

A Bayesian approach for estimating and checking block designs in
agricultural experiments

by

Josefina Lacasa

B.S., Universidad de Buenos Aires, 2020

A REPORT

submitted in partial fulfillment of the
requirements for the degree

MASTER OF SCIENCE

Department of Statistics
College of Arts and Sciences

KANSAS STATE UNIVERSITY
Manhattan, Kansas

2023

Approved by:

Major Professor
Trevor J. Hefley

Copyright

Josefina Lacasa

2023

Abstract

Spatial blocking is a common technique used in designed agricultural experiments. Spatial blocking, however, requires two important assumptions: 1) the spatial blocks can be clearly delineated within a field by an expert; and 2) the area within each block is homogeneous. Statistical analyses of blocked agricultural experiments often show spatially correlated residuals, suggesting one of the assumptions mentioned above was not met. We propose a model for estimating block membership with data, as an alternative method to account for spatial effects, while preserving the traditional designed experiment framework. We embed a classification and regression tree within a Bayesian statistical model, to estimate block membership. We illustrate possible applications of this approach with some of the most typical scenarios we have encountered in our applied experience in agricultural research with four synthetic data sets and two field data sets. Our hybrid Bayesian-Machine Learning approach can serve one of two purposes: validating the originally designed block layout, or estimating the spatial block layout. Thus, this model provides researchers with a flexible tool for analyzing designed agricultural experiments.

Keywords: Generalized Linear Model, Experimental design, Randomized Block Design.

Table of Contents

Table of Contents	iv
List of Figures	vi
List of Tables	vii
Acknowledgements	vii
Dedication	viii
0.1 Introduction	1
0.2 Methods	3
0.2.1 Simulated Data	4
0.2.2 Data example I: Wheat resistance to Hessian fly damage – Binomial distribution	7
0.2.3 Data example II. Corn yield trial data example – Normal distribution	7
0.3 Results	7
0.3.1 Simulated Data	7
0.3.2 Data example I: Wheat resistance to Hessian fly damage	8
0.3.3 Data example II: Corn yield trial data	8
0.4 Discussion	8
0.5 Conclusions	12
Bibliography	13

A	Relevance of blocks in designed experiments in agriculture	16
B	R code	18
B.1	Sampler for Data example I (Binomial distribution)	18
B.2	Sampler for Data example II (Normal distribution)	22

List of Figures

1	Simulated data	6
2	Results. Estimates for Data example I	9
3	Results. Estimates for Data example II	10
A.1	Relevance of blocks in designed experiments in agriculture	17

List of Tables

1 Summary of simulated datasets. 5

Acknowledgments

The completion of this report, my Master's itself, and all the personal/professional growth that came with it, would have been impossible without the people that helped me along the way.

I am grateful to my Statistics mentor, Trevor Hefley, for his patient advice and guidance throughout the years; and for being courageous enough to trust in me ever since my undergrad days.

I am grateful to my Agronomy mentor, Ignacio Ciampitti, for his trust and for selflessly supporting each one of my ideas.

I am grateful to Christopher Vahl for generously assisting me on my advisory committee.

I am grateful to my labmates, especially Adrian's early advice in 2019, "make sure you study a lot of statistics"!

I am grateful to my fellow students in the Departments of Statistics and Agronomy, for exchanging ideas and making class more enjoyable.

I am grateful to Frau Frank, my high school Math teacher, for introducing me to Bayes' Theorem.

I am grateful to my friends and family for their constant support.

I am grateful to Victoria, who encouraged me to attend graduate school.

I am grateful to Juan for feeding my brain by being my best book recommendation algorithm.

I am grateful to Mariana for pushing me to my best self in school and in life.

Last, I am deeply grateful to my parents, who always made sure there was a math puzzle book under the Christmas tree. Thank you for shaping my curiosity.

Dedication

To Rolfi

0.1 Introduction

Spatial blocking is a widely used technique in designed agricultural experiments. Spatial blocks were introduced to agricultural experiments in the early 20th century as a surrogate for spatial effects ([Verdooren 2020](#)). The concept of spatial blocks consists of dividing the field into distinct areas with similar soil fertility or topography, the blocks, that are expected to modulate plant growth. The notion of spatial blocks is present in several types of experimental designs, like randomized complete block designs (RCBD), incomplete block designs (IBD), split-plot designs, strip plot designs, among others. The primary benefit of blocked experimental designs is an increased precision in the estimation of treatment effects ([Milliken and Johnson 2009](#)), which is normally the primary goal of agricultural experiments. The scientific community in agriculture has adopted spatial blocks so extensively that in 2022, all peer-reviewed publications in a top Agronomy journal ([EJA 2023](#)) included some version of spatial blocks in their experimental designs ([Appendix A](#)). Blocking, however, requires two important assumptions: 1) the spatial blocks can be clearly delineated within a field; and 2) the area within each block is approximately homogeneous ([Mead 1990](#)).

