Statistica Sinica Preprint No: SS-2022-0003	
Title	Synthesis of Order-of-Addition Models
Manuscript ID	SS-2022-0003
URL	http://www.stat.sinica.edu.tw/statistica/
DOI	10.5705/ss.202022.0003
Complete List of Authors	Robert W. Mee
Corresponding Authors	Robert W. Mee
E-mails	rmee@utk.edu

Statistica Sinica: Preprint
doi:10.5705/ss.202022.0003

Synthesis of Order-of-Addition Models

Robert W. Mee

University of Tennessee

Abstract: Numerous models have been proposed for experiments that vary the order in which components are added, or steps are performed. Some models have a linear effect of each component's position in the sequence, and others have added quadratic effects and product terms. Kriging models based on a component's position have also been proposed. Before models based on each component's position were introduced, models based on the relative position of each pair of components were popular. Here, we connect these models and provide lack-of-fit tests to assist with model selection and the interpretability of the parameters. Three examples from the literature illustrate the need for these varied models.

Key words and phrases: Component-position model, experimental design, pairwise order model, sequence effects.

1. Introduction

The literature on the design and analysis of experiments has recently expanded to include experiments that vary the order of steps or components. Typifying such *order-of-addition* experiments, Voelkel and Gallagher (2019) describe a series of experiments involving up to six components of a new premium automotive clearcoat. The primary response optimized is the clearcoat's viscosity, which affects the coating's desired smooth appearance. Each experiment varied the sequencing for adding components to

Statistica Sinica: Preprint
doi:10.5705/ss.202022.0003

the mixture, while holding the mixture proportions fixed. Studying the sequencing of components distinguishes these from traditional mixture experiments. The literature on mixture experiments that vary the proportions of components is extensive, from the seminal work of (Scheffe, 1958) to that of Cornell (2002), among others. However, prior to 2019, we know of just one proceedings article (**?**) that discusses models for order-of-addition data. For a review of the history of order-of-addition experiments, see Lin and Peng (2019, Section 1) and Voelkel (2019b).

Order-of-addition experiments are not limited to mixture applications. Many other situations involve ordering problems, such as web search bid-for-placement ad orders, drug treatment sequencing, job scheduling, and so on. In many such applications, experiments can be used to identify sequence effects.

Van Nostrand (1995) proposed using pairwise order (PWO) factors, a model that has become popular in the literature; see Voelkel (2019a,b), Voelkel and Gallagher (2019), Lin and Peng (2019), Mee (2020), Chen, Peng, and Lin (2021), Tsai (2021), Wang and Mee (2022), and Zhao, Lin, and Liu (2022). If m components are to be sequenced, then estimating the PWO model requires that we explore at least $1 + m(m-1)/2$ of the $m!$ possible sequences. Peng, Mukerjee, and Lin (2019) proposed a PWO model with diminishing effects for pairwise factors as the separation in the pair's positions in the sequence increases; Mee (2020) proposed adding PWO factor interactions to account for sequencing effects not accounted for by pairwise main effects alone. Voelkel and Gallagher (2019) found that Van Nostrand's original PWO model sufficed for each of the clearcoat experiments. In fact, in the experiment discussed

Statistica Sinica: Preprint
doi:10.5705/ss.202022.0003

later involving six clearcoat components, a model with just three of the 15 PWO factors seems satisfactory, with $R^2 \approx 92\%$.

Recently, a second class of models has been proposed, based on each component's position in the sequence. Yang, Sun, and Xu (2020) proposed a component-position (CP) model for order-of-addition using categorical (nominal) explanatory variables. Stokes and Xu (2022) proposed three CP models using quantitative variables, rather than nominal variables, and Xiao and Xu (2021) proposed two Kriging models in which distance is based on the components' positions in each sequence. The full CP model with nominal variables requires estimating $1 + (m-1)^2$ parameters, and even more if interactions are added. In contrast, the three Stokes and Xu models require only m, $2m-1$, and $(m+2)(m-1)/2$ parameters, respectively. The Kriging models of Xiao and Xu (2021) require only $m+2$ and $3m+2$ parameters, respectively. Certainly, parsimony is an advantage, provided the models fit satisfactorily.

The full $m!$ design is optimal for PWO models (Peng, Mukerjee, and Lin, 2019) and CP models (Yang, Sun, and Xu, 2020; Stokes and Xu, 2022). However, fractional designs are needed for $m \geq 5$, because the full $m!$ design is prohibitively large. Voelkel (2019a) defined order-of-addition orthogonal arrays (OofA-OAs), which have the same normalized information matrix as the full design for Van Nostrand's PWO model. This work improved greatly on Van Nostrand's early attempts to construct smaller designs for the PWO model. Yang, Sun, and Xu (2020) proposed component orthogonal arrays (COAs), which are optimal for their CP model and for two Stokes and Xu (2022) CP models. COAs, also known as permutation orthogonal arrays (Wang, Xu, and Ding

Statistica Sinica: Preprint
doi:10.5705/ss.202022.0003

(2020)), have run sizes restricted to a multiple of $m(m-1)$. Although we focus on modeling and inference, some results presented later assume that an estimation is based on one of these optimal designs.

The remainder of this paper is organized as follows. In Section 2, we review all parametric PWO and CP models in the literature, and provide the covariance matrix of the least squares estimators, assuming an optimal (OofA-OA or COA) design. In Section 3, we show how the Stokes and Xu CP models are nested within other CP and PWO models. This nesting leads to lack-of-fit tests, which show whether a fitted model is satisfactory, aiding model selection and interpreting the parameters. In Section 4, we present three examples to illustrate the insights from Sections 2 and 3, highlighting the need for different models. Section 5 concludes the paper, helping readers to understand and use order-of-addition models to analyze experiments.

2. Review of Order-of-Addition Models

Suppose we have m components, denoted by $1, \ldots, m$. Let $\mathbf{S} = [s_{ij}]$ denote an $n \times m$ matrix defining the n sequences of an experiment. That is, s_{ij} identifies the component appearing in position j of sequence i. Let $\mathbf{O} = [o_{ij}]$ denote the corresponding $n \times m$ matrix, where o_{ij} is the position of component j in sequence i. Sections 2.1–2.6 define the parametric models that have been proposed for order-of-addition data, beginning with the PWO models. Section 2.7 summarizes the Kriging models of Xiao and Xu (2021).

Statistica Sinica: Preprint
doi:10.5705/ss.202022.0003

2.1 Van Nostrand's main-effects PWO model

The PWO model of Van Nostrand (1995) is based on $m(m-1)/2$ pairwise order factors, $P_{(jk)}$ $(1 \le j < k \le m)$. Define $\mathbf{P} = [p_{i(jk)}]$, where, for $i = 1, \ldots, n$ and $1 \le j < k \le m$,

$$p_{i(jk)} = \begin{cases} 1 & \text{if } o_{ij} < o_{ik}, \\ -1 & \text{if } o_{ij} > o_{ik}. \end{cases}$$

We order the columns of $\mathbf{P}$ as $P_{(12)}$, $P_{(13)}$, $\ldots$, $P_{(1m)}$, $P_{(23)}$, $\ldots$, $P_{((m-1)m)}$. The main-effects PWO model is

$$\mathbf{Y} = \mathbf{1}_n \alpha_0 + \mathbf{P}\alpha + \mathbf{e}, \tag{2.1}$$

where α_0 and $\alpha = [\alpha_{(12)}, \alpha_{(13)}, \ldots, \alpha_{(1m)}, \alpha_{(23)}, \ldots, \alpha_{((m-1)m)}]^T$ are unknown parameters, $\mathbf{e}$ is a vector of independent and identically distributed (i.i.d.) normal random variables with mean zero and variance σ^2, and $\mathbf{Y}$ is an $n \times 1$ vector of observed responses. For OofA-OAs, the normalized information matrix is given in Peng, Mukerjee, and Lin (2019, Theorem 2) and Schoen and Mee (2021, Equation 4); these designs are D-, A-, and I-optimal for model (2.1).

Peng, Mukerjee, and Lin (2019) proposed a PWO model with tapering, where the elements of $\mathbf{P}$ are replaced with $p_{i(jk)}c(h_{i(jk)})$, where $h_{i(jk)} = |o_{ij} - o_{ik}|$ and $c(h)$ is a decreasing nonlinear function of h, such as $c(h) = 1/h$ or C^{h-1}, for some constant $0 \le C < 1$. For $C = 0$, the pairwise effect only affects the response when the components are adjacent, that is, when $|o_{ij} - o_{ik}| = 1$. Piepho and Williams (2021) proved that linear tapering $(c(h) = m - h)$ results in a model equivalent to the standard PWO

Statistica Sinica: Preprint
doi:10.5705/ss.202022.0003

model (2.1).

