frontiers

FieldSimR: An R package for simulating plot data in multi-environment field trials

Christian R. Werner [1,2*], **Dorcus C. Gemenet** [1,2] **and Daniel J. Tolhurst** [3*]

[1] *Accelerated Breeding Initiative (ABI), Consultative Group of International Agricultural Research (CGIAR), Texcoco, Mexico*
[2] *International Maize and Wheat Improvement Center (CIMMYT), Texcoco, Mexico*
[3] *The Roslin Institute and Royal (Dick) School of Veterinary Studies, University of Edinburgh, Easter Bush, Midlothian, United Kingdom*

Correspondence*:
Christian R. Werner
c.werner@cgiar.org

Daniel J. Tolhurst
uow.tolhurst@gmail.com

Language style note: the authors prefer the article to be formatted in British English.

ABSTRACT

This paper presents a general framework for simulating plot data in multi-environment field trials with one or more traits. The framework is embedded within the R package FieldSimR, whose core function generates plot errors that capture global field trend, local plot variation, and extraneous variation at a user-defined ratio. FieldSimR's capacity to simulate realistic plot data makes it a flexible and powerful tool for a wide range of improvement processes in plant breeding, such as the optimization of experimental designs and statistical analyses of multi-environment field trials. Therefore, FieldSimR provides crucial functionality that is currently missing in other software for simulating plant breeding programmes. The paper includes an example simulation of a field trial to evaluate a set of 100 maize hybrids for two traits across three environments. To demonstrate FieldSimR's value as an optimisation tool, the simulated data set is then used to compare eight spatial models for their capacity to accurately predict the maize hybrids' genetic values and to reliably estimate the variance parameters of interest.

Keywords: Simulation, Spatial variation, Plot error, Multi-environment field trials

1 INTRODUCTION

This paper presents a general framework for simulating plot data in multi-environment field trials with one or more traits. The framework is embedded within the R package FieldSimR, whose core function generates plot errors that capture global field trend, local plot variation, and extraneous variation. FieldSimR's capacity to simulate realistic plot data makes it well-suited to a wide range of improvement processes in plant breeding, such as the optimization of experimental designs and statistical analyses of multi-environment field trials. It is also well-suited to a range of education purposes, such as teaching the principals of spatial modelling and analysing multi-environment trial data.

Plant breeding programmes continuously evaluate, select and release improved genotypes in order to meet the complex and dynamic requirements of different customer groups, including farmers, processors and end-users (Covarrubias-Pazaran et al., 2022). The resources required to compare different improvement strategies in the field, however, can quickly exceed the practical possibilities of a plant breeding programme. Often, multiple factors must be evaluated simultaneously over several years or even decades to identify an optimised breeding strategy. This requires a pragmatic approach to identify profitable long-term strategies in plant breeding programmes.

Simulation is a fast and cost-efficient tool for comparing different breeding strategies over time (Gaynor et al., 2021). Interestingly, this is not a new concept. Simulations have been utilised by plant and animal breeders for almost a century, beginning with the application of the Breeder's equation (Lush, 1937), a form of deterministic simulation to predict genetic gain based on selection intensity, selection accuracy, genetic variance, and generation interval. However, only recently, with the availability of modern computers and flexible software have breeders and researchers been granted access to more powerful stochastic simulation for optimising entire breeding programmes across multiple generations. Currently available software includes QU-GENE (Podlich and Cooper, 1998), ADAM-plant (Liu et al., 2019), and ChromaX (Younis et al., 2023), as well as the R packages Selection Tools (Frisch, 2023), and AlphaSimR (Gaynor et al., 2021). These software applications can be used, for example, to compare different crossing and selection strategies over time. They lack, however, the functionality to simulate realistic plot data in multi-environment field trials. This capacity is necessary to evaluate the impact of different experimental designs, multi-environment testing strategies, and statistical analyses on the performance of a breeding programme.

The motivation to simulate realistic plot data has stemmed from the importance of spatial variation in plant breeding field trials (see, for example, Wilkinson et al., 1983; Besag and Kempton, 1986; Cullis and Gleeson, 1991; Rodríguez-Álvarez et al., 2018; Piepho et al., 2022). Spatial variation occurs naturally in field trials laid out as a two-dimensional lattice of plots (Gogel et al., 2023), and can account for more than 50% of the total phenotypic variation. Spatial variation can be broadly categorised as either global trend, local variation or extraneous variation (Gilmour et al., 1997). Global trend occurs on a large scale across the field, such as large scale moisture and fertility gradients (Green et al., 1985). Local variation occurs on a small scale between neighbouring plots. It may reflect small scale changes in soil composition (trend) or random error (noise), such as measurement error and within-plot variability (Besag, 1977). Conversely, extraneous variation is predominately induced during the conduct of the trial, and as a result is often aligned with the columns and rows. It may reflect management practices, such as serpentine harvesting and spraying, multi-plot seeders that sow multiple plots simultaneously, or inaccurate trimming resulting in unequal plot lengths (Stefanova et al., 2009). The complexity and importance of spatial variation dictate the need for a framework to simulate plot errors that capture the main components of spatial variation.

FieldSimR is an R package for simulating plot errors in multi-environment field trials that comprise global and local trend, random error, and extraneous variation. It also provides compatibility with AlphaSimR to simulate plot phenotypes. This makes FieldSimR a powerful tool for a wide range of improvement processes, such as:

- Comparing spatial modelling approaches, e.g., separable autoregressive processes, tensor product penalised splines, and nearest neighbour adjustments.
- Comparing experimental designs for single- and multi-environment studies, e.g., fully-replicated designs, p-rep designs and sparse testing designs.

- Comparing approaches for analysing multi-environment trial data, e.g., random regressions and factor analytic models.

The paper is arranged as follows. The Methods section presents the theoretical framework for simulating plot errors, which are constructed by combining spatial error, random error, and extraneous error components at a user-defined ratio. The Results and Discussion section introduces an example simulation of a field trial to evaluate a set of 100 maize hybrids for two traits across three environments. To demonstrate FieldSimR's value as an optimisation tool, the simulated data set is then used to compare eight spatial models for their capacity to accurately predict the maize hybrids' genetic values and reliably estimate the variance parameters of interest.

2 METHODS

This section presents the framework in FieldSimR for simulating plot errors in multi-environment field trials. FieldSimR generates plot errors by combining spatial error, random error and extraneous error components at a user-defined ratio. The simulation framework is initially developed for a single trait and then generalised for multiple traits.

2.1 Framework for simulating sinlge-trait plot errors in multi-environment field trials

Assume a single-trait multi-environment trial (MET) dataset comprises p environments with n plots in total, where $n = \sum_{j=1}^{p} n_j$ and n_j is the number of plots in environment j. Also assume that each environment is laid out as a two-dimensional lattice of plots such that $n_j = c_j \times r_j$, where c_j and r_j are the number of columns and rows, respectively. The n-vector of plot errors is then given by $\boldsymbol{\varepsilon} = \left(\boldsymbol{\varepsilon}_1^\top, \ldots, \boldsymbol{\varepsilon}_p^\top\right)^\top$, where $\boldsymbol{\varepsilon}_j$ is the n_j-vector of plot errors for environment j (ordered as rows within columns). The vector $\boldsymbol{\varepsilon}_j$ comprises the main components of spatial variation, i.e. global and local trend, random error, and extraneous variation.