However, statistical analyses of blocked agricultural experiments often show spatially correlated residuals ([Pugh et al. 2019](#); [Brownie et al. 1993](#); [Stroup 2012](#)), suggesting that one or both of the critical assumptions mentioned above were not met. Unaddressed spatial effects can result in loss of accuracy and precision in estimating treatment effects ([Stroup et al. 1994](#)), and thus misleading results ([Stroup et al. 1994](#); [Stroup 2002](#)). For example, agricultural experiments that test differences in plant traits among genotypes (i.e., breeding trials) typically present spatially correlated residuals ([Brownie and Gumpertz 1997](#)). Spatially-correlated residuals usually arise when testing a large number of treatments [≥ 100 , [Piepho et al. \(2006\)](#)] in large blocks ([Borges et al. 2019](#)). Meanwhile, a RCBD assumes that each homogeneous area in space (i.e., each block) is large enough to fit all treatments. However, in practice, RCBD experiments with many treatments may result in large blocks, which are prone to violate the assumption of spatial homogeneity within blocks ([Gusmão](#)

1986; Stroup 2012). Thus, experiments with large blocks often present spatially correlated residuals and may compromise conclusions (Stroup et al. 1994).

We hypothesize that the traditional framework of spatial blocks in designed experiments is appropriate for most field experiments, as proposed initially by R. A. Fischer (1925), but that blocks may be frequently mis-specified. Thus, we propose a model for estimating block membership and block effects with data, for checking the original block layout. We consider a major challenge to estimate block effects from data, that is the difficulty to estimate non-stationary spatial effects (i.e., where abrupt changes in space are found, like block effects in Figure 1a). More precisely, we incorporate classification and regression trees (CART) to account for spatial dependence. The CART is a popular, easily interpretable Machine Learning (ML) algorithm that performs well in scenarios with non-stationarity (Gramacy 2020). Bayesian approaches using treed priors are adequate for prediction purposes (Chipman et al. 2003). Likewise, ML algorithms have been embedded into Bayesian models for experimental data that combine interpretable parameters with ML flexibility to model spatial effects (Shaby and Fink 2012).

The main goal of this report is to introduce a model that can validate the block layout of a designed experiment by estimating block membership and effect of each experimental unit, independently of the original design. We develop a Bayesian model with an embedded CART that estimates block membership, while preserving the traditional designed experiment framework. We illustrate the benefits of our hybrid Bayesian-Machine Learning model (herein termed Bayes-ML) simulating data for four typical scenarios we have encountered in our applied experience in agricultural research. We further elaborate our Bayes-ML model with two field datasets and characterize possible applications.

0.2 Methods

We propose a hybrid Bayesian-Machine Learning approach that can identify spatial trends in the data and classify them into groups that resemble experimental blocks. Our proposed approach embeds a CART algorithm into a Bayesian hierarchical model. We illustrate this approach to simulated binary data, showcasing the model performance for different scenarios that may occur in designed experiments. Then, we elaborate our Bayes-ML approach with two data examples.

We conceptualize our model under the generalized linear model (GLM) framework (Stroup 2012). Under this framework, a model for a blocked experiment can be generally described as

$$y_i \sim [y_i | \mu_i, \psi] \quad (1)$$

$$g(\mu_i) = \mathbf{x}_i' \boldsymbol{\beta} + \eta_i, \quad (2)$$

for $i = 1, \dots, n$ and where y_i is the i th observation that arises from a distribution of the exponential family, denoted by $[\cdot]$. The parameters μ_i and ψ are the expected value and dispersion of y_i , respectively. The function $g(\cdot)$ is the link function (e.g., probit or logit). Its inverse, $g^{-1}(\cdot)$, transforms the linear predictor $\mathbf{x}_i' \boldsymbol{\beta} + \eta_i$ to guarantee that it has the same support as μ_i . In the linear predictor, $\mathbf{x}_i \equiv (x_{0,i}, \dots, x_{p-1,i})$, indicates the treatment of the i th observation, $\boldsymbol{\beta} \equiv (\beta_1, \dots, \beta_p)'$ contains treatment effects, and η_i indicates the block effect or spatial effect of the i th observation. Estimating the parameter $\boldsymbol{\beta}$ is usually the major goal in an agricultural experiment, while an accurate estimation of η_i is convenient to obtain accurate and precise estimates of $\boldsymbol{\beta}$.

Now, note that a vast number of discussions regarding statistical modeling of designed experiments have dealt with modeling elections of the block effects $\boldsymbol{\eta} \equiv (\eta_1, \dots, \eta_n)'$. Our study introduces an alternative model for $\boldsymbol{\eta}$. At the i th observation level, the traditional framework introduced by Fischer (1925) defines $\eta_i = \mathbf{z}_i' \boldsymbol{\tau}$, where $\mathbf{z}_i$ is a vector that indicates

the block membership, and $\boldsymbol{\tau} \equiv (\tau_1, \dots, \tau_k)'$ contains the block effects for all k plots. There are currently two mainstream frameworks to model block effects in $(\boldsymbol{\tau})$. The older framework models fixed block effects, while a newer version models $\boldsymbol{\tau}$ as a random variable arising from a Normal distribution: $\boldsymbol{\tau} \sim N(0, \sigma_\tau^2)$. Models for random block effects can also be expressed as $\boldsymbol{\eta} \sim N(0, \sigma_\tau^2 \mathbf{Z}\mathbf{Z}')$. Research in statistics has deployed many efforts in the ongoing discussion about the most appropriate modeling framework. Further efforts have been deployed to model $\boldsymbol{\eta}$ borrowing methods from spatial statistics (Brownie et al. 1993). In this study, we use the GLM framework in (1) and model $\boldsymbol{\eta}$ in (2) with a tree-generating process as an alternative modeling framework.