2.2 PWO model with interactions

Mee (2020) proposed adding interactions to the main-effects PWO model (2.1). Two-factor interactions (2FIs) that share a common component are called triplet interactions. Voelkel (2019a) refers to $P_{(ij)} * P_{(ik)}$ and $P_{(ik)} * P_{(jk)}$ as *synergistic* triplet interactions, because they sum to $m!/3$ in the full design, and refers to $P_{(ij)} * P_{(jk)}$ as an *antagonistic* triplet interaction, because it sums to $-m!/3$. To avoid linear dependencies, the triplet PWO model includes at most two of the three interactions from each subset of three components, because $P_{(ij)} * P_{(ik)} + P_{(ik)} * P_{(jk)} - P_{(ij)} * P_{(jk)} = \mathbf{1}_n$. The triplet PWO model is

$$\mathbf{Y} = \mathbf{1}_n\alpha_0 + \mathbf{P}\alpha + \mathbf{T}\tau + \mathbf{e}, \tag{2.2}$$

where $\tau = [\tau_{12*13}, \ldots, \tau_{(m-2)m*(m-1)m}]^T$.

Let $\mathbf{U}$ denote the $n \times 3\binom{m}{4}$ matrix corresponding to all $P_{(ij)} * P_{(kl)}$ for $i < j$, $k < l$, and i, j, k, l all distinct. The 2FI PWO model is

$$\mathbf{Y} = \mathbf{1}_n\alpha_0 + \mathbf{P}\alpha + \mathbf{T}\tau + \mathbf{U}\psi + \mathbf{e}, \tag{2.3}$$

where $\psi = [\psi_{12*34}, \ldots, \psi_{(m-3)(m-2)*(m-1)m}]^T$. The full 2FI PWO model has $1 + \binom{m}{2} + 2\binom{m}{3} + 3\binom{m}{4}$ parameters.

Statistica Sinica: Preprint
doi:10.5705/ss.202022.0003

2.3 Nominal CP model

Yang, Sun, and Xu (2020) proposed a CP model using indicator variables z_{kp} that equal one if component k is in the pth position, and zero otherwise. Let $\mathbf{Z}$ denote an $n \times (m-1)^2$ matrix, with the columns of $\mathbf{Z}$ ordered as $(kp) = 11, 12, \ldots, 1(m-1), 21, 22, \ldots, 2(m-1), 31, \ldots, (m-1)(m-1)$, and let $\mathbf{Z}^*$ denote the $n \times m^2$ matrix with columns $(kp) = 11, 12, \ldots, 1m, 21, 22, \ldots, 2m, 31, \ldots, mm$. Their CP model (with baseline constraints) is

$$\mathbf{Y} = \mathbf{1}_n\gamma_0 + \mathbf{Z}\gamma + \mathbf{e}, \tag{2.4}$$

where γ_0 and $\gamma = [\gamma_{11}, \gamma_{12}, \ldots, \gamma_{1(m-1)}, \gamma_{21}, \ldots, \gamma_{(m-1)(m-1)}]^T$ are unknown parameters. Yang, Sun, and Xu (2020) also proposed using a stepwise regression with all m^2 columns of $\mathbf{Z}^*$ as candidates, as well as possible interactions $z_{kp}z_{k'p'}$, if necessary.

The CP model (2.4) can be fit in several ways. One can compute the sum of squares (SS) attributable to the kth component $\{z_{k1}, \ldots, z_{k(m-1)}\}$ $(k = 1, \ldots, m-1)$, by including the columns of $\mathbf{O}$ as categorical variables. Alternatively, one can compute the SS attributable to the pth position in the sequence $\{z_{1p}, \ldots, z_{(m-1)p}\}$ $(p = 1, \ldots, m-1)$, by including the columns of $\mathbf{S}$ as categorical variables. The first approach seems less useful, because it disregards the meaningful ordering of the levels of the columns of $\mathbf{O}$. Each method compares the included components (or positions) with the omitted one. Thus, when we omit the last position, we expect the earliest positions to be the most significant. For example, when Wang, Xu, and Ding (2020) omit the last position of

Statistica Sinica: Preprint
doi:10.5705/ss.202022.0003

$m = 5$, they find positions 1 and 2 to be the most significant. If the first position is omitted instead of the last one, positions 5 and 4 are the most significant. Model (3.4) shows a third (better) way to partition the SS for model (2.4).

2.4 The first-order CP model

The Stokes and Xu models are based on the columns of $\mathbf{O}$, after centering and scaling. Define $c_0 = (m+1)/2$, $c_1 = (12/(m^2-1))^{1/2}$, and $\mathbf{B}^* = c_1(\mathbf{O} - c_0\mathbf{J})$, where $\mathbf{J}$ is an $n \times m$ matrix of ones. The matrix $\mathbf{B}^*$ is a centered and scaled version of $\mathbf{O}$, such that each row sums to zero and has sum of squares of m; that is, the squared elements of $\mathbf{B}^*$ average one. Denote the columns of $\mathbf{B}^*$ by $\mathbf{x}_1, \ldots, \mathbf{x}_m$. Because $\mathbf{x}_1 + \ldots + \mathbf{x}_m$ is a vector of zeros, we must drop one column when constructing the models. Let d denote the deleted column and $\mathbf{B}^{(-d)}$ denote the resulting $n \times (m-1)$ matrix. Stokes and Xu (2022) drop the mth column of $\mathbf{B}^*$. Thus, for convenience, we denote $\mathbf{B}^{(-m)}$ as $\mathbf{B}$, without a superscript. That is, $\mathbf{B} = [\mathbf{x}_1, \ldots, \mathbf{x}_{m-1}]$.

The Stokes and Xu first-order component-position (FOCP) model is

$$\mathbf{Y} = \mathbf{1}_n\beta_0 + \mathbf{B}\beta_1 + \mathbf{e}, \tag{2.5}$$

where $(\beta_0, \ \beta_1^T) = (\beta_0, \beta_{11}, \ldots, \beta_{1(m-1)})$ is a vector of m unknown parameters. The model matrix for model (2.5) is $\mathbf{X_{FOCP}} = [\mathbf{1}_n, \mathbf{B}] = [\mathbf{1}_n, \mathbf{x}_1, \ldots, \mathbf{x}_{m-1}]$. Schoen and Mee (2021) proved that for the full $m!$ design, the normalized information matrix is

$$\mathbf{M}_{FOCP} = \mathbf{X}_{FOCP}^T\mathbf{X}_{FOCP}/m! = diag\{1, \frac{1}{m-1}(m\mathbf{I}_{m-1} - \mathbf{J}_{m-1})\}, \tag{2.6}$$

Statistica Sinica: Preprint
doi:10.5705/ss.202022.0003

and the corresponding precision matrix is $\mathbf{M}_{FOCP}^{-1} = \text{diag}\{1, \mathbf{V}\}$, where

$$\mathbf{V} = (1 - m^{-1})(I_{m-1} + J_{m-1}). \tag{2.7}$$

Given the form of $\mathbf{V}$, the following lemma is easily proved.

Lemma 1. *For any n-run design with normalized information matrix* $\mathbf{M}_{FOCP}$ *for model (2.5), the standard errors for the least squares estimators* $\hat{\beta}_{1i}$ *and* $\hat{\beta}_{1i} - \hat{\beta}_{1j}$ *are both* $SE = \sigma(VIF/n)^{1/2}$*, where the variance inflation factor (VIF) is equal to* $2(1 - m^{-1})$.

Stokes and Xu (2022) showed that the full $m!$ sequence design is D-optimal. Thus, the D-efficiency of a design for model (2.5) can be assessed relative to the the maximum scaled determinant $D_{FOCP} = det(\mathbf{M}_{FOCP})^{1/m} = m^{1-2/m}/(m-1)^{1-1/m}$.

Stokes and Xu's example 1 is an experiment with all 24 sequences for $m = 4$. The first-order model has fitted values that do not depend on which column is omitted, but the coefficients change dramatically. Table 1 gives the least squares estimates, depending on whether column 1, 2, 3, or 4 of $\mathbf{B}^*$ is omitted. Stokes and Xu fit the last of these four models and drop the estimate for component 1, because its p-value (0.062) exceeds 0.05. The null hypothesis for this test is that components 1 and 4 do not differ in terms of their position's effect on the mean response. By Lemma 1, the standard error for each $\hat{\beta}_{1j}$ and each $\hat{\beta}_{1j} - \hat{\beta}_{1j'}$ is estimated by $\hat{\sigma}/4 = 0.915$. Note that all $\binom{m}{2}$ tests of differences can be performed from a single model.

The numerical values of the coefficients for a single component vary between mod-

Statistica Sinica: Preprint
doi:10.5705/ss.202022.0003

els, because the effect of each component is compared with that of the omitted one. When all estimated coefficients are positive and statistically significant, as for the model based on omitting $\mathbf{x}_3$, to maximize the response, the omitted component should appear first; that is, the mean response is made greater by increasing the position number for all other components. Similarly, when all the coefficients are negative, as they are when we omit $\mathbf{x}_2$, the omitted component should appear last. If a coefficient is zero, then this component and the omitted one are similar in terms of position effect.

