The International Society of Precision Agriculture presents the
15th International Conference on
Precision Agriculture
26–29 JUNE 2022
Minneapolis Marriott City Center | Minneapolis, Minnesota USA

Where to Put Treatments for On-Farm Experimentation

Davood Poursina[a*]

B. Wade Brorsen[a]

[a] Department of Agricultural Economics, Oklahoma State University, Stillwater, OK, U.S.A

*davood.poursina@okstate.edu

A paper from the Proceedings of the
15th International Conference on Precision Agriculture
June 26-29, 2022
Minneapolis, Minnesota, United States

Abstract.

Conducting on-farm experiments have become more cost-efficient with new technology. The spatial correlations in on-farm experiments will affect optimal experimental designs. In this paper, we address the problem of allocating the optimal locations of treatment levels for a fixed number of replications. Pseudo-Bayesian D optimal designs are obtained for an experiment to collect the data for estimating a non-differentiable production function, linear plateau (LP), with spatially varying coefficients (SVC). Obtained optimal designs are more efficient than classic designs like Latin square, strip plot, and random allocation when estimating the SVC model. The optimal designs were robust to misspecification of the assumed true values for the variance parameters.

Keywords.
Locally D-optimal Design, On-Farm Experimentation, Linear Plateau Model, Spatially Varying Coefficients.

Introduction

On-farm experimentation has become of interest due to advancements in technology. These experiments are not as costly as before, as current machinery can allocate different levels of treatment to specific plots. The main goal of this kind of experiment is to obtain a site-specific nutrient level. One unanswered question for on-farm experimentation is how the treatments should be allocated in the first place such that the appropriate model can be estimated precisely. Poursina and Brorsen (2022) obtained the nearly Ds-Optimal allocation design of the experiment for the linear in parameters model with SVC. They showed that optimal allocations are more informative than the standard designs such as strip plots or random assignment. They use the Ds-Optimality criterion that maximizes the determinant of the Fisher information matrix for a subset of parameters. In some cases, however, the linear plateau (LP) with spatially varying coefficients (SVC) is an appropriate modeling scheme, and Poursina and Brorsen's linear in parameters assumption that greatly simplifies the calculations would not apply.

Xiaofei, et al. (2021) made a Monte Carlo simulation over several classic experimental designs. They showed that the randomly assigned location to treatment levels is not very informative in many cases. They use the quadratic plateau model for simulating data and concluded that blocking can increase the information gained from an experiment, and as a result, gained profit for these kinds of designs increases.

The LP model creates two problems. Firstly, since it is a non-linear non-differentiable model, the information matrix cannot be derived directly from the likelihood function of the model. Secondly, the Fisher information matrix depends on the model's unknown parameters. This paper uses a two-step approximation to obtain the Fisher information matrix. The LP model is approximated with a differentiable model at the first step based on the assumption that the true value of the optimal nitrogen levels is known. Then this function is linearized to find the Fisher information matrix. We employ the pseudo-Bayesian optimal design approach for the second problem that considers the parameters' best initial guess and distribution for the optimal nitrogen value to aid our design in being robust against misspecification of the parameters. We also examine the robustness of the obtained design against misspecification in the parameters' true values.

There is a vast literature on experimental design for agronomy purposes (Casler, 2015; Clewer and Scarisbrick, 2013). The main goal of experimental design is to select the treatment levels and allocate frequencies for each treatment level such that the production function can be estimated with the most possible precision (Hanrahan and Lu, 2006). The classical optimal design usually considers the independence of the observations; however, this assumption is questionable in most agricultural experiments.

Fast algorithm and computational power make the spatially varying coefficients models feasible for large data sets in a practical situation (Gelfand, et al., 2003; Mu, et al., 2018; Murakami, et al., 2019). Selecting treatment levels and replications is essential in these models, but treatment locations also play a vital role.

Atkinson and Haines (1996) obtained the optimal design for the LP model. They showed that this design has three treatment levels at zero, optimal value, and one point on the plateau. Brorsen and Richter (2012) obtained the optimal design of the experiment for the stochastic LP model. To find the optimal nitrogen value variance, they consider two estimation methods, the linearized response method, and the Monte Carlo simulation. They assumed that there is one small value for the treatment, one at the plateau, and then search for the last design point.

Ng'ombe and Brorsen (2019) considered the Bayesian sampling system to overcome the problem of optimal design for the stochastic LP model over several years. They concluded that conducting experiments on a small portion of the field for up to 6 years is optimal based on an economic point of view.

