SOFTWARE RELEASE

GBS – Genetics, Breeding, and Statistics: Software for experimental design and analysis in plant breeding

Crop Breeding and Applied Biotechnology
26(2): e557026213, 2026
Brazilian Society of Plant Breeding.
Printed in Brazil
http://dx.doi.org/10.1590/1984-70332026v26n2s28

Tiago de Souza Marçal[1]✉ **and Vinícius Quintão Carneiro**[1]

Abstract: *GBS is a free software package featuring an interactive, user-friendly interface designed to simplify routine tasks in the design and data analysis of plant breeding experiments. Available at https://github.com/VQCarneiro/GBS, it supports multiple experimental designs and enables analyses using fixed, random, or mixed models.*

Keywords: *Design of field experiments, mixed models, quantitative genetics, biometry*

INTRODUCTION

Data collection and data analysis are two major steps in modern scientific research. In this context, during the first half of the 20th century, scientists Ronald A. Fisher and Frank Yates were pioneers in establishing the principles of experimentation and in developing several experimental designs, as well as their corresponding analytical approaches (Fisher 1925, Fisher 1926, Yates 1936a, b, Yates 1940a, b). Although the theoretical principles developed by these scientists are widely applied across several scientific fields, their original motivation was agricultural experimentation with a focus on cultivar development. During the second half of the 20th century, several other scientists introduced and/or disseminated new experimental designs and analytical approaches to address the emerging demands of agricultural experimentation (Federer et al. 1956, Henderson et al. 1959, Cochran and Cox 1957, Patterson and Thompson 1971, Patterson and Williams 1976, Williams et al. 1986, Federer and Wright 1988, Oliveira 1990, Gomes and Garcia 1991, Gilmour et al. 1995, Johnson and Thompson 1995, Meyer 1996, Gilmour et al. 1997).

Despite the advances achieved so far, the design and analysis of experiments applied to plant breeding, especially those involving incomplete block designs, still represent a major challenge for many researchers. In this context, the development of practical solutions in this area remains a current demand (Resende 2016, Bhering 2017). A growing trend toward the use of mixed-model methodologies for experimental analysis is noteworthy. This approach combines linear mixed models (Henderson et al. 1959) with residual maximum likelihood (Patterson and Thompson 1971) to perform complex statistical analyses. These tools have greater flexibility in handling the daily challenges of plant breeding. However, given the complexity of these analyses, the use of computationally efficient software with a user-friendly interface becomes essential in routine procedures in plant breeding. In this context, GBS (Genetics, Breeding, and

***Corresponding author:**
E-mail: tiago.marcal@ufla.br

Scientific Editor:
Luiz Antônio dos Santos Dias

Received: 18 March 2026
Accepted: 03 May 2026
Published: 13 May 2026

[1] Universidade Federal de Lavras, Departamento de Biologia, Trevo Rotatório Professor Edmir Sá Santos, s/n, 37203-202, Lavras, MG, Brazil

Statistics) software emerges as a versatile and intuitive tool for the design and analysis of experiments in quantitative genetics and plant breeding.

The GBS software is a user-friendly tool that integrates the versatility of Python and R (Python Software Foundation 2025, R Core Team 2025). In this sense, GBS enables the execution of complex statistical analyses using a mixed-model approach in a simple and intuitive way, allowing users to devote greater effort to interpreting results and making decisions, thereby contributing to democratization of access to this technology. Many of these analyses are available in software such as R and ASReml; however, these platforms do not provide a user-friendly and interactive graphical interface. In addition, ASReml software is not freely available.

Currently, the GBS software provides procedures for the design and analysis of the main experimental frameworks used in plant breeding, including both single and multi-environment trials. These procedures accomodate complex (co)variance structures and variable numbers of replications and treatments, enabling the estimation of genetic and non-genetic parameters used in quantitative genetics and plant breeding. GBS supports the analysis of experiments with missing plots and provides comprehensive output for inferences on fixed and random effects. To ensure high computational efficiency, all analytical procedures exploit the sparsity of matrix models to optimize memory allocation and accelerate matrix operations (Bates et al. 2025, Zammit-Mangion et al. 2025).

Thus, the purpose of this paper is to provide an overview of the functional capabilities and theoretical foundations of the GBS software.