Table 1: Estimates and p-values for Stokes and Xu's example 1 based on $\mathbf{B}^{(-d)}$

Model Matrix	$\hat{\beta}_0$	$\hat{\beta}_{11}$	$\hat{\beta}_{12}$	$\hat{\beta}_{13}$	$\hat{\beta}_{14}$
$[\mathbf{1}_n, \mathbf{B}^{(-1)}]$	45.2167	0.0000	3.8377	-3.7426	1.8056
			(4.5e-4)	(5.7e-4)	(0.062)
$[\mathbf{1}_n, \mathbf{B}^{(-2)}]$	45.2167	-3.8377	0.0000	-7.5803	-2.0320
		(4.5e-4)		(6.8e-8)	(0.038)
$[\mathbf{1}_n, \mathbf{B}^{(-3)}]$	45.2167	3.7426	7.5803	0.0000	5.5482
		(5.7e-4)	(6.8e-8)		(6.3e-6)
$[\mathbf{1}_n, \mathbf{B}^{(-4)}]$	45.2167	-1.8056	2.0320	-5.5482	0.0000
		(0.062)	(0.038)	(6.3e-6)	
Mean	45.2167	-0.4752	3.3625	-4.2178	1.3305

To improve the interpretability of the estimates, rather than dropping one column of $\mathbf{B}^*$, one might use the model matrix $[\mathbf{1}_n, \mathbf{B}^*]$, and impose the constraint that the coefficients for $\mathbf{x}_1, \ldots, \mathbf{x}_m$ sum to zero. That is, model (2.5) is reparameterized as

$$\mathbf{Y} = \mathbf{1}_n\beta_0^* + \mathbf{B}^*\beta_1^* + \mathbf{e}, \quad \text{where } \beta_{11}^* + \ldots + \beta_{1\mathrm{m}}^* = 0. \tag{2.8}$$

The constrained least squares solution is equivalent to averaging the estimates across the m different models, as shown in the last line of Table 1. Because each of the m models gives the same predicted values, the average of the m models based on $\mathbf{B}^{(-i)}$

Statistica Sinica: Preprint
doi:10.5705/ss.202022.0003

$(i = 1, \ldots, m)$ does as well. Furthermore, because the coefficients across the m models appear in $\pm$ pairs, the mean model's coefficients must sum to zero. The individual coefficients β_{1j}^* in model (2.8) are not estimable functions; rather, only contrasts of these are estimable.

2.5 The pure quadratic CP model

If we add squared terms to the linear effects of model (2.5), we have Stokes and Xu's quadratic component-position (QCP) model

$$\mathbf{Y} = \mathbf{1}_n\beta_0 + \mathbf{B}\beta_1 + \mathbf{Q}\beta_2 + \mathbf{e}, \tag{2.9}$$

where $(\beta_0,\ \beta_1^T, \beta_2^T) = (\beta_0, \beta_1^T, \beta_{21}, \ldots, \beta_{2(m-1)})$ is vector of $2m-1$ unknown parameters, and the columns of $\mathbf{Q}$ are $Q_i = (c_2/c_1)(\mathbf{x}_i^2 - 1)$, for $i = 1, \ldots, m-1$, with $c_2 = (15/(m^2-4))^{1/2}$. The model matrix is $\mathbf{X}_{QCP} = [\mathbf{1}_n, \mathbf{B}, \mathbf{Q}]$. For the full $m!$ design, the normalized information matrix of the quadratic model is

$$\mathbf{M}_{QCP} = \frac{\mathbf{X}_{QCP}^T\mathbf{X}_{QCP}}{m!} = diag\{1, \frac{1}{m-1}(m\mathbf{I}_{m-1} - \mathbf{J}_{m-1}),\ \frac{1}{m-1}(m\mathbf{I}_{m-1} - \mathbf{J}_{m-1})\}. \tag{2.10}$$

Lemma 2. *For any n-run design with normalized information matrix $\mathbf{M}_{QCP}$ for model (2.9), the precision matrix $\mathbf{M}_{QCP}^{-1}$ is $diag(1, \mathbf{V}, \mathbf{V})$, where $\mathbf{V}$ is as defined in (2.7). As a result, the standard errors for the least squares estimators $\hat{\beta}_{1i}$, $\hat{\beta}_{2i}$, $\hat{\beta}_{1i} - \hat{\beta}_{1j}$, and $\hat{\beta}_{2i} - \hat{\beta}_{2j}$ are all $SE = \sigma(VIF/n)^{1/2}$, where $VIF = 2(1 - m^{-1})$.*

Statistica Sinica: Preprint
doi:10.5705/ss.202022.0003

Because the normalized information matrices for $\mathbf{B}$ and $\mathbf{Q}$ are identical and $\mathbf{B}^T\mathbf{Q} = 0$ for the full $m!$ design, the determinant of (2.10) is $m^{2m-4}/(m-1)^{2m-2}$, which is the square of the determinant for (2.6), and $A_{QCP} = \text{trace}(\mathbf{M}_{QCP}^{-1}) = 4(m + m^{-1}) - 7$.

2.6 The second-order CP model

If we add linear-by-linear interactions to the pure quadratic model (2.9), the result is Stokes and Xu's second-order (SO) CP model

$$\mathbf{Y} = \mathbf{1}_n\beta_0 + \mathbf{B}\beta_1 + \mathbf{Q}\beta_2 + \mathbf{R}\delta + \mathbf{e}, \tag{2.11}$$

where the columns of $\mathbf{R}$ are the $(m-1)(m-2)/2$ interactions $\mathbf{x}_i\mathbf{x}_j$ $(1 \le i < j \le m-1)$. The columns of $[\mathbf{1}_n, \mathbf{B}, \mathbf{Q}, \mathbf{R}]$ involve one linear dependency, even for the full $m!$ design. Stokes and Xu (2022) eliminate the last column of $\mathbf{Q}$ to avoid this dependency, and so take the model matrix to be $\mathbf{X}_{SO} = [\mathbf{1}_n, \mathbf{B}, \mathbf{Q}^{\{-(m-1)\}}, \mathbf{R}]$, where $\mathbf{Q}^{\{-(m-1)\}}$ denotes the first $m-2$ columns of $\mathbf{Q}$. Because the columns of $\mathbf{R}$ are correlated with both the intercept column and with the pure quadratic columns, the block diagonal structure of $\mathbf{X}_{SO}^T\mathbf{X}_{SO}$ separates into only two blocks, one for the columns of $\mathbf{B}$, and a second for the remaining columns of $\mathbf{X}_{SO}$.

Appendix A presents formulae for $\mathbf{M}_{SO}$. No simple closed-form formula has been found for the determinant or inverse of $\mathbf{M}_{SO} = \mathbf{X}_{SO}^T\mathbf{X}_{SO}/m!$. For this model with $p = 1 + (m-1) + (m-2) + (m-1)(m-2)/2 = (m-1)(m+2)/2$ parameters, the maximum values of $det(\mathbf{M}_{SO})^{1/p}$, computed using the full $m!$ design (Stokes and Xu, 2022, Theorem 3), are 0.7759, 0.6912, 0.6704, 0.6694, 0.6758, 0.6849, 0.6950, and

Statistica Sinica: Preprint
doi:10.5705/ss.202022.0003

0.7052, for $m = 3, \ldots, 10$, respectively. We can use these values to compute the relative D-efficiency of designs for estimating model (2.11).

Instead of dropping Q_{m-1}, consider the advantages of dropping the $\mathbf{x}_{m-2}\mathbf{x}_{m-1}$ interaction:

- Dropping $\mathbf{x}_{m-2}\mathbf{x}_{m-1}$ nests the quadratic model (2.9) more naturally in the SOCP model, which helps interpretability.
- Dropping $\mathbf{x}_{m-2}\mathbf{x}_{m-1}$ results in smaller variances for the intercept and many SO parameter estimates.

To illustrate the second advantage, Table 2 gives the values of v such that the variance of the least squares estimates is $v\sigma^2/n$ for any design with the same normalized information matrix as the full $m!$ design for $m = 4$, 6, and 8. In every case, the trace is reduced by at least 10%. In general, the interaction estimates have two variances. When dropping Q_{m-1}, the interactions $\mathbf{x}_i\mathbf{x}_j$ $(i < j \leq m-2)$ have the larger variance and $\mathbf{x}_i\mathbf{x}_{m-1}$ $(i = 1, \ldots, m-2)$ have the smaller variance in Table 2. When dropping $\mathbf{x}_{m-2}\mathbf{x}_{m-1}$, Q_{m-2}, Q_{m-1} and interactions involving $\mathbf{x}_{m-2}$ and/or $\mathbf{x}_{m-1}$ have the smaller variance.