The papers mentioned above do not consider the potential spatial behavior in the model's parameters. The spatial behavior in the model parameters changes the amount of information obtained from the data. In addition, the location of the treatment levels affects the information, treatment level, and the number of replications. Hence, we should consider the locations of the level of the treatment when we find the optimal experimental design for estimating SVC models.

The obtained optimal designs are for a 4×4 square. We consider four equally weighted levels of nitrogen equal to 20, 50, 100, and 150 and uniform prior distribution on the optimal nitrogen value between 90 to 110. We adjusted the method suggested by Poursina and Brorsen (2022) to find pseudo–Bayesian Ds optimal designs. They showed that their method is feasible for any shape of the field. Optimal designs are far more informative than the standard experimental designs and do not impose any extra cost on the application system (assuming that the applicator can switch nitrogen levels between plots). Inefficient designs like strip plots made sense when machinery could not easily apply nutrients and seeds at different levels within a field. Completely random designs only have 50% efficiency on average. We can learn much more from our experiments simply by putting more care into designing the location of each treatment.

Spatially Varying Coefficient Linear Plateau Model and Information Matrix

The LP model is widely used in agricultural applications. Hermesch, et al. (1998) used this model to describe pig growth. Ouedraogo and Brorsen (2014) use the Bayesian method to estimate the optimal nitrogen value in an LP model. Tembo, et al. (2008) employ the stochastic LP model to describe the crop production function response to nitrogen. They also mentioned that the plateau term in this model might vary across the field and year. Poursina and Brorsen (2021) consider the spatial behavior in both the intercept and the plateau parts. They considered three different spatial correlation matrices and showed that using the precision matrix can improve the run time in the Bayesian Kriging method.

Assume that the model is

$$y_i = \min(\beta_{0i} + \beta_{1i} x_i, P_i) + \epsilon_i \tag{1}$$

where y_i shows the yield in each location, β_{0i}, and β_{1i} are the location intercept and slope, respectively and P_i is the plateau value. Let also assume that $\epsilon \sim N(0, \sigma^2 I)$; and

$$f(\boldsymbol{\beta}_i(s)|\boldsymbol{\Psi}_i) \sim N(\mathbf{0}, \boldsymbol{\Psi}_i) \tag{2}$$

where $\boldsymbol{\Psi}_i$ is the covariance matrix that describes the spatial behavior of the parameters; hence all the spatial behavior can be captured by the spatial behavior in the parameters. We assume that the Bayesian method is used to estimate the parameters in (1) and the asymptotic variance of the parameters can be estimated by the inverse of the Fisher information matrix.

The information matrix for the linear/nonlinear model

$$y = \eta(x, \beta) + \epsilon \tag{3}$$

is equal to

$$I = \frac{\partial \eta(x,\beta)}{\partial \beta} \, \Omega^{-1} \left(\frac{\partial \eta(x,\beta)}{\partial \beta}\right)^T \tag{4}$$

where $\frac{\partial \eta(x,\beta)}{\partial \beta}$ is the partial derivatives of the production function with respect to the parameters, and Ω is the variance-covariance matrix of the response variable y. Based on equation (4), we can calculate the information matrix, however, we need one more step here since the production function is not differentiable.

The $\min(x, y) = \frac{1}{2}(x + y - |x - y|)$ so the LP model can be written as

$$\min(a+bx, plateau) = \frac{1}{2}(a+bx+plateau-|a+bx-plateau|).$$

Hence, if the value of the optimal nitrogen value is known priorri, then the derivative of the linear plateau is

$$\begin{bmatrix} 1 & x & 0 \\ 1 & x & 0 \\ \vdots & \vdots & \vdots \\ & \overset{before\ x^*}{\longleftrightarrow} & \\ 0 & 0 & 1 \\ 0 & 0 & 1 \end{bmatrix}.$$

After linearizing the LP model and integrating out the parameters, the variance of the vector y can be approximated by

$$\Omega = \left(\boldsymbol{D}_1\boldsymbol{\Psi}_0\boldsymbol{D}_1 + \boldsymbol{D}_x\boldsymbol{\Psi}_1\boldsymbol{D}_x + \boldsymbol{D}_p\boldsymbol{\Psi}_2\boldsymbol{D}_p + \tau^2\boldsymbol{I}\right) \quad (5)$$

where $\boldsymbol{D}_1, \boldsymbol{D}_x$ and $\boldsymbol{D}_p$ are the diagonal matrices with diagonal elements of columns of the $\frac{\partial\eta(x,\beta)}{\partial\beta}$.

