

## Two-Level Factorial Design with Circular Response: Model and Analysis

Alyaa Roshdy Zahran  
*Cairo University*

*Abstract:* Since late thirties, factorial analysis of a response measured on the real line has been well established and documented in the literature. No such analysis, however, is available for a response measured on the circle (or sphere in general), despite the fact that many designed experiments in industry, medicine, psychology and biology could result in an angular response. In this paper a full factorial analysis is presented for a circular response using the Spherical Projected Multivariate Linear model. Main and interaction effects are defined, estimated and tested. Analogy to the linear response case, two new effect plots: Circular-Main Effect and Circular-Interaction Effect plots are proposed to visualize main and interaction effects on circular responses.

*Key words:* Circular-Interaction plot, Circular-Main Effect plot, EM algorithm, main and interaction effects, Spherical Projected Multivariate Linear model.

### 1. Introduction

Factorial designs are widely used in experiments involving several factors where it is necessary to investigate the joint effects of the factors on a response variable. These joint effects include either the sole effect of each factor (main) or any interaction between two or more factors. The analysis of factorial designs is well established for a response variable that is measured on the real line. Ideas of this analysis were developed in the late thirties by Fisher in 1935 and Yates in 1937 [Hinkelmann and Kempthorne, 1994, p. 350]. Since then this topic was a standard chapter in every Design of Experiments textbook; among these are: Fisher (1960), Kuehl (1994), Hinkelmann and Kempthorne (1994), Myers and Montgomery (2002), Montgomery (2009).

Little attention has been paid, however, to the analysis of factorial designs when the response is measured on a circle (or sphere in general). Two approaches to circular data analysis are generally used in the literature: the intrinsic and the

embedding approach. In intrinsic approach, circular data are analyzed in its angular form (polar coordinates with a unit radius), in which angles are considered random variables on the unit circle that follow some directional probability distributions. Most commonly, the von Mises distribution is assumed which plays a key role in statistical inference similar to the role of the normal distribution on the real line. Underwood and Chapman (1985) developed a multi-way analysis of a circular response for crossed designs as an extension of Watson-Williams likelihood ratio test that had been initially developed for the one-way analysis. However, a simulation study of Anderson and Wu (1995) showed that this extension does not suit factorial designs because of the difficulty of identifying what the interaction terms measure and the possibility of getting negative sum of squares for these terms. In the embedding approach, the angles are represented by unit vectors (Cartesian representation). Harrison, Kanji and Gadsden (1986) and Harrison and Kanji (1988) used this approach to develop an ANOVA type approach to circular data. Considering this geometric aspect when decomposing the variance yields sums of squared of norms that would never be negative. This decomposition works only for problems where location and dispersion effects are studied simultaneously [Anderson and Wu, 1995]. Anderson and Wu (1995) proposed a likelihood ratio test statistic for testing the effects from a factorial design on the location of a circular response. They used the angular rotations occurring when moving from one level of one factor to its other level at fixed level of the other factor to evaluate the two-factor interaction effect on the location. If these angular rotations are not the same in magnitude or direction, then there is evidence of interaction. For higher order interaction, a two level surrogate factor is defined such that all combinations of odd high levels are considered as one level and all the combinations of even high levels constitute the second level. If the angular rotation resulting when moving from the high level of this surrogate factor to its low level is zero, then there is no interaction. This zero angular rotation concept is also used to evaluate any main effect. However no effect estimation was given in their paper, while merely relative importance of the significant effects was used as an approach to rank effects. Anderson and Wu (1995) justified the lack of model by intractability reasons.

In the context of regression analysis of circular response on real-line valued independent variables, remarkable link functions were offered to map the real line onto the unit circle (see Gould, 1969, Johnson and Wehrly, 1978, and Fisher and Lee, 1992). However, all of these models suffer from computational difficulties because of either the existing of a multimodal likelihood or an irregular likelihood where the maximum might occur on a very narrow peak [Presnell *et al.*, 1998]. In addition, for factorial design, main and interaction effects are hard to be identified and separated from each other [Anderson and Wu, 1995].

