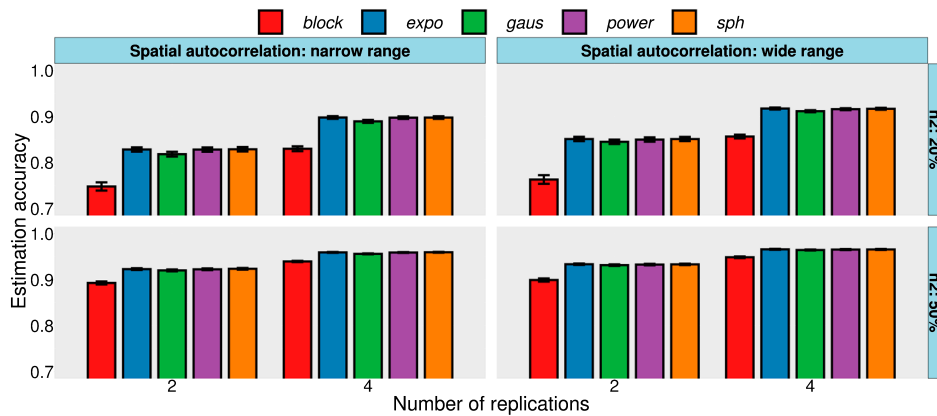


Fig. 2. Effects of replications on the estimation accuracy of genotypic values (one-step method). The simulation study was performed with different conditions for three factors: replication number (two or four), spatial autocorrelation range (narrow or wide), and heritability (20% or 50%). Here, the variogram model (exponential) and random noise (20%) were fixed. For each simulation, genotypic values were estimated with five statistical models. The estimation accuracy is the mean value of 50 simulations. Error bars show the standard error. h^2 : heritability. *block*, block model; *expo*, exponential model; *gaus*, Gaussian model; *power*, power model; *sph*, spherical model.

The results of simulations based on the exponential variogram model are shown in **Fig. 2**. The effects of h^2 and replication number on the estimation accuracy of genotypic values were evident. With 50% h^2 and four replications, the estimation accuracy was high using all one-step models. The estimation accuracy generally decreased when either condition was changed (20% h^2 or two replications). Moreover, the spatial autocorrelation range affected the estimation of genotypic values. The estimation accuracy increased when the spatial autocorrelation range was wide. In particular, the effect of spatial autocorrelation range was observed

with 20% h^2 and two replications.

With any condition of these factors, the four spatial kernel models always represented higher estimation accuracy than the *block* model (**Fig. 2**). With 20% h^2 , the spatial kernel models were evidently superior to the *block* model. Meanwhile, with 50% h^2 , the difference in estimation accuracy between the *block* and spatial kernel models was relatively small. There were no evident differences in accuracy among the four spatial kernel models.

In simulations based on the Gaussian and spherical variogram models, the influences of h^2 , replication number, and

spatial autocorrelation range on genotypic value estimation were similar to simulations based on the exponential variogram model (Supplemental Fig. 2). Likewise, the spatial kernel models had higher estimation accuracy than the *block* model. The differences among the spatial kernel models were small. The influences of spatial variance size, missing plots, and estimation method were also similar among simulations based on different variogram models (data not shown). Therefore, we only describe the results of simulations based on the exponential variogram model in the following sections.

Influences of spatial variance size

To determine if spatial analysis depends on the size of spatial variance, we performed simulations with different spatial variance sizes under 20% h^2 . The estimation accuracy of genotypic values using the spatial kernel models was positively related to the size of spatial variance (Supplemental Fig. 3). In contrast, the estimation accuracy using the *block* model was similar upon using different spatial variance sizes.

Robustness of the one-step method with missing plots

Compared with simulations without missing plots, the estimation accuracy of genotypic values decreased with missing plots in all one-step models (Fig. 3). The missing plot rate directly affected the estimation accuracy. The estimation accuracy decreased to a much greater extent when 50% of the plots were missing than when 20% of the plots were missing. Similarly, h^2 affected the estimation accuracy; estimation accuracy was lower with 20% h^2 than with 50% h^2 .

Under the narrow spatial autocorrelation range, the difference in estimation accuracies under different missing plot rates between the *block* and spatial kernel models was small. In contrast, under a wide spatial autocorrelation range, the missing plot rate negatively affected the estima-

tion accuracy of the *block* model to a greater extent than that of the spatial kernel models.