Optimal Design

The design of experiments has a rich literature that started from the early 20th century (Smith, 1918). However, the theory of optimal design was proven in papers by Kiefer (1974). Assume that the experimenters can run N experiments. The theory of optimal design deals with selecting, not necessarily distinct, N treatment level to collect the data. In classic design of experiment, independency are usually assumed for fitting the production function. When the data are spatially correlated, the location of the experiments is also added to the selecting variables. Hence, we want to maximize the information gained from the experiment by selecting the treatment and their locatoins. In statistics literature, the standard criterion is a D-optimal design that maximizes the determinant of the Fisher information matrix. Maximizing the Fisher information matrix's determinant is equivalent to minimizing the volume of the confidence ellipsoid of the estimated parameters. So, the main goal for the experimental design is

$$\max_{\xi} \phi(Y, \boldsymbol{\theta}, \xi)$$

where Y is the response variable, $\boldsymbol{\theta}$ is the vector of parameters, ξ shows the experimental design and the ϕ is the selected criterion.

For the linear models, the Fisher information matrix does not depend on the unknown model parameters. So, a closed form for these models can be derived. However, the Fisher information matrix for non-linear models and SVC models does depend on the model parameters. Hence, chiken and egg situation is occured. Wethe main goal is to find the optimal design to estimate the model parameters, but the design itself depends on these parameters. One solution for this situation is to assume the parameters are known and find the locally optimal designs (Chernoff, 1953; Yang and Stufken, 2012). An alternative solution is Bayesian method. Bayesian method consider a prior distribution on the unknwon parameters and maximize the expected value of the Fisher information matrix (Chaloner and Verdinelli, 1995; Dette and Neugebauer, 1997). In the following section, the robustness of these locally optimal designs are investigated against the misspecification for true value of the parameters.

We consider two spatial covariance matrices: SAR and Exponential. The Exponential covariance function is

$$\mathrm{cov}\left(\beta_r(s_i), \beta_r(s_j)\right) = \sigma_r^2 \exp\left(-\frac{d_{ij}}{\rho_r}\right) \ , r = 0,1,2 \quad (6)$$

where d_{ij} is the distance between location i and j, σ shows the sill, and the ρ is the effective spatial range. SAR covariance function is

$$\boldsymbol{\Sigma}_r^{SAR} = \sigma_r^2((\boldsymbol{I} - \rho_r \boldsymbol{W}^{*\prime})(\boldsymbol{I} - \rho_r \boldsymbol{W}^{*}))^{-1} \ , r = 0,1,2 \quad (7)$$

where σ_r^2 is the common variance for the r'th parameters, $\boldsymbol{W}^*$ is the row standardized contiguity matrix, and ρ_r shows the amount of spatial dependence.

Application and Results

Assume that the level of treatment and number of replications are fixed, and we want to select the treatment location in a 4×4 field. The information matrix given in (4) depends on the unknown parameters of the spatial variance as well as the optimal nitrogen value. We consider the pseudo-Bayesian method to overcome this problem.For the pseudo-Bayesian method, we assume that true values of the covariance parameters are known (locally optimal), and uniform prior distribution between 90 and 110 on the optimal nitrogen values. With this prior distribution, we can be sure that there is at least one point in the plateau and one point in the optimal nitrogen distribution range.

Suppose we want to allocate 4 equally weighted treatment levels in a 4×4 square. There are more than 63 million possible permutations that need 8.1 gigabytes of RAM. All the possible permutation are calculated by the RcppAlgos package in R (Wood, 2020). We assume that the true values of the spatial behavior are known for all matrices and equal to $\rho_1 = 0.8$, $\sigma_1^2 = 20$, $\rho_N = 0.9$, $\sigma_N^2 = 10$, $\rho_p = 0.8$, and $\sigma_p^2 = 30$ for the SAR and these values are $4, 0.1, 3, 0.5, 10$, and 1 for exponential, respectively. The value of $\sigma_\epsilon^2 = 1$ for both variance matrices. Figures 1 and 2 show the optimal allocation for the SAR and Exponetial.