Presnell *et al.* (1998) emphasized that the embedding approach is more useful

in modeling directional data on real-line valued variables. They proposed the Spherically Projected Multivariate Linear (SPML) model for circular response. The EM algorithm or Newton-Raphson algorithm could be used in the estimation phase. Presnell *et al.* noted that Newton-Raphson did converge faster than the EM algorithm but the latter is easier to program. In addition they indicated that SPML model could work well for designed experiments, but did not show any analysis in that context.

In brief, up to date no workable model is known for the analysis of factorial designs with circular response. This results in the lack of complete analysis for factorial designs for circular responses. In this paper, the SMPL model is used to analyze factorial designs with circular response. Utilizing the SPML model, enables the author to offer a complete factorial analysis (estimation, testing and interpretation of the effects) for a circular response in a factorial setup. In addition and analogy to the effect-plots for responses measured on the real line, two new effect-plots are introduced to complete the factorial analysis for circular responses. Definitions of main and interaction effects under SPML model are introduced in Section 3. In addition, two proposed effect plots: Circular-Main Effect (C-ME) and Circular-Interaction Effect (C-IE) plots are introduced. Effect estimation and testing are presented in Section 4. An illustrative example is given in Section 5. Finally, Section 6 concludes the paper. The following section gives the notations of circular data and briefly reviews the SPML model.

## 2. Notations and the SPML Model for Circular Data

Let  $\theta_i, i = 1, \dots, n$ , be a random sample of circular data. Then  $\mathbf{u}_i = ([c_i, s_i])'$ , is a unit random vector pointing in the direction of the angle  $\theta_i$  with a mean direction unit vector of  $E(\mathbf{u}_i)/\rho$ , where  $c_i = \cos(\theta_i)$  and  $s_i = \sin(\theta_i), i = 1, \dots, n$ ,  $\rho = \|E(\mathbf{u}_i)\|$  and  $\|\cdot\|$  represents the length of the vector. The mean resultant vector  $\bar{\mathbf{u}} = [\sum_i c_i/n, \sum_i s_i/n]'$  is pointing in the direction of the overall mean direction (average direction of the sample)  $\bar{\theta}$  while its length,  $\bar{R} = \|\bar{\mathbf{u}}\|$ , is used as a concentration measure. The circular variance is  $(1 - \bar{R})$ . When data points are more concentrated around the mean, the circular variance is close to zero. A close to one value is associated with data that are evenly spread around the circle.

The idea of the embedding approach is that distributions on  $S^{p-1}$  can be obtained by radial projection of distributions on the  $\mathfrak{R}^p$ . If  $\mathbf{y}$  is a vector in  $\mathfrak{R}^2$ , where  $\Pr(\mathbf{y} = 0) = 0$ , then the distribution of  $\mathbf{y}/\|\mathbf{y}\|$  is a random point on the unit circle. Under the assumption that  $\mathbf{y}$  is bivariate normal ( $\mathcal{N}_2$ ) random vector with mean vector  $\mu$  and variance covariance matrix  $\Sigma$ , Mardia (1972) proved that the unit vector  $\mathbf{y}/\|\mathbf{y}\|$  is distributed as projected normal ( $\mathcal{PN}_2$ ) with mean vector  $\mu$  and variance covariance matrix  $\Sigma$  [Mardia and Jupp, 1999, p.

46]. Assuming  $\mathbf{u}_i = \mathbf{y}_i / \|\mathbf{y}_i\|$ , Presnell *et al.* (1998) developed the SPML model, where  $\mathbf{u}_i \sim \mathcal{PN}_2(\boldsymbol{\mu}_i, \Sigma)$ ,  $\boldsymbol{\mu}_i$  is the mean direction vector that is pointing in the direction of the angle  $\varpi_i$  and its length serves as the concentration parameter. Further, SPML model assumes that the mean direction vector is linked linearly to  $k$  covariates as follows:  $\boldsymbol{\mu}_i = B' \mathbf{x}_i$  where  $\boldsymbol{\mu}'_i = (\mu_i^{\cos}, \mu_i^{\sin})$ ,  $B = (\boldsymbol{\beta}^{\cos}, \boldsymbol{\beta}^{\sin})$ ,  $\mathbf{x}'_i = (1, x_{i1}, \dots, x_{ik})$  and each  $\boldsymbol{\beta}$  is of order  $(k + 1)$ . Note that the dispersion depends on the covariates since it is a function of the concentration parameter  $\eta_i = \|\boldsymbol{\mu}_i\|$ . The SPML model parameters are only identifiable under  $|\Sigma| = 1$ . Presnell *et al.* (1998) used the identity matrix as the variance covariance matrix and showed that  $\mathcal{PN}(\mu_i, I)$  is a good approximation of the von Mises distribution. Using the EM algorithm, the SPML parameters could be estimated. Presnell *et al.* (1998) indicated that the predicted angle