FieldSimR constructs the vector of plot errors as the sum of three terms:

$$\boldsymbol{\varepsilon}_j = \mathbf{s}_j + \mathbf{r}_j + \mathbf{e}_j, \tag{1}$$

where $\mathbf{s}_j$ is a vector of errors that capture spatial trend, $\mathbf{r}_j$ is a vector of random errors, and $\mathbf{e}_j$ is a vector of errors that capture extraneous variation. The errors in $\mathbf{s}_j$ and $\mathbf{e}_j$ are hereafter referred to as the spatial and extraneous errors, respectively. All terms are simulated as mutually independent with zero means and variance components given by $\sigma^2_{s_j}$, $\sigma^2_{r_j}$, and $\sigma^2_{e_j}$, respectively. The total plot error variance is then given by $\sigma^2_{\varepsilon_j} = \sigma^2_{s_j} + \sigma^2_{r_j} + \sigma^2_{e_j}$.

2.1.1 Spatial error

The errors in $\mathbf{s}_j$ capture both global and local trend, such as large scale fertility gradients (Green et al., 1985) and small scale changes in soil composition (Gilmour et al., 1997). The vector $\mathbf{s}_j$ is generated in FieldSimR using either bivariate interpolation (Akima, 1978) or a separable autoregressive process (Box and Jenkins, 1970).

Bivariate interpolation is implemented through the `interp` function in the R package interp (Gebhardt et al., 2023), which applies piece-wise linear interpolation across the two-dimensional lattice of plots. An example field array with spatial trend generated using bivariate interpolation is presented in Figure 1. The field array comprises $c_j = 10$ columns and $r_j = 20$ rows for $n_j = 200$ plots in total. The field spans

80 m long in the column direction and 40 m wide in the row direction, with rectangular plots 8 m long by 2 m wide (Figure 1a). There are two square blocks aligned in the column direction ("side-by-side"), with 100 plots in each block. Four interpolation (knot) points are placed outside the four corners of the field, which prevents continuity issues that occur at the interpolation boundary. The z-values at these points were sampled from a standard normal distribution, with $z = 2.56, 1.08, 0.43$, and -2.56 for this example (clockwise from top left). The continuous array between the knot points is then interpolated, which produces a smooth continuous surface across the lattice of plots (Figure 1b). A single error value is assigned to each plot by averaging over the continuous surface within each plot (Figure 1c). The error values are then scaled to the defined spatial error variance for each environment, $\sigma^2_{s_j}$. This produces the vector of spatial errors, $\mathbf{s}_j$.

The complexity of spatial trend can be controlled in FieldSimR by setting the number of additional knot points sampled inside the field array. By altering the complexity, users can explicitly change the ratio of global to local trend. The example in Figure 1 has no additional knot points besides those at the four corners, so the simulated spatial error predominately captures global trend with minimal or no local trend. Three additional examples are presented in Figure 2, which have 5, 10, and 50 knot points, respectively. The knot points are sampled from a continuous uniform distribution defined by all points in the continuous array. This means that more than one knot point can be sampled for each plot. The position of the knot points and corresponding z-values are presented in Supplementary Figure S1, which displays the smooth continuous surface for the examples in Figure 2.

The examples demonstrate FieldSimR's capacity to generate global and local trend, as well as within-plot variability. Increasing the complexity will generate more local trend relative to global trend, up to a point where the errors capture minimal or no trend (i.e., only noise). At this point, numerous knot points may be sampled for each plot which may further increase the amount of within-plot variability. By default, FieldSimR sets the number of knot points to half the maximum of the number of columns and rows. For example, the default complexity for a field trial with 20 columns and 10 rows is given by `max(20,10)/2 = 10` knot points (see, for example, Figure 2b). This generally provides a good ratio of global to local trend, but users are encouraged to alter the complexity as required.

Trellis plots for the three examples are presented in Supplementary Figure S2. These plots also demonstrate that various ratios of global to local trend can be generated by altering the `complexity`. For example, the first plot demonstrates a gradual decrease in spatial error as the row number increases, which is a classical sign of global trend in field trials. In contrast, the last plot demonstrates more small-scale fluctuations between neighbouring columns and rows, which is a sign of local trend. It is important to note that bivariate interpolation is a smoothing function, rather than a stochastic process, so the errors are not simulated as random variables.

The separable autoregressive process simulates spatial errors as random variables based on a stochastic variance matrix. Separable autoregressive processes explicitly model spatial dependence (correlation) between neighbouring plots, rather than a smooth continuous surface across the field. In this case, FieldSimR simulates the vector of spatial errors as:

$$\mathbf{s}_j \sim \mathrm{N}\big(\mathbf{0}, \sigma^2_{s_j}\mathbf{S}_j\big), \tag{2}$$

where $\sigma^2_{s_j}$ is the spatial error variance and $\mathbf{S}_j$ is the stochastic variance matrix, which is constructed as:

$$\mathbf{S}_j = \mathbf{\Sigma}_{\mathbf{c}_j}\big(\rho_{c_j}\big) \otimes \mathbf{\Sigma}_{\mathbf{r}_j}\big(\rho_{r_j}\big), \tag{3}$$

where ρ_{c_j} is the column autocorrelation parameter with $c_j \times c_j$ correlation matrix $\boldsymbol{\Sigma}_{\mathbf{c}_j}$ and ρ_{r_j} is the row autocorrelation parameter with $r_j \times r_j$ correlation matrix $\boldsymbol{\Sigma}_{\mathbf{r}_j}$. FieldSimR has the capacity to generate errors based on a separable first order autoregressive process (AR1). Note that, in contrast to bivariate interpolation, the autoregressive process is not based on plot dimensions, since they are implicitly modelled through ρ_{c_j} and ρ_{r_j} (see Gilmour et al., 1997). This approach allows users to implement estimates of ρ_{c_j} and ρ_{r_j} previously obtained from empirical analyses of field trial data.

The ratio of global to local trend can be controlled by altering the column and row autocorrelation parameters. Decreasing the autocorrelation parameters will effectively increase the complexity of the spatial trend, in the sense that more local trend will be generated relative to global trend, up to a point where the errors capture minimal or no trend (i.e., only noise). This occurs when the autocorrelation parameters are set to zero. Three examples are presented in Supplementary Figure S3, which show spatial trend generated using a separable first order autoregressive process with (a) $\rho_c = 0.7$ and $\rho_r = 0.9$, (b) $\rho_c = 0.5$ and $\rho_r = 0.7$, and (c) $\rho_c = 0.3$ and $\rho_r = 0.5$. The theoretical and sample variograms for these examples are presented in Supplementary Figure S4. These examples demonstrate the stochastic nature of the spatial errors generated based on autoregressive processes.

The methods above for generating global and local trend will be well-suited to most applications. However, some users may want to explicitly set the amount of global and local trend without fine-tuning the complexity or the autocorrelation parameters. In this case, users may simulate trend as the sum of two components, i.e., global trend (with no to low complexity) and local trend (with moderate to high complexity or low to moderate autocorrelations). This is left to the discretion of the user.

2.1.2 Random error

The errors in $\mathbf{r}_j$ capture local variation that is not trend, such as measurement error and intrinsic variability within the plots (Besag, 1977; Wilkinson et al., 1983). FieldSimR simulates the vector of random errors as:

$$\mathbf{r}_j \sim \mathrm{N}\big(\mathbf{0}, \sigma^2_{r_j}\mathbf{I}_{n_j}\big), \tag{4}$$

where $\sigma^2_{r_j}$ is the random error variance and $\mathbf{I}_{n_j}$ is an identity matrix of order n_j.

