

Copula-based robust optimal block designs

A. Rappold¹ | W.G. Müller¹  | D.C. Woods²

¹Institute of Applied Statistics, Johannes Kepler University Linz, Linz, Austria

²Southampton Statistical Sciences Research Institute, University of Southampton, Southampton, UK

Correspondence

W.G. Müller, Institute of Applied Statistics, Johannes Kepler University Linz, 4040 Linz, Austria.
Email: werner.mueller@jku.at

Funding information

Upper Austrian Government, Grant/Award Number: LIT-2017-4-SEE-001; Austrian Science Fund (FWF), Grant/Award Number: I 3903-N32; UK Engineering and Physical Sciences Research Council, Grant/Award Number: EP/J018317/1

Abstract

Blocking is often used to reduce known variability in designed experiments by collecting together homogeneous experimental units. A common modeling assumption for such experiments is that responses from units within a block are dependent. Accounting for such dependencies in both the design of the experiment and the modeling of the resulting data when the response is not normally distributed can be challenging, particularly in terms of the computation required to find an optimal design. The application of copulas and marginal modeling provides a computationally efficient approach for estimating population-average treatment effects. Motivated by an experiment from materials testing, we develop and demonstrate designs with blocks of size two using copula models. Such designs are also important in applications ranging from microarray experiments to experiments on human eyes or limbs with naturally occurring blocks of size two. We present a methodology for design selection, make comparisons to existing approaches in the literature, and assess the robustness of the designs to modeling assumptions.

KEYWORDS

Binary response, equivalence theorem, generalized linear model, marginal model, pseudo-Bayesian D -optimality

1 | INTRODUCTION AND MOTIVATION

Statistical design of experiments underpins much quantitative work in the biological, physical, and engineering sciences, providing a principled approach to the efficient allocation of (typically sparse) experimental resources to address the aims of the study. Often, experiments aim to understand a process by modeling discrete data, for example, arising from the observation of a binary or count response. For completely randomized experiments, assuming homogeneous experimental units, a generalized linear model (GLM) may provide an appropriate description, and there has been much research into the construction of optimal and efficient designs for multifactor GLMs.^{1-3,4} See the work of Atkinson and Woods⁵ for a comprehensive review.

When heterogeneous experimental units can be grouped into more homogeneous groups, or blocks, accounting for this grouping can improve the precision of inferences made from the experimental data. Methods to find block designs for discrete data have recently been proposed by Woods and vande Ven,⁶ Niaparast and Schwabe,⁷ and Waite and Woods,⁸ among

.....
This is an open access article under the terms of the Creative Commons Attribution License, which permits use, distribution and reproduction in any medium, provided the original work is properly cited.

© 2019 The Authors Applied Stochastic Models in Business and Industry Published by John Wiley & Sons Ltd.

others. Two modeling paradigms have been adopted in the design literature: conditional models where the joint distribution of the data is derived by explicitly including block-specific random effects (eg, generalized linear mixed models⁹) and marginal models, where the dependence structure of the data is specified separately from the marginal distribution of each response (eg, with parameters estimated via generalized estimating equations [GEEs]¹⁰). For the linear model, these two modeling approaches coincide. In this paper, we find optimal designs under a marginal modeling approach when the intrablock dependence structure is defined via a copula. Such models are particularly appropriate when block effects are not of interest in themselves and the aim of the experiment is to understand the effects of treatment factors averaged across blocks. Optimal designs for marginal models using alternative definitions of the dependence structure have been found by other works.^{11–13} Defining dependence via a copula model has the advantages of providing a flexible dependence modeling separate to the marginal probability models, and a more interpretable approach to defining the degree of dependence via commonly used measures; see Section 2 and, in particular, Section 2.3.

Although our methods can be generalized to arbitrary block sizes, we focus on the important special case of experiments with blocks of size two (see the work of Godolphin¹⁴). Such blocks occur routinely in microarray experiments^{15,16} and in experiments on people, for example, with eyes or arms as experimental units.¹⁷ Practical motivation for our work comes from a materials science experiment. In Section 3, we find designs appropriate for aerospace materials testing experiments similar to those performed by our collaborators at the UK Defence Science and Technology Laboratory. The aim of these experiments is to compare the thermal properties of a set of novel materials against a reference material. In particular, one aim is to assess the probability of failure due to the exposure to extreme (high) temperatures. The experiment is performed using an arc jet to heat material samples, which are held in one of six “wedges,” each of which holds a pair of samples on a strut attached to a circular carousel (see Figure 1). Hence, the experiment can be considered as a block design with six blocks, each containing two units. In the particular experiment considered here, six materials were tested, a reference and five novel samples. A variety of measures are made on each tested sample, including a visual inspection of quality to assess material failure, which leads to a binary (pass/fail) response. It is this response for which we find optimal designs.

In common with most nonlinear models, the performance of a given design for a copula-based GLM model may depend on the values of the model parameters that define both the marginal model and the dependence structure. If strong prior information is available, then locally optimal designs can be sought for given values of the model parameters. Otherwise, Bayesian¹⁸ or maximin¹⁹ approaches can be adopted. In common with much of the recent literature on designs for GLMs, we find optimal designs robust to the values of the model parameters via a pseudo-Bayesian approach^{20(ch. 18)} with a classical quantity for design performance averaged with respect to a prior distribution on the parameters. Here, we adopt variants of *D*-optimality for design selection.