In two conditions with the same plot numbers (i.e., the same sample size), the spatial kernel models with no missing plots showed better performance in the estimation of genotypic values (Supplemental Fig. 4).

Comparison between one-step and two-step methods

To compare the performances of the one-step and two-step methods in the estimation of genotypic values, we examined their estimation accuracy (i.e., the correlation coefficient) and root mean square error (RMSE) under the same conditions. There was no difference in estimation accuracy based on the correlation coefficient of the two methods under any condition (statistical model, h^2 , and spatial autocorrelation range) (Fig. 4A). In contrast, the one-step methods had smaller RMSE values than the two-step methods (Fig. 4B). Using the two-step methods, the estimated genotypic values could be more severely shrunk than using the one-step methods (Supplemental Fig. 5).

QTL detection power of the one-step method

Using the spatial kernel models developed in this study, spatial and marker effects were simultaneously estimated. In the simulation study, the QTL detection power was calculated by analyzing the estimated marker variances of simulated QTLs (Fig. 5A). h^2 , spatial autocorrelation range, and replication number directly affected QTL detection power and the false-positive rate (Fig. 5B, 5C). Under all conditions, the spatial kernel models showed higher QTL detection power and a lower false-positive rate than the *block* model.

Spatial analysis of real sorghum data

To examine whether the spatial kernel models can capture unknown field heterogeneity in real data, we employed the spatial kernel models to a sorghum dataset obtained

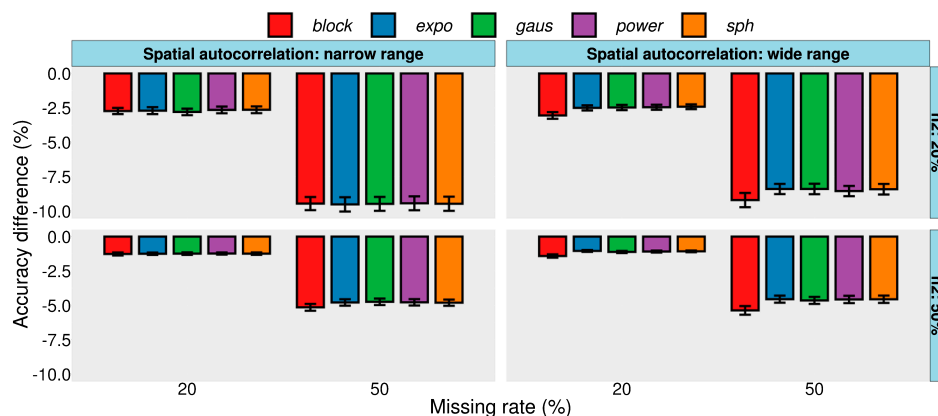


Fig. 3. Robustness of the one-step method in the presence of missing plots. We examined two plot missing rates (20% or 50%). Missing plots were randomly introduced across four blocks. Here, the variogram model (exponential), replication number (four), and random noise (20%) were fixed. Genotypic values were estimated with the one-step method. The estimation accuracy of genotypic values (%) was calculated and compared between the presence or absence of missing plots. The accuracy difference is the mean value of 50 simulations. Error bars show the standard error. h^2 : heritability. *block*, block model; *expo*, exponential model; *gaus*, Gaussian model; *power*, power model; *sph*, spherical model.