2.1.3 Extraneous error

The errors in $\mathbf{e}_j$ capture extraneous variation predominately induced during the conduct of the trial, such as serpentine harvesting or spraying and unequal plot dimensions (Gilmour et al., 1997; Stefanova et al., 2009). This type of variation is assumed to be aligned exclusively with the columns and rows of the trial. FieldSimR constructs the vector of extraneous errors as the sum of two terms:

$$\mathbf{e}_j = \mathbf{Z}_{\mathbf{c}_j}\mathbf{e}_{\mathbf{c}_j} + \mathbf{Z}_{\mathbf{r}_j}\mathbf{e}_{\mathbf{r}_j} \tag{5}$$

where $\mathbf{e}_{\mathbf{c}_j}$ is the vector of column errors with $n_j \times c_j$ design matrix $\mathbf{Z}_{\mathbf{c}_j}$ and $\mathbf{e}_{\mathbf{r}_j}$ is the vector of row errors with $n_j \times r_j$ design matrix $\mathbf{Z}_{\mathbf{r}_j}$. The design matrices are given by $\mathbf{Z}_{\mathbf{c}_j} = \mathbf{I}_{c_j} \otimes \mathbf{1}_{r_j}$ and $\mathbf{Z}_{\mathbf{r}_j} = \mathbf{1}_{c_j} \otimes \mathbf{I}_{r_j}$. The column and row errors are simulated as:

$$\mathbf{e}_{\mathbf{c}_j} \sim \mathrm{N}\big(\mathbf{0}, \sigma^2_{e_{c_j}}\mathbf{I}_{c_j}\big) \quad \text{and} \quad \mathbf{e}_{\mathbf{r}_j} \sim \mathrm{N}\big(\mathbf{0}, \sigma^2_{e_{r_j}}\mathbf{I}_{r_j}\big), \tag{6}$$

where $\sigma^2_{e_{c_j}}$ is the column error variance with identity matrix $\mathbf{I}_{c_j}$ and $\sigma^2_{e_{r_j}}$ is the row error variance with identity matrix $\mathbf{I}_{r_j}$. The column and row error variances are set based on whether column and/or row errors are simulated, such that $\sigma^2_{e_j} = \sigma^2_{e_{c_j}} + \sigma^2_{e_{r_j}}$.

FieldSimR has the capacity to generate extraneous errors based on zig-zag or random ordering across neighbouring columns and rows. The zig-zag ordering is achieved by alternating positive and negative values between neighbouring columns and rows. The two examples in Figure 3 demonstrate the two types of extraneous variation. The first example demonstrates a zig-zag pattern, with the extraneous errors in odd row numbers being consistently higher than those in even row numbers (mean of +0.37 compared to -0.37). This type of non-stationarity is a classical sign of extraneous variation attributed to systematic management practices, such as serpentine harvesting and spraying. The second example demonstrates a more stochastic pattern in which the errors may be attributed to random processes, such as inaccurate plot trimming resulting in unequal plot dimensions. Interested users may also manipulate the above functionality to simulate intraplot competition, typically observed as a negative correlation between neighbouring rows (Durban et al., 2001; Stringer et al., 2011).

2.1.4 Total error

FieldSimR constructs the total plot errors in Equation 1 by combining the spatial errors with the random and extraneous errors according to a user-defined ratio. The desired ratio is applied by setting the proportions of spatial error and extraneous error, with the remaining proportion assigned to random error. By default, FieldSimR sets the proportion of spatial error to 0.5 and extraneous error to 0, resulting in a random error proportion of 0.5.

2.2 Extension to multiple traits

FieldSimR has the capacity to simulate correlated plot errors across multiple traits. The correlation matrix between traits can be set for the spatial, random and extraneous errors, respectively. By default, FieldSimR fits a separable correlation structure between traits and environments (Bančič et al., 2023), but note that different error variances can be set for different environment-within-trait combinations. It is also important to note that when bivariate interpolation is used, the correlation matrix for the spatial error is applied to the z-values at the knot points, not the spatial errors themselves. This is because the spatial errors generated by bivariate interpolation do not have an assumed covariance structure.

3 RESULTS AND DISCUSSION

FieldSimR is an R package for simulating plot errors that comprise global and local trend, random error, and extraneous variation. This functionality makes FieldSimR a powerful tool for a wide range of improvement processes, such as the comparison of different spatial modelling approaches. This section demonstrates the simulation and analysis of a field trial which evaluates 100 maize hybrids for two traits across three environments. In the first part, FieldSimR is used to simulate multi-environment plot errors, genetic values and phenotypes for the 100 maize hybrids. In the second part, eight spatial models are compared for their ability to accurately predict the true genetic values of the simulated hybrids and to reliably estimate the true variance parameters of interest.

3.1 Simulation example

Consider a scenario in which 100 maize hybrids are evaluated for grain yield (t/ha) and plant height (cm) in a field trial across three environments. The simulation of the maize phenotypes with FieldSimR involves three steps:

1. Simulation of plot errors.
2. Simulation of genetic values.
3. Simulation of phenotypes by combining the plot errors with the genetic values.

3.1.1 Simulation of plot errors

Plot errors for grain yield and plant height were simulated assuming independence between traits and environments. Environments 1 and 2 comprised two blocks each, while environment 3 comprised three blocks. The blocks were aligned in the column direction ("side-by-side") and comprised 5 columns and 20 rows for 100 plots in each block. The plots were 8 m long in the column direction by 2 m wide in the row direction.

To obtain target heritabilities of $H^2 = 0.3$ for grain yield and $H^2 = 0.5$ for plant height in all three environments, the total error variances for the two traits were defined relative to their genetic variances as described in the Supplementary Script S10. The simulated plot errors comprised spatial error, random error, and extraneous error terms. The spatial error was simulated using bivariate interpolation with complexity set to 10 and proportion of spatial error variance set to 0.4 in all three environments. The extraneous error was simulated using zig-zag ordering across neighbouring rows. The proportion of extraneous error variance was set to 0.2 in all three environments. This resulted in a proportion of random error variance given by `1 - (0.4 + 0.2) = 0.4`.

```
error_df <- field_trial_error(n_envs = 3,
                              n_traits = 2,
                              n_blocks = c(2,2,3),
                              n_cols = c(10,10,15),
                              n_rows = 20,
                              block_dir = "col",
                              var_R = c(0.20, 0.28, 0.14, 15.1, 8.5, 11.7),
                              spatial_model = "Bivariate",
                              complexity = 10,
                              plot_length = 8,
                              plot_width = 2,
                              prop_spatial = 0.4,
                              prop_ext = 0.2,
                              ext_dir = "row",
                              ext_ord = "zig-zag",
                              return_effects = TRUE)
```

The simulated spatial errors, extraneous errors, random errors, and the total plot errors stored in `error_df` are presented in Figure 4.

3.1.2 Simulation of genetic values

Genetic values for grain yield and plant height across three environments were simulated based on an unstructured model for genotype-by-environment (GxE) interaction. The simulation was done in AlphaSimR (Gaynor et al., 2021), using FieldSimR's wrapper functions `unstr_asr_input()` and `unstr_asr_output()`. The R code can be found in Supplementary Script S10. The simulated genetic values can be directly accessed through the package's example data frame `df_gv_unstr`, which was used to simulate phenotypes, as described below. The simulated genetic values for trait 1 in environment 1 are presented in Figure 4.

In addition to the unstructured model, FieldSimR provides wrapper functions for simulating genetic values based on a compound symmetry model for GxE interaction. Alternatively, users can provide their own set of genetic values, e.g. through simulation or previously obtained from empirical analyses.

3.1.3 Simulation of phenotypes

Phenotypes for grain yield and plant height were simulated by combining the simulated plot errors with the genetic values stored in FieldSimR's example data frame `df_gv_unstr`. The genotypes were randomly allocated to plots according to a randomised complete block design (RCBD).

```
pheno_df <- make_phenotypes(gv_df = df_gv_unstr,
                            error_df = error_df$plot_df,
                            randomise = TRUE)
```

The phenotypes are presented together with the plot errors and genetic values in Figure 4. Note that FieldSimR does not provide functionality to generate experimental designs other than an RCBD. Users are encouraged to generate alternative experimental designs externally, e.g. with R packages such as agricolae (de Mendiburu, 2023), odw (Butler, 2021), and DiGGer (Coombes, 2020).