The remainder of the paper is organized as follows. In Section 2, we introduce the statistical models we employ, including copulas, and develop design methods for blocked experiments. An illustrative comparison is made to previous design approaches based on GEEs using an example from the work of Woods and vande Ven.⁶ In Section 3, we demonstrate and

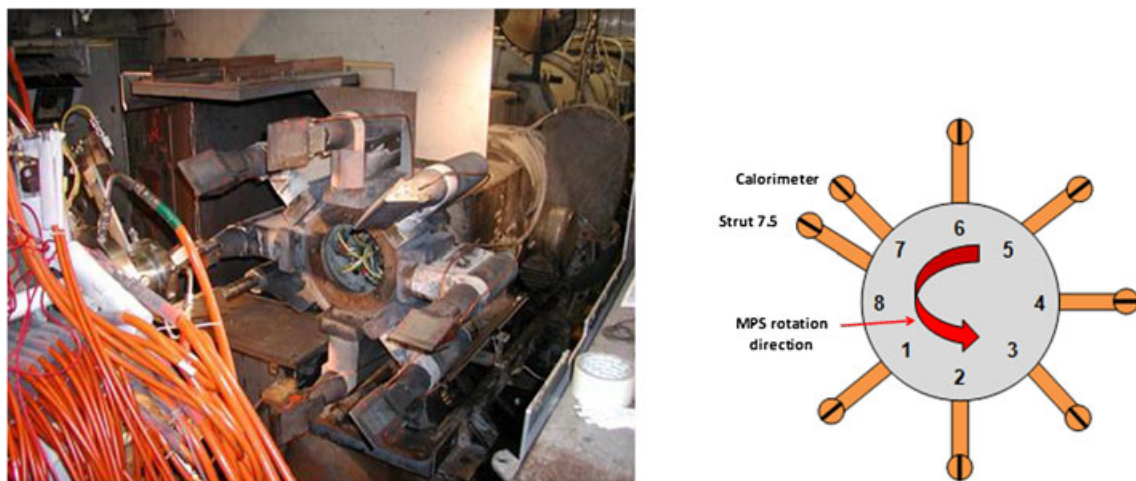


FIGURE 1 Arc jet carousel, struts, and “wedges” (left) and schematic (right). In addition to the six wedges for holding material samples, the carousel had two further wedges used for temperature measurement [Colour figure can be viewed at wileyonlinelibrary.com]

assess our methods via application to the materials testing example. In particular, we show how prior information on the parameters influences the choice of optimal design. We provide a brief discussion and some areas for future work in Section 4.

2 | DESIGNS FOR COPULA-BASED MARGINAL MODELS

Suppose the experiment varies m treatment factors, $\mathbf{x}^T = (x_1, \dots, x_m)$, and the experiment has b blocks of size k ; throughout, our examples will assume $k = 2$. The j th unit in the i th block receives treatment $\mathbf{x}_{ij}^T = (x_{1ij}, \dots, x_{mij})$ ($i = 1, \dots, b; j = 1, \dots, k$) and realizes observation Y_{ij} . The \mathbf{x}_{ij} are chosen from a (typically standardized) compact design space \mathcal{X} , which could also be discrete and are not necessarily distinct. Independence of observations $Y_{ij}, Y_{i'j'}$, for $i, i' = 1, \dots, b; j, j' = 1, \dots, k$, is assumed across blocks ($i \neq i'$), but we allow dependence within a block ($i = i'$), which we describe via a copula model.

2.1 | Statistical modeling via copulas

The problem of specifying a probability model for dependent random variables Y_{i1}, \dots, Y_{jk} can be simplified by expressing the corresponding k -dimensional joint distribution $\mathbf{F}_{Y_{i1}, \dots, Y_{jk}}$ in terms of marginal distributions $F_{Y_{i1}}, \dots, F_{Y_{jk}}$, and an associated k -copula (or dependence function) C defined as follows (cf the work of Nelsen²¹).

Definition 1. A k -copula is a function $C : [0, 1]^k \rightarrow [0, 1]$, $k \geq 2$, with the following properties.

1. (*Uniform margins*) for every $\mathbf{u} \in [0, 1]^k$, if at least one coordinate of \mathbf{u} is 0, then

$$C(\mathbf{u}) = 0,$$

and if all coordinates of \mathbf{u} are 1 except u_i , then

$$C(\mathbf{u}) = u_i.$$

2. (*k -increasing*) for all $\mathbf{a}, \mathbf{b} \in [0, 1]^k$ such that $\mathbf{a} \leq \mathbf{b}$,

$$V_C([\mathbf{a}, \mathbf{b}]) \geq 0,$$

where V_C is the measure induced by C on $[0, 1]^k$.

The connection between a copula and a joint probability distribution is given by Sklar's theorem,²² which affirms that for every k -dimensional joint distribution $F_{Y_{i1}, \dots, Y_{jk}}$ with marginal distributions $F_{Y_{i1}}, \dots, F_{Y_{jk}}$, there exists a k -copula C , defined as in Definition 1, such that

$$\mathbf{F}_{Y_{i1}, \dots, Y_{jk}}(y_1, \dots, y_k) = C(F_{Y_{i1}}(y_1), \dots, F_{Y_{jk}}(y_k)), \quad (1)$$

for all $y_1, \dots, y_k \in \mathbb{R}$. Conversely, if C is a k -copula and F_{Y_1}, \dots, F_{Y_k} are distribution functions, then the function F_{Y_1, \dots, Y_k} given by (1) is a joint distribution with marginals F_{Y_1}, \dots, F_{Y_k} . The copula C may not be unique for discrete margins; however, the practical limitations for statistical purposes are little; cf the work of Genest and Nešlehová.²³