In summary, our approach poses an additional alternative for modeling η_i in (2) with a CART that is a function of space, i.e., the experimental plot coordinates. Our Bayes-ML approach estimates $\boldsymbol{\beta}$ and $\boldsymbol{\eta}$ simultaneously. The CART estimates $\boldsymbol{\eta}$ and defines the block membership and their effect size. We use the Gibbs sampler described by Shaby and Fink (2012), and further expanded by Mohankumar and Hefley (2022). Details on R code for the Gibbs sampler is provided in Appendix B.

0.2.1 Simulated Data

We simulated binary data in a spatial domain $D = [0, 1] \times [0, 2]$ that was divided into a grid of 10 rows and 16 columns of equally-sized cells (i.e., experimental plots). This grid mimicks the spatial structure of a designed experiment in agriculture. The i th plot's location can be described by the coordinates its centroid. Each plot was assigned a probability of success p_i using (2) with a probit link function. Then, binary data were drawn from a Bernoulli distribution, $y_i \sim \text{Bernoulli}(p_i)$, for the i th plot with probability of success p_i . To generate each $\mathbf{x}_i$ vector in (2), we randomly assigned the plot's treatment with 0 or 1, meaning 'control' or 'treated'. The treatment effect was set to 0.5, i.e., $\beta_1 = 0$ and $\beta_2 = 0.5$.

We simulated four scenarios with different spatial patterns for η_i (Table 1). The first scenario describes a spatial process that follows the assumptions in one of the most standard

Scenario	Spatial dependence pattern
Block assumptions met	Block effect with 4 blocks
Mis-specified blocks	Block effect with 8 blocks
No block effect	No spatial dependence pattern
Smooth spatial covariance	Exponential covariance function

Table 1: Summary of simulated datasets.

designs in agricultural experiments: the field is divided into four North-to-South blocks. In the first scenario, both assumptions mentioned above are met: 1) the spatial blocks can be clearly delineated within a field; and 2) the area within each block is approximately homogeneous. Thus, we named this first scenario “Block assumptions met”. We named the other three scenarios relative to the first one, envisioning a situation where a field scientist receives a protocol to carry out an experiment with a RCBD (i.e., scenario 1), but observes different ‘true’ spatial dependence patterns. The second scenario, “Mis-specified blocks”, illustrates a situation with homogeneous areas in space (assumption 2 holds), but where blocks were mis-specified (assumption 1 does not hold). In other words, scenario 2 is compatible with traditional experimental design, but where blocks were mis-specified in the original block layout. The third scenario, “No block effect”, describes a situation with no apparent spatial variation. The fourth scenario, “Smooth spatial covariance”, describes a situation where spatial blocks mis-represent the true smooth spatial process (assumption 1 and 2 do not hold). We generated the smooth spatial process with an exponential covariance function. All scenarios are likely to happen in designed experiments in agriculture.

We fitted our hybrid Bayes-ML model to the synthetic datasets presented in Table 1.