$$\widehat{\varpi}_i = \left[ \tan^{-1} \left( \frac{x'_i \hat{\beta}_{\sin}}{x'_i \hat{\beta}_{\cos}} \right) + \pi I(x'_i \hat{\beta}_{\cos} < 0) \right] \bmod 2\pi$$

is approximately normal distributed with mean  $\varpi_i$  and approximate variance

$$\frac{1}{\|\hat{\boldsymbol{\mu}}\|^4} (-\hat{\mu}_{\sin} \mathbf{x}'_i, \hat{\mu}_{\cos} \mathbf{x}'_i) [-\ddot{i}(\hat{B})]^{-1} \begin{pmatrix} -\hat{\mu}_{\sin} \mathbf{x}'_i \\ \hat{\mu}_{\cos} \mathbf{x}'_i \end{pmatrix},$$

where  $-\ddot{i}(\hat{B})$  is the observed Fisher information matrix associated with SPML model. The log likelihood function of the SPML model is

$$l(\beta) = -\frac{1}{2} \sum_{i=1}^n \boldsymbol{\mu}'_i \boldsymbol{\mu}_i + \sum_{i=1}^n \psi(u'_i \boldsymbol{\mu}_i) - n \log(2\pi),$$

where  $\psi(t) = \log[1 + (t\Phi(t)/\phi(t))]$ ,  $\phi(t)$  and  $\Phi(t)$  are the standard normal density function and the standard normal distribution function, respectively. Presnell *et al.* (1998) proposed using the asymptotic likelihood ratio test to compare nested models; and the Akaike Information and/or Bayes Information Criterion (AIC and/or BIC) to select the best model among non-nested models.

### 3. Factorial Designs, SPML Model and SPML-Effect Model

The radial projection of the circular data facilitates the use of multivariate regression, where the response variables are the Cartesian coordinates of the angles. Using SPML model in the context of factorial design enables us to define, estimate and test the main and interaction effects in a unified manner.