Owing to Sklar's theorem, parametric families of copulas represent a powerful tool to describe the joint relationship between dependent random variables. Selecting the appropriate dependence within an assumed parametric copula family reduces to the selection of copula parameters, which correspond, for example, to a specific measure of association for the modeled random variables. Assuming Y_{i1}, \dots, Y_{ik} are continuous random variables with associated copula $C(\cdot; \alpha)$, one measure of association proposed by Joe²⁴ is given by

$$\tau_k = \frac{1}{2^{k-1} - 1} \left\{ 2^k \int_{[0,1]^k} C(\cdot; \alpha) dC(\cdot; \alpha) - 1 \right\}. \quad (2)$$

Equation (2) is a generalized version of Kendall's τ and, hence, establishes a correspondence between a scalar copula parameter α and the degree of dependence. More details and properties of this quantity, and another more traditional measure of concordance can be found in the work of Genest et al.²⁵

2.2 | Design of experiments for copula models

In common with most work on optimal design of experiments, we base our criterion on the Fisher information matrix (FIM), the inverse of which provides an asymptotic approximation to the variance-covariance matrix of the maximum likelihood estimators of the model parameters.

Let $\zeta_i = (\mathbf{x}_{i1}, \dots, \mathbf{x}_{ik}) \in \mathcal{X}^k$ denote the k treatment vectors assigned to the units in block i ($i = 1, \dots, b; j = 1, \dots, k$). We will work within a class of normalized block designs defined as

$$\xi = \left\{ \begin{array}{c} \zeta_1, \dots, \zeta_n \\ w_1, \dots, w_n \end{array} \right\}, \quad 0 < w_i \leq 1, \quad \sum_{i=1}^n w_i = 1,$$

with $n \leq b$ distinct (support) blocks. As defined, bw_i must be an integer and it represents the replication of the i th support block ($i = 1, \dots, n$). Without loss of generality, we assume the first n blocks in the design corresponding to ζ_1, \dots, ζ_n , with the remaining $b - n$ blocks being replicates. We relax the assumption that bw_i is the integer to find the so-called approximate or continuous designs; see also the works of Cheng²⁶ and Waite and Woods.⁸ Let Ξ denote the space of all possible designs of this form.

Denote the vector of responses from the i th block as

$$\mathbf{Y}_i = (Y_{i1}, \dots, Y_{ik})^T, \quad i = 1, \dots, b,$$

with corresponding expectation vector

$$\boldsymbol{\eta}_i = [\eta(\mathbf{x}_{i1}; \boldsymbol{\beta}), \dots, \eta(\mathbf{x}_{ik}; \boldsymbol{\beta})]^T,$$

where $\eta(\cdot; \cdot)$ is a known function and $\boldsymbol{\beta} = (\beta_1, \dots, \beta_r)^T$ is a vector of unknown parameters requiring estimation. Denote the marginal distribution function for the j th entry in the block as $F_{Y_{ij}}(y_{ij}; \mathbf{x}_{ij}, \boldsymbol{\beta}), j = 1, \dots, k$, and denote the joint distribution, derived via a copula transformation, for the k responses in the i th block as $C(F_{Y_{i1}}, \dots, F_{Y_{ik}}; \boldsymbol{\alpha})$, where $\boldsymbol{\alpha} = (\alpha_1, \dots, \alpha_l)^T$ are unknown (copula) parameters.

The FIM $M(\zeta_i; \boldsymbol{\gamma})$ for the i th block is an $(r + l) \times (r + l)$ matrix with vw th element

$$M(\zeta_i; \boldsymbol{\gamma})_{vw} = \mathbf{E} \left(-\frac{\partial^2}{\partial \gamma_v \partial \gamma_w} \log c_{\mathbf{Y}_i}(\boldsymbol{\eta}_i, \boldsymbol{\alpha}) \right), \quad (3)$$

where $\boldsymbol{\gamma} = (\gamma_1, \dots, \gamma_{r+l})^T = (\beta_1, \dots, \beta_r, \alpha_1, \dots, \alpha_l)^T$ and

$$c_{\mathbf{Y}_i}(\boldsymbol{\eta}_i, \boldsymbol{\alpha}) = \frac{\partial^k}{\partial y_{i1} \dots \partial y_{ik}} C(F_{Y_{i1}}, \dots, F_{Y_{ik}}; \boldsymbol{\alpha})$$

is the joint density function represented through a copula C in accordance with Equation (1). The FIM for an approximate block design ξ is then given by

$$M(\xi; \boldsymbol{\gamma}) = \sum_{i=1}^n w_i M(\zeta_i; \boldsymbol{\gamma}).$$

An optimal design ξ^* maximizes a scalar function $\psi\{M(\xi; \boldsymbol{\gamma})\}$ of the information matrix. Previous work on optimal designs for copulas has focused on finding completely randomized locally optimal designs for multivariate responses, which can be considered as a block design where every unit within a block must receive the same treatment. We generalize these methods to allow different treatments for each unit within each block. Denman et al²⁷ found D -optimal designs for a bivariate response ($k = 2$) that maximized $\psi^D\{M(\xi; \boldsymbol{\gamma})\} = \det M(\xi; \boldsymbol{\gamma})$, and Perrone and Müller²⁸ developed a corresponding equivalence theorem. These methods were extended to the local D_A -criterion and, as a special case, for the D_s -criterion in the work of Perrone et al.²⁹ Other relevant uses of design of experiments in copula models are those by Deldossi et al³⁰ and Durante and Perrone,³¹ but until now, all relied on the availability of a single “best guess” vector of parameter values. It is well recognized that, for many nonlinear models, optimal designs for particular values of the model parameters may be very inefficient under different values; see the work of Woods et al¹ for the case of scalar response GLMs.