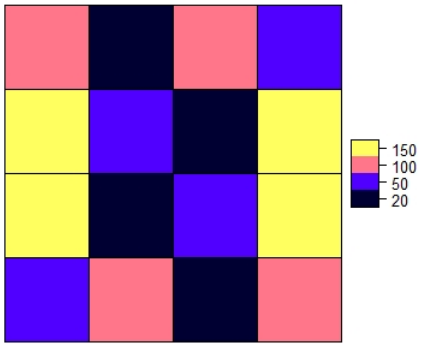

Figure 1. Best allocation for EXP covariance

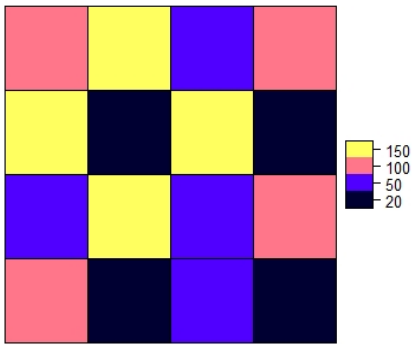

Figure 2. Best allocation for SAR covariance

In order to investigate the efficiency of the obtained designs, we use relative efficiency used by

Poursina and Talebi (2014). Relative efficiency of two design $\xi_1$and ξ_2 can be calculated by

$$E = \frac{|M(\xi_1,\theta)|}{|M(\xi_2,\theta)|}. \quad (8)$$

Since for classic designs like Latin square, strip plot, and randomly assigned designs, there is more than one allocation possible, we consider the average efficiency of these designs over several possible conditions. Table 1 illustrates the efficiency of the classic designs to the optimal allocation of the locations.

Table 1. Efficiency of designs for 16 Locations

Design of experiment	Number of designs	Average Efficiency for SAR	Average Efficiency for Exponential	Maximum Efficiency SAR	Maximum Efficiency Exponential
Latin Square	576	71.29	61.81	0.94	0.91
Randomly Assigned	1000	40.85	51.01	0.48	0.96
Strip plot	24	37.59	28.89	0.56	0.51

In this paper, we consider the pseudo-Bayesian method. Hence the true values of the variance parameters are assumed to be known. Another important issue that should be addressed here is the robustness of obtained designs against the misspecification of the variance parameters. We consider two different scenarios here. Firstly, the misspecification is not very severe and the $\boldsymbol{\theta}_{real} = \frac{3}{4}\boldsymbol{\theta}_{assumed}$ or $\boldsymbol{\theta}_{real} = \frac{4}{3}\boldsymbol{\theta}_{assumed}$. In another scenario, we consider the severe misspecification where $\boldsymbol{\theta}_{real} = \frac{1}{10}\boldsymbol{\theta}_{assumed}$ or $\boldsymbol{\theta}_{real} = 10\boldsymbol{\theta}_{assumed}$. In both scenarios, we change the parameters such that we have more variance or less information about the field. Table 2 shows the results of the robustness check against the misspecification.

Table 2. Robustness of Nearly Optimal Designs Against Misspecification

Model	Parameter	Efficiency	Parameter	Efficiency
SAR	$\rho_1 = 0.6$	1	$\rho_1 = 0.08$	0.96
	$\rho_N = 0.675$	1	$\rho_N = 0.09$	1
	$\rho_p = 0.6$	1	$\rho_p = 0.08$	1
	$\tau_1 = 15$	1	$\tau_1 = 2$	0.95
	$\tau_N = 7.5$	1	$\tau_N = 1$	1
	$\tau_p = 22.5$	1	$\tau_p = 3$	1
	$\sigma_\epsilon = 1.33$	1	$\sigma_\epsilon = 10$	1
Exponential	$\rho_1 = 5.33$	1	$\rho_{1=}40$	1
	$\rho_N = 4$	1	$\rho_N = 30$	1
	$\rho_p = 6.66$	1	$\rho_p = 50$	1

$\sigma_1 = 0.13$	1	$\sigma_1 = 1$	0.99
$\sigma_N = 0.66$	1	$\sigma_N = 5$	1
$\sigma_p = 1.33$	1	$\sigma_p = 10$	1
$\sigma_\epsilon = 1.33$	1	$\sigma_\epsilon = 10$	1

Conclusion and Disscusion

By increasing the computational power and faster algorithms, SVC models become more and more feasible. In this paper, we consider a non-differentiable production function (LP). The treatment levels and the replication for them are predetermined by the researchers' budget. The optimal location for the treatment levels is found based on the D optimal criterion that maximizes the determinant of the Fisher information matrix for the LP model with spatially varying coefficients. Current technology lets us apply different levels of treatment without extra cost. So, finding the optimal location for the treatments helps to increase the amount of information gained from an experiment without imposing an extra cost on the project. The obtained designs are far more informative than the classical designs like Latin square, strip plot, and random designs. These designs also are robust against the misspecification of the parameters, but the true functional form of the covariance function should be known in the first place.