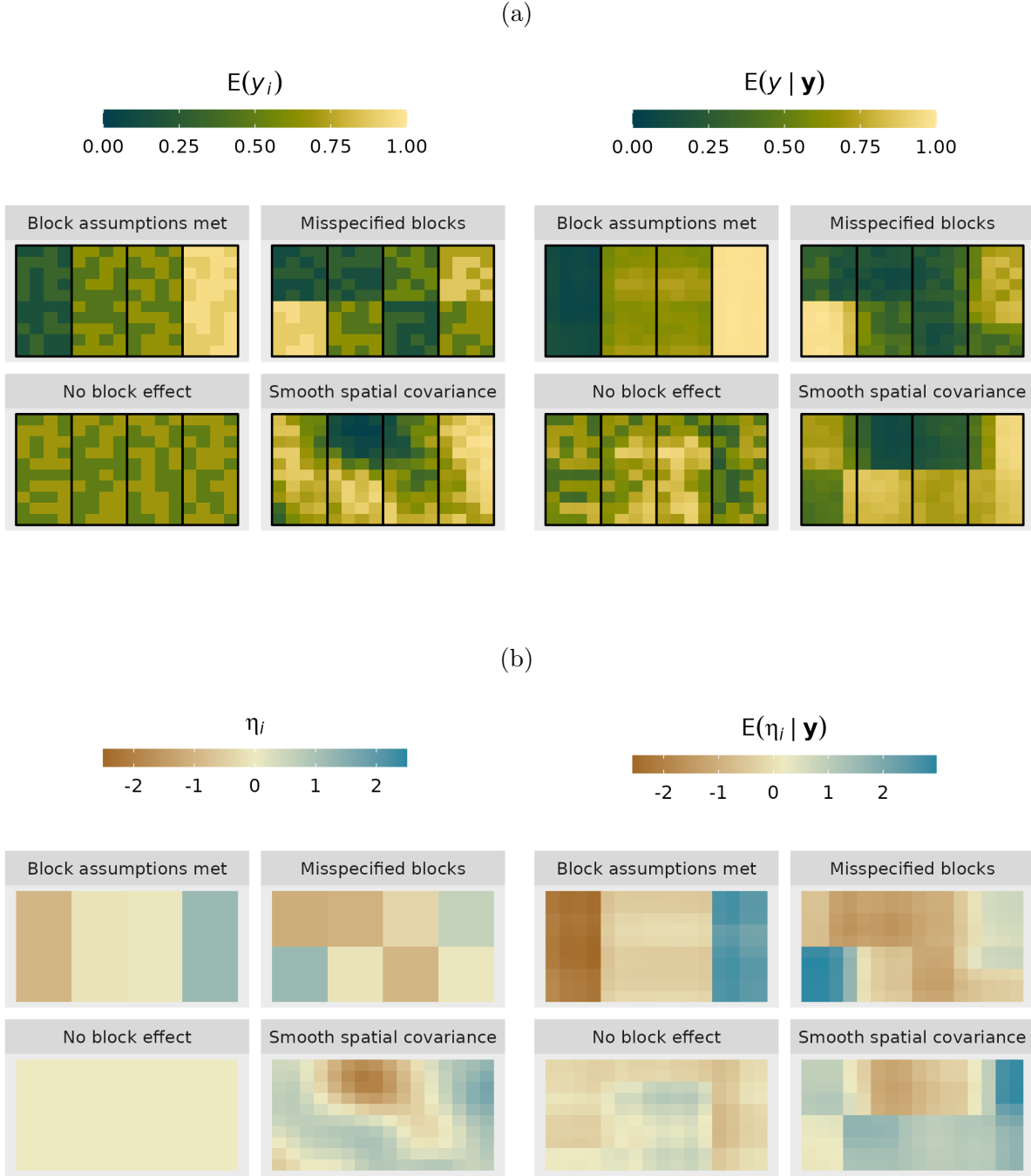


Figure 1: **Simulated data.** Left panels show simulated data and right panels show estimates, for both y (a) and η (b). The names of the scenarios are relative to the first one with 4 vertical blocks, considered a standard design in agronomy. In (a), the black contour lines indicate the block layout for a standard randomized complete block design with four North-to-South blocks.

0.2.2 Data example I: Wheat resistance to Hessian fly damage – Binomial distribution

We elaborate our Bayes-ML model using a dataset with information about wheat resistance to Hessian fly damage from [Gotway and Stroup \(1997\)](#). The objective of the experiment was to evaluate wheat varieties regarding their tolerance to Hessian fly, i.e., the probability of a variety to show plant damage. The data follow a Binomial distribution in (1) and in (2), $g(\cdot)$ is a logit link function. The dataset represents an experiment conducted on a grid of 8 rows by 8 columns of plots, each 3.7×3.7 m. The blocks were delineated by dividing the 8×8 square into four 4×4 quarters (2). For each variety j and i th block, a sample of plants was collected and each plant was classified as damaged or healthy.

0.2.3 Data example II. Corn yield trial data example – Normal distribution

We further elaborate our Bayes-ML model for a corn yield trial data set, as published in [Brownie and Gumpertz \(1997\)](#). The data can be described as normally distributed in (1) with an identity link function in (2). The corn yield trial was divided into a grid of 18×11 plots, numbered in a serpentine fashion. Blocks were delineated by 8×3 incomplete blocks (Figure 3a).

0.3 Results

0.3.1 Simulated Data

Overall, the Bayes-ML model was able to find spatial patterns, in order to verify whether the original block layout was appropriate. We tested a wide range of spatial dependence patterns that we have encountered in our applied agricultural research (Figure 1). The

Bayes-ML model was especially good identifying abrupt changes in space, like the case for mis-specified blocks. Our results show that the Bayes-ML model can verify cases where block assumptions are met, but also indicate that blocks are mis-specified (i.e., assumption 1 is violated but 2 holds) or that the spatial trends may not be synthesized with spatial blocks.

0.3.2 Data example I: Wheat resistance to Hessian fly damage

The original designed block layout of the experiment testing wheat resistance to Hessian fly damage did not match the estimated blocks (Figure 2). The estimates for η_i show roughly four irregularly-sized and irregularly-shaped quadrants. The 95% credible intervals for all wheat variety effects (i.e., all elements in β) included the point estimates of the model with spatial correction in [Gotway and Stroup \(1997\)](#). The interpretability of the block effect shown in Figure 2 is limited, due to the inverse probit function affecting it. Nonetheless, the spatial trend shows that the blocks were mis-specified.

0.3.3 Data example II: Corn yield trial data

The original block layout did not match the estimated blocks in the corn yield trial. While the experiment in hand did have an IBD, the results suggest an IBD with a different block number and layout (Figure 3). The original design included 8 blocks, mostly in East-to-West, but our results indicated approximately 12 irregularly-sized and irregularly-shaped North-to-South blocks. The block effect was relatively large, namely about a 10% of the expected value for yield.