To overcome this dependence on assumed parameter values, here we adopt a pseudo-Bayesian approach for constructing block designs. Adopting this approach provides a more robust approach to design than the localized methods provided by Perrone et al.²⁹ Furthermore, our primary interest is typically in s meaningful linear combination of the parameters. Such combinations can be defined as elements of the vector $A^T \boldsymbol{\gamma}$, where A^T is an $s \times (r + l)$ matrix of rank $s < (r + l)$.

If $M(\xi; \boldsymbol{\gamma})$ is nonsingular, the variance-covariance matrix of the maximum likelihood estimator of $A^T \boldsymbol{\gamma}$ is proportional to $A^T \{M(\xi; \boldsymbol{\gamma})\}^{-1} A$. Hence, we define a *robust D_A -optimal block design* ξ^* as the design that maximizes

$$\Psi^D(\xi; G, A) = \int_{\Gamma} \log \det [A^T \{M(\xi; \boldsymbol{\gamma})\}^{-1} A]^{-1} dG(\boldsymbol{\gamma}), \tag{4}$$

where $G(\boldsymbol{\gamma})$ is a proper prior distribution function for $\boldsymbol{\gamma}$ and $\Gamma \subset \mathbb{R}^{r+l}$ is the support of G . See also the work of Woods and vande Ven.⁶

Most often the main interest is in an $s < (r+l)$ -dimensional subset of the parameters. In such a case, a *robust D_s -optimal block design* can be found by maximizing

$$\Psi^D(\xi; G) = \int_{\Gamma} \log \det \{M_{11} - M_{12} M_{22}^{-1} M_{12}^T\} dG(\boldsymbol{\gamma}), \tag{5}$$

following the partition of the information matrix as

$$M(\xi; \boldsymbol{\gamma}) = \begin{pmatrix} M_{11} & M_{12} \\ M_{12}^T & M_{22} \end{pmatrix}.$$

Here, M_{11} is the $(s \times s)$ partition related to the parameters of interest. This criterion follows as a special case of the D_A -criterion with $A^T = (I_s \ 0_{s \times (r+l-s)})$, where I_s is the $s \times s$ identity matrix and $0_{s \times (r+l-s)}$ is the $s \times (r+l-s)$ zero matrix.

We evaluate a design ξ via its *Bayesian efficiencies* under a given criterion, relative to an appropriate reference design ξ^* (see, for example, the work of Waite³²). Under robust D_s -optimality, this efficiency is given by

$$\text{eff}(\xi, \xi^*) = \left(\frac{\exp \int_B \log \det [M_{11}(\xi, \boldsymbol{\gamma}) - M_{12}(\xi, \boldsymbol{\gamma}) M_{22}^{-1}(\xi, \boldsymbol{\gamma}) M_{12}^T(\xi, \boldsymbol{\gamma})] dF(\boldsymbol{\gamma})}{\exp \int_B \log \det [M_{11}(\xi^*, \boldsymbol{\gamma}) - M_{12}(\xi^*, \boldsymbol{\gamma}) M_{22}^{-1}(\xi^*, \boldsymbol{\gamma}) M_{12}^T(\xi^*, \boldsymbol{\gamma})] dF(\boldsymbol{\gamma})} \right)^{1/s}.$$

We find designs that maximize (4) and (5) numerically using a version of the Fedorov-Wynn algorithm,^{33,34} as implemented in R package `docopulae`.³⁵

The optimality of a block design ξ^* under the robust D_A -criterion, regardless of how it was found, can be assessed via application of the following Kiefer-Wolfowitz-type equivalence theorem. The proof is similar to that for completely randomized experiments with multivariate response; see the work of Perrone et al²⁹ for the locally optimal design case.

Theorem 1. *The following properties are equivalent:*

1. ξ^* is D_A -optimal;
2. for every $\zeta \in \mathcal{X}^k$,

$$\int_B \text{tr} [M(\xi^*; \boldsymbol{\gamma})^{-1} A (A^T M(\xi^*; \boldsymbol{\gamma})^{-1} A)^{-1} A^T M(\xi^*; \boldsymbol{\gamma})^{-1} M(\zeta; \boldsymbol{\gamma})] dG(\boldsymbol{\gamma}) \leq s;$$

3. over all $\xi \in \Xi$, the design ξ^* minimizes the function

$$\max_{\zeta \in \mathcal{X}^k} \int_B \text{tr} [M(\xi^*, \boldsymbol{\gamma})^{-1} A (A^T M(\xi^*, \boldsymbol{\gamma})^{-1} A)^{-1} A^T M(\xi^*, \boldsymbol{\gamma})^{-1} M(\zeta; \boldsymbol{\gamma})] dG(\boldsymbol{\gamma}),$$

where Ξ is the set of all possible block designs.

2.3 | Comparative example

We demonstrate robust optimal block designs for copula models using a simple example from the work of Woods and vande Ven,⁶ which allows comparison to the designs found by those authors for a GEE model. We again stress that the copula approach allows us to explicitly specify this dependence in terms of interpretable quantities (eg, generalized Kendall's τ), unlike the GEE model under which the dependence is only specified implicitly.

We find robust designs for a single-factor log-linear regression model assuming Poisson marginal distributions and quadratic linear predictor, implying $\log\{\eta(\mathbf{x}; \boldsymbol{\beta})\} = \beta_0 + \beta_1 x + \beta_2 x^2$. The prior distribution G is uniform on the parameter space $[-1, 1] \times [4, 5] \times [0.5, 1.5]$. In line with our motivating example, we assume blocks of size $k = 2$ and intrablock dependence defined according to one of the following bivariate copula functions.

1. *Product copula*, which represents the independence case,

$$C(u_1, u_2) = u_1 u_2,$$

with generalized Kendall's τ of $\tau_2 = 0$.

