

Web Application for Spatial Modelling of Field Trials

Aplicación Web para la Modelación Espacial de Ensayos de Campo

JOHAN APARICIO^{1-2,a}, DANIEL ARIZA-SUAREZ^{2,b}, BODO RAATZ^{2,c}

¹ESCUELA DE ESTADÍSTICA, UNIVERSIDAD DEL VALLE, CALI, COLOMBIA

²BEAN PROGRAM, INTERNATIONAL CENTER FOR TROPICAL AGRICULTURE (CIAT), PALMIRA, COLOMBIA

Abstract

In plant improvement experimental designs are used to decompose the total phenotypic variance observed in field experiments into at least two components: A *genetic* and a *non-genetic* component that is attributable to the spatial variation or environment. Recently, new methodologies for the modelling of spatial trends have been published using the arrangement of the experimental units in the field. These methodologies have shown an improvement in the prediction of the genetic potential of evaluated genotypes. However, the use of these tools may be limited because of the cost to access a licensed product and/or the requirement to be familiar with the language and environment that was used for their implementation. This, in turn limits the data analysis efficiency for decision making. These limitations led to the development of **Mr.Bean**, an easy-to-access, practical and user-friendly tool that integrates the spatial modelling capabilities of **SpATS**, the graphical versatility of **plotly** and the interactive and simple construction approach offered by **Shiny** for the development of Web applications. This tool incorporates descriptive analyses, measures of dispersion and centralization, graphical visualization for comparing multiple variables, the adjustment of mixed models with or without spatial components and the identification of outlier data. All these capabilities are aimed at plant breeders and in general people working with agricultural field data to make precise decisions more quickly.

Key words: Agriculture, Web-application, Field Trials, Mixed Models, Shiny, Spatial Statistics.

^aUniversidad del Valle. E-mail: J.Aparicio@cgiar.org

^bCIAT. E-mail: D.Ariza@cgiar.org

^cCIAT. E-mail: B.Raatz@cgiar.org

Resumen

En los diseños experimentales utilizados en el mejoramiento de plantas es típicamente de interés descomponer la varianza total (fenotípica) en al menos dos componentes: Un componente *genotípico* y un componente *no genético* asociado al ambiente o al espacio. Recientemente se han publicado nuevas metodologías para el modelamiento de las tendencias espaciales utilizando el arreglo de las unidades experimentales en el campo. Estas metodologías han demostrado una mejora en la predicción del potencial genético de los genotipos evaluados. Sin embargo, el uso de estas herramientas puede verse limitado por el costo para acceder a ellas, y/o por la necesidad de estar familiarizado con los lenguajes de programación y el entorno en el que fueron implementadas. Esto a su vez puede limitar la eficiencia en el análisis de datos para la toma de decisiones. Estas dificultades motivaron el desarrollo de *Mr.Bean*, una herramienta de análisis de ensayos de campo de fácil acceso, práctica y amigable que integra las capacidades de modelamiento espacial de *SpATS*, la versatilidad gráfica de *Plotly* y el enfoque interactivo y de sencilla construcción que ofrece *Shiny* para el desarrollo de aplicaciones Web. Esta herramienta incorpora análisis descriptivos, medidas de dispersión y centralización, visualización para la comparación de variables, ajuste de modelos mixtos con y sin componentes espaciales e identificación de datos atípicos. Todas estas opciones están destinadas a los mejoradores de plantas y público en general que quiera analizar los datos de experimentos en campo para tomar decisiones más precisas y de forma más rápida.

Palabras clave: Agricultura, Aplicación web, Shiny, Estadística Espacial, Modelos Mixtos, Ensayos de campo.

1. Introduction

Predicting the genetic potential of new genotypes is one of the most important tasks in plant improvement. To accomplish this, several genotypes are tested in field trials to assess their performance in regards to certain traits of interest. These trials are usually characterized by some unavoidable spatial heterogeneity. These differences in the field can be associated with the intrinsic characteristics of the soil (slope, water retention, etc.) and the management of the trial for its establishment (mechanization, irrigation, fertilizer application, localized attacks of pests and diseases, etc.) (Isik et al. 2017). These circumstances promote the generation of localized patterns of similarity among adjacent experimental units in the field, reducing the uniformity of the whole trial. For this reason, the experimental designs commonly used in plant improvement allow to separate out the genotypic from the environmental variability.

To model the genotypic and environmental components in a trial, different statistical procedures make use of the mixed linear models (MLM) approach. Some of these procedures can incorporate a component to model the spatial variation in the agricultural field trials. This component can improve predictions of the genetic potential of the genotypes that are being tested. In that sense, several approaches have been proposed trying to correct for the spatial heterogeneity in the field (e.g. Cullis & Gleeson (1991), Currie & Durban (2002), Piepho & Williams

(2010), Robbins et al. (2012), Rodriguez-Alvarez et al. (2018)). These approaches can be classified as those using spatial variance-covariance structures and those using smoothing techniques to model spatial trends (Rodriguez-Alvarez et al. 2018). However, the implementation of these approaches into end-user tools is limited either because of the cost for having access to licensed products or the requirement to be familiar with the programming environment that was used for their implementation. This work reports the development of *Mr.Bean*, a web application with a friendly graphical user interface (GUI) to analyze field data using MLMs with or without spatial correction. For this, *Mr.Bean* combines the powerful statistical capabilities offered by several R packages, including *SpATS* and *lme4*, with the interactive features and flexibility offered by *Shiny* and *plotly*, aiming to offer the end-user an alternative way to analyze field data with no previous experience on the R programming environment.