0.4 Discussion

We provide a hybrid Bayesian-Machine Learning modeling framework that poses a new alternative for modeling blocks with a tree-generating process. The Bayes-ML model is based

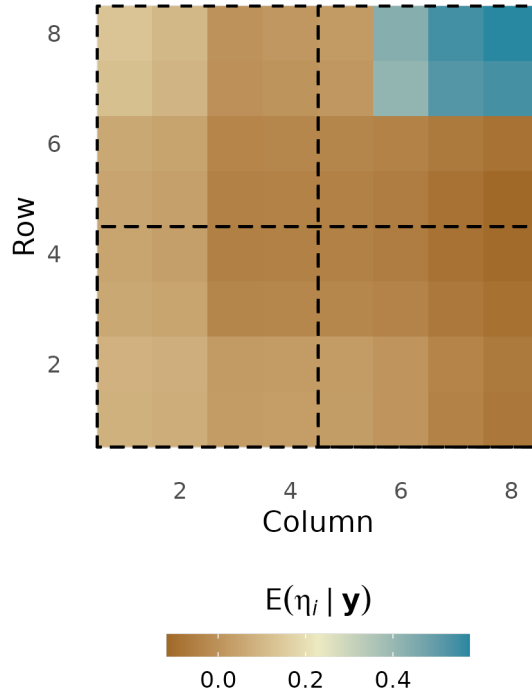


Figure 2: **Data example I.** Designed blocks (black dashed lines) versus estimated block effects (colors) for the first example data set, describing an experiment studying wheat resistance to Hessian fly damage as published in [Gotway and Stroup \(1997\)](#).

on the widely used GLM framework, combined with the ideas in [Shaby and Fink \(2012\)](#), and serves the dual purpose of validating the originally designed block layout, and estimating the spatial block layout. This study may be put in context as a flexible alternative for modeling blocks in designed experiments, within the GLM framework. While the foundational concepts for modeling designed experiments in agriculture were established early in the 20th century, additional features have been integrated steadily down the road ([Verdooren 2020](#)). In fact, many efforts (like this study) have been devoted to modeling the block effects in $\boldsymbol{\eta}$. For example, some of the most common approaches model $\boldsymbol{\eta}$ as ‘fixed’ or ‘random’ effects (borrowing the terminology from the traditional Frequentist framework). Indeed, further variations to modeling $\boldsymbol{\eta}$ can be conceptualized as including spatial blocks *and* smooth spatial effects ([Piepho et al. 2006](#)). For example, methods borrowed from spatial statistics (e.g.,

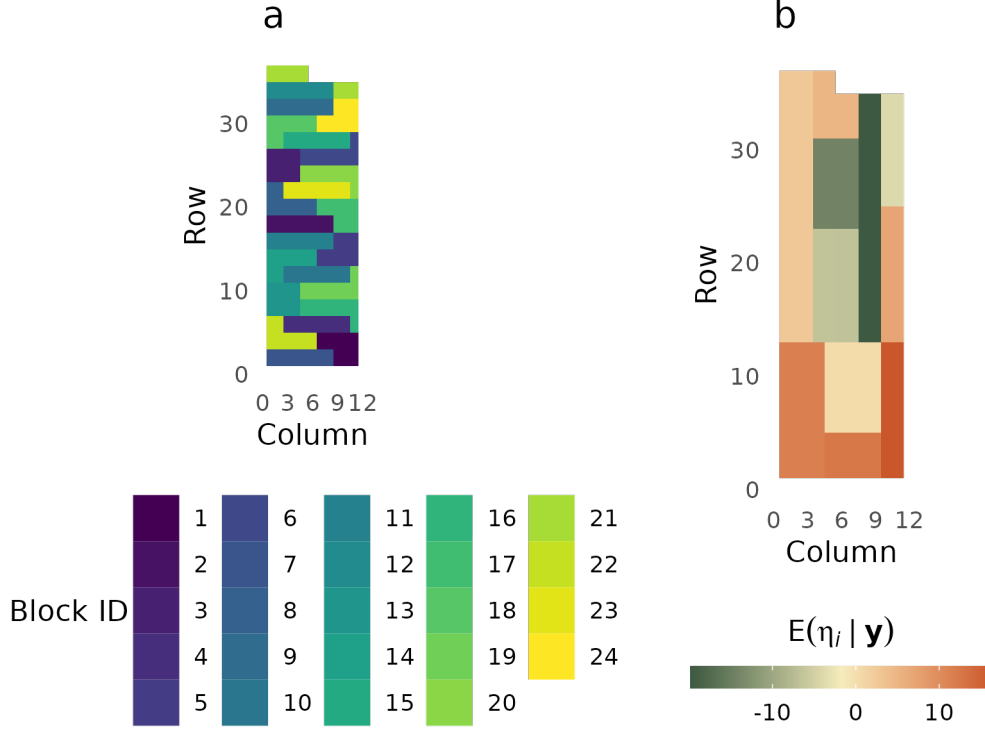


Figure 3: **Data example II.** Original designed experiment blocks (colors in a) versus block effects η_i given the data $\mathbf{y}$ (colors in b) for a corn yield trial as published by [Brownie and Gumpertz \(1997\)](#).