2. *Clayton copula*,

$$C_\alpha(u_1, u_2; \alpha) = [\max(u_1^{-\alpha} + u_2^{-\alpha} - 1, 0)]^{-\frac{1}{\alpha}},$$

with $\alpha \in (0, +\infty)$ and generalized $\tau_2 = \frac{\alpha}{\alpha+2}$.

3. *Gumbel copula*,

$$C_\alpha(u_1, u_2; \alpha) = \exp\left(-[(-\ln u_1)^\alpha + (-\ln u_2)^\alpha]^{\frac{1}{\alpha}}\right),$$

with $\alpha \in [1, +\infty)$ and generalized $\tau_2 = \frac{\alpha-1}{\alpha}$.

The first copula is chosen for reference purposes; the latter two represent opposing dependencies in the tails (lower-tail dependence for the Clayton versus upper-tail dependence for the Gumbel). To isolate the effect of the copula structure from the strength of the dependence, we set α for each copula such that the values for Kendall's τ coincide at three levels $\tau_2 = \epsilon, 1/10, 1/3$, respectively. Here, $\epsilon = 10^{-9} > 0$ is a small number to approximate the zero case but avoids singularity issues. Note that, as blocks always consists of pairs of points, here the design space for each block is $[-1, 1]^2$.

To find robust D -optimal designs, objective function (4) was evaluated using quadrature.³⁶ Optimal designs under the Clayton and Gumbel copulas are shown in Figure 2 and demonstrate that increasing the generalized dependence (ie, increasing τ_2) leads to designs placing more weight on support blocks with points on the edge of the design space. All the designs display a “mirror-image” structure, with all design points having $x > 0$. These features are common in designs for Poisson regression (see the work of Russell et al⁴). The designs found under the Gumbel copula tend to include more support blocks, but the pattern in the changes to these blocks as τ_2 is increased is similar for both copulas.

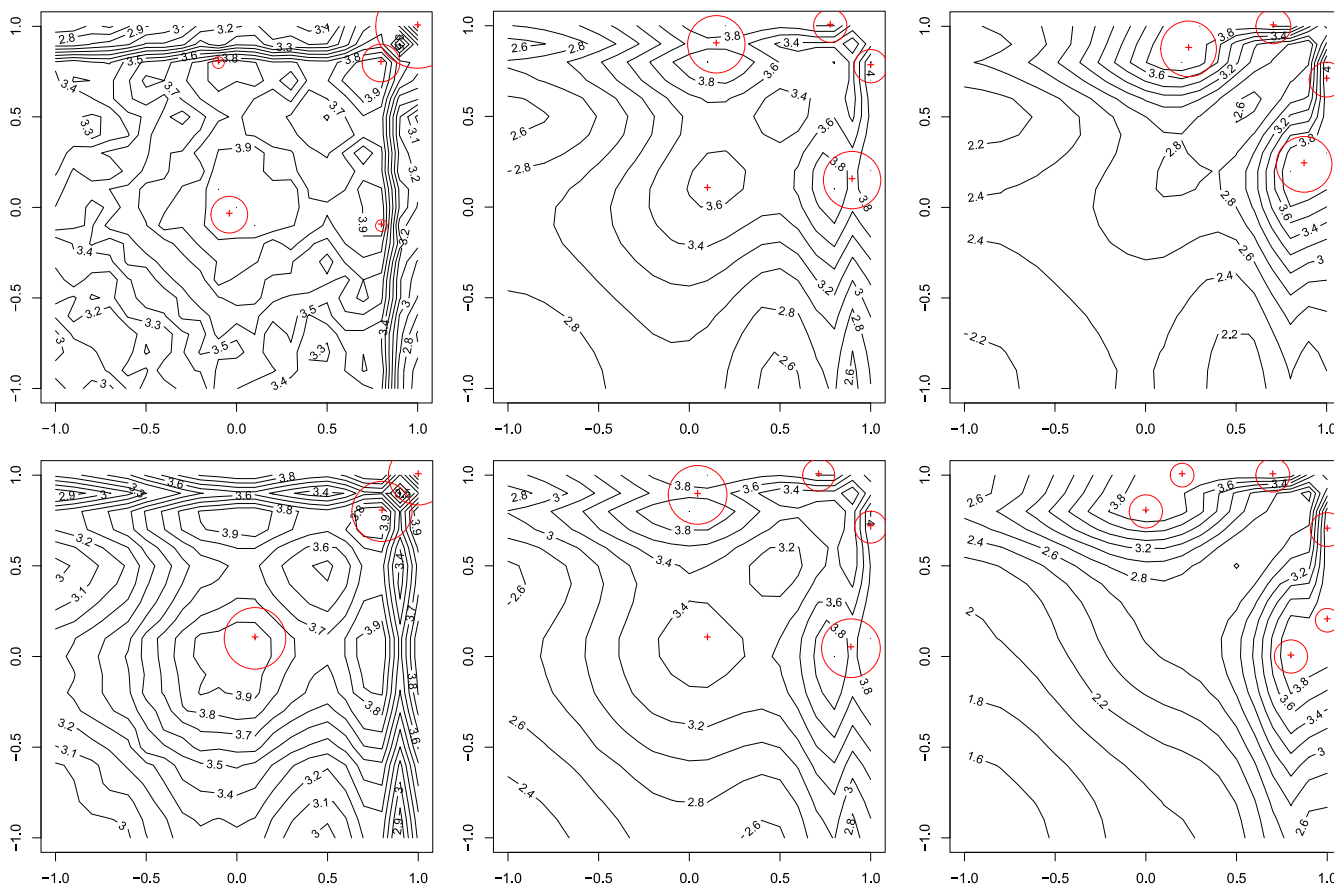


FIGURE 2 Optimal designs for the comparative example. (Crosses) design points, circles diameters are proportional to design weight); the axes correspond to each design point within a block. (Rows) Clayton and Gumbel copula; column levels $\tau_2 = \epsilon, 1/10, 1/3$ [Colour figure can be viewed at wileyonlinelibrary.com]