3.2 Comparison of spatial models

The comparison of spatial models is demonstrated using the simulated grain yield data in environment 1. A sequential approach was adopted for model fitting following Gilmour et al. (1997), with global trend and extraneous variation diagnosed using the sample variogram and accounted for using fixed and random model terms. This resulted in eight different spatial models, including a baseline model, three models with a separable first order autoregressive (AR1) process, two models with a tensor product penalized spline (TPS), and two models implementing nearest neighbour (NN) adjustments (Table 1). All models were fitted using ASReml-R (Butler et al., 2018) or SpATS (Rodríguez-Álvarez et al., 2018), and are summarised in Table 1.

The spatial models were evaluated in three ways (Table 2):

1. The prediction accuracy was calculated using Pearson's correlation coefficient (r) between the simulated true genetic values and the predicted values.
2. The model fit was assessed using the residual maximum likelihood ratio test (REMLRT) and the Akaike information criterion (AIC).
3. The reliability was calculated as the bias between the simulated true genetic variance parameter and the estimated parameter.

Note that the expected prediction accuracy for the data set is 0.68, based on the simulation parameters. Also note that the REMLRT is based on the positive variance parameter approach of Stram and Lee (1994) and the AIC is based on the full log-likelihood approach of Verbyla (2019), which can compare models with different fixed effects. Typical experimental design and data checks were performed prior to model fitting (Supplementary Figure S5).

3.2.1 Baseline model

The analyses commenced by fitting a baseline linear mixed model, which included random genotype and block effects and an independent (ID) error model (Table 1). This model reflects a classical complete block analysis that assumes independent genotypes, blocks and residuals. The estimated genetic variance was $\hat{\sigma}^2_g = 0.02$, which was substantially lower than the true value of 0.09 (bias = 0.07; Tables 2 and 3). The accuracy of the baseline model was also lower than the expected accuracy ($r = 0.65$ compared to 0.68; Table 2).

3.2.2 Separable first order autoregressive processes

The sequence of models 1-3 comprises three variants of a model that implemented a random genotype effect, block effects and a separable first order autoregressive (AR1) process (Table 1). The separable AR1 process represents a random process which assumes correlated residuals in two dimensions, i.e., in column and row direction (Martin, 1990; Cullis and Gleeson, 1991).

Model 1 did not include the ID error model. The estimated genetic variance was 0.07, which provided a much better estimate of the true value than the baseline model (bias = 0.02; Tables 2 and 3). Model 1 also provided a significantly better fit than the baseline model in terms of LRT ($p < 0.0001$) and AIC (-94.4 compared to -66.7), and a substantially higher prediction accuracy (0.72), as shown in Table 2. The estimated column and row autocorrelations were $\hat{\rho}_c = 0.51$ and $\hat{\rho}_r = 0.23$ (Table 3).

Model 2 reintroduced the ID term, which acted as an uncorrelated random error component (Besag, 1977). It provided a significantly better fit than Model 1, and was also more accurate (Table 2). The estimated column and row autocorrelations were $\hat{\rho}_c = 0.95$ and $\hat{\rho}_r = 0.87$, which were substantially higher than in Model 1 (Table 3). This indicates that the AR1 process captured (highly correlated) spatial trend, while the ID term captured the remaining random error. The sample variogram in Figure 5a shows a zig-zag pattern between neighbouring rows, with consistently higher semivariances for odd displacements compared to even displacements (also see Supplementary Figure S6). The sill of the variogram shows that the semivariances do not fall within the sample quantiles (Figure 5c). This is a classical sign of extraneous variation attributed to systematic practices, which matches the extraneous error simulated in this dataset.

Model 3 fitted a fixed and random row term to model this extraneous variation, following Gilmour et al. (1997). The fixed term was coded as 1 for odd row numbers and 2 for even row numbers (Stefanova et al., 2009). The significance of the fixed term was assessed using a Wald F-test with denominator degrees of freedom ($p < 0.001$; Kenward and Rogers, 1997). The estimated variances in Model 3 were $\hat{\sigma}^2_s = 0.09$, $\hat{\sigma}^2_r = 0.08$, and $\hat{\sigma}^2_{e_r} = 0.05$, which closely matched the true values (Table 3). The estimated column autocorrelation decreased to $\hat{\rho}_c = 0.68$ compared to Model 2. Model 3 provided a significantly better fit than Model 2, and was also more accurate (Table 2). The sample variogram in Figure 5b no longer shows a zig-zag pattern. Instead, a discontinuity is shown at 0 displacement, reflecting the random error variance, followed by a gradual incline in the column direction. This type of non-stationarity is a sign of global trend in the column direction, which matches the spatial error simulated in this dataset. However, the sill of the variogram shows that the semivariances fall within the sample quantiles (Figure 5d). The observed

non-stationarity is, therefore, an artefact of the correlated AR1 process, rather than global trend requiring further remediation.

3.2.3 Tensor product penalised splines

The sequence of models 4-5 used a tensor product penalised spline (TPS), fitted using the SpATS package in R (Rodríguez-Álvarez et al., 2018). A cubic B-spline basis was used with 6 knots in the column direction and 12 knots in the row direction, following Velazco et al. (2017). The TPS included fixed column, row, and interaction terms as well as five random spline components.

Model 4 was more accurate and showed a better fit than the baseline model (Table 2), but was less accurate than any of three models implementing the separable AR1 process. Like for Model 2, the sample variogram shows a zig-zag patterns (Supplementary Figure S7), indicating that a better model fit could be obtained by including random row terms.

Model 5 is an extension of Model 4 to include random column and row terms. This model is equivalent to the SpATS approch of Rodríguez-Álvarez et al. (2018). Model 5 was more accurate and provided a significantly better fit than Model 4 (Table 2). It was, however, less accurate than the best model using the separable AR1 process (Model 3), despite having five more model parameters (Table 1). The sample variogram of Model 5 no longer shows a zig-zag pattern (Supplementary Figure S8).

3.2.4 Nearest neighbour adjustments

The sequence of models 6-7 implemented nearest neighbour (NN) adjustments to the phenotypes (Papadakis, 1937; Bartlett, 1978). The adjustments were obtained by averaging over neighbouring plots using the mvngGrAd package in R (Technow, 2015). The grids used for Models 6 and 7 are shown in Supplementary Figure S9. Both models were fitted in ASReml-R, with model terms equivalent to the baseline model (Table 1). Model 6 was more accurate than Model 7, but both models were less accurate than Model 3 (Table 2) and proved insufficient to effectively capture local trend. The estimated variances ranged from $\sigma^2_g = 0.06$ to 0.08 and $\sigma^2_r = 0.14$ to 0.17 (Table 3). Note that the model fit criteria in Table 2 cannot compare models with different (adjusted) phenotypes, so that the final model was selected based on the ratio of genetic to total phenotypic variance.

4 CONCLUDING REMARKS

FieldSimR's capacity to simulate realistic plot errors that capture global field trend, local plot variation, and extraneous variation makes it a flexible and powerful tool for various improvement processes in plant breeding. In contrast to real-world experimental data, FieldSimR enables the efficient and comprehensive assessment of trial designs and analysis models on a large scale, across an extensive array of scenarios. Furthermore, it allows for an unbiased comparison of designs and models for their capacity to generate accurate predictions of genetic values and to reliably estimate variance parameters of interest, as the true values are defined by the user and, therefore, are known.