Piepho and Williams (2021) parameterize model (2.11) by dropping the intercept rather than dropping a column of $\mathbf{Q}$ or $\mathbf{R}$. This has the advantage of model symmetry for components $1, \ldots, m-1$, but it dramatically alters the meaning of the components of β_2 and δ, making model reduction more complicated. Piepho and Williams (2021) also explore third-order CP models analogous to the special cubic and full cubic Scheffe

Statistica Sinica: Preprint
doi:10.5705/ss.202022.0003

Table 2: Variance factors v for the least squares estimates of the SOCP model (2.11) parameters

	$m=4$		$m=6$		$m=8$	
	If one drops:		If one drops:		If one drops:	
	Q_3	$\mathbf{x}_2\mathbf{x}_3$	Q_5	$\mathbf{x}_4\mathbf{x}_5$	Q_7	$\mathbf{x}_6\mathbf{x}_7$
β_0	3.5841	1.9615	12.059	7.8056	27.175	19.342
β_{1i}'s	3 @ 1.5000	3 @ 1.5000	5 @ 1.6667	5 @ 1.6667	7 @ 1.7500	7 @ 1.7500
β_{2i}'s		1 @ 2.8846		3 @ 3.7407		5 @ 4.1413
	2 @ 1.9615	2 @ 1.9615	4 @ 2.5556	2 @ 2.5556	6 @ 2.8370	2 @ 2.8370
δ_{ij}'s	1 @ 4.5072		6 @ 5.1143	3 @ 4.8611	15 @ 5.4355	10 @ 5.1359
	2 @ 3.0649	2 @ 2.8846	4 @ 3.4939	6 @ 3.2407	6 @ 3.7235	10 @ 3.4239
Trace($\mathbf{M}^{-1}$)	22.644	19.038	75.276	66.500	160.32	143.57

mixture models (Cornell, 2002), but report that these extensions are not useful for the examples they considered.

2.7 Parsimonious Kriging CP models

Xiao and Xu (2021) proposed using Kriging models for order-of-addition experiments, where the covariance is a function of the CP-scaled distances $h_l = |o_{il} - o_{i'l}|/m$ ($l = 1, \ldots, m$). The assumed model for the ith observation, assuming no blocking, is

$$y_i = \mu + Z(\mathbf{o}_i) + e_i,$$

where μ is an unknown mean, $\mathbf{o}_i$ is the ith row of $\mathbf{O}$, and $Z(\mathbf{o}_i)$ is a Gaussian process

Statistica Sinica: Preprint
doi:10.5705/ss.202022.0003

with zero mean and stationary covariance function defined as

$$cov(Z(\mathbf{o}_i), Z(\mathbf{o}_{i'})) = \eta^2 \prod_{l=1}^{m} K(h_l : \theta_l).$$

Xiao and Xu (2021) chose the Matern kernel with parameter 5/2 for their examples, resulting in the following covariance function (when $i \neq i'$):

$$cov(Z(\mathbf{o}_i), Z(\mathbf{o}_{i'})) = \eta^2 \prod_{l=1}^{m} [\{1 + \sqrt{5}\theta_l h_l + 5(\theta_l h_l)^2/3\} exp(-\sqrt{5}\theta_l h_l)], \tag{2.12}$$

where $\theta_1, \ldots, \theta_m$ are unknown parameters, along with η^2 and μ. If there is blocking, μ is replaced by a separate parameter for each block. For $i = i'$, we add σ^2 to (2.12).

The simple universal Kriging (UK) model described above has, for each component l, the same fixed scaled distance $h_l = 1/m$ between adjacent positions. This constraint is relaxed in their mapping-based (MUK) model, where distance is defined as $h_l = |g_l(o_{il}) - g_l(o_{i'l})|/m$ $(l = 1, \ldots, m)$; Xiao and Xu (2021) use an incomplete beta function for g_l. The MUK model has $3m + 2$ parameters, because each g_l now depends on two additional parameters. For further details, see (Xiao and Xu, 2021, Section 3.2).

3. Connections Between Models

We now explore the connections between the PWO models (2.1), (2.2), and (2.3) and the parametric CP models (2.4), (2.5), (2.9), and (2.11).

Statistica Sinica: Preprint
doi:10.5705/ss.202022.0003

3.1 The FOCP model and the main-effects PWO model

Schoen and Mee (2021) showed that the column space of $\mathbf{B}$ is nested within the column space of $\mathbf{P}$. Specifically, for $j = 1, \ldots, m$, $\mathbf{x}_j = 0.5c_1(\sum_{i=j+1}^{m} P_{(ji)} - \sum_{i=1}^{j-1} P_{(ij)})$. Nested column spaces and the usual (i.i.d. normal) assumptions for $\mathbf{e}$ imply that we can perform a lack-of-fit test for the hypotheses H_0: FOCP model (2.5) is true versus H_a: PWO model (2.1) is true, with test statistic

$$F(\text{FOCP model vs. PWO model}) = \frac{[SSE(2.5) - SSE(2.1)]/[dfE(2.5) - dfE(2.1)]}{SSE(2.1)/dfE(2.1)}, \tag{3.1}$$

where $SSE(i)$ and $dfE(i)$ denote the error sum of squares and the error degrees of freedom, respectively, for model (i).

The difference between the number of parameters for models (2.1) and (2.5) is equal to $1 + m(m-3)/2$, which is the number of linearly independent constraints on α implied by (2.5). Suppose we have a design for which $\mathbf{P}$ has full column rank, and define the $m(m-1)/2 \times m - 1$ matrix $\mathbf{C}_m$ such that $\mathbf{P}\mathbf{C}_m = \mathbf{B}$. The matrix $\mathbf{C}_m$ can be defined recursively as $\mathbf{C}_3 = (3/8)^{1/2}[-1 \;\; 1; -1 \;\; 0; 0 \;\; -1]$ and, for $m = 4, 5, \ldots,$

$$\mathbf{C}_m = [3/(m^2-1)]^{1/2} \begin{Bmatrix} -\mathbf{1}_{m-2} & \mathbf{I}_{m-2} \\ -1 & \mathbf{0}_{m-2}^T \\ \mathbf{0}_{(m-1)(m-2)/2} & [((m-1)^2-1)/3]^{1/2}\mathbf{C}_{m-1} \end{Bmatrix}, \tag{3.2}$$

where for any integer u, $\mathbf{1}_u$ and $\mathbf{0}_u$ denote column vectors of ones and zeros, respectively,

Statistica Sinica: Preprint
doi:10.5705/ss.202022.0003

and $\mathbf{I}_u$ denotes the identity matrix of dimension u. If the FOCP model (2.5) is true, then $\alpha_0 = \beta_0$ and $\alpha = \mathbf{C}_m\beta$. For $m = 3$,

$$\alpha_{(12)} = (3/8)^{1/2}(-\beta_{11} + \beta_{12}), \quad \alpha_{(13)} = -(3/8)^{1/2}\beta_{11}, \quad \alpha_{(23)} = -(3/8)^{1/2}\beta_{12},$$

which implies the constraint $\alpha_{(23)} = \alpha_{(13)} - \alpha_{(12)}$. For $m > 3$, model (2.5) imposes the constraints

$$\alpha_{(jk)} = \alpha_{(1k)} - \alpha_{(1j)} \qquad (2 \le j < k \le m). \tag{3.3}$$

Thus, the FOCP model implies that the PWO coefficients are additive in the sense that, for example, $\alpha_{(12)} + \alpha_{(23)} + \alpha_{(34)} = \alpha_{(14)}$.

Conversely, suppose the main-effects PWO is the true model. What is the expected value of the least squares estimator $\hat{\beta}_1$? This depends on the design; for the full design:

$$E(\hat{\beta}_{1i}) = 0.5c_1(\sum_{j<i} \alpha_{(ji)} - \sum_{j>i} \alpha_{(ij)} - \sum_{j<m} \alpha_{(jm)}) \qquad (i = 1, \ldots, m-1).$$

For example, consider $i = 1$, for which $E(\hat{\beta}_{11}) = -0.5c_1(\sum_{j>1} \alpha_{(1j)} + \sum_{j<m} \alpha_{(jm)})$. Thus, $E(\hat{\beta}_{11})$ is proportional to the negative sum of all PWO coefficients involving component 1 and component m, including $\alpha_{(1m)}$ twice.

3.2 The QCP model and the nominal CP model

For $m = 3$, the QCP model (2.9) and the nominal CP model (2.4) are equivalent, whereas for $m > 3$, model (2.9) is nested in model (2.4). This is easiest to see if we

Statistica Sinica: Preprint
doi:10.5705/ss.202022.0003

extend the quadratic model by adding orthogonal polynomials for powers $3, \ldots, m-1$. For $q = 1, 2, \ldots, m-1$, let $\mathbf{O}^{\{q\}}$ denote the orthogonal polynomials of power q for the first $m-1$ columns of the order design matrix $\mathbf{O}$. The matrices $\mathbf{B}$ and $\mathbf{Q}$ defined earlier correspond to $\mathbf{O}^{\{1\}}$ and $\mathbf{O}^{\{2\}}$, respectively. Then, we rewrite model (2.4) as

$$\mathbf{Y} = \mathbf{1}_n\beta_0 + \mathbf{B}\beta_1 + \mathbf{Q}\beta_2 + \sum_{q=3}^{m-1} \mathbf{O}^{\{q\}}\beta_q + \mathbf{e}, \tag{3.4}$$

where $\beta_p = (\beta_{p1}, \ldots, \beta_{p(m-1)})^T$. This model reparameterizes the nominal CP model parameters γ_{i1},γ_{i2}, ..., $\gamma_{i(m-1)}$ in terms of $\beta_{1i}, \beta_{2i}, \ldots, \beta_{(m-1)i}$, for component $i = 1, \ldots, m-1$. We can test the adequacy of the QCP model (2.9), assuming the nominal CP model (2.4) is the true model; using the notation introduced by equation (3.1),

$$F(\text{QCP model vs. nominal CP model}) = \frac{[SSE(2.9) - SSE(2.4)]/[(m-3)(m-1)]}{SSE(2.4)/dfE(2.4)}.$$

Alternatively, for $m > 4$, we can partition the CP model sum of squares into the contribution for each power. We illustrate this in Section 4.1, Table 4.