TABLE 1 D -efficiencies under various copula models for design (6), which was found assuming a generalized estimating equation (GEE) model

Independence	Clayton, $\tau_2 = \epsilon$	Clayton, $\tau_2 = 1/3$	Gumbel, $\tau_2 = \epsilon > 0$	Gumbel, $\tau_2 = 1/3$
96.48%	89.85%	84.41%	95.55%	92.96%

For reference purposes, the optimal design using the independence copula, ie, an optimal design assuming no block effect, was evaluated. It showed little difference to setting the nominal level for $\tau_2 = 0$ for a particular copula. In particular, the D -efficiencies (with respect to the reference design assuming no block effect) for the Clayton and Gumbel model were 96.3% and 99.7%, respectively. This efficiency expectedly decreases as the association within the block increases, for $\tau_2 = 1/3$; for instance, it is already down to 65.0% and 61.3%, respectively.

In the work of Woods and vande Ven,⁶ robust D -optimal designs were found under the same Poisson marginal models and prior distribution but with the dependence described using a GEE approach with an exchangeable correlation matrix and pairwise working correlation of 0.5. The optimal design found was given by

$$\xi^* = \left\{ \begin{array}{ccc} (.03, 1) & (1, .60) & (-.40, .78) \\ .355 & .310 & .335 \end{array} \right\}. \quad (6)$$

That is, for example, the first support block is $\zeta_1 = (0.03, 1)$. This design is somewhat different in structure to the copula designs, without the same mirror structure. Quantitatively, the comparison shows the efficiencies under various scenarios given in Table 1. Surprisingly, the design from the work of Woods and vande Ven⁶ seems to be most compatible with an independence assumption.

3 | APPLICATION TO THE MATERIALS EXAMPLE

In this section, we return to the materials testing example to find and assess designs for comparing six materials in block of size two under a variety of modeling assumptions. The measured response is binary, with each material sample either passing or failing a visual check. We label the five novel materials as “treatments,” with the reference material considered as a control. Marginally, we assume a logistic regression to model the differences between materials set up as

$$Y_{ij} \sim \text{Bernoulli} \{ \eta(\mathbf{x}_{ij}; \boldsymbol{\beta}) \}; \quad \eta(\mathbf{x}_{ij}; \boldsymbol{\beta}) = \text{expit} \left(\beta_0 + \sum_{l=1}^5 \beta_l x_{ijl} \right),$$

where $\text{expit}(u) = 1/\{1 + \exp(-u)\}$, Y_{ij} is the binary response from the i th unit in the j th block ($i = 1, 2; j = 1, \dots, b$), $\eta(\mathbf{x}_{ij}; \boldsymbol{\beta})$ is the associated probability of success, x_{ijl} is an indicator variable taking the value 1 if the i th unit in the j th block was assigned treatment l ($l = 1, \dots, 5$) and 0 otherwise, and β_0, \dots, β_5 are unknown parameters to be estimated. Here, β_0 is the logit for the reference material, with β_l being the difference in expected response, on the logit scale, between the reference material and the l th novel material or treatment.

The choice of copula and the strength of intrablock association makes little difference to the design selected. However, assuming different marginal models and adopting a local or pseudo-Bayesian approach has a strong impact on the designs. The impact of the marginal model here is not surprising, as the degree of dependence between binary random variables is also strongly determined by their marginal distributions.^{37(ch. 7)} Example designs for the Gumbel copula are shown in Figure 3. Note that the numbers 1, \dots , 6 must be understood as nominal labels.

With a null marginal model, ie, $\boldsymbol{\beta}^T = (0, 0, 0, 0, 0, 0)$, when the response variance is constant, the locally D -optimal design contains all material combinations, excluding those blocks containing replicates of a single treatment. This design would also be optimal under a linear model with constant error variance. For different assumed parameter vectors, for example $\boldsymbol{\beta}^T = (0, -1, 2, -3, 4, -5)$, the optimal design contains only a few distinct treatment and treatment control combinations, with differing weights; here, (1,2), (3,4), (4,5), and (5,6) are selected. The (pseudo)-Bayesian approach, assuming a continuous uniform prior on $[-1, 1]$ for each β_l ($l = 0, \dots, 5$) yields designs with unequal weights spread across all material combinations. Changing to a continuous uniform prior on the space $[-1, 1] \times [-2, 0] \times [1, 3] \times [-4, -2] \times [3, 5] \times [-6, -4]$, so centered on $\boldsymbol{\beta}^T = (0, -1, 2, -3, 4, -5)$, adjusts the weighting of the support blocks to give more emphasis on comparing treatments 2 and 4 and treatments 3 and 5. These pairs of treatments have coefficients with differences to the control with the same sign. This last prior distribution is representative of the type of materials study that might be conducted in practice, with differing prior beliefs about the size and direction of the difference between the expected response from each treatment and the reference material.