DESCRIPTION

The GBS software is designed for the Windows operating system and is freely available to the scientific community at https://github.com/VQCarneiro/GBS, where additional information about the program is also provided. All software files are packaged in a compressed file designated *GBS.rar*, which, after being downloaded, must be extracted to the C drive (C:) of the computer. Once the extraction is complete, the graphical interface can be accessed through the executable file *GBS.exe*.

The software works exclusively with ".csv" files and requires the computer to be configured to use standard data formatting symbols. This configuration can be accessed by opening the Start menu and then selecting the Settings option (Figure 1). In the corresponding window, the "Recommended" option must be selected for both "Decimal Symbol" and

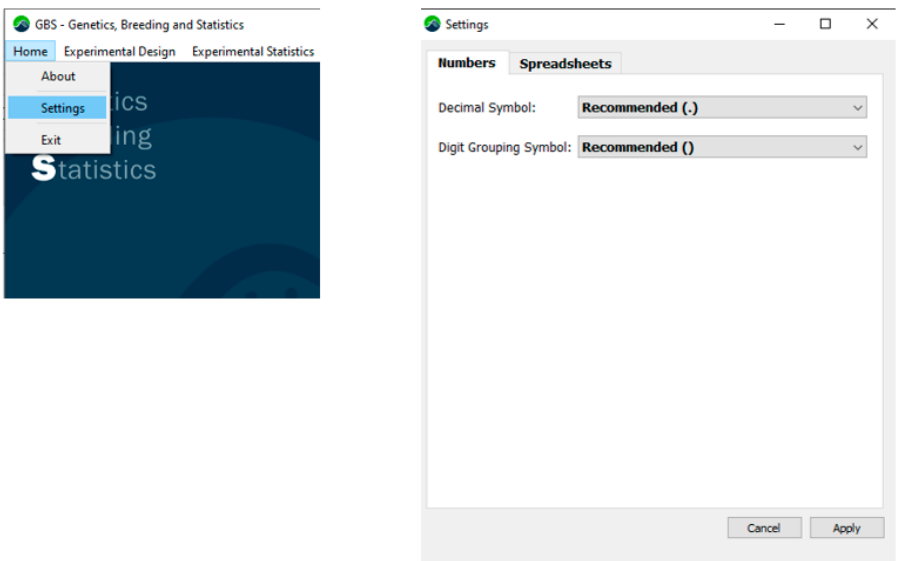

Figure 1. Start menu and system option settings of the GBS software.

"Digit Grouping Symbol" in the Numbers tab, as well as for "List Separator" in the Spreadsheets tab. After completing this configuration, the software will be ready for use.

AVAILABLE PROCEDURES

The procedures implemented in the GBS software are organized into two main sections: "Experimental Design" and "Experimental Statistics" (Figures 2 and 3). The tutorial and case studies are available at https://github.com/VQCarneiro/GBS/tree/main/Tutorial.