3.3 The QCP model and the triplet PWO model

The QCP model (2.9) is nested in the triplet PWO model (2.2).

Lemma 3. *The ith column of* $\mathbf{Q}$, Q_i, *is equal to* $0.5c_1c_2(L_i + 0.5(m-1) - 2c_1^{-2})$, *where*

$$L_i = \sum_{j=1}^{i-2}\sum_{k=j+1}^{i-1} P_{(ji)} * P_{(ki)} - \sum_{j=1}^{i-1}\sum_{k=i+1}^{m} P_{(ji)} * P_{(ik)} + \sum_{j=i+1}^{m-1}\sum_{k=j+1}^{m} P_{(ij)} * P_{(ik)}.$$

Statistica Sinica: Preprint
doi:10.5705/ss.202022.0003

Here, L_i is the linear combination of $(m-1)(m-2)/2$ interaction columns, with "+" indicating i's synergistic triplet interactions and "−" indicating i's antagonistic interactions. Thus, $\beta_{2i} \neq 0$ implies the presence of triplet interaction effects involving component i. If the QCP model (2.9) is true, the symmetry with which the $(m-1)(m-2)/2$ interaction columns appear in L_i implies $(m^2-1)(m-3)/3$ constraints on the τ parameters of the triplet PWO model; see Appendix B for details.

3.4 The SOCP model and the 2FI PWO model

The SOCP model (2.11) is nested in the 2FI PWO model (2.3). To show that the 2FIs in $\mathbf{U}$ are necessary to achieve nested models, the product of the first two columns of $\mathbf{B}$ is

$$\mathbf{x}_1\mathbf{x}_2 = \frac{c_1^2}{2}[\frac{m-3}{2}+\sum_{j=3}^{m}(P_{(12)}*P_{(2j)}-P_{(12)}*P_{(1j)})+\frac{1}{2}\sum_{j=3}^{m-1}\sum_{k=j+1}^{m}(P_{(1j)}*P_{(2k)}+P_{(1k)}*P_{(2j)})]. \tag{3.5}$$

Without the extra $3\binom{m}{4}$ interactions in $\mathbf{U}$ of model (2.3), the triplet PWO model (2.2) still accounts for most of the explanatory power of the SOCP model. We can use the trace correlation (Hopper, 1959) to quantify this measure. For the full $m!$ design, let $\mathbf{X}_T = [\mathbf{1}_n,\ \mathbf{P},\ \mathbf{T}]$ denote the model matrix for (2.2), and $\mathbf{X}_{SO}$ denote the full-rank model matrix for (2.11) obtained by deleting one column from $\mathbf{Q}$ or $\mathbf{R}$. Then,

$$\bar{r}^2 = \text{trace}\{(\mathbf{X}_{SO}^T\mathbf{X}_{SO})^{-1}\mathbf{X}_{SO}^T\mathbf{X}_T(\mathbf{X}_T^T\mathbf{X}_T)^{-1}\mathbf{X}_T^T\mathbf{X}_{SO}\}/\{(m-1)(m+2)/2\} \tag{3.6}$$

Statistica Sinica: Preprint
doi:10.5705/ss.202022.0003

gives the proportion of variation in the SOCP model (2.11) that is accounted for by the triplet PWO model (2.2). Table 3 shows this proportion for $m = 3, ..., 8$. Thus, for $m \geq 4$, the triplet PWO model cannot explain all the variation that can be modeled with the SOCP model, but the proportion missed is small. The triplet model (2.2) already has $(m^3 - 3m^2 - m + 6)/3$ more parameters than model (2.11); to make the models fully nested requires adding $3\binom{m}{4}$ additional interactions.

Table 3: Overlap between the SOCP model and the triplet PWO model

	Number of parameters for:			*% of SO-CP's variation
m	SO-CP(2.11)	Triplet-PWO(2.2)	2FI-PWO(2.3)	explainable by Triplet-PWO
3	5	6	6	100.00
4	9	15	18	98.72
5	14	31	46	97.58
6	20	56	101	96.67
7	27	92	197	95.93
8	35	141	351	95.34

* % of variation based on trace correlation (3.6)

3.5 Summary

Figure 1 summarizes the nesting of the parametric models discussed in Sections 3.1–3.4. The FOCP and QCP models are the most parsimonious, with a maximum of m and $2m - 1$ parameters, respectively.

4. Examples

Here, we use the connections between these models to better understand three examples. The first example, already discussed in Section 2.4, has also been analyzed

Statistica Sinica: Preprint
doi:10.5705/ss.202022.0003

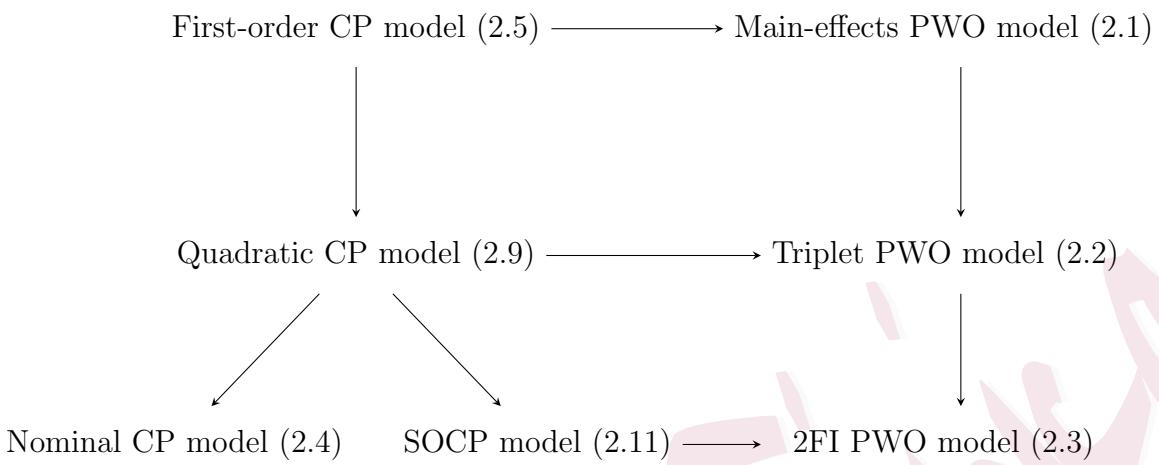

Figure 1: Nesting relationship between order-of-addition models.

by Mee (2020) and Yang, Sun, and Xu (2020); the design is a full 4!, and involves four chemotherapeutic drugs. The second example also appears in Mee (2020), Yang, Sun, and Xu (2020), and Stokes and Xu (2022); it involves five chemotherapeutic drugs, investigated in an experiment with two blocks of size 20. In each case, the response is % cell inhibition, coded by subtracting 30. In Examples 1 and 2, we use the uncoded % cell inhibition values provided by Wang, Xu, and Ding (2020) as the response. Wang, Xu, and Ding (2020) also provide the names and dose levels of the chemotherapeutics, and describe multiple experiments involving changes in both order and dose. The third example, discussed in Section 4.3, involves six clearcoat components, and has also been analyzed by Voelkel and Gallagher (2019).

Statistica Sinica: Preprint
doi:10.5705/ss.202022.0003

4.1 Example 1 ($m = 4$, $n = 24$)

The following four chemotherapeutics are studied: 1: paclitaxel; 2: doxorubicin; 3: mitoxantrone; 4: cisplatin. Wang, Xu, and Ding (2020) perform the full 4! design, so we may fit any of the models discussed in Sections 2 and 3. Mee (2020) found the main-effects PWO model satisfactory, with no need for interaction terms. Yang, Sun, and Xu (2020) found the nominal CP model satisfactory, and fit the reduced model using just four terms:

$$\widehat{Inhibit\%} = 82.85 - 10.86z_{21} - 4.41z_{22} - 9.18z_{34} - 6.09z_{41}. \tag{4.1}$$

The simpler FOCP model (2.5) also suffices, because it exhibits no lack-of-fit vis-a-vis the main-effects PWO model or the nominal CP model. The lack-of-fit test (3.1) is

$$F(\text{FOCP model vs. PWO model}) = \frac{[(267.87 - 211.54)/(20 - 17)]}{211.54/17} = \frac{18.78}{12.44} = 1.51,$$

with p-value $Pr(F_{3,17} > 1.51) = 0.25$. Thus, the data are consistent with the parsimonious first-order model and its constraints (3.3) on the pairwise order effects.