nearest neighbors analyses, random spatial effects, or polynomial functions) were incorporated into agricultural experiment analyses in the 1970s ([Verdooren 2020](#)). The novelty of our approach is that it combines the flexibility of ML with interpretable results and Bayesian uncertainty estimates. Hybrid models that blend ML flexibility with transparent regression techniques have shown great potential for advancing scientific discoveries ([Karpatne et al. 2017](#); [McCormick et al. 2021](#)). Moreover, the tree generating process is compatible with blocking concepts and thus, with the traditional designed experiment framework in agriculture. Such compatibility of the original designed experiment framework in agriculture further enhances its practical applicability. The Bayes-ML approach can serve one of two purposes: validating the originally designed block layout, or estimating the spatial block layout. This flexibility provides researchers with a tool for analyzing designed agricultural

experiments. In the remaining section, we will discuss both validation and estimation purposes, limitations, and further research opportunities.

Our results show that the Bayes-ML framework can be safely implemented to check the originally designed block layout. Recall that the assumptions behind blocked experiments are 1) the spatial blocks can be clearly delineated within a field; and 2) the area within each block is approximately homogeneous. Even though homogeneous areas (i.e., blocks) may exist in the field (i.e., assumption 2 holds), mis-specified blocks are common in agricultural experiment (Stroup et al. 1994; Gusmão 1986; Gotway and Stroup 1997; Brownie and Gumpertz 1997). Stroup (2012) highlights that one important reason for mis-specified blocks is ‘convenience blocking’, that is blocking for practical convenience more than soil fertility or topography. An additional reason for mis-specified blocks may be seasonal effects that drive spatial patterns (Martinez-Feria and Basso 2020). Thus, the Bayes-ML approach may aid researchers with a tool to assess the spatial dependence patterns in their data, and let them decide about the appropriateness of the block layout. We argue that scenarios with mis-specified blocks may benefit from updating their block layout.

Moreover, the Bayes-ML model estimates block layout for data with different patterns of spatial dependence. Our innovative approach poses a solution for cases with non-stationary data and is compatible with traditional experimental design. While most spatial corrections involve borrowing methods from spatial statistics that assume relatively smooth variations in space (i.e., stationary data), the tree-generating process we propose is capable of detecting non-traditional spatial dependence patterns like abrupt changes in space (i.e., non-stationary). In our applied experience, we have found these scenarios may arise in cases with different previous crops or management practices in the same experimental field. Nonetheless, it is worth highlighting that there may be efficiency losses in cases where RCBD or IBD assumptions hold. Replacing block effects for spatial covariance can signify a loss of efficiency when block assumptions actually hold (Stroup 2002). Likewise, although our model successfully estimates block membership, a CART may not be the most efficient method for

estimating treatment effects for field experiments where block assumptions hold. When the block assumptions hold, the Bayes-ML framework may perform best as a preliminary model checking tool, followed by data analysis under the traditional experimental design framework. Optimal inference will ultimately depend on the user deciding to use the Bayes-ML approach for model fitting or model checking.

Future research may build up on the concept of estimating block membership by further developing the tree-generating process in the Bayes-ML model. We have presented a case that uses a simple algorithm to estimate block effects, where neither block size or shape had any restrictions. However, inference from β under too flexible conditions could be limited and thus, further options for modeling the tree-generating process should be explored. Indeed, inference limitations produced by our post-blocking approach may be addressed by incorporating elements of the designed experiment into the tree priors. Avenues for research include (i) designing the tree-generating process in terms of size and shape, (ii) including designed blocks as prior information. We envision future models that can integrate data from experimental design with observed data, to make inference on the treatments of interest.

0.5 Conclusions

We provided a novel approach to check and estimate spatial blocks from data. We use a hybrid Bayesian-Machine Learning model that estimates block membership with a Bayesian model with an embedded CART. Overall, our hybrid Bayesian-Machine Learning approach can serve one of two purposes: validating the originally designed block layout, or estimating the spatial block layout. Thus, this model provides researchers with a flexible tool for analyzing agricultural designed experiments. Future research may be directed towards improving the tree-generating process model.

Bibliography

(2023). European Journal of Agronomy.

Borges, A., González-Reymundez, A., Ernst, O., Cadenazzi, M., Terra, J., and Gutiérrez, L. (2019). Can Spatial Modeling Substitute for Experimental Design in Agricultural Experiments? *Crop Science*, 59(1):44–53.

Brownie, C., Bowman, D. T., and Burton, J. W. (1993). Estimating Spatial Variation in Analysis of Data from Yield Trials: A Comparison of Methods. *Agronomy Journal*, 85(6):1244–1253.

Brownie, C. and Gumpertz, M. L. (1997). Validity of Spatial Analyses for Large Field Trials. *Journal of Agricultural, Biological, and Environmental Statistics*, 2(1):1–23.

Chipman, H. A., George, E. I., and McCulloch, R. E. (2003). Bayesian Treed Generalized Linear Models.

Fischer, R. A. (1925). *Statistical Methods for Research Workers*. Oliver and Boyd, London, UK, 1 edition.