Consider a  $2^2$  factorial design with interaction, SPML model is

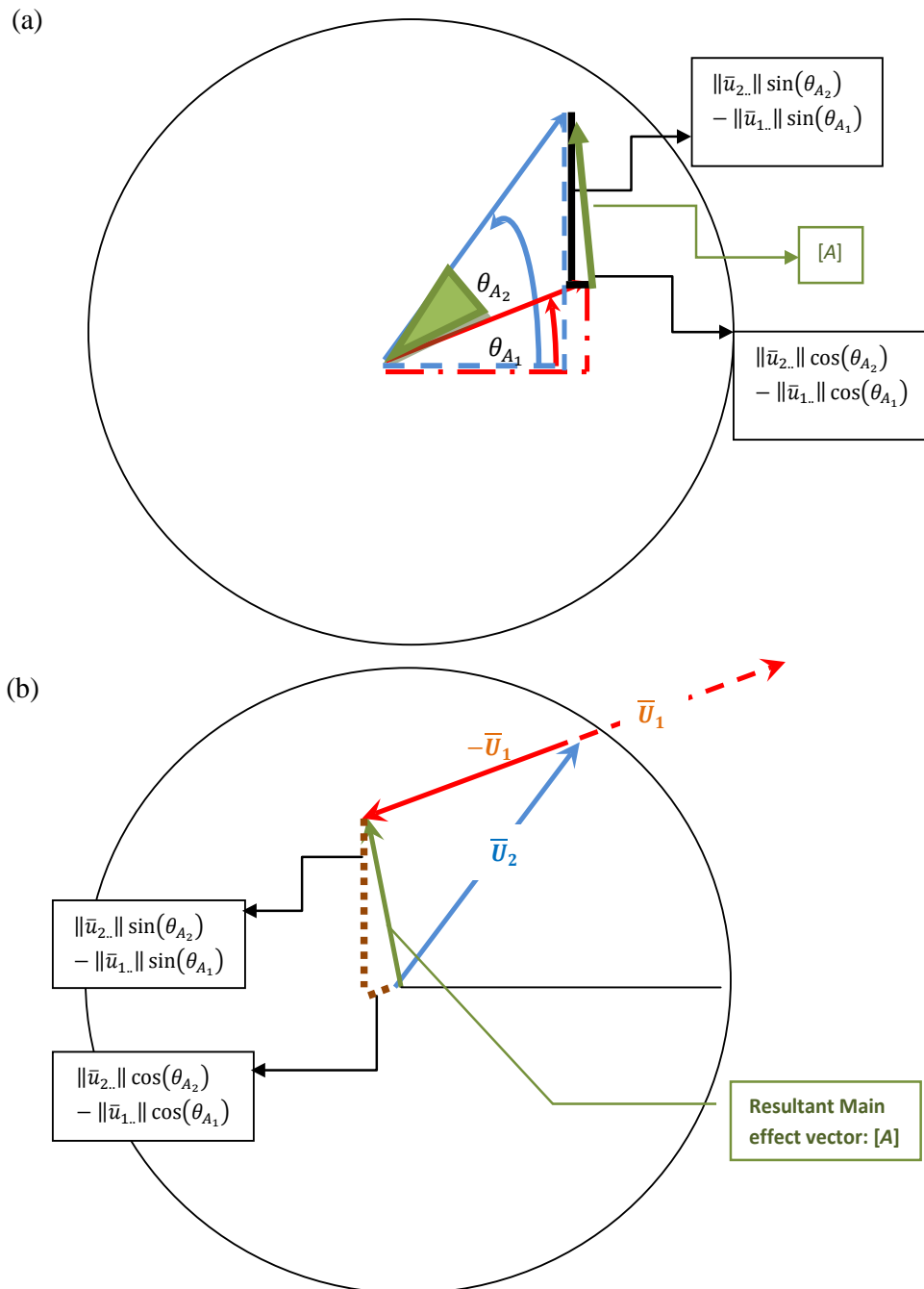
$$\boldsymbol{\mu}'_{ijk} = \boldsymbol{\varepsilon}_{ijk} + [\beta_0^{\cos} + x_{ik}\beta_A^{\cos} + x_{jk}\beta_B^{\cos} + x_{ik}x_{jk}\beta_{AB}^{\cos}, \beta_0^{\sin} + x_{ik}\beta_A^{\sin} + x_{jk}\beta_B^{\sin} + x_{ik}x_{jk}\beta_{AB}^{\sin}], \quad (3.1)$$

where  $u_{ijk}$  is the unit vector associated with angle  $\theta_{ijk}$ , which is the  $k^{\text{th}}$  replicate that is measured at the  $i^{\text{th}}$  level of factor  $A$ , and the  $j^{\text{th}}$  level of factor  $B$ ,  $i = 1, 2$ ,  $j = 1, 2$ , and  $k = 1, \dots, r$  ( $r$  the number of replicates). The  $x$ 's are coded variables that are set at  $-1$  if the corresponding factor is at its low (1) level and  $1$  if it is at its high (2) level.  $\varepsilon_{ijk} \sim \mathcal{PN}(\mathbf{0}, I)$  is an error row vector of order 2 associated with  $\theta_{ijk}$  while  $\beta$ 's are the suitable intercept and slope parameters associated with factor  $A$  and  $B$  in the sine and cosine equations (Cartesian representation). The SPML model decomposes the main (or interaction) effect into two sub-effects associated with the Cartesian representation. The SPML slope parameters can be multiplied by two to be analogous to an effect of a two-unit change in a factor changing from  $-1$  to  $1$ . The main/interaction effect is then the combined effect of its sub-effects that will affect the direction of the mean response vector and/or its length.