2. Mr.Bean app

Mr.Bean presents a dashboard GUI that seeks to streamline data analyses and decision making from agricultural field experiments. It can be easily accessed from <https://beanteam.shinyapps.io/MrBean/>. These are the modules that constitute *Mr.Bean*.

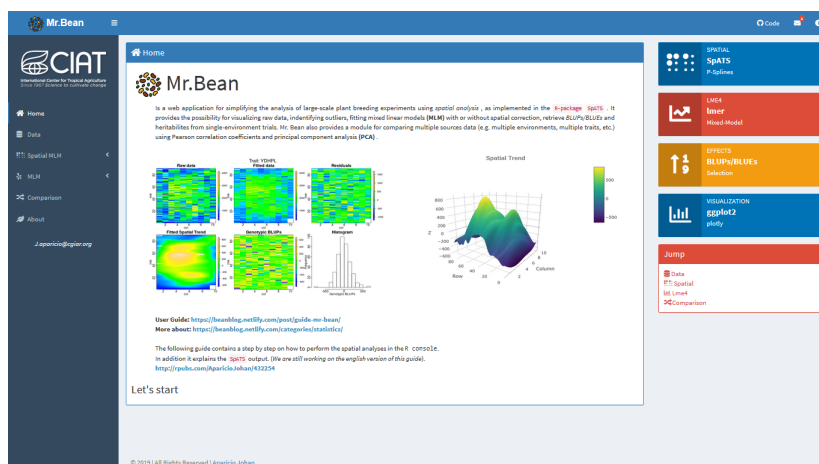


FIGURE 1: Mr.Bean's home page

Data: This module was designed to load and view the raw data to be analyzed. It is prepared to receive a data set that follows the tidy format, in which every variable has a single column and every observation has a single assigned row (check Wickham et al. (2014) for a detailed explanation). This table can be uploaded in any of the “csv” (comma-separated values), “tsv” (tab-separated values) “xlsx” (Excel) or “xls” (Excel 97-2003) formats. Once the data set has been uploaded, this module provides a quick view of the data for navigation (sorting, filtering, pagination). This module also provides some measures of dispersion and

the possibility to visually compare different qualitative and quantitative variables using box plots and two-dimensional scatter plots.

Spatial MLM: This module contains the core functionality of *Mr.Bean*. Here, the user can fit a MLM with spatial correction as implemented in *SpATS*. *SpATS* is an attractive alternative to classical analyses of field trials that model spatial variation as correlated noise (Rodriguez-Alvarez et al. 2018). It uses two-dimensional smoothing surfaces with penalized splines to model the spatial trends within the MLM framework. A helper button guides the user to specify the parameters of the *SpATS* model. The sub-module “*SpATS*” is used to fit and get a summary of the model, visualize the spatial trends in the field, compare variance components of the model and to get an estimate of the broad-sense heritability (as proposed by Rodriguez-Alvarez et al. (2018)). The sub-module “*BLUPs/BLUEs*” returns the predicted/estimated values (see Piepho et al. (2008)) and displays a Cleveland dot plot that ranks the genotypic values for the variable of interest. Finally, this module provides a section to identify outlier observations either from the raw data or from the analysis of residuals. The former, uses the median absolute deviation (MAD) or the interquartile range (IQR) methods to identify outliers. The latter uses the assumption that residuals from the model follow a normal distribution with mean 0 and uses a 99% confidence interval to identify outlier data that fall beyond the range of ± 3 standard deviations from the mean. This sub-module contributes to the data cleaning process.

MLM: It is common to find that agricultural field trials do not have information related with the spatial arrangement of experimental units. For this reason, *Mr.Bean* also offers the possibility to fit an MLM with no spatial correction using *lme4*. In the context of field trials for plant improvement, some experimental designs are regularly used, including completely randomized designs (CRD), randomized complete block designs (RCBD) or resolvable lattice incomplete block designs. *Mr.Bean* comes bundled with the models to easily analyze these designs. In addition, it offers the possibility to specify any other model formula using the *lme4* syntax, which is very similar to the regular mathematical notation for specifying linear models (Bates et al. 2015). This module was built under the same scheme as the module “*Spatial MLM*”. It provides the significance of the fixed effects in the model using the F statistic, variance components and its importance using the likelihood-ratio test, an estimate of the broad-sense heritability (as proposed by Cullis et al. (2006)) and some regularly used criteria to compare different fitted models.