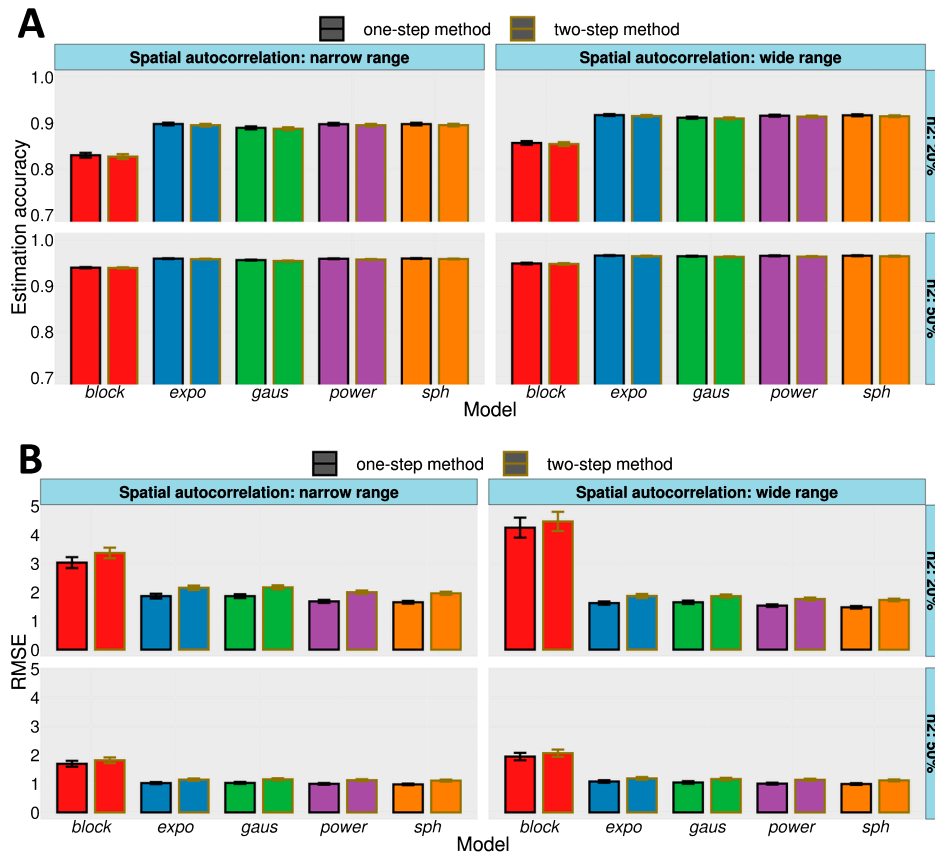


Fig. 4. Comparison of the estimation accuracy (A) and root mean square error (B) of genotypic values between the one-step and two-step methods. Here, the variogram model (exponential), replication number (four), and random noise (20%) were fixed. The estimation accuracy is the mean value of 50 simulations. Error bars show the standard error. RMSE: root mean square error, h^2 : heritability. *block*, block model; *expo*, exponential model; *gaus*, Gaussian model; *power*, power model; *sph*, spherical model.

from a field experiment with two blocks (A and B). In this experiment, two important traits of bioenergy sorghum, namely culm juice Brix and culm length, were examined. The kernel averaging approach was comparable to the single spatial kernel model for the estimation of genotypic values in the simulation study (Supplemental Fig. 6). Therefore, we employed the kernel averaging approach to investigate the sorghum data. We calculated the correlation coefficients between true phenotypic values from real data and predicted phenotypic values from five CV rounds to evaluate the spatial kernel models. The spatial kernel models explained the phenotypic variations more accurately than the *block* model did (Fig. 6A). Although the spatial kernel models except *sph* exhibited similar correlation coefficients, the *gaus* model was the optimal model for both Brix and culm length. The spatial effects estimated by the *gaus* model are depicted in Fig. 6B. For both traits, the spatial effects changed continuously along the X-axis over the field. In particular, high field heterogeneity was observed within block B. Manhattan plots were used to present marker variance estimated by the *block* and *gaus* models (Fig. 6C). No distinct biases between the two models were observed in terms of the estimation of marker variance, although the estimates were slightly different (Fig. 6D).

Discussion

Field heterogeneity can occur due to various natural factors and artifacts, creating spatial variations. Eccleston and Chan (1998) proposed a model to detect field heterogeneity, while considering several sources of spatial variations, such as the effects of rows and columns as well as other smooth trends. In fact, appropriate models for individual trials must be evaluated due to the complexity of field heterogeneity. Stefanova *et al.* (2009) emphasized that diagnostic investigations are necessary to select appropriate models.

In the present study, four spatial kernel models (*expo*, *gaus*, *power*, and *sph*) were developed and their performances were examined. These spatial kernel models were always superior to the classical *block* model in the simulation studies. Although specific factors in agronomical field experiments greatly affected the estimation accuracy of genotypic values, the robustness of the developed spatial kernel models was evident, specifically under unfavorable conditions (low heritability and few replications) (Fig. 2). In the spatial kernel models, a decrease in the estimation accuracy due to missing plots was also suppressed to some extent, particularly under a wide autocorrelation range