FieldSimR has been extensively deployed as part of the Excellence in Breeding (EiB) initiative to provide guidance on the improvement of field trial design and analysis strategies across numerous CGIAR breeding programmes.

Table 1. Linear mixed models fitted to the simulated maize breeding dataset, Part 1: Summary of model terms.

		Fixed			Random					Residual	
Model	Terms	Col	Row	Col:Row	Hybrid	Block	Col	Row	Spline	AR1	ID
baseline	ID				✓	✓					✓
1	AR1				✓	✓				✓	
2	AR1 + ID				✓	✓				✓	✓
3	AR1 + ID + Frow + Rrow		✓		✓	✓		✓		✓	✓
4	TPS + ID	✓	✓	✓	✓	✓			✓		✓
5	TPS + ID + Rcol + Rrow	✓	✓	✓	✓	✓	✓	✓	✓		✓
6	NN_0 + ID				✓	✓					✓
7	NN_1 + ID				✓	✓					✓

Presented for each model are the fixed, random, and residual terms. All models also include an overall mean. Model 5 is equivalent to the SpATS approach of Rodrıguez-Alvarez et al. (2018). The grids used in the NN adjustments are presented in Supplementary Figure 8. Note: AR1 - separable first order autoregressive process; ID - independent error term; TPS - tensor product penalised spline; NN - nearest neighbour adjustment.

Table 2. Linear mixed models fitted to the simulated maize breeding dataset, Part 2: Model selection criteria.

Model	Fixed	Vars	-2 loglik	REMLRT	AIC	avsed	Accuracy	Bias
baseline	1	3	-70.4		-66.7	0.274	0.65	0.07
1	1	5	-99.1	< 0.0001	-94.4	0.260	0.72	0.02
2	1	6	-113.7	< 0.0001	-103.7	0.249	0.74	0.02
3	**2**	**7**	**-133.4**		**-126.6**	**0.247**	**0.76**	0.02
4	4	8	-110.8		-93.8	0.226	0.69	0.05
5	**4**	**10**	**-135.7**	$<$**0.0001**	**-116.7**	**0.245**	**0.72**	0.02
6	**1**	**3**	**-103.8**		**-102.0**	**0.279**	**0.71**	0.01
7	1	3	-94.2		-94.7	0.270	0.70	0.03

Presented for each model are the number of fixed effects and variance parameters, residual deviance, REMLRT, AIC, average standard error of difference (avsed), accuracy and the bias. The selected models are distinguished with ***bold font***. Note: The REMLRT is applied sequentially and cannot compare models with different fixed effects. Models 6 and 7 are not comparable because the phenotypes have been adjusted.

REFERENCES

Akima H (1978) A Method of Bivariate Interpolation and Smooth Surface Fitting for Irregularly Distributed Data Points. ACM Transactions on Mathematical Software 4:148–159, URL https://doi.org/10.1145/355780.355786

Bančič J, Ovenden B, Gorjanc G, Tolhurst DJ (2023) Genomic selection for genotype performance and stability using information on multiple traits and multiple environments. Theoretical and Applied Genetics 136:104, URL https://doi.org/10.1007/s00122-023-04305-1

Bartlett BM (1978) Nearest neighbour models in the analysis of field experiments. Journal of the Royal Statistical Society Series B (Methodological), 40:147–158, URL https://doi.org/10.1111/j.2517-6161.1978.tb01657.x

Besag J (1977) Errors-In-Variables Estimation for Gaussian Lattice Schemes. Journal of the Royal Statistical Society Series B (Methodological) 39:73–78, URL https://doi.org/10.1111/j.2517-6161.1977.tb01607.x

Table 3. Linear mixed models fitted to the simulated maize breeding dataset, Part 3: REML estimates of variance parameters.

	Hybrid	Block	Col	Row	Spline					AR1			ID
Model	$\hat{\sigma}^2_g$	$\hat{\sigma}^2_b$	$\hat{\sigma}^2_{e_c}$	$\hat{\sigma}^2_{e_r}$	$\hat{\sigma}^2_{c_1}$	$\hat{\sigma}^2_{r_1}$	$\hat{\sigma}^2_{c_2}$	$\hat{\sigma}^2_{r_2}$	$\hat{\sigma}^2_{cr}$	$\hat{\sigma}^2_s$	$\hat{\rho}_c$	$\hat{\rho}_r$	$\hat{\sigma}^2_r$
baseline	0.02	0.06											0.19
1	0.07	0.00								0.20	0.51	0.23	
2	0.07	0.00								0.40	0.95	0.87	0.09
3	**0.07**	**0.00**		**0.01**						**0.10**	**0.75**	**0.89**	**0.08**
4	0.04	0.05			0.03	0.03	0.69	0.00	0.07				0.16
5	**0.07**	**0.06**	**0.00**	**0.05**	**0.04**	**0.02**	**0.70**	**0.00**	**0.08**				**0.09**
6	**0.08**	**0.00**											**0.14**
7	0.06	0.00											0.17
True	0.09			0.04						0.08			0.08

The selected models are distinguished with ***bold font***. The true values are given in the final row, with bias given in parentheses. Note: The j indices on the variance parameters have been removed for brevity.

Besag J, Kempton RA (1986) Statistical Analysis of Field Experiments Using Neighbouring Plots. Biometrics 42:231–251, URL https://doi.org/10.2307/2531047

Box G, Jenkins G (1970) Time Series Analysis: Forecasting and Control. Holden-Day, San Francisco, URL https://doi.org/10.1057/9781137291264_6

Butler D (2021) odw: Generate optimal experimental designs. URL https://mmade.org/optimaldesign/, r package version 2.1.2

Butler DG, Cullis BR, Gilmour AR, Gogel BJ, Thompson R (2018) ASReml-r reference manual version 4. VSN international URL https://vsni.co.uk/software/asreml-r

Coombes N (2020) DiGGer: DiGGer design generator under correlation and blocking. URL http://nswdpibiom.org/austatgen/software, r package version 1.0.5

Covarrubias-Pazaran G, Gebeyehu Z, Gemenet D, Werner C, Labroo M, Sirak S, Coaldrake P, Rabbi I, Kayondo SI, Parkes E, Kanju E, Mbanjo EGN, Agbona A, Kulakow P, Quinn M, Debaene J (2022) Breeding Schemes: What Are They, How to Formalize Them, and How to Improve Them? Frontiers in Plant Science 12, URL https://doi.org/10.3389/fpls.2021.791859

Cullis BR, Gleeson GN (1991) Spatial Analysis of Field Experiments - An Extension to Two Dimensions. Biometrics 47:1449–1460, URL https://doi.org/10.2307/2532398

Durban M, Currie ID, Kempton RA (2001) Adjusting for Fertility and Competition in Variety Trials. Journal of Agricultural Science 136:129–140, URL https://doi.org/10.1017/S0021859601008541

Frisch M (2023) Selection tools. URL http://population-genetics.uni-giessen.de/~software/

Gaynor RC, Gorjanc G, Hickey JM (2021) AlphaSimR: an R package for breeding program simulations. G3 11, URL https://doi.org/10.1093/g3journal/jkaa017

Gebhardt A, Bivand R, Sinclair D (2023) interp: Interpolation Methods. URL https://CRAN.R-project.org/package=interp, r package version 1.1-4

Gilmour AR, Cullis BR, Verbyla AP (1997) Accounting for Natural and Extraneous Variation in the Analysis of Field Experiments. Journal of Agricultural, Biological, and Environmental Statistics 2:269–293, URL https://doi.org/10.2307/1400446

Gogel B, Welham S, Cullis BR (2023) Empirical comparison of time series models and tensor product penalised splines for modelling spatial dependence in plant breeding field trials. Frontiers in Plant Science 13, URL https://doi.org/10.3389/fpls.2022.1021143