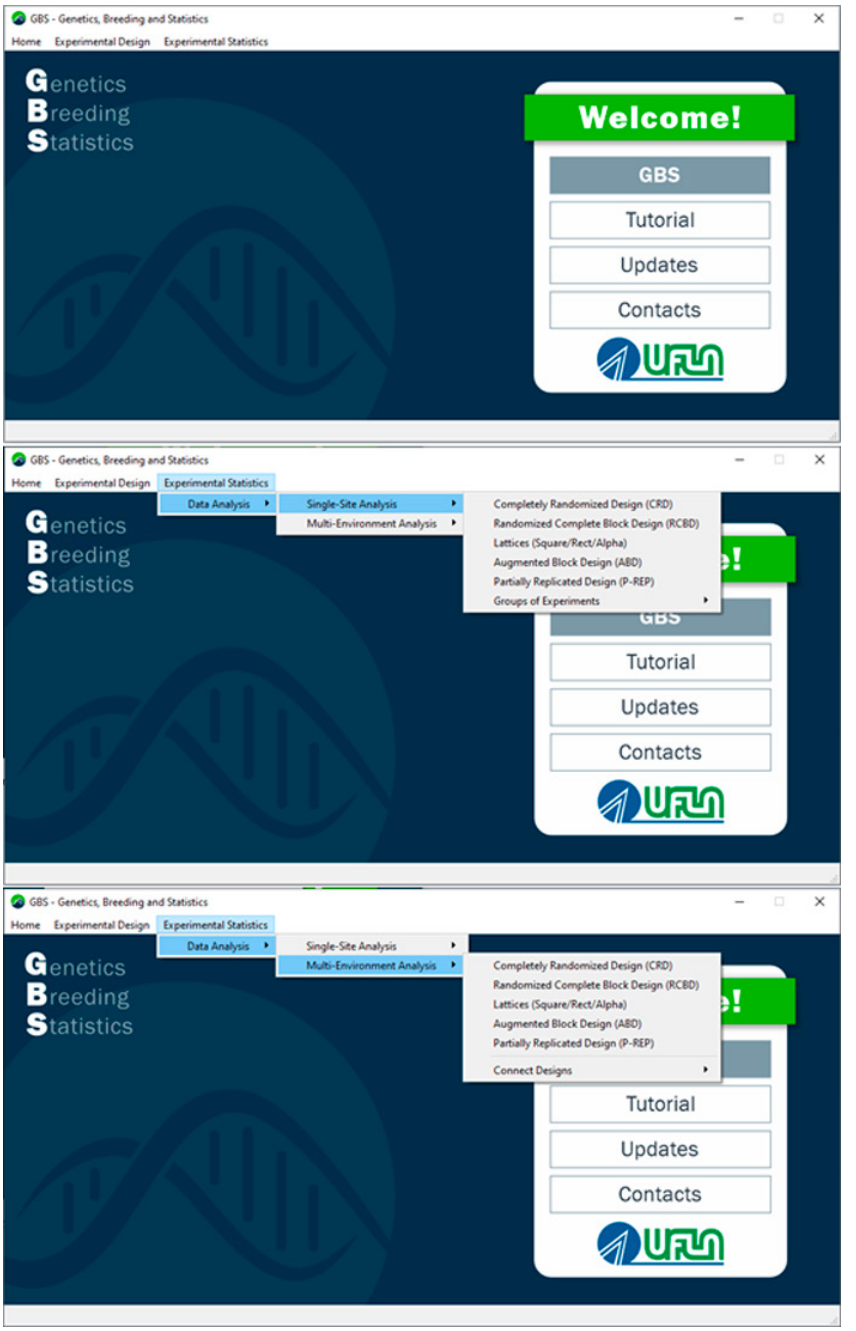

Figure 2. Graphical interface and procedures for individual and joint analyses implemented in the GBS software.

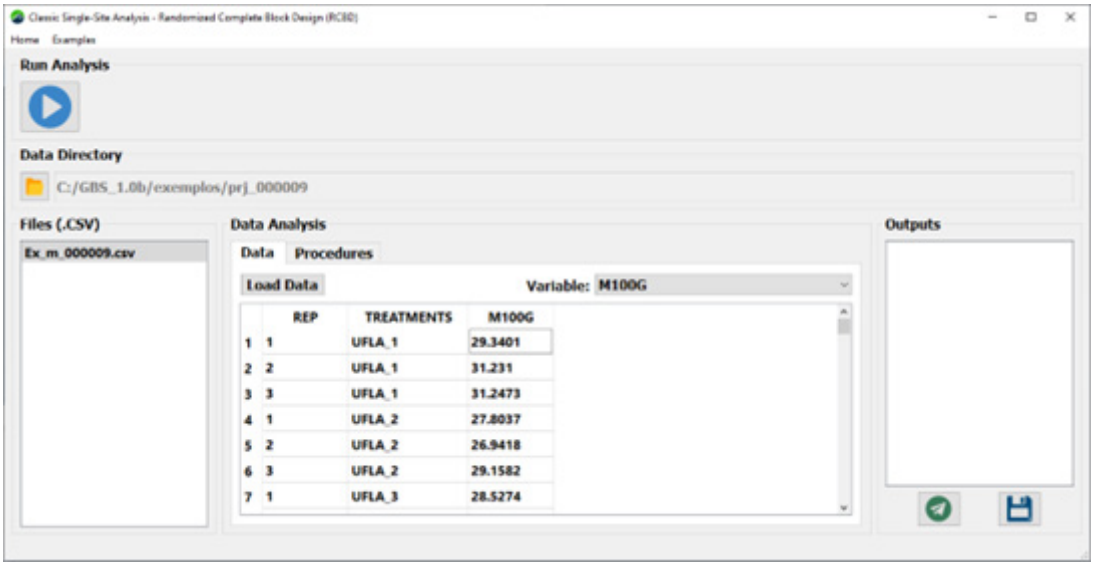

Figure 3. Standard graphical interface of the GBS software, illustrated by the analysis procedure for a randomized complete block design (RCBD).