Figure 1(a) shows high and low fitted mean response vectors of factor  $A$ ,  $\bar{\mathbf{u}}_{2..}$  and  $\bar{\mathbf{u}}_{1..}$  and a possible effect of factor  $A$ . The main effect is then the change that occurs in the direction and/or the length of the fitted mean response vector when changing factor  $A$  from high to low averaged over all levels of factor  $B$  and all its replicates. This main effect is a joint location-dispersion effect, where the location effect is the angular difference between the two angles  $\theta_{A_2}$  and  $\theta_{A_1}$  associated with  $\bar{\mathbf{u}}_{2..}$  and  $\bar{\mathbf{u}}_{1..}$ , while the dispersion effect is the difference in the length of these two vectors. Differences that occur in the cosine and sine coordinates when changing from high to low construct two important sine and cosine coordinates which define the resultant main-effect vector  $[A] = \bar{\mathbf{u}}_{2..} - \bar{\mathbf{u}}_{1..}$ . Figure 1(b) shows the proposed Circular-Main Effect (C-ME) plot. The graph depends on graphing the two mean vectors with the low-mean vector translated to begin from the endpoint of the high-mean vector and mirrored/reflected (graphed in the opposite direction of its normal direction) to reflect the fact that it is multiplied by a negative sign. There will be no main effect if  $[A] = \mathbf{0}$ , i.e., both coordinates of the resultant main effect vector are zeros. In Figure 1(b) a nonzero  $A$  main effect is present. To have a zero  $A$  main effect both its slopes in SPML has to be zero (see Appendix 1). In Appendix 2, possible main effect scenarios are shown.

Regarding the interaction effect, it is hard to visualize the SPML slopes of sub-interaction effects in a graph like Figure 1(a), but it could be visualized in a plot similar to C-ME plot. Figure 2 depicts the Circular-Interaction Effect (C-IE) plot that shows the four fitted mean response vectors that are involved in AB interaction effect [for more details see Appendix 3]. In the proposed C-IE plot no interaction occurs if the resultant vector of the four vectors is the zero-vector. Figure 2 shows a nonzero interaction.

Analogy to the factorial design in the linear response, the simple effect of a factor is the mean response change that occurs when moving from the high level