Green P, Jennison C, Seheult A (1985) Analysis of Field Experiments by Least Squares Smoothing. Journal of the Royal Statistical Society Series B (Methodological) 47:299–315, URL https://doi.org/10.1111/j.2517-6161.1985.tb01358.x

Kenward MG, Rogers JH (1997) Small Sample Inference for Fixed Effects from Restricted Maximum Likelihood. Biometrics 53:983–997, URL https://doi.org/10.2307/2533558

Liu H, Tessema BB, Jensen J, Cericola F, Andersen JR, Sørensen AC (2019) ADAM-Plant: A Software for Stochastic Simulations of Plant Breeding From Molecular to Phenotypic Level and From Simple Selection to Complex Speed Breeding Programs. Frontiers in Plant Science 9, URL https://doi.org/10.3389/fpls.2018.01926

Lush JL (1937) Animal Breeding Plans. Ames: Iowa State University Press

Martin RJ (1990) The use of time-series models and methods in the analysis of agricultural field trials. Communications in Statistics - Theory and Methods 19:55–81, URL https://doi.org/10.1080/03610929008830187

de Mendiburu F (2023) agricolae: Statistical Procedures for Agricultural Research. URL https://CRAN.R-project.org/package=agricolae, r package version 1.3-6

Papadakis JS (1937) Methode statistique pour des experiences sur champ. bulletin scientifique Institut d'Amelioration des Plantes a Thessaloniki 23:13–29

Piepho HP, Boer MP, Williams ER (2022) Two-dimensional P-spline smoothing for spatial analysis of plant breeding trials. Biometrical Journal 64:835–857, URL https://doi.org/10.1002/bimj.202100212

Podlich DW, Cooper M (1998) QU-GENE: a simulation platform for quantitative analysis of genetic models. Bioinformatics 14(7):632–653, URL https://doi.org/10.1093/bioinformatics/14.7.632

Rodríguez-Álvarez MX, Boer MP, van Eeuwijk FA, Eilers PHC (2018) Correcting for spatial heterogeneity in plant breeding experiments with P-splines. Spatial Statistics 23:52–71, URL https://doi.org/10.1016/j.spasta.2017.10.003

Stefanova KT, Smith AB, Cullis BR (2009) Enhanced Diagnostics for the Spatial Analysis of Field Trials. Journal of Agricultural, Biological, and Environmental Statistics 14:392–410, URL https://doi.org/10.1198/jabes.2009.07098

Stram DO, Lee JW (1994) Variance Components Testing in the Longitudinal Mixed Effects Setting. Biometrics 50:1171–1177, URL https://doi.org/10.2307/2533455

Stringer JK, Cullis BR, Thompson R (2011) Joint Modeling of Spatial Variability and Within-Row Interplot Competition to Increase the Efficiency of Plant Improvement. Journal of Agricultural, Biological, and Environmental Statistics 16:269–281, URL https://doi.org/10.1007/s13253-010-0051-5

Technow F (2015) R package mvngGrAd: moving grid adjustment in plant breeding field trials

Velazco JG, Rodríguez-Álvarez MX, Boer MP, Jordan DR, Eilers PHC, Malosetti M, van Eeuwijk FA (2017) Modelling spatial trends in sorghum breeding field trials using a two-dimensional P-spline mixed model. Theoretical and Applied Genetics 130:1375–1392, URL https://doi.org/10.1007/s00122-017-2894-4

Verbyla AP (2019) A note on model selection using information criteria for general linear models estimated using REML. Australian & New Zealand Journal of Statistics 61:39–50, URL https://doi.org/10.1111/anzs.12254

Wilkinson GN, Eckert SR, Hancock TW, Mayo O (1983) Nearest Neighbour (NN) Analysis of Field Experiments. Journal of the Royal Statistical Society Series B (Methodological) 45:151–211, URL https://doi.org/10.1111/j.2517-6161.1983.tb01240.x

Younis OG, Turchetta M, Ariza Suarez D, Yates S, Studer B, Athanasiadis IN, Krause A, Buhmann JM, Corinzia L (2023) ChromaX: a fast and scalable breeding program simulator. bioRxiv) URL https://doi.org/10.1101/2023.05.29.542709

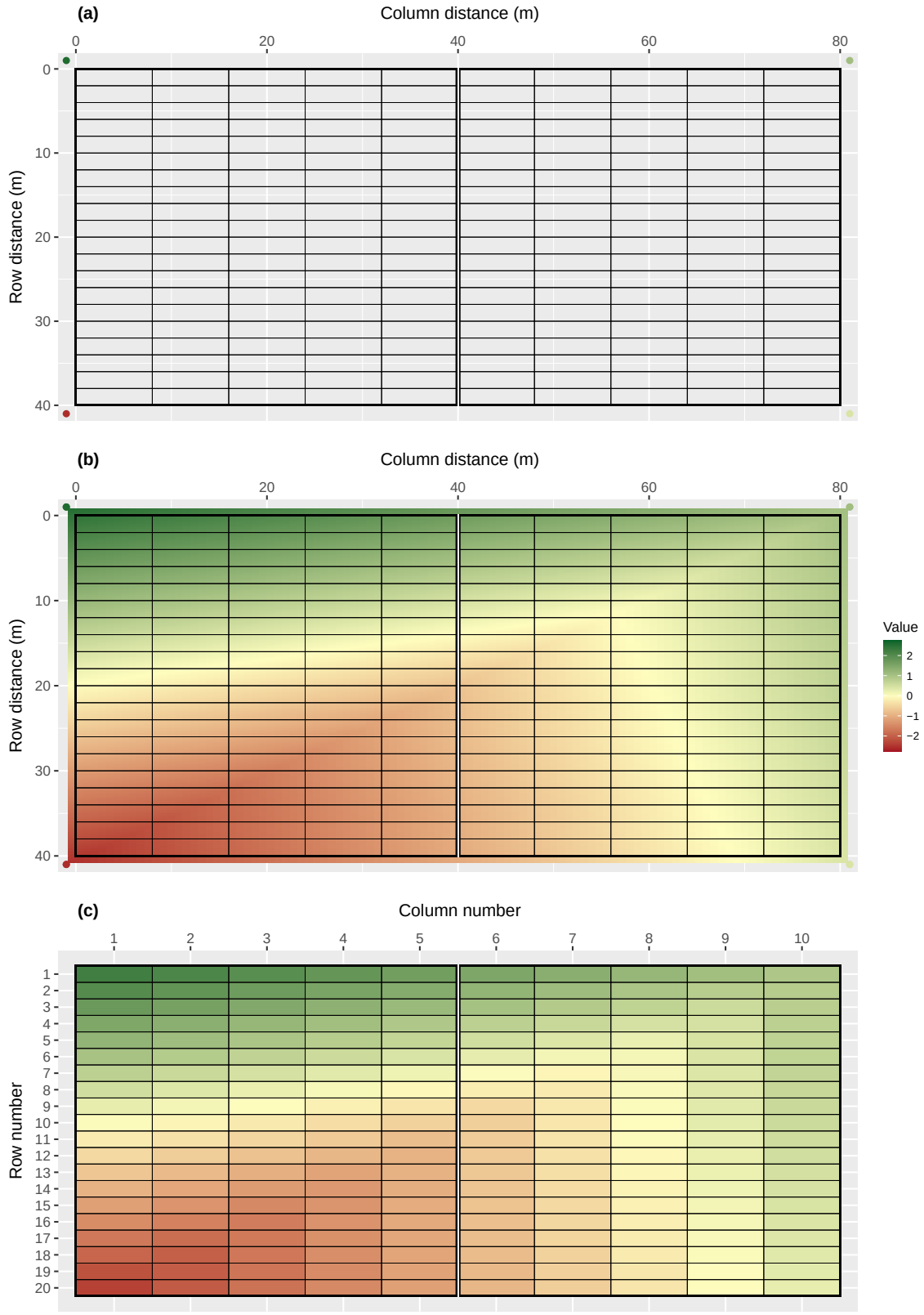

Figure 1. Demonstration of how FieldSimR generates spatial errors using bivariate interpolation: (a) the two-dimensional lattice of plots is constructed with four interpolation (knot) points placed outside the four corners, (b) the continuous array between the knot points is interpolated using bivariate interpolation, which produces a smooth continuous surface, (c) a single error value is assigned to each plot by averaging over the continuous surface within each plot.