Table 4 partitions the variation explained by model (2.4) into linear, quadratic, and cubic portions. Neither the quadratic nor the cubic terms account for the systematic variation in $Inhibit\%$. Once again, the FOCP model proves consistent with the data. Furthermore, the FOCP model is easier to interpret than the reduced nominal CP model (4.1). Based on the signs of the terms in (4.1), to maximize this fitted response,

Statistica Sinica: Preprint
doi:10.5705/ss.202022.0003

component 2 should not be in either of the first two positions, and component 3 (4) should not be in position 4 (1); seven of the 24 sequences satisfy these four conditions. The FOCP model's estimates, reported in Table 1, are more straightforward. The model indicates a preference for component 3 in position 1 and component 2 in position 4; the only remaining ambiguity is whether to place component 1 or 4 in position 2. If one retains the term contrasting component 1's position with that of component 4 (with p-value 0.062), the optimal sequence is 3142 (mitoxantrone, paclitaxel, cisplatin, doxorubicin).

Table 4: Example 1: Tests for $H_0: \beta_i = \mathbf{0}$, $(i = 1, 2, 3)$

Source	df	Sum of squares	Mean Square	F	p-value
CP model	9	1066.19	118.47	8.43	0.0003
Linear	3	994.94	331.65	23.61	9.7e-6
Quadratic	3	51.43	17.14	1.22	0.34
Cubic	3	19.81	6.60	0.47	0.71
Error	14	196.63	14.045		
Total	23	1262.81			

4.2 Example 2 ($m = 5$, $n = 40$)

The following five chemotherapeutics are studied: 1: paclitaxel; 2: doxorubicin; 3: mitoxantrone; 4: cisplatin; 5: etoposide. Example 2 is more challenging to analyze, because neither the main-effects PWO model nor the nominal CP model adequately account for these data. Mee (2020, Table 6) shows that the main-effects PWO model is not satisfactory, and that some triplet terms are needed. Yang, Sun, and Xu (2020) found that the nominal CP model fits better than the main-effects PWO model, but that two interactions are also needed for the CP model. Stokes and Xu (2022) found

Statistica Sinica: Preprint
doi:10.5705/ss.202022.0003

that their SOCP model was preferred to their FOCP and QCP models. Table 5 shows a sequential partitioning of the sum of squares for the SOCP model.

Table 5: Example 2: Partitioning sum of squares for the SOCP model

Source	df	Sum of squares	Mean Square	F	p-value
SO model	14	1412.9176			
Blocks	1	166.1785			
First-order | Blocks	4	873.0682	218.267	16.81	8.3e-7
Quadratic | First-order	4	121.6122	30.403	2.34	0.0827
Second-order | Quadratic	5	252.0587	50.412	3.88	0.0097
Error	25	324.6562	12.986		
Total	39	1737.5737			

Is the SOCP model adequate? Adding pure cubic terms to the SO model is not useful (F = 9.17/13.71 = 0.67; p = 0.62). Similarly, PWO terms added to the SOCP model are not significant (F = 2.70/16.24 = 0.17; p = 0.98). These nonsignificant results lend support to the SOCP model, which has $R^2 = 81.3\%$. Stokes and Xu (2022) use the AIC to select a reduced model with only two interactions and one quadratic term. For ease of interpretation, we express their model in terms of the elements of $\mathbf{O}$, with possible values 1,. . . , 5:

$$\widehat{Inhibit\%} = 60.09 \pm 2.04 + 0.18o_{i2} - 2.24o_{i3} + 0.92o_{i4} - 2.28o_{i5} \tag{4.2}$$

$$+ 0.63(o_{i3} - 3)^2 + 0.91(o_{i2} - 3)(o_{i5} - 3) - 0.82(o_{i3} - 3)(o_{i5} - 3).$$

The first-order terms indicate that components 3 and 5 should appear early, and that component 4 should appear last. The quadratic term for component 3 strengthens the benefit of having component 3 first, and the 3∗5 interaction indicates that if component

Statistica Sinica: Preprint
doi:10.5705/ss.202022.0003

3 is first, the need to have component 5 early is reduced. The sequence 35214 has the highest $\widehat{Inhibit}\%$: 61.7; there are five other sequences with $\widehat{Inhibit}\% > 60.7$, all beginning with component 3 (mitoxantrone): 31524, 35124, 32514, 35241, and 31452.

Consider now a PWO model with triplet interactions, selected using a stepwise regression with the AIC; this is the same model selection procedure used by Stokes and Xu (2022) and Yang, Sun, and Xu (2020) to obtain their models. The selected triplet PWO model is

$$\begin{aligned}\widehat{Inhibit}\% = 50.47 \pm 2.57 + 0.84P_{(12)} - 0.85P_{(13)} - 1.92P_{(15)} - 1.68P_{(25)} \\ + 2.95P_{(34)} + 0.03P_{(35)} + 2.19P_{(12)} * P_{(15)} - 2.67P_{(13)} * P_{(15)} + 1.68P_{(34)} * P_{(35)}.\end{aligned} \tag{4.3}$$

(This model differs from that of Yang, Sun, and Xu (2020, Equation 10) because they allowed interactions from $\mathbf{U}$; that is, they fit terms from (2.3) rather than (2.2).) The main effects in (4.3) indicate that sequences with "3 before 4" and "5 before 1 before 2"' are desired. Without the interactions, whether component 3 or 5 comes first matters little, given the 0.03 coefficient for $P_{(35)}$. However, the last interaction term produces a strong preference for "3 before 5." The terms involving $P_{(15)}$ are $(-1.92 + 2.19P_{(12)} - 2.67P_{(13)}) * P_{(15)}, and$ sequences with "1 before 2" and "3 before 1" make this $2.94P_{(15)}$; thus, component 5 should come after component 1, reversing the order indicated by $\hat{\alpha}_{(15)} = -1.92$.

Now, we compare the favored sequences for these two models. For the reduced triplet model (4.3), the four sequences with the order "3512" (and 4 anywhere after 3) have the same highest predicted response. For the reduced SOCP model (4.2), the

Statistica Sinica: Preprint
doi:10.5705/ss.202022.0003

four sequences with the highest predicted response all begin with component 3 and end with component 4. Thus, only the sequence 35124 scores among the best four for both models. These models do agree that a main-effects model does not suffice, because the effect of chemotherapeutic 5's position depends on the positions of chemotherapeutics 2 and 3.

Yang, Sun, and Xu (2020, equation (11)) proposed a reduced nominal CP model with two interactions for these data. For their model, all six sequences that begin "35xxx" have the same maximum predicted response; the six sequences with "3x5xx" have the next best predicted response.

Xiao and Xu (2021) analyze these data using Kriging models. There is considerable inconsistency between the predicted best sequences according to these various models; only the sequence 31524 is among the best for all models. For the fitted MUK model with 18 estimated parameters, Xiao and Xu (2021) reported that only five of the 10 sequences with the largest predicted reported were among the 80 sequences not in the experiment. In contrast, the parametric models consistently predict a higher percentage of the best sequences among the two-thirds of sequences not in the experiment. Of the top 12 sequences for the reduced nominal CP model (reduced SOCP model), 8 (7) were not in the experiment; for the reduced triplet PWO model, 9 of 13 sequences with the top predicted responses were not in the experiment. This difference between the predictions with the highest (or lowest) response for parametric and Gaussian process models is likely to occur in other applications, especially when $n/m!$ is small.

Statistica Sinica: Preprint
doi:10.5705/ss.202022.0003

4.3 Example 3 ($m = 6$, $n = 24$)

Voelkel and Gallagher (2019) describe three order-of-addition experiments seeking to increase the shear thinning behavior of automotive coatings. The response variable is the log of the ratio of low shear viscosity (LSV) to high shear viscosity (HSV). Here, a higher LSV means the coating will not separate when stored, and a lower HSV means the coating will level well when sprayed, producing a smoother finish. Their third experiment involved the order of the following six components: 1) primary binder resin; 2) secondary binder resin; 3) flow and leveling additive; 4) rheology modifier #1; 5) crosslinking resin; and 6) rheology #2. The data for this 24-run experiment appear in Voelkel (2019c). The full main-effects PWO model (2.1) has $R^2 = 95.3\%$ (RMSE = 0.047), whereas a model with just three of the 15 main-effect terms has $R^2 = 91.8\%$ (RMSE = 0.039):

$$\widehat{log_{10}(LSV/HSV)} = 0.454 - 0.120P_{(14)} - 0.024P_{(15)} + 0.037P_{(16)}. \tag{4.4}$$

Consider now the various CP models. We cannot estimate the full model (2.4), because $n < 25$, but we can estimate each of the models of Stokes and Xu (2022). The FOCP model (2.5) has $R^2 = 61.8\%$ (RMSE = 0.089). Only two terms are statistically significant, $\hat{\beta}_1 = 0.056$ and $\hat{\beta}_4 = -0.069$, indicating that the primary binder resin should be last and the rheology modifier #1 should be first. This is consistent with (4.4). However, if we assume model (2.1) is the true model, model (2.5) has significant lack-of-fit, per the test in equation (3.1): $F = 0.0126/0.0022 = 5.64$; $p = 0.011$. Before

Statistica Sinica: Preprint
doi:10.5705/ss.202022.0003

investigating this lack-of-fit, we consider the QCP and SOCP models. Models (2.9) and (2.11) have RMSE = 0.096 and 0.110, respectively, both larger than that of the FOCP model. Thus, adding higher-order functions of component position does not explain the order effects here.