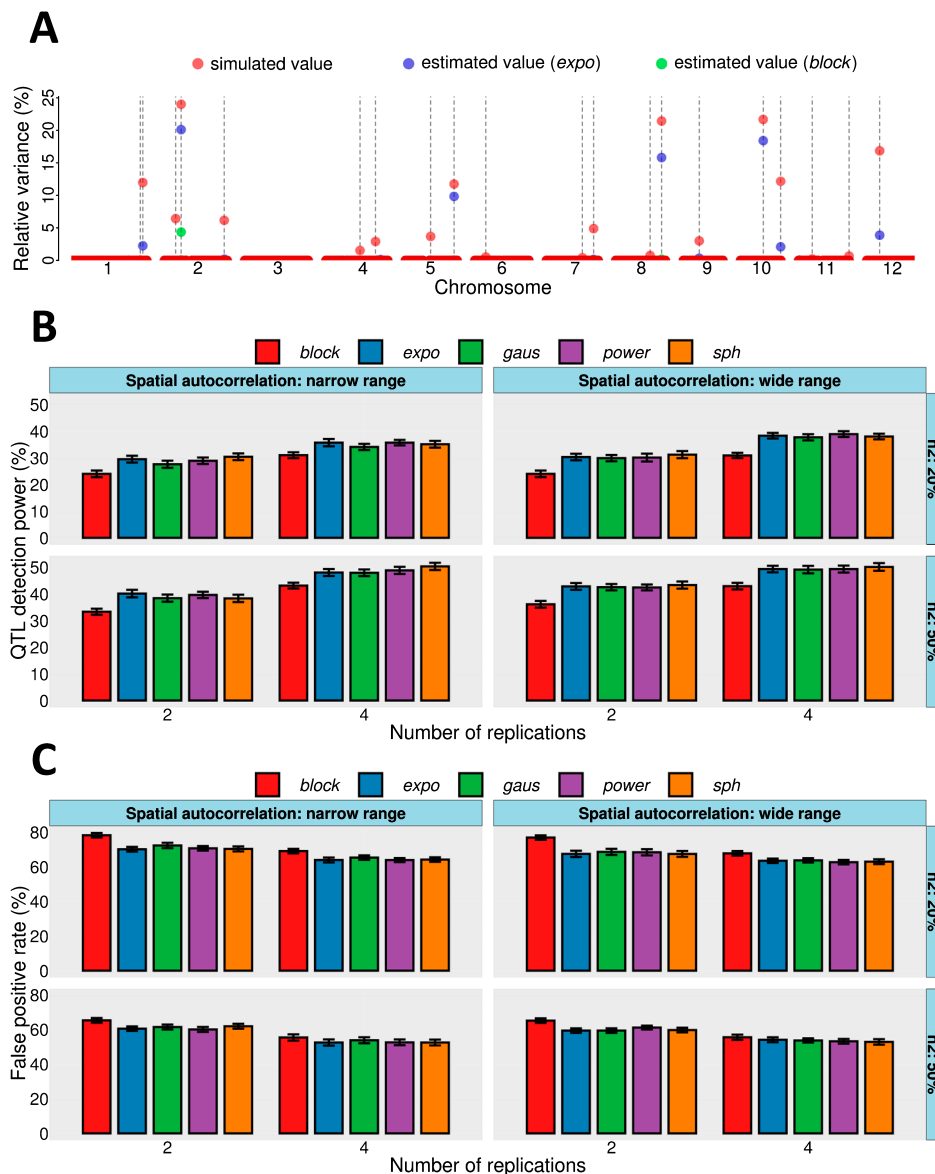


Fig. 5. QTL detection power of the one-step method. (A) Estimation of marker variances in a simulation study. QTLs were randomly located on 20 markers across chromosomes (vertical dashed lines). Simulated (red circles) and estimated (blue, exponential model; green, block model) marker variances are shown as the relative values to total genetic variance. (B) QTL detection power of the one-step method. (C) False-positive rate for the QTL study. The QTL detection power and false-positive rate are the mean values of 50 simulations. Error bars show the standard error. h^2 : heritability. *block*, block model; *expo*, exponential model; *gaus*, Gaussian model; *power*, power model; *sph*, spherical model.

(Fig. 3). Therefore, the developed spatial kernel models appear to be more advantageous in the presence of high field heterogeneity within a block. Such a situation is often observed in fields with large blocks or block misalignments (Stroup *et al.* 1994). In plant evaluation trials with numerous genotypes (varieties/lines), an experimental design with no replications might be schemed (Wu *et al.* 2013). The developed spatial kernel models can capture field heterogeneity even under these conditions, which cannot be simply solved using the classical *block* model.