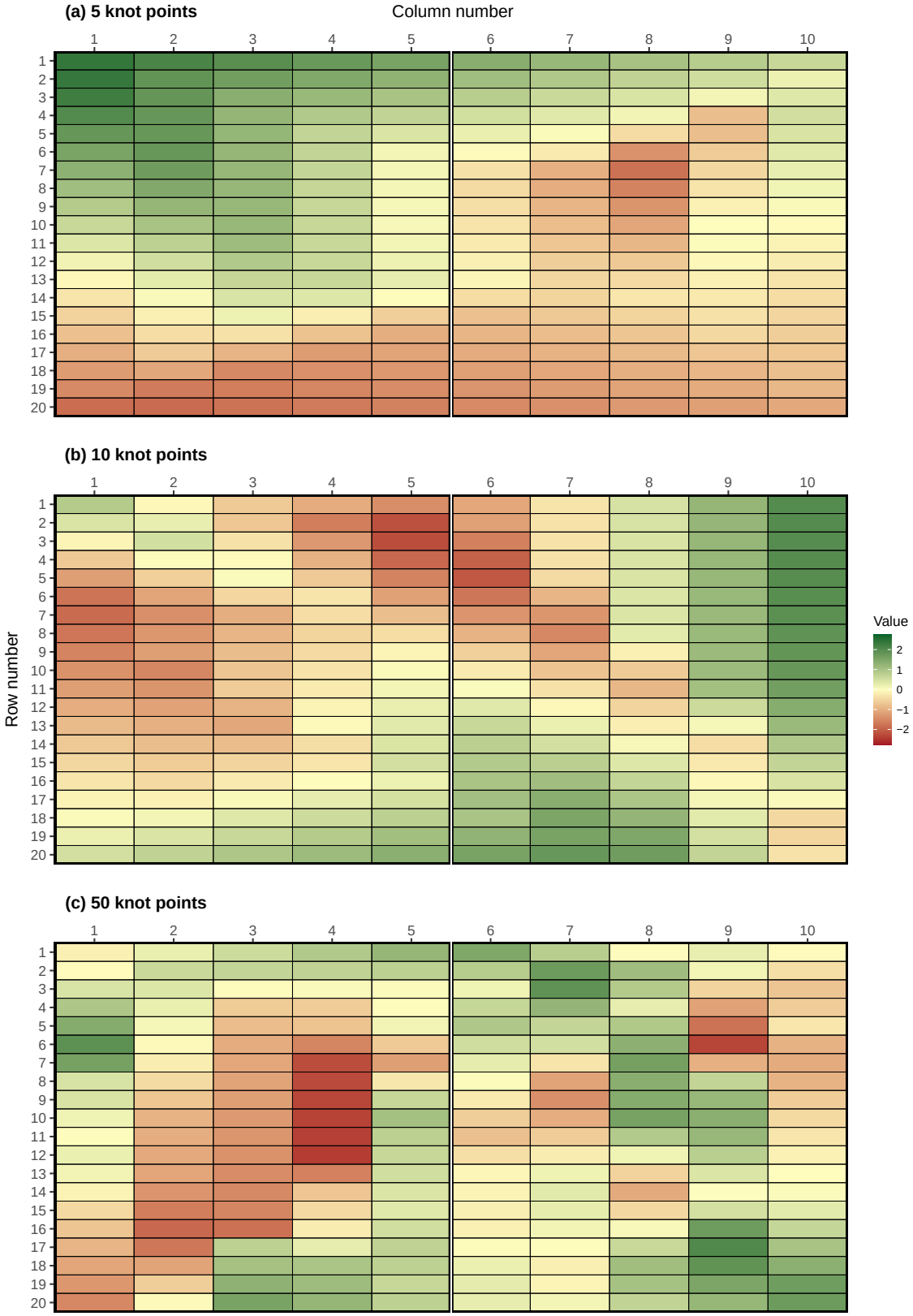

Figure 2. Examples of spatial errors generated using bivariate interpolation with (a) 5, (b) 10, and (c) 50 knot points. These options are set using `complexity = 5`, `10`, and `50`. The coordinates of the knot points are presented in Supplementary Figure 2.

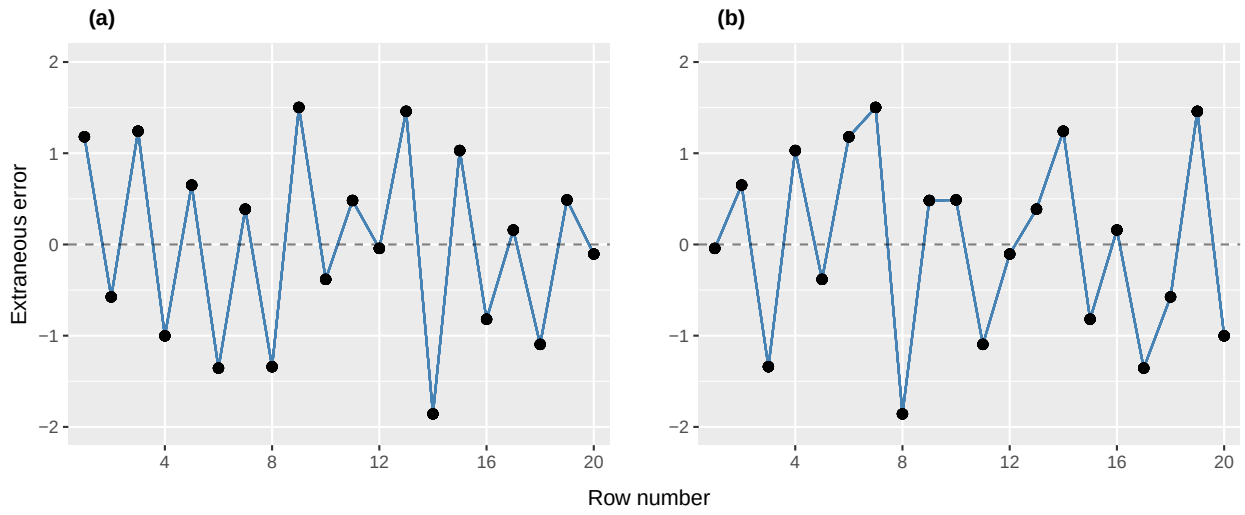

Figure 3. Examples of extraneous errors generated using (a) zig-zag or (b) random ordering.