Experimental design

The GBS software provides a comprehensive experimental design module accomodating the main designs used in plant breeding, including the completely randomized design (CRD), randomized complete block design (RCBD), augmented block design (ABD), and square lattice designs, both balanced and partially balanced, with or without checks allocated within the incomplete blocks. For experiments conducted across multiple locations, GBS enables the simultaneous generation of multiple experimental designs, thereby simplifying the researcher's workflow. These designs can be exported directly to Excel and Word files.

Experimental statistics

The "Experimental Statistics" section features a robust and comprehensive module for analyzing both single and multi-environment trials under CRD, RCBD, ABD, P-REP, and lattice designs. For each procedure, examples are provided to help users correctly organize their data files. Some case-study applications are also available on the GBS GitHub page. It is noteworthy that the analyses implemented in this section can handle missing observations, which must be coded as "NA" in the data file.

Single-trial analysis

All procedures allow treatment effects to be specified as either fixed or random. By default, GBS performs statistical tests and generates scatter plots to assist researchers in diagnosing residuals. For each procedure, the software provides files containing the analysis results, along with the appropriate tests and estimates of statistical parameters, such as the overall mean, the coefficient of variation, and the average standard deviation of pairwise mean contrasts. GBS also generates files with the estimated treatment means and the standard deviations of pairwise contrasts. When treatment effects are considered fixed, GBS offers the option to apply the Tukey and Scott-Knott tests, both at a 5% significance level. Conversely, when treatment effects are considered random, GBS estimates both generalized heritability and accuracy (Cullis et al. 2006, Piepho and Möhring 2007), in addition to providing the Empirical Best Linear Unbiased Predictions (EBLUPs) of the treatments and their respective accuracies (Smith et al. 2005).

Multi-environment trial analysis

The GBS includes a multi-environment trial analysis module for all the previously described designs, supporting both fixed and random effects of treatments and environments. The heterogeneity of residual variances and incomplete

block variances can be modeled, as well as the heterogeneity of row and column variances in the case of the P-REP design. In addition, GBS enables estimation ofinteraction effects and the average performance of treatments across different environments, while providing outputs similar to those from single-trial analyses. It should be noted that the multi-environment trial analysis procedures are equipped to assess the estimability of the linear combinations used to estimate means and to allow the analysis of sequential experiments with a variable number of treatments as a result of the selection process.

Connecting different experimental designs

In the multi-environment analysis section, procedures are also available that enable experiments conducted under different experimental designs to be jointly analyzed, a common situation in plant breeding (Resende 2016). This approach may be necessary, as different experimental designs are often adopted throughout the different phases of plant breeding programs, especially due to changes in the number of treatments resulting from the selection process. GBS provides procedures for combining RCBD or ABD experiments with lattice experiments underscenarios with either fixed-effect or random-effect treatments.

Algorithm implemented in GBS software

The GBS software estimates variance components using the residual maximum likelihood (REML) method, implemented through the Average Information (AI) algorithm. The AI algorithm was originally proposed by Johnson and Thompson (1995) and later formally generalized by Gilmour et al. (1995). The rationale behind this algorithm is to obtain an information matrix (based on the second-order partial derivatives of the residual log-likelihood) composed of matrix traces that are computationally simpler to evaluate. To achieve this, Johnson and Thompson (1995) and Gilmour et al. (1995) used the average of the observed (Hessian) and expected information matrices. Meyer (1996) extended this idea and presented a computationally efficient approach to obtain the AI matrix from the Cholesky factorization of expanded mixed model equations (MME), including working variables, in order to exploit the benefits of sparsity in the MME.

Convergence is assessed based on the magnitude of the change in the residual log-likelihood and the proximity of the score vector elements (first-order partial derivatives of the residual log-likelihood) to zero. In general, the model is considered to have converged when both the change in the log-likelihood and the largest absolute value among the score vector elements are below 0.001. It is worth noting that the AI algorithm often produces parameter estimates outside the parameter space during the iterative steps. To address this, the constraints described by Gilmour (2019) were incorporated into the GBS software.