We simulated field heterogeneity based on three variogram models: exponential, Gaussian, and spherical (Fig. 1, Supplemental Fig. 1). These variogram models correspond

to the covariance function of three spatial kernel models: *expo*, *gaus*, and *sph*. Nevertheless, we did not observe the advantage of corresponding spatial kernel models (e.g., the *expo* kernel models under simulations based on the exponential variogram model). In addition to variogram models used in simulation studies, the degrees of heritability and spatial variance might influence the superiority of spatial kernel models in different ways (Elias *et al.* 2018).

For the missing plots, spatial kernel models also had the advantage over the *block* model (Fig. 3). The effect of missing plots was not only due to the decrease in plot number (i.e., sample size) but also other factors (Supplemental Fig. 4). Field heterogeneity might increase in a large field

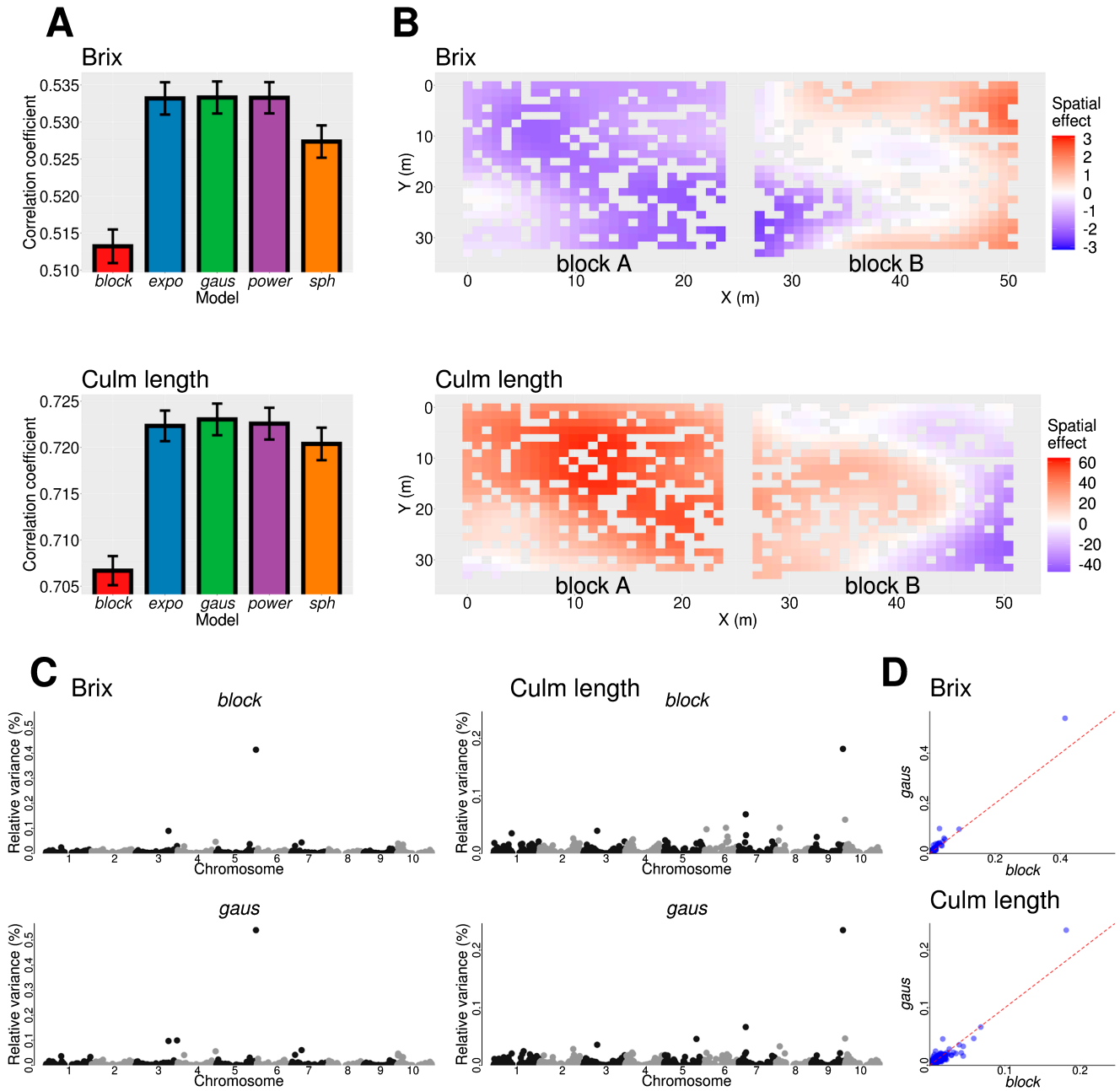


Fig. 6. Spatial analysis of experimental field data of bioenergy sorghum. (A) Decision of the optimal model based on five-fold cross-validation (5-CV). The mean values of 20 simulations are shown with the standard errors. The optimal model was *gaus* for both Brix and culm length, respectively. (B) The estimated spatial effects with the *gaus* model. Missing plots are shown by the background color. (C) Manhattan plots of estimated relative marker variances. Marker variances estimated by the *block* (upper panels) and *gaus* (lower panels) models are shown. (D) Comparison of estimated marker variances between the *block* (X-axis) and *gaus* (Y-axis) models. *block*, block model; *expo*, exponential model; *gaus*, Gaussian model; *power*, power model; *sph*, spherical model.

with missing plots in comparison with that in a small field with no missing plots even if the same sample size is maintained. Additionally, the replication number may be unbalanced for a genotype in fields with missing plots. Our results show the potential of the spatial kernel models in experiments with many missing plots.

Spatial analysis has been incorporated into GS indirectly (Bernal-Vasquez *et al.* 2014) as well as directly (Elias *et al.*

2018). Our results showed that the differences between the one-step (direct) and two-step (indirect) methods were small for the estimation accuracy of genotypic values based on the correlation coefficient in field simulations (Fig. 4A). The estimated genotypic values from the two-step methods, however, had larger RMSE than the estimated values of the one-step methods due to the shrinkage effect (Fig. 4B, Supplemental Fig. 5). Nevertheless, the spatial kernel models

always outperformed the *block* model across the one-step and two-step methods, supporting the use of spatial analysis for GS.