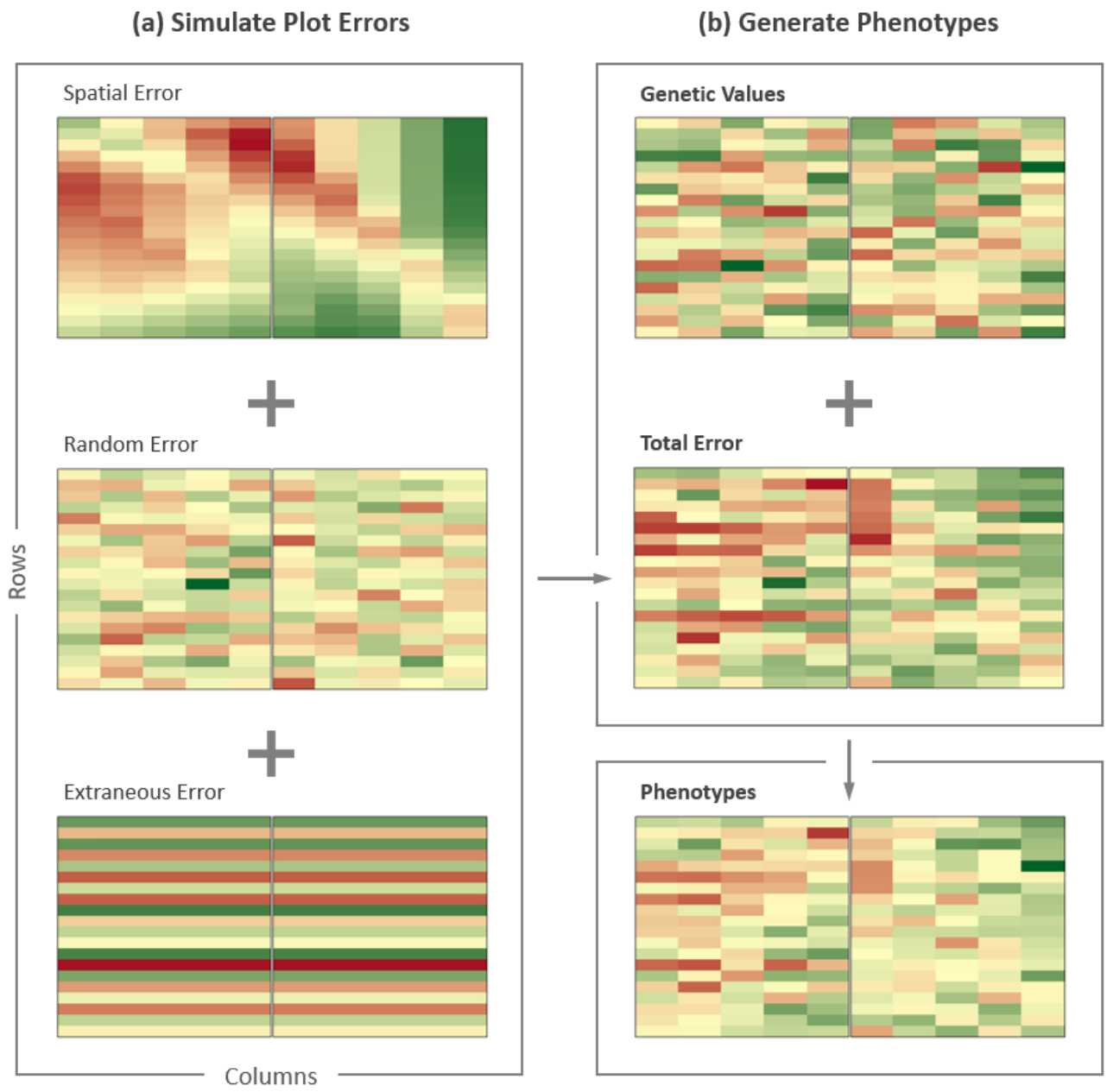

Figure 4. Demonstration of how FieldSimR generates phenotypes: (a) the plot errors are constructed by combining the spatial errors with the random and extraneous errors at a user-defined ratio, (b) the phenotypes are generated by combining the plot errors with the true genetic values obtained from AlphaSimR. A randomised complete block design is used to allocate genotypes to plots.

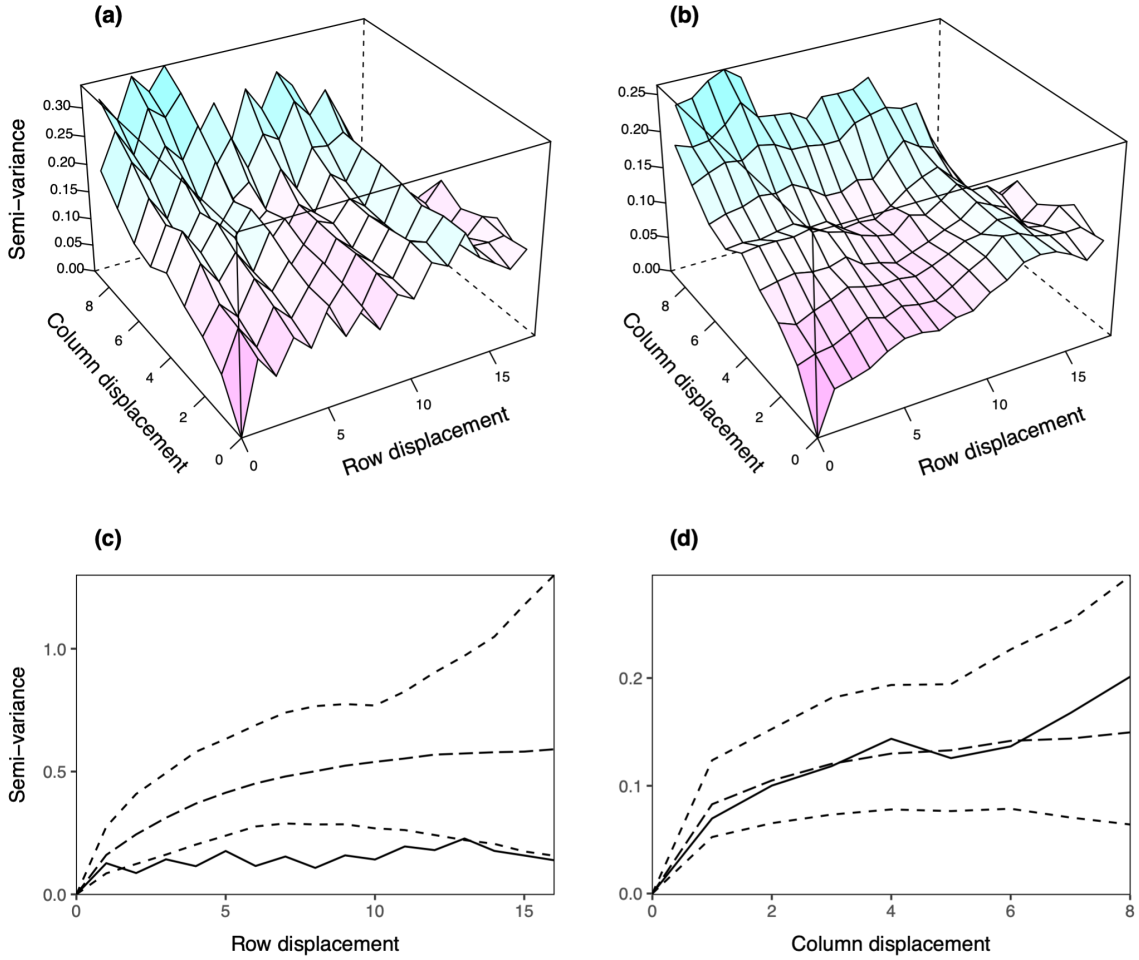

Figure 5. Sample variograms for the AR1 based models fitted to the simulated plot data in Figure 4: (a) Model 2: AR1+ID and (b) Model 3: AR1+ID+Frow+Rrow. The corresponding variogram sills are presented in (c) for Model 2 and (d) for Model 3. Note: Only semi-variances based on more than 30 pairs are shown. AR1 - separable first order autoregressive process; ID - independent error term, Frow - fixed row term; Rrow - random row term.