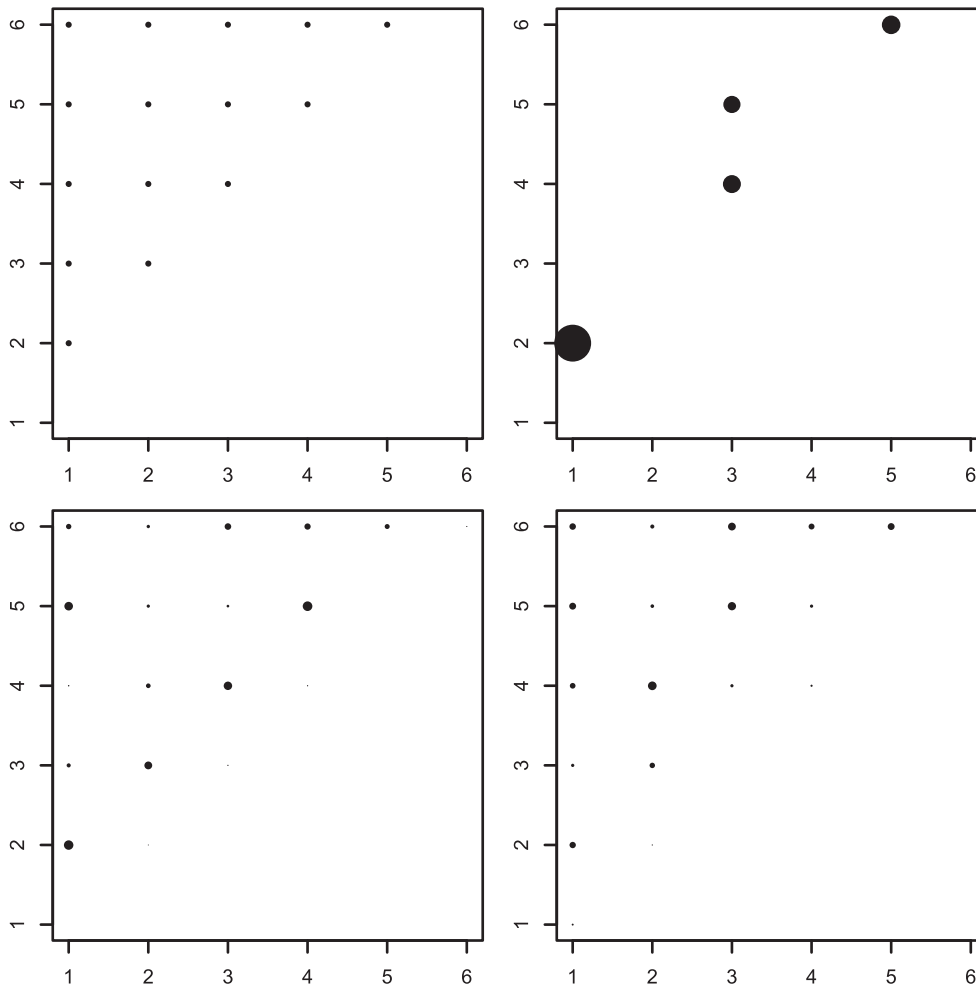


FIGURE 3 Optimal designs for the materials testing example assuming a Gumbel copula with $\tau_2 = 0.33$ (circles areas are proportional to design weight); the axes correspond to the material labels. (Rows) local and pseudo-Bayesian. (Columns) assumed parameters or prior mean of $\beta^T = (0, 0, 0, 0, 0, 0)$ and $\beta^T = (0, -1, 2, -3, 4, -5)$, respectively

4 | DISCUSSION

The modeling of block effects by copulas seems to be a natural choice and allows for elegant separation of the block and the marginal effects. Experimental designs for such models are now readily calculable. The pseudo-Bayesian D_A -optimality criterion was added to the R package `docopulae` version 0.4 (see the work of Rappold³⁵) with the functions `wDsensitivity` and `wDefficiency`, both relying on a prespecified quadrature scheme for evaluation of the integrals. In this paper, we have concentrated on finding designs to estimate the complete parameter vector but the implementation provides flexibility for checking for symmetry, model discrimination, etc, as investigated in the work of Perrone et al.²⁹

Our examples are confined to the case $k = 2$. While there is no theoretical necessity for that, it is difficult to specify high-dimensional parametric copulas with a sufficient range of dependence, for details, see the excellent survey of Nikoloulopoulos.³⁸ However, work on this issue would go well beyond the scope of this paper. It might also be interesting to contrast our findings with some known analytic results for blocks of size two as, for example, given in the work of Cheng,²⁶ where a Gaussian copula is implicitly assumed.

ACKNOWLEDGEMENTS

We are grateful to Keith Warburton and Rob Ashmore from the UK Defence Science and Technology Laboratory for providing details of the materials testing example. W.G. Müller would like to acknowledge the hospitality of the Southampton Statistical Sciences Research Institute during his sabbatical, when this research was initiated. He was partially supported through project grants LIT-2017-4-SEE-001 funded by the Upper Austrian Government, and I 3903-N32 by the Austrian

Science Fund (FWF). D.C. Woods was partially supported through Fellowship EP/J018317/1 from the UK Engineering and Physical Sciences Research Council. We are also grateful to the associate editor and two referees, whose comments have lead to a considerable improvement of this paper.