In the two-step method, large records derived from multi-environmental trials (METs) can be integrated into a single dataset (Bernal-Vasquez *et al.* 2014). However, the influence of shrinkage effects due to two-step methods needs to be considered for METs. In contrast, the one-step method cannot be easily implemented in METs. However, Elias *et al.* (2018) proposed the use of this model as the first analytical step of METs. The GS model proposed by Lopez-Cruz *et al.* (2015) estimates the genotype–environment interaction and can be theoretically extended with spatial kernels, which may be an alternative for applying the one-step method in METs.

We applied the spatial GS models for direct QTL mapping (Fig. 5). The spatial kernel models outperformed the *block* model in terms of QTL detection power and false-positive rate, representing the effectiveness of spatial analysis in QTL mapping. Ho *et al.* (2002) incorporated spatial analysis into QTL mapping to reduce the effects of field heterogeneity. In QTL mapping, an experimental design with no or few replications is often employed to increase the number of lines, because the sample size directly affects QTL detection power. However, the detection power might be lowered in the presence of high field heterogeneity. To retain high QTL detection power in heterogeneous fields, spatial analysis should be incorporated into QTL mapping.

Finally, we validated the performance of the spatial kernel models using real data of bioenergy sorghum (Ishimori *et al.* 2020). We focused on the following three questions: (1) can the kernel averaging approach explain phenotypic variations better than the *block* model?; (2) can real field heterogeneity be successfully captured by spatial kernel models?; (3) are the results based on spatial kernel models (e.g., GWAS) different from the results based on the *block* model?

In the real sorghum dataset, the kernel averaging approach could predict phenotypic variations of both tested traits more accurately than the *block* model (Fig. 6A). We expected that the spatial kernel models would be advantageous when field heterogeneity was high within each block. The estimated spatial effects suggested that field heterogeneity was mainly spread between the columns (i.e., along X-axis), with both decreasing (culm length) and increasing (Brix) trends, particularly in block B (Fig. 6B). Anisotropic bias along rows or columns is often caused by extraneous variations (Gilmour *et al.* 1997). Therefore, irrigation irregularity may have been larger within rows than between rows, although other extraneous stresses might have also contributed to heterogeneity. However, the distribution of the estimated spatial effects seemed to be different from the normal distribution that we expected in this study (Fig. 6B). If extraneous variations specific to a field were expected, reasonable explanatory variables might be incorporated into statistical models (e.g., the effect of columns for irrigation

irregularity in this study).