Inference for fixed effects in GBS software

For fixed-effects models, analysis of variance is widely used to decompose the total variation into components attributable to the model effects. Based on this information, Snedecor's F-test is applied to verify the existence of contrasts among the levels of the factor under study. In contrast, for linear mixed models, the Wald test is used to assess the significance of the fixed effects (Kenward and Roger 1997, Kuznetsova et al. 2017). Asymptotically, this test follows a chi-squared (χ^2) distribution with p degrees of freedom. Alternatively, an F approximation (referred to as the Wald F) can be obtainedwith ν_{DFN} degrees of freedom for the numerator and ν_{DFD} degrees of freedom for the denominator (Kenward and Roger 1997, Kuznetsova et al. 2017). However, due to the multiple variance components in a linear mixed model, the denominator degrees of freedom (ν_{DFD}) cannot be obtained in a straightforward manner. Thus, ν_{DFD} can be approximated using the Satterthwaite or Kenward-Roger methods (Kenward and Roger 1997, Kuznetsova et al. 2017). Further details on the estimation of ν_{DFD} can be found in Kuznetsova et al. (2017).

Due to its lower computational cost, the Satterthwaite method was implemented in the GBS software to estimate the ν_{DFD}. It is also noteworthy that GBS produces Type III analyses of variance (Kuznetsova et al. 2017), meaning that all sums of squares and, consequently, the F approximations are computed marginally. Notably, in linear mixed models, calculating sums of squares for fixed effects is more challenging because of the increased complexity of the (co)variance structure of the **y** vector. Even so, the estimator proposed by Bates et al. (2015), based on the Cholesky factorization, can be used for this purpose. In cases involving heterogeneous variances, the computation of sums of squares was extended from the expanded mixed-model equation system described by Bates and Debroy (2004).

Inference for random effects in GBS software

Unlike fixed-effects models, in linear mixed models the significance of the variance components of random effects is assessed using the likelihood ratio test (LRT) (Resende et al. 2016). This test approximately follows a chi-squared (χ^2) distribution with ν degrees of freedom, where ν is the difference in the number of parameters between the full and reduced models. Based on these comparisons, GBS evaluates the significance of the variance components in all procedures.

CONCLUDING REMARKS

Over many years of work in plant breeding, the authors have identified several practical challenges related to experimental design and data analysis. Based on this accumulated experience, user-friendly software was developed to democratize access to these technologies for researchers in the field. GBS is free software designed to address the main demands of genetic breeding programs by integrating the best functionalities of widely used mixed-model analysis software. These functionalities includecomplete outputs for fixed and random effects, modeling of residual (co)variance structures, and implementation of mean comparison and grouping tests, even under unbalanced data. Thus, GBS is software developed by breeders, for breeders, with the goal of simplifying the routine tasks involved in the design and data analysis of experiments in plant breeding.

ACKNOWLEDGMENTS

The authors thank the funding agencies CAPES, CNPq, and FAPEMIG for the financial support provided throughout their academic training and subsequent research projects, which continue to inspire innovations in plant breeding. The authors also express their gratitude to God for the gift of life, constant inspiration, strength, and the opportunity to have come this far. They also thank their mentors José Eustáquio de Souza Carneiro, Pedro Crescêncio Carneiro, Adésio Ferreira, Magno Antônio Patto Ramalho, Marcos Deon Vilela Resende, Cosme Damião Cruz, and Arthur Gilmour for their inspiration and valuable insights on agricultural experimentation, plant breeding, quantitative genetics, and biometrics. In addition, the authors thank their colleagues José Henrique Soler Guilhen and Roberto Henrique de Lima Ribeiro, whose suggestions and insights inspired theoretical and visual aspects of the software.

CREDIT STATEMENT

Both authors participated in the development and implementation of the algorithms for model fitting, as well as in their integration into the software graphical interface. In addition, both authors contributed to the writing of the manuscript.

REFERENCES

Juliatti FC and Silva SA (2001) Antracnose: *Colletotrichum gloesporioides* Bates DM and Debroy S (2004) Linear mixed models and penalized least squares. **Journal of Multivariate Analysis 91**: 1-17.

Bates DM, Mächler M, Bolker B and Walker S (2015) Fitting linear mixed-effects models using lme4. **Journal of Statistical Software 67**: 1-48.

Bates DM, Maechler M, Jagan M, Davis TA, Karypis G, Riedy J, Oehlschlägel J and R Core Team (2025) Matrix: Sparse and dense matrix classes and methods (manual). Available at <https://cran.r-project.org/web/packages/Matrix/Matrix.pdf>. Accessed on November 21, 2025.

Bhering LL (2017) Rbio: A tool for biometric and statistical analysis using the R platform. **Crop Breeding and Applied Biotechnology 17**: 187-190.