Gotway, C. A. and Stroup, W. W. (1997). A Generalized Linear Model Approach to Spatial Data Analysis and Prediction. *Journal of Agricultural, Biological, and Environmental Statistics*, 2(2):157–178.

Gramacy, R. B. (2020). *Surrogates: Gaussian process modeling, design, and optimization for the applied sciences*. Chapman and Hall/CRC.

Gusmão, L. (1986). Inadequacy of blocking in cultivar yield trials. *Theoretical and Applied Genetics*, 72(1):98–104.

- Karpatne, A., Atluri, G., Faghmous, J. H., Steinbach, M., Banerjee, A., Ganguly, A., Shekhar, S., Samatova, N., and Kumar, V. (2017). Theory-Guided Data Science: A New Paradigm for Scientific Discovery from Data. *{IEEE} Transactions on Knowledge and Data Engineering*, 29(10):2318–2331.
- Martinez-Feria, R. A. and Basso, B. (2020). Unstable crop yields reveal opportunities for site-specific adaptations to climate variability. *Scientific Reports*, 10(1):2885.
- McCormick, R. F., Truong, S. K., Rotundo, J., Gaspar, A. P., Kyle, D., van Eeuwijk, F., and Messina, C. D. (2021). Intercontinental prediction of soybean phenology via hybrid ensemble of knowledge-based and data-driven models. *in silico Plants*, 3(1):diab004.
- Mead, R. (1990). *The design of experiments: statistical principles for practical applications*. Cambridge university press.
- Milliken, G. A. and Johnson, D. E. (2009). *Analysis of messy data volume 1: designed experiments*, volume 1. CRC Press.
- Mohankumar, N. M. and Hefley, T. J. (2022). Using machine learning to model nontraditional spatial dependence in occupancy data. *Ecology*, 103(2):1–9.
- Piepho, H. P., Büchse, A., and Truberg, B. (2006). On the use of multiple lattice designs and α -designs in plant breeding trials. *Plant Breeding*, 125(5):523–528.
- Pugh, N. A., Morgan, C. L. S., Schnell, R., Horn, K., Pietsch, D., and Rooney, W. L. (2019). A statistical evaluation of replicated block designs and spatial variability in sorghum performance trials. *Journal of Crop Improvement*, 33(4):551–566.
- Shaby, B. A. and Fink, D. (2012). Embedding black-box regression techniques into hierarchical Bayesian models. *Journal of Statistical Computation and Simulation*, 82(12):1753–1766.

- Stroup, W. W. (2002). Power analysis based on spatial effects mixed models: A tool for comparing design and analysis strategies in the presence of spatial variability. *Journal of Agricultural, Biological, and Environmental Statistics*, 7(4):491–511.
- Stroup, W. W. (2012). *Generalized linear mixed models: modern concepts, methods and applications*. CRC press.
- Stroup, W. W., Baenziger, P. S., and Muiltze, D. K. (1994). Removing Spatial Variation from Wheat Yield Trials: A Comparison of Methods. *Crop Science*, 34(1):crop-sci1994.0011183X003400010011x.
- Verdooren, L. R. (2020). History of the Statistical Design of Agricultural Experiments. *Journal of Agricultural, Biological and Environmental Statistics*, 25(4):457–486.

Appendix A

Relevance of blocks in designed experiments in agriculture

Publications appearing in European Journal of Agronomy ([EJA 2023](#)) in 2022 were screened for the experimental design of original research papers. A total of 219 papers were retrieved, and 98 were dropped for containing unclear information.

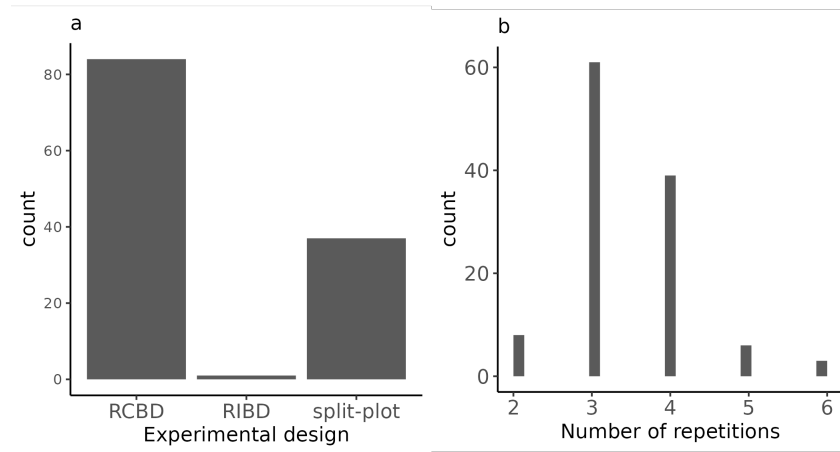


Figure A.1: Experimental designs reported in peer-reviewed publications in a top journal in agronomy (European Journal of Agronomy) published between January 1st and December 31st of 2022: randomized complete block design (RCBD), randomized incomplete block design (RIBD) and split-plot designs (a). Amount of repetitions of the same publications (b).