The effects of spatial modeling on QTL mapping were not distinct in our real dataset (Fig. 6C, 6D). Unfortunately, we could not verify the results because true QTLs in the real dataset were unknown. Nevertheless, spatial analysis can be used in trial designs without replications (Ho *et al.* 2002). To offer precise insight into the usefulness of spatial kernel models, additional studies on real datasets are required. Furthermore, the influence of the range of the hyperparameter in the kernel averaging approach should be tested with each real dataset. Although the kernel averaging approach might be an alternative to the optimization of the hyperparameter (Supplemental Fig. 6), a reasonable range for the hyperparameter needs to be defined (Pérez and de los Campos 2014).

Spatial analysis is particularly crucial for ensuring the precision of selection in plant breeding trials (Sripathi *et al.* 2017). However, the appropriateness of spatial models should be carefully verified, because they can bias the result (Stefanova *et al.* 2009). Nevertheless, differences in accuracy among the four spatial kernel models (*expo*, *gaus*, *power*, and *sph*) were not discussed in detail in the present study. The appropriate spatial model may differ with the extent of field heterogeneity in each trial (Richter and Kroschewski 2012). Therefore, multiple spatial models should be examined for each field trial. To compare parameters estimated from different spatial kernels directly, the proposed method by Legarra may be necessary (Legarra 2016). An appropriate spatial model is anticipated to help understand the causes of field heterogeneity which could otherwise lead to inaccurate interpretations of agronomical trials.

Author Contribution Statement

MI, TT, NT, and HI designed the study; MI performed the simulation study; MI, HT, MF, HK-K, JY, NT, and HI contributed to the acquisition of the real sorghum dataset; MI analyzed the real sorghum dataset; MI interpreted results and drafted the manuscript; NT and HI supervised the study. All authors read and approved the final manuscript.

Acknowledgments

The research was funded by CREST, Japan Science and Technology Agency, Japan (research subject: “Novel techniques of tailor-made breeding for energy crop improvement using high-throughput genotyping”) and partially supported by Japan Society for the Promotion of Science KAKENHI (Grant 17H01457). The authors appreciate the technical assistance of the members of the Laboratory of Biometry and Bioinformatics and the Laboratory of Plant Molecular Genetics at the University of Tokyo.