ORCID

W.G. Müller  <https://orcid.org/0000-0002-3564-766X>

REFERENCES

1. Woods DC, Lewis SM, Eccleston JA, Russell KG. Designs for generalized linear models with several variables and model uncertainty. *Technometrics*. 2006;48:284-292.
2. Dror HA, Steinberg DM. Robust experimental design for multivariate generalized linear models. *Technometrics*. 2006;48:520-529.
3. Dror HA, Steinberg DM. Sequential experimental designs for generalized linear models. *J Am Stat Assoc*. 2008;103:288-298.
4. Russell KG, Woods DC, Lewis SM, Eccleston JA. D-optimal designs for Poisson regression models. *Statistica Sinica*. 2009;19:721-730.
5. Atkinson AC, Woods DC. Designs for generalized linear models. In: Dean AM, Morris MD, Stufken J, Bingham DR, eds. *Handbook of Design and Analysis of Experiments*. Boca Raton, FL: Chapman & Hall/CRC; 2015.
6. Woods DC, vande Ven P. Blocked designs for experiments with correlated non-normal response. *Technometrics*. 2011;53(2):173-182. <https://doi.org/10.1198/TECH.2011.09197>
7. Niaparast M, Schwabe R. Optimal design for quasi-likelihood estimation in Poisson regression with random coefficients. *J Stat Plan Inference*. 2013;143:296-306.
8. Waite TW, Woods DC. Designs for generalized linear models with random block effects via information matrix approximations. *Biometrika*. 2015;102:677-693.
9. Breslow NE, Clayton DG. Approximate inference in generalized linear mixed models. *J Am Statist Assoc*. 1993;88(421):9-25.
10. Liang KY, Zeger SL. Longitudinal data analysis using generalized linear models. *Biometrika*. 1986;73(1):13-22.
11. Hughes-Oliver JM. Optimal designs for nonlinear models with correlated errors. *Lect Notes Monogr Ser*. 1998;34:163-174.
12. Atkinson AC, Ucinski D. Experimental design for time-dependent models with correlated observations. *Stud Nonlinear Dyn Econom*. 2004;8(2). Article No. 13.
13. vande Ven PM, Woods DC. Optimal blocked minimum-support designs for non-linear models. *J Stat Plan Inference*. 2014;144:152-159.
14. Godolphin J. Designs with blocks of size two and applications to microarray experiments. *Ann Stat*. 2018;46:2775-2805.
15. Bailey RA. Designs for two-colour microarray experiments. *J R Stat Soc Ser C Appl Stat*. 2007;56(4):365-394.
16. Kerr KF. Optimality criteria for the design of 2-color microarray studies. *Stat Appl Genet Mol Biol*. 2012;11(1). Article No. 10.
17. David O, Kempton RA. Designs for interference. *Biometrics*. 1996;52(2):597-606.
18. Overstall AM, Woods DC. Bayesian design of experiments using approximate coordinate exchange. *Technometrics*. 2017;59:458-470.
19. King J, Wong W-K. Minimax D-optimal designs for the logistic model. *Biometrics*. 2000;56:1263-1267.
20. Atkinson AC, Donev AN, Tobias RD. *Optimum Experimental Design, With SAS*. 2nd ed. Oxford, UK: Oxford University Press; 2007.
21. Nelsen RB. *An Introduction to Copulas*. 2nd ed. New York, NY: Springer; 2007. *Springer Series in Statistics*.
22. Sklar A. Fonctions de répartition à n dimensions et leurs marges. *Publications de l'Institut de Statistique de Paris*. 1959;8:229-231.
23. Genest C, Nešlehová J. A primer on copulas for count data. *ASTIN Bull J Int Actuar Assoc*. 2007;37:475-515.
24. Joe H. Multivariate concordance. *J Multivar Anal*. 1990;35(1):12-30.
25. Genest C, Nešlehová J, Ben Ghorbal N. Estimators based on Kendall's tau in multivariate copula models. *Aust N Z J Stat*. 2011;53(2):157-177.
26. Cheng C-S. Optimal regression designs under random block-effects models. *Statistica Sinica*. 1995;5:485-497.
27. Denman NG, McGree JM, Eccleston JA, Duffull SB. Design of experiments for bivariate binary responses modelled by Copula functions. *Comput Stat Data Anal*. 2011;55(4):1509-1520.
28. Perrone E, Müller WG. Optimal designs for copula models. *Statistics*. 2016;50(4):917-929.
29. Perrone E, Rappold A, Müller WG. D_s -optimality in copula models. *Stat Methods App*. 2016;26(3):403-418.
30. Deldossi L, Osmetti SA, Tommasi C. Optimal design to discriminate between rival copula models for a bivariate binary response. *TEST*. 2018;28:147-165.
31. Durante F, Perrone E. Asymmetric copulas and their application in design of experiments. In: Saminger-Platz S, Mesiar R, eds. *On Logical, Algebraic, and Probabilistic Aspects of Fuzzy Set Theory*. Vol. 336. Cham, Switzerland: Springer International Publishing; 2016:157-172.
32. Waite TW. Singular prior distributions and ill-conditioning in Bayesian D-optimal design for several nonlinear models. *Statistica Sinica*. 2018;28:505-525.
33. Wynn HP. The sequential generation of D-optimal experimental designs. *Ann Math Stat*. 1970;41:1655-1664.
34. Fedorov VV. The design of experiments in the multiresponse case. *Theory Probab Appl*. 1971;16:323-332.
35. Rappold A. docopulae: optimal designs for copula models. R Package Version 0.3.5. 2018.

36. Gotwalt CM, Jones BA, Steinberg DM. Fast computation of designs robust to parameter uncertainty for nonlinear settings. *Technometrics*. 2009;51:88-95.
37. Joe H. *Multivariate Models and Dependence Concepts*. Boca Raton, FL: Chapman & Hall/CRC; 1997.
38. Nikoloulopoulos AK. Copula-based models for multivariate discrete response data. In: Jaworski P, Durante F, Härdle WK, eds. *Copulae in Mathematical and Quantitative Finance*. Berlin, Germany: Springer Berlin Heidelberg; 2013:231-249. *Lecture Notes in Statistics*; vol. 213.

How to cite this article: Rappold A, Müller WG, Woods DC. Copula-based robust optimal block designs. *Appl Stochastic Models Bus Ind*. 2019;1–10. <https://doi.org/10.1002/asmb.2469>