Appendix B

R code

B.1 Sampler for Data example I (Binomial distribution)

```
#load packages  
library(extrafont)  
library(tidyverse)  
library(truncnorm)  
library(rpart)  
library(MASS)  
library(raster)  
library(RColorBrewer)  
library(rasterVis)  
library(grid)  
library(mosaic)  
library(mgcv)
```

```

library(scico)

#create MCMC function
mcmc_rt <- function(data, m_draws) {
  logit <- function(p){
    log(p/(1-p))
  }

  logit.inv <- function(z){
    exp(z)/(1+ exp(z))
  }

  y <- data$y

  #create design matrix with 1s and 0s
  mm <- lm(y ~ trt - 1, data = data) %>% model.matrix()
  n <- dim(mm)[1]
  X <- cbind(rep(1, n), mm)
  p <- dim(X)[2]

  # s_1 and s_2 are coordinates
  s_1 <- data$s_1
  s_2 <- data$s_2
  df_pred <- data.frame(s_1 = s_1, s_2 = s_2)

```

```

# create output space
samples.beta <- matrix(0, p, m_draws + 1)
samples.Ey <- matrix(, m_draws + 1, n)
samples.block <- matrix(, m_draws + 1, n)

# set initial values
beta.start <- as.vector(c(1, coef( lm(y ~ trt , data = data) )))
sigma2.f <- .5
sigma2.beta <- 1
samples.beta[, 1] <- beta.start
beta <- samples.beta[, 1]

samples <- matrix(m_draws + 1, 1)
samples[1, ] <- c(1) #Starting values for sigma2
sigma2_f <- .5 #Prior variance for beta

for (k in 1:m_draws) {
  if(k==1){n <- length(y); f.hat <- 10}

  # sample z
  z <- rtruncnorm(n,
                  a = ifelse(y == 0, -Inf, 0),
                  b = ifelse(y == 1, Inf, 0),
                  mean =
                      ## treatment effect ##

```

```

X %*% beta +

## block effect ##

f.hat,

sd = 1)

# sample beta
temp.mn <- as.vector(solve(t(X) %*%
                        X + 1/sigma2.beta) %*% t(X) %*% (z - f.hat))
temp.var <- solve(t(X) %*% X + 1/sigma2.beta)
beta <- rmvn(1, temp.mn, temp.var)

# sample eta (here called f)
f <- rnorm(n, (sigma2.f/(sigma2.f + 1)) * (z - X %*% beta),
           (sigma2.f/(sigma2.f + 1))^0.5)

# sample eta.star (here called f.star)
f.star <- rpart(f ~ s_1 + s_2, method = 'anova')
f.hat <- predict(f.star)

block.eff <- predict(f.star, df_pred)

# save
samples.Ey[k+1,] <- pnorm(X %*% beta + block.eff)
samples.block[k+1,] <- block.eff
samples.beta[, k+1] <- beta

```

```

    if (k%%100==0) print(paste(k, "/", m_draws))
  }

  post <- list(samples.Ey,
               samples.block,
               samples.beta)
  names(post) <- c('Ey', 'block', 'beta')
  return(post)
}

```

B.2 Sampler for Data example II (Normal distribution)

```

library(extrafont)
library(tidyverse)
library(truncnorm)
library(rpart)
library(MASS)
library(raster)
library(RColorBrewer)
library(rasterVis)
library(grid)
library(mosaic)
library(mgcv)

```

```

library(scico)

# load data
y <- df_brownie$YIELD
# set treatment IDs
X <- model.matrix(lm(YIELD ~ factor(ENT) , data = df_brownie))
s_1 <- df_brownie$ROW
s_2 <- df_brownie$COL
#dataframe for predictions
df.pred <- data.frame(s_1 = s_1,s_2 = s_2)

# preliminary
# data features
n <- dim(X)[1]
p <- dim(X)[2]
m.draws <- 3000
n.pred <- dim(df.pred)[1]

# output space
samples.beta <- matrix(0, p, m.draws + 1)
samples.Ey <- matrix(, m.draws + 1, n.pred)
samples.block <- matrix(, m.draws + 1, n.pred)

# set initial values
f.hat <- 10
beta.start <- coef(lm(YIELD ~ factor(ENT) , data = df_brownie))

```



```

sigma2.f <- 2
sigma2.beta <- 1
samples.beta[, 1] <- beta.start
beta <- samples.beta[, 1]

set.seed(98234)
for (k in 1:m.draws) {

  # sample beta
  temp.mn <- as.vector(solve(t(X) %*% X +
1/sigma2.beta) %*% t(X) %*% (y - f.hat))
  temp.var <- solve(t(X) %*% X + 1/sigma2.beta)
  beta <- rmvn(1, temp.mn, temp.var)

  # sample eta
  f <- rnorm(n, (sigma2.f/(sigma2.f + 1)) *
(y - X %*% beta), (sigma2.f/(sigma2.f + 1))^0.5)

  # sample eta.star
  f.star <- rpart(f~s_1+s_2,method='anova')
  f.hat <- predict(f.star)

  block.eff <- predict(f.star, df.pred)

  # save
  samples.Ey[k+1,] <- X %*% beta + block.eff

```

```

    samples.block[k+1,] <- block.eff
    samples.beta[, k+1] <- beta

    if(k%%100==0)print(paste(k, "/", m.draws))
  }

```