Literature Cited

- Bernal-Vasquez, A.-M., J. Möhring, M. Schmidt, M. Schönleben, C.-C. Schön and H.-P. Piepho (2014) The importance of phenotypic data analysis for genomic prediction—a case study comparing different spatial models in rye. *BMC Genomics* 15: 646.
- Browning, S.R. and B.L. Browning (2007) Rapid and accurate haplotype phasing and missing-data inference for whole-genome association studies by use of localized haplotype clustering. *Am. J. Hum. Genet.* 81: 1084–1097.
- Cullis, B.R. and A.C. Gleeson (1991) Spatial analysis of field experiments—an extension to two dimensions. *Biometrics* 47: 1449–1460.
- de los Campos, G., D. Gianola, G.J.M. Rosa, K.A. Weigel and J. Crossa (2010) Semi-parametric genomic-enabled prediction of genetic values using reproducing kernel Hilbert spaces methods. *Genet. Res.* 92: 295–308.
- Dilla-Ermita, C.J., E. Tandayu, V.M. Juanillas, J. Detras, D.N. Lozada, M.S. Dwiyantri, C.V. Cruz, E.G.N. Mbanjo, E. Ardales, M.G. Diaz *et al.* (2017) Genome-wide association analysis tracks bacterial leaf blight resistance loci in rice diverse germplasm. *Rice (N Y)* 10: 8.
- Duarte, J.B. and R. Vencovsky (2005) Spatial statistical analysis and selection of genotypes in plant breeding. *Psequi. Agropecu. Bras.* 40: 107–114.
- Eccleston, J. and B. Chan (1998) Design algorithms for correlated Data. *In: Payne, R. and P. Green (eds.) COMPSTAT98 Proceedings in Computational Statistics*, Springer-Verlag, Heidelberg, pp. 41–52.
- Elias, A.A., I. Rabbi, P. Kulakow and J.-L. Jannink (2018) Improving genomic prediction in cassava field experiments using spatial analysis. *G3 (Bethesda)* 8: 53–62.
- Fisher, R.A. (1935) *The design of experiments*. Oliver and Boyd, Edinburgh, p. 252.
- Gilmour, A.R., B.R. Cullis and A.P. Verbyla (1997) Accounting for natural and extraneous variation in the analysis of field experiments. *J. Agric. Biol. Environ. Stat.* 2: 269–293.
- Gräler, B., E. Pebesma and G. Heuvelink (2016) Spatio-temporal interpolation using gstat. *R J.* 8: 204–218.
- Gusmão, L. (1986) Inadequacy of blocking in cultivar yield trials. *Theor. Appl. Genet.* 72: 98–104.
- Ho, J.C., S.R. McCouch and M.E. Smith (2002) Improvement of hybrid yield by advanced backcross QTL analysis in elite maize. *Theor. Appl. Genet.* 105: 440–448.
- Ishimori, M., T. Hattori, K. Yamazaki, H. Takanashi, M. Fujimoto, H. Kajiya-Kanegae, J. Yoneda, T. Tokunaga, T. Fujiwara, N. Tsutsumi *et al.* (2020) Impacts of dominance effects on genomic prediction of sorghum hybrid performance. *Breed. Sci.* 70: 605–616.
- Lado, B., I. Matus, A. Rodríguez, L. Inostroza, J. Poland, F. Belzile, A. del Pozo, M. Quincke, M. Castro and J. von Zitzewitz (2013) Increased genomic prediction accuracy in wheat breeding through spatial adjustment of field trial data. *G3 (Bethesda)* 3: 2105–2114.
- Legarra, A. (2016) Comparing estimates of genetic variance across different relationship models. *Theor. Popul. Biol.* 107: 26–30.
- Lopez-Cruz, M., J. Crossa, D. Bonnett, S. Dreisigacker, J. Poland, J.-L. Jannink, R.P. Singh, E. Autrique and G. de los Campos (2015) Increased prediction accuracy in wheat breeding trials using a marker \times environment interaction genomic selection model. *G3 (Bethesda)* 5: 569–582.
- Meuwissen, T.H.E., B.J. Hayes and M.E. Goddard (2001) Prediction of total genetic value using genome-wide dense marker maps. *Genetics* 157: 1819–1829.
- Pérez, P. and G. de los Campos (2014) Genome-wide regression and prediction with the BGLR statistical package. *Genetics* 198: 483–495.
- Piepho, H.-P., J. Möhring, M. Pflugfelder, W. Hermann and E.R. Williams (2015) Problems in parameter estimation for power and AR(1) models of spatial correlation in designed field experiments. *Commun. Biometry Crop Sci.* 10: 3–16.
- R Core Team (2019) *R: A language and environment for statistical computing*. R Foundation for Statistical Computing, Vienna, Austria.
- Richter, C. and B. Kroschewski (2012) Geostatistical models in agricultural field experiments: investigations based on uniformity trials. *Agron. J.* 104: 91–105.
- Sripathi, R., P. Conaghan, D. Grogan and M.D. Casler (2017) Spatial variability effects on precision and power of forage yield estimation. *Crop Sci.* 57: 1383–1393.
- Stefanova, K.T., A.B. Smith and B.R. Cullis (2009) Enhanced diagnostics for the spatial analysis of field trials. *J. Agric. Biol. Environ. Stat.* 14: 392–410.
- Stroup, W.W., P.S. Baenziger and D.K. Mulitze (1994) Removing spatial variation from wheat yield trials: a comparison of methods. *Crop Sci.* 34: 62–66.
- Vieira, S.R., J.L. Hatfield, D.R. Nielsen and J.W. Biggar (1983) Geostatistical theory and application to variability of some agronomical properties. *Hilgardia* 51: 1–75.
- Wu, J., K. Bondalapati, K. Glover, W. Berzonsky, J.N. Jenkins and J.C. McCarty (2013) Genetic analysis without replications: model evaluation and application in spring wheat. *Euphytica* 190: 447–458.
- Wu, T., D.E. Mather and P. Dutilleul (1998) Application of geostatistical and neighbor analyses to data from plant breeding trials. *Crop Sci.* 38: 1545–1553.