What does explain the lack-of-fit of the FOCP model? In (3.3), we observed that the first-order Stokes and Xu model assumes $\alpha_{(jk)} = \alpha_{(1k)} - \alpha_{(1j)}$, for all $2 \leq j < k \leq m$. Thus, one might define individual lack-of-fit terms as

$$LOF_{jk} = P_{(jk)} - P_{(1k)} + P_{(1j)} \text{ for all } 2 \leq j < k \leq m. \tag{4.5}$$

Adding all $\binom{m-1}{2}$ LOF_{jk} columns to model (2.5) creates a model that is a reparameterization of the PWO model (2.1). A simple analysis strategy to assess departures from model (2.5) is to include all terms for the FOCP model, and then to use forward selection to add the LOF_{jk} terms. Doing so for this example, we add the following three terms with naive p-values < 0.05: in order, LOF_{46} (p = 0.002), LOF_{24} (p = 0.004), and LOF_{34} (p = 0.049). These indicate that the constraints on the PWO coefficients $\alpha_{16} = \alpha_{12} + \alpha_{26}$ and $\alpha_{14} = \alpha_{12} + \alpha_{24} = \alpha_{13} + \alpha_{34}$ required by the FOCP model are contradicted by the data.

Figure 2a reveals how $\log_{10}$(LSV/HSV) is not affected by the position of component 4 once we control for the pairwise order P_{14}. Finally, if we include $\mathbf{x}_i$ $(i = 1, \ldots, 6)$ and the 15 $P_{(ij)}$ factors as eligible terms, and perform forward selection, the first five terms to enter are PWO factors; when $\mathbf{x}_3$ enters in step 6, its p-value = 0.28. These

Statistica Sinica: Preprint
doi:10.5705/ss.202022.0003

results indicate the importance of pairwise orders, especially relative to component 1, the primary binder resin.

Building on the ideas of Peng, Mukerjee, and Lin (2019), Voelkel and Gallagher (2019) considered several tapered PWO models. The largest extra tapering parameter found was for $P_{(14)}$, but it was not statistically significant ($p = 0.094$). Figure 2b shows some evidence of the $P_{(14)}$ effect diminishing when component 4 precedes component 1 as the distance increases, but none when component 4 follows component 1. Finally, Voelkel and Gallagher (2019) commented that component 6 is less sensitive to order effects, and so imparts shear thinning to all samples. In Section 5, we show how to detect whether a component has any effect.

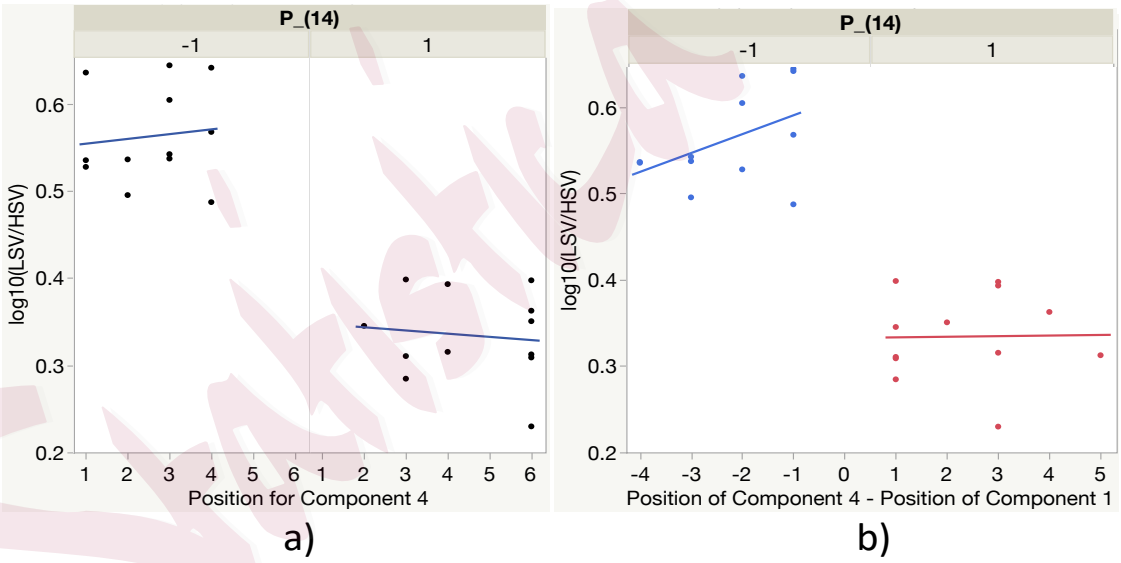

Figure 2: a) $\log_{10}$(LSV/HSV) versus O_4 by $P_{(14)}$; b) $\log_{10}$(LSV/HSV) versus $O_4 - O_1$

5. Discussion

This paper synthesizes the parametric CP models and the pairwise order models,

Statistica Sinica: Preprint
doi:10.5705/ss.202022.0003

showing their connections and identifying lack-of-fit tests. Building on the results of Schoen and Mee (2021) for the FOCP model, we proivde closed-form expressions for the information matrix for the QCP and SOCP models, provided that the design has the same normalized information matrix as that of the full $m!$ design. The FOCP model by Stokes and Xu (2022) is nested in both the main-effects PWO model (2.1) and the nominal CP model (2.4). Example 1 illustrated the FOCP model's advantage in terms of interpreting which sequences are best. Example 2 illustrated a case requiring that the interactions fit the data satisfactorily. Section 3.4 discussed connections between CP and PWO models with interactions, and Example 2 illustrated the potential similarity of the PWO models with triplet interactions and the Stokes and Xu (2022) SOCP model (2.11). Example 3 illustrated how to use single degree-of-freedom lack-of-fit terms (4.5) to identify whether the simplifying assumptions (3.3) of the FOCP model are contradicted by the data. When the FOCP model has significant lack-of-fit vis-a-vis the main PWO model, the example in Section 4.3 showed how to include single degree of freedom lack-of-fit terms to determine which PWO parameters are needed. These examples show why it is useful to have a variety of order-of-addition models.

Although we focus on parametric order-of-addition models, Example 2 afforded an opportunity to compare the Kriging CP models of Xiao and Xu (2021) with the parametric CP models. Whether the Kriging models or the parametric models better identify the best sequences based on fractional designs is worthy of further investigation. As mentioned at the end of Example 2, it is relevant to compare alternative models in terms of their ability to identify the sequence that is best. A methodology for multiple

Statistica Sinica: Preprint
doi:10.5705/ss.202022.0003

comparisons with the best (MCB) inference is left to future research.

Order-of-addition experiments with fixed levels are ill-equipped to determine whether a component has any effect at all. This question seems to require experiments in which a step is omitted or a dose level is set to zero. One might conceive of designs with runs that drop one component. This approach, and the larger class of experiments that vary both the sequence and the dose, are important areas of ongoing research on design construction and modeling; see Wang, Xu, and Ding (2020), Rios, Winker, and Lin (2022), and Jiang and Zhang (2022).

Acknowledgments

The author gratefully acknowledges the help of Eric Schoen for his helpful comments and suggestions, and a referee for suggesting Figure 1 and other improvements. I also thank Dr. Chunyan Wang for her help with LaTex.

A. Appendix: Second-order CP model information matrix

Here we present the normalized information matrix for the two versions of the second-order model presented in Section 2.6. First, consider the model matrix arranged

Statistica Sinica: Preprint
doi:10.5705/ss.202022.0003

as $\mathbf{X}_{SO} = [\mathbf{1}_n, \mathbf{B}, \mathbf{Q}, \mathbf{R}]$. For the full $n = m!$ design,

$$\mathbf{X}_{SO}^T \mathbf{X}_{SO}/n = \frac{1}{n} \begin{bmatrix} 1 & 0 & 0 & \mathbf{1}_n^T \mathbf{R} \\ 0 & \mathbf{B}^T\mathbf{B} & 0 & 0 \\ 0 & 0 & \mathbf{Q}^T\mathbf{Q} & \mathbf{Q}^T\mathbf{R} \\ \mathbf{R}^T\mathbf{1}_n & 0 & \mathbf{R}^T\mathbf{Q} & \mathbf{R}^T\mathbf{R} \end{bmatrix}. \tag{A.1}$$

From Section 2.5, we know that $\mathbf{B}^T\mathbf{B}/n = \mathbf{Q}^T\mathbf{Q}/n = (m-1)^{-1}(m\mathbf{I}_{m-1} - \mathbf{J}_{m-1})$, and that $\mathbf{R}^T\mathbf{1}_n/n = -(m-1)^{-1}\mathbf{1}_{(m-1)(m-2)/2}$.