Comparison: This module was designed to compare different variables with a quantitative response. In the context of plant breeding experiments, it is common to compare the behavior of one or multiple traits in one or multiple trials. In that sense, the input data set should follow the tidy or long format (check Wickham et al. (2014)). The analyses in this module offer a graphical display of Pearson’s correlation coefficients and a principal components analysis (PCA).

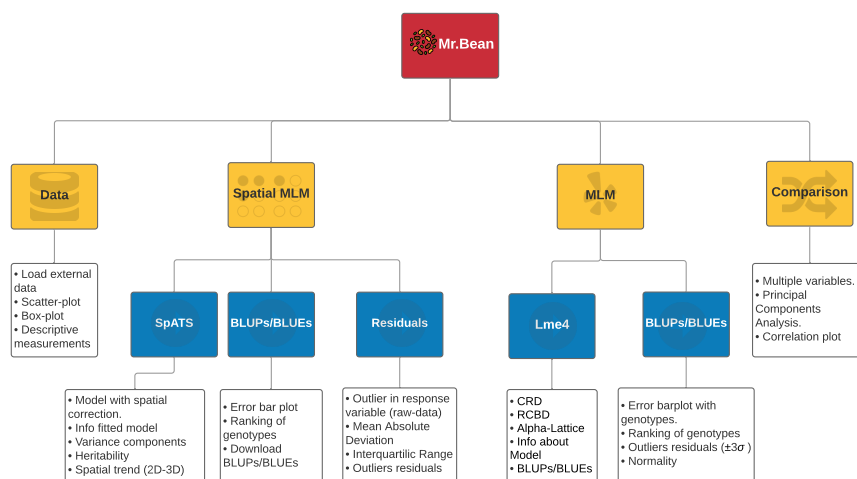


FIGURE 2: Schematic representation of Mr.Bean's modules

3. Implementation

Mr.Bean was developed in R using the package Shiny (R Core Team 2018, Chang et al. 2019), which offers an elegant and powerful web framework to create applications using R with no previous experience using HTML, CSS or JavaScript. Some additional extensions like ShinyJS, shinyWidgets and ShinyBS were also used to improve the interactive experience. Mr.Bean uses the packages SpATS (Rodriguez-Alvarez et al. 2018) and lme4 (Bates et al. 2015) for fitting MLM with or without spatial corrections respectively. The module “MLM” also uses the packages LmerTest and Broom Mixed to calculate additional information about the fitted model. For data management and table visualization the packages dplyr and DT were used. The graphical representations used in Mr.Bean use ggplot, GGally, GGmisc and plotly. Finally, the ‘Comparison’ module uses the packages factoextra and FactomineR for PCA and graphical visualization.

4. Conclusion

Mr.Bean is a novel tool that allows the user to analyze field trials in a fast and efficient way. It incorporates the spatial modelling capability provided by SpATS, offering a useful alternative for decision making in a plant improvement program.

Future work includes developing this tool as a desktop application that can be easily installed on a regular operating system, allowing the user to run the application with no internet connection.

References

- Bates, D., Mächler, M., Bolker, B. & Walker, S. (2015), ‘Fitting linear mixed-effects models using lme4’, *Journal of Statistical Software* **67**(1), 1–48.
- Chang, W., Cheng, J., Allaire, J., Xie, Y. & McPherson, J. (2019), *shiny: Web Application Framework for R*. R package version 1.3.2.
*<https://CRAN.R-project.org/package=shiny>
- Cullis, B. & Gleeson, A. (1991), ‘Spatial analysis of field experiments-an extension to two dimensions’, *Biometrics* pp. 1449–1460.
- Cullis, B. R., Smith, A. B. & Coombes, N. E. (2006), ‘On the design of early generation variety trials with correlated data’, *Journal of agricultural, biological, and environmental statistics* **11**(4), 381.
- Currie, I. D. & Durban, M. (2002), ‘Flexible smoothing with p-splines: a unified approach’, *Statistical Modelling* **2**(4), 333–349.
- Isik, F., Holland, J. & Maltecca, C. (2017), *Genetic data analysis for plant and animal breeding*, Springer.
- Piepho, H., Möhring, J., Melchinger, A. & Büchse, A. (2008), ‘Blup for phenotypic selection in plant breeding and variety testing’, *Euphytica* **161**(1-2), 209–228.
- Piepho, H. & Williams, E. (2010), ‘Linear variance models for plant breeding trials’, *Plant breeding* **129**(1), 1–8.
- R Core Team (2018), *R: A Language and Environment for Statistical Computing*, R Foundation for Statistical Computing, Vienna, Austria.
*<https://www.R-project.org/>
- Robbins, K., Backlund, J. & Schnelle, K. (2012), ‘Spatial corrections of unreplicated trials using a two-dimensional spline’, *Crop science* **52**(3), 1138–1144.
- Rodriguez-Alvarez, M. X., Boer, M. P., van Eeuwijk, F. A. & Eilers, P. H. (2018), ‘Correcting for spatial heterogeneity in plant breeding experiments with p-splines’, *Spatial Statistics* **23**, 52–71.
- Wickham, H. et al. (2014), ‘Tidy data’, *Journal of Statistical Software* **59**(10), 1–23.