Cochran WG and Cox GM (1957) **Experimental designs**. John Wiley and Sons, New York, 617p.

Cullis BR, Smith AB and Coombes NE (2006) On the design of early generation variety trials with correlated data. **Journal of Agricultural, Biological, and Environmental Statistics 11**: 381-393.

Federer WT (1956) Augmented (hoonuiaku) designs. **Hawaiiwan Planter's Record 55**: 191-208.

Federer WT and Wright J (1988) Construction of lattice square designs. **Biometrical Journal 30**: 77-85.

Fisher RA (1925) **Statistical methods for research workers**. Edinburgh, Oliver and Boyd, 362p.

Fisher RA (1926) The arrangement of field experiments. **Journal of the Ministry of Agriculture 33**: 503-515.

Gilmour AR (2019) Average information residual maximum likelihood in practice. **Journal of Animal Breeding and Genetics 136**: 262-272.

Gilmour AR, Cullis BR and 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.

Gilmour AR, Thompson R and Cullis BR (1995) Average information REML: an efficient algorithm for variance parameter estimation in linear mixed models. **Biometrics 51**: 1440-1450.

Gomes FP and Garcia CH (1991) Experimento sem látice: planejamento e

análise por meio de "pacotes" estatísticos. **Série Técnica IPEF 7**: 1-69.

Henderson CR, Kempthorne O, Searle SR and von Krosigk CM (1959) The estimation of environmental and genetic trends from records subject to culling. **Biometrics 15**: 192-218.

Johnson DL and Thompson R (1995) Restricted maximum likelihood estimation of variance components for univariate animal models using sparse matrix techniques and average information. **Journal of Dairy Science 78**: 449-456.

Kenward MG and Roger JH (1997) Small sample inference for fixed effects from restricted maximum likelihood. **Biometrics 53**: 983-997.

Kuznetsova A, Brockhoff, PB, Christensen and Rune HB (2017) lmerTest package: tests in linear mixed effects models. **Journal of Statistical Software 82**: 1-26.

Meyer K (1996) An "average information" Restricted Maximum Likelihood algorithm for estimating reduced rank genetic covariance matrices or covariance functions for animal models with equal design matrices. **Genetics Selection Evolution 28**: 23-49.

Oliveira AC (1990) Experimentos em reticulado quadrado com alguns tratamentos comuns adicionados em cada bloco: análise com recuperação da informaçãointerblocos. **Pesquisa Agropecuária Brasileira 25**: 289-298.

Patterson HD and Thompson R (1971) Recovery of inter-block information when block sizes are unequal. **Biometrika 58**: 545-554.

Patterson HD and Williams ER (1976) A new class of resolvable incomplete block designs. **Biometrika 63**: 83-92.

Piepho HP and Möhring J (2007) Computing heritability and selection response from unbalanced plant breeding trials. **Genetics 177**: 1881-1888.

Python Software Foundation (2025) Version: 3.11. Available at <https://www.python.org/>. Accessed on November 21, 2025.

R Core Team (2025) R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna. Available at <https://www.R-project.org/>. Accessed on November 21, 2025.

Resende MDVD (2016) Software Selegen-REML/BLUP: a useful tool for plant breeding. **Crop Breeding and Applied Biotechnology 16**: 330-339.

Smith AB, Cullis BR and Thompson R (2005) The analysis of crop cultivar breeding and evaluation trials: an overview of current mixed model approaches. **The Journal of Agricultural Science 143**: 449-462.

Williams ER, Ratcliff D and Van Ewijk PH (1986) The analysis of lattice square designs. **Journal of the Royal Statistical Society: Series B (Methodological) 48**: 314-321.

Yates F (1936a) A new method of arranging variety trials involving a large number of varieties. **The Journal of Agricultural Science 26**: 424-455.

Yates F (1936b) Incomplete randomized blocks. **Annals of Eugenics 7**: 121-140.

Yates F (1940a) Lattice squares. **The Journal of Agricultural Science 30**: 672-687.

Yates F (1940b) The recovery of inter-block information in balanced incomplete block designs. **Annals of Eugenics 10**: 317-325.

Zammit-Mangion A, Davis TA, Amestoy P, Duff I and Reid J (2025) sparseinv: Computation of the sparse inverse subset (manual). Available at <https://cran.r-project.org/web/packages/sparseinv/sparseinv.pdf>. Accessed on November 21, 2025.