Funding

This work was supported by the A.J. and Susan Jacques Chair, the Oklahoma Agricultural Experiment Station, and USDA National Institute of Food and Agriculture [Hatch Project number OKL03170].

References

Atkinson, A.C., and L.M. Haines. (1996). "14 designs for nonlinear and generalized linear models." Handbook of statistics 13:437-475

Brorsen, B.W., and F.G.-C. Richter. (2012). "Experimental designs for estimating plateau-type production functions and economically optimal input levels." Journal of Productivity Analysis 38:45-52

Casler, M.D. (2015). "Fundamentals of experimental design: Guidelines for designing successful experiments." Agronomy Journal 107:692-705

Chaloner, K., and I. Verdinelli. (1995). "Bayesian experimental design: A review." Statistical Science:273-304

Chernoff, H. (1953). "Locally optimal designs for estimating parameters." The Annals of Mathematical Statistics:586-602

Clewer, A.G., and D.H. Scarisbrick. (2013). Practical statistics and experimental design for plant and crop science: John Wiley & Sons.

Dette, H., and H.-M. Neugebauer. (1997). "Bayesian D-optimal designs for exponential regression models." Journal of Statistical Planning and Inference 60:331-349

Gelfand, A.E., H.-J. Kim, C. Sirmans, and S. Banerjee. (2003). "Spatial modeling with spatially varying coefficient processes." Journal of the American Statistical Association 98:387-396

Hanrahan, G., and K. Lu. (2006). "Application of factorial and response surface methodology in modern experimental design and optimization." Critical Reviews in Analytical Chemistry 36:141-151

Hermesch, S., K. Egbert, and J. Eissen. (1998). "Description of a growth model: The linear-

plateau model." AGBU, Wageningen Agricultural University, Animal Breeding and Genetics Group, Wageningen, The Netherlands

Kiefer, J. (1974). "General equivalence theory for optimum designs (approximate theory)." The annals of Statistics:849-879

Mu, J., G. Wang, and L. Wang. (2018). "Estimation and inference in spatially varying coefficient models." Environmetrics 29:e2485

Murakami, D., B. Lu, P. Harris, C. Brunsdon, M. Charlton, T. Nakaya, and D.A. Griffith. (2019). "The importance of scale in spatially varying coefficient modeling." Annals of the American Association of Geographers 109:50-70

Ng'ombe, J.N., and B.W. Brorsen. (2019). "Bayesian optimal dynamic sampling procedures for on-farm field trials." Paper presented at Agricultural & Applied Economics Association Annual Meeting. Atlanta, GA, USA.

Ouedraogo, F.B., and B.W. Brorsen. (2014). "Bayesian estimation of optimal nitrogen rates with a non-normally distributed stochastic plateau function." Paper presented at SAEA. Dallas, Texas, USA.

Poursina, D., and B.W. Brorsen. (2022). "Nearly Ds optimal assigned location design for a linear model with spatially varying coefficients." Paper presented at SAEA. New Orleans, Louisiana, USA.

Poursina, D., and W. Brorsen. (2021). "Site-specific nitrogen recommendation: Using Bayesian Kriging with different correlation matrices." Paper presented at AAEA. Austin, Tx, USA.

Poursina, D., and H. Talebi. (2014). "Modified D-optimal design for logistic model." Journal of Statistical Computation and Simulation 84:428-437

Smith, K. (1918). "On the standard deviations of adjusted and interpolated values of an observed polynomial function and its constants and the guidance they give towards a proper choice of the distribution of observations." Biometrika 12:1-85

Tembo, G., B.W. Brorsen, F.M. Epplin, and E. Tostão. (2008). "Crop input response functions with stochastic plateaus." American Journal of Agricultural Economics 90:424-434

Wood, J., (2020), "RcppAlgos: high performance tools for combinatorics and computational mathematics." https://cran.r-project.org/web/packages/RcppAlgos/index.html

Xiaofei, L., M. Taro, and B. David. (2021). "Economic performances of trial design methods in on-farm precision experimentation: A Monte Carlo evaluation"Unpublished".

Yang, M., and J. Stufken. (2012). "Identifying locally optimal designs for nonlinear models: A simple extension with profound consequences." The annals of Statistics 40:1665-1681