Lemma 4. *Let $c_{(ij)*(kl)}$ denote an element of $\mathbf{R}^T\mathbf{R}/n$ corresponding to the cross-product of $\mathbf{x}_i\mathbf{x}_j$ and $\mathbf{x}_k\mathbf{x}_l$, Then the diagonals and off-diagonals of $\mathbf{R}^T\mathbf{R}/n$ are*

$$c_{(ij)*(ij)} = (m^3 - 1.8m^2 - m + 4.2)/(m^3 - m^2 - m + 1) \tag{A.2}$$

$$c_{(ij)*(jk)} = (-m^2 + 1.6m + 4.2)/(m^3 - m^2 - m + 1) \tag{A.3}$$

$$c_{(ij)*(kl)} = (3m + 4.2)/(m^3 - m^2 - m + 1). \tag{A.4}$$

Lemma 5. *Let $c_{(ij)*Q_k}$ denote an element of $\mathbf{R}^T\mathbf{Q}/n$ corresponding to the cross-product*

Statistica Sinica: Preprint
doi:10.5705/ss.202022.0003

of $\mathbf{x}_i\mathbf{x}_j$ *and* $Q_k = (c_2/c_1)(\mathbf{x}_k^2 - 1)$. *Then*

$$c_{(ij)*Q_k} = \begin{cases} -\frac{\sqrt{0.8}}{m-1}\sqrt{\frac{m^2-4}{m^2-1}} & \text{if } i = k \text{ or } j = k, \\ \\ \frac{4}{m-1}\sqrt{\frac{m+2}{5(m^2-1)(m-2)}} & \text{otherwise.} \end{cases} \tag{A.5}$$

The formulae (A.2)–(A.5) were determined with the aid of symbolic computation.

Since $\mathbf{X}_{SO}$ does not have full column rank, one must drop a column from $\mathbf{Q}$ or a column from $\mathbf{R}$. In Section 2.6, we argued for dropping a column from $\mathbf{R}$.

B. Appendix: Constraints from the quadratic CP model on the triplet PWO parameters

Here we elucidate the constraints imposed by the quadratic CP model (2.9) on the triplet model's τ parameters. Lemma 3 shows that each Q_i is a linear combination of $\binom{m-1}{2}$ triplet interaction terms, one for each {ijk} triplet: $P_{(ij)} * P_{(ik)}$ if $i < j < k$; $P_{(ji)} * P_{(ik)}$ if $j < i < k$; $P_{(ji)} * P_{(ki)}$ if $j < k < i$.

Assume that model (2.9) is the true model. The coefficient of Q_i, β_{2i}, implies a coefficient of $\pm 0.5 c_1 c_2 \beta_{21}$ to $\binom{m-1}{2}$ triplet interactions. For instance, for $i = 2$, the $m - 2$ antagonistic interactions $P_{(12)} * P_{(2k)}$ $(3 \le k \le m)$ have coefficient $-0.5 c_1 c_2 \beta_{21}$, while the $\binom{m-2}{2}$ synergistic interactions $P_{(2j)} * P_{(2k)}$ $(3 \le j < k \le m)$ have coefficient $0.5 c_1 c_2 \beta_{21}$. If we write the triplet model in terms of the $2\binom{m}{2}$ synergic interactions,

Statistica Sinica: Preprint
doi:10.5705/ss.202022.0003

then we replace each antagonistic interaction term using the equality $P_{(ji)} * P_{(ik)} = P_{(ji)} * P_{(jk)} + P_{(jk)} * P_{(ik)} - \mathbf{1}_n$. For the full triplet model with all synergistic interactions, the quadratic model implies the constraints

$$\frac{\tau_{ij \cdot ik}}{0.5 c_1 c_2} = \beta_{2i} - \beta_{2j} \; \forall \;\; 1 \leq i < j < m; \;\; k = j+1, \ldots, m$$

$$\frac{\tau_{ik \cdot jk}}{0.5 c_1 c_2} = \beta_{2k} - \beta_{2j} \; \forall \;\; 1 < j < k \leq m; \;\; i = 1, \ldots, j-1,$$

where we take the convention that $\beta_{2m} = 0$. These constraints imply an additive structure for sets of PWO interaction coefficients. For example, for $m \geq 5$,

$$\frac{\tau_{12 \cdot 15} + \tau_{23 \cdot 25} + \tau_{34 \cdot 35}}{0.5 c_1 c_2} = (\beta_{21} - \beta_{22}) + (\beta_{22} - \beta_{23}) + (\beta_{23} - \beta_{24}) = (\beta_{21} - \beta_{24}) = \frac{\tau_{14 \cdot 15}}{0.5 c_1 c_2}.$$

References

Chen, J. B., Peng, J. Y., and Lin, D. K. J. (2021). A statistical perspective on non-deterministic polynomial-time hard ordering problems: Making use of design for order-of-addition experiments. *Comput. Ind. Eng.* **162**, DOI10.1016/j.cie.2021.107773

Cornell, J. A. (2002). *Experiments with Mixtures: Designs, Models, and the Analysis of Mixture Data*, Wiley.

Hopper, J. W. (1959). Simultaneous equations and canonical correlation theory. *Econo-*

metrica **27**, 245–256.

Jiang, H. and Zhang, C. (2022). Construction of full order-of-addition generalization simplex-centroid designs by the directed graph approach. *Mathematics* **10**, p.443. https://doi.org/10.3390/math10030443.

Lin, D. K. J. and Peng, J. (2019). Order-of-addition experiments: A review and some new thoughts. *Qual. Eng.* **31**, 41–59.

Mee, R. W. (2020). Order-of-addition modeling. *Statist. Sinica* **30**, 1543–1559.

Peng, J., Mukerjee, R. and Lin, D. K. J. (2019). Design of order-of-addition experiments. *Biometrika* **106**, 683–694.

Piepho, H.-P. and Williams, E. R. (2021). Regression models for order-of-addition experiments. *Biometrical J.*, DOI: 10.1002/bimj.202100048.

Rios, N., Winker, P. and Lin, D. K. J. (2022). TA algorithms for D-optimal OofA Mixture designs. *Comput. Statist. Data Anal.* **168**, 107411.

Scheffe, H. (1958). Experiments with mixures, *J. Roy. Statist. Soc. Ser. B* **20**, 344–360.

Schoen, E. D. and Mee, R. W. (2021). Order of addition orthogonal arrays to elucidate the sequence effects of treatments. Preprint.

Stokes, Z. and H. Xu. (2022). A position-based approach for design and analysis of order-of-addition experiments. *Statist. Sinica*, **32**, 1467-1488.

Tsai, S.-F. (2021). Generating optimal order-of-addition designs with flexible run sizes. *J. Statist. Plann. Inference* **218**, 147–163.

Van Nostrand, R. C. (1995). *Design of experiments where the order of addition is important.* ASA Proceedings of the Section on Physical and Engineering Sciences, 155–160. Alexandria, VA: American Statistical Association.

Voelkel, J. G. (2019a). The design of order-of-addition experiments. *J. Qual. Technol.* **51**, 230–241. *Qual. Eng.* **51**, 230–241

Voelkel, J. G. (2019b). Discussion on 'Order-of-addition experiments: A review and some new thoughts.' *Qual. Eng.* **31**, 60–63.

Voelkel, J. G. (2019c). "Data from: Order of addition experiments: five designs and three datasets," Mendeley Data, v1. http://dx.doi.org/10.17632/8vyvrkd7tk.1.

Voelkel, J. G. and Gallagher, K. P. (2019). The design and analysis of order-of-addition experiments: An introduction and case study. *Qual. Eng.* **31**, 627–638.

Wang, A., Xu, H. and Ding, X. (2020). Simultaneous optimization of drug combination dose-ratio-sequence with innovative design and active learning. *Adv. Ther.* **3**, 1900135. https://doi.org/10.1002/adtp.201900135

Wang, C. and Mee, R.W. (2022). Saturated and supersaturated order-of-addition designs. *J. Statist. Plann. Inference* **219**, 204–215.

Xiao, Q. and Xu, H. (2021). A mapping-based universal Kriging model for order-of-

Statistica Sinica: Preprint
doi:10.5705/ss.202022.0003

addition experiments in drug combination studies. *Comput. Statist. Data Anal.* **157**, 107155.

Yang, J. F., Sun, F. S. and Xu, H. (2020). A component-position model, analysis and design for order-of-addition experiments. *Technometrics*, **63**, 212–224.

Zhao, Y., Lin, D. K. J. and Liu, M.-Q. (2022). Optimal designs for order-of-addition experiments. *Comput. Statist. Data Anal.* **165**, 107320.

Department of Business Analytics and Statistics, University of Tennessee, Knoxville, Tennessee 37996, United States

E-mail: rmee@utk.edu