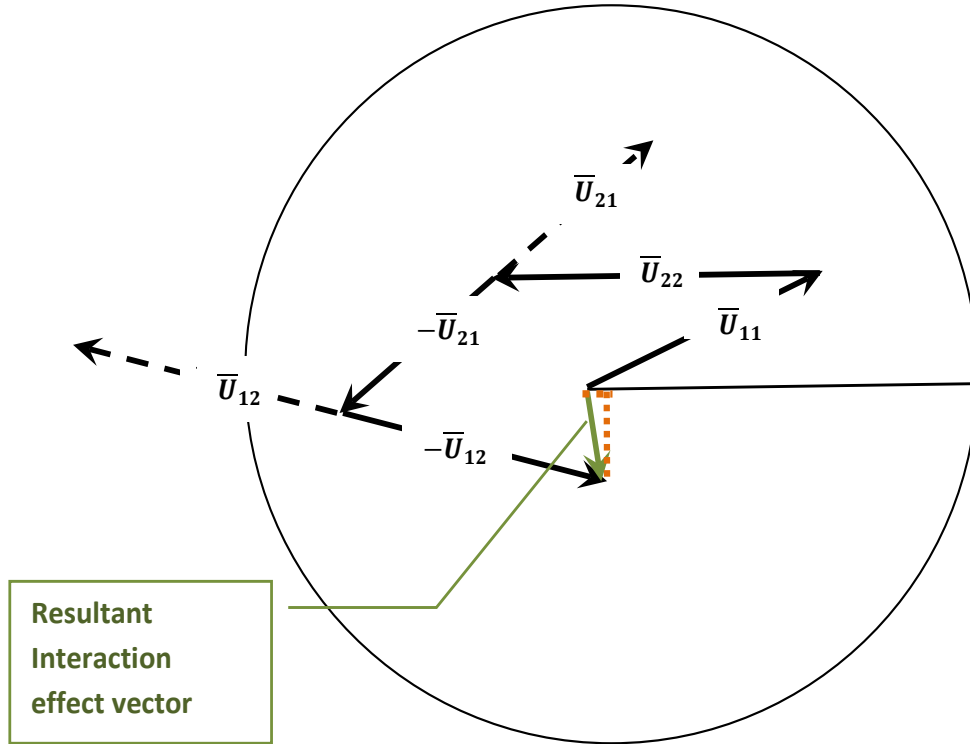


Figure 2: Proposed C-IE plot of  $AB$

of the factor to its low level at one level of the other factor (or at a one combination of levels of other factors) in the experiment. A sub-main effect is then the average of the simple effects of one factor at different levels of other factor (or combinations of other factors) in the experiment. While a sub-interaction effect exists if the simple effect of one factor depends on the level of other factor (or on the combinations of other factors) in the experiment. Using the interaction definition of Hinkelmann (2004) in the linear response case, the sub-interaction is the “average difference” between the simple effects of one factor at levels of other factor (or combinations of other factors) in the experiment. Higher order interactions are definable also in terms of simple effects of the lower-factor interactions (see example in Appendix 3).

Analogy to the linear response case, the SPML-effect model could be defined as

$$\mathbf{u}'_{ijk} = \boldsymbol{\varepsilon}_{ijk} + \left[ \mu_{\cdot}^{\cos} + \tau_{A_i}^{\cos} + \tau_{B_j}^{\cos} + \tau_{AB_{ij}}^{\cos}, \mu_{\cdot}^{\sin} + \tau_{A_i}^{\sin} + \tau_{B_j}^{\sin} + \tau_{AB_{ij}}^{\sin} \right],$$

where  $\tau_A$ ,  $\tau_B$  and  $\tau_{AB}$  represent the effects of  $A$ ,  $B$  and their interaction, respec-

tively, at the different levels  $i = 1, 2, j = 1, 2$ , whereas  $\mu.$  is the grand mean.  $\tau_A$  ( $\tau_B$ ) is the deviation of  $\mu_i.$  ( $\mu.j$ ) from the grand mean, whereas  $\tau_{AB}$  is the difference of the deviation of  $\mu_{ij}$  mean from the grand mean and the sum of  $\tau_{Ai}$  and  $\tau_{Bj}$  at suitable  $i$  and  $j$  levels.

#### 4. Estimation and Testing Factor Effects in SPML Model

Estimation and testing of the SPML parameters is documented in Presnell *et al.* (1998). In the factorial design case and as a result of the orthogonality of the  $X$  columns, we have  $X'X = 2^f r I_p$ , where  $f$  is the number of factors in the experiment and  $p = k + 1$  is the number of parameters in one equation. This simplifies the SPML likelihood estimation equations as following:

$$B_{p \times 2} = \frac{1}{2^f r} X'_{p \times n} M_n U_{n \times 2},$$

where

$$M_n = \text{diag}\{\psi(\mathbf{u}'_1 B' \mathbf{x}_1), \dots, \psi(\mathbf{u}'_n B' \mathbf{x}_n)\}, \quad \psi(t) = t + \frac{\Phi(t)}{\Phi(t) + t\phi(t)},$$

$$U_{n \times 2} = \begin{pmatrix} \mathbf{u}'_1 \\ \vdots \\ \mathbf{u}'_n \end{pmatrix},$$

$\mathbf{u}'_i = (\cos(\theta_i), \sin(\theta_i))$ ,  $\Phi(t)$  is the standard normal distribution form and  $\phi(t)$  is the standard normal density form.

To test the significance of a main or interaction effect, the asymptotic likelihood ratio test is used. Generally in factorial designs, interaction effects are tested first, if it was significant the main effects of the factors involved in this interaction are not to be tested.

#### 5. Illustrative Example

The effect of four two-level factors ( $A$ : location of a butt weld to the flywheel-fixed or random,  $B$ : flywheel radius grade- low or high,  $C$ : flywheel thickness grade-high or low,  $D$ : size of the counter weight attached at 0-low or high) on the angle where a corrective weight should be attached to balance the engine part is under study. Each combination of the factors is replicated ten times. The data set is published in Anderson and Wu (1996). A MINITAB14 macro is written to fit the SPML model and to calculate the likelihood ratio test. The macro programs the EM algorithm and graphs the C-ME and C-IE plots.

Figure 3 shows C-ME plot of the main effects considered in fitted full  $2^4$  FD model. It is clear how factor  $B$  and  $D$  affect mainly the dispersion of the



response, while factor C affects both location and dispersion. On the other hand, main effect of factor A is relatively much smaller than the other main effects and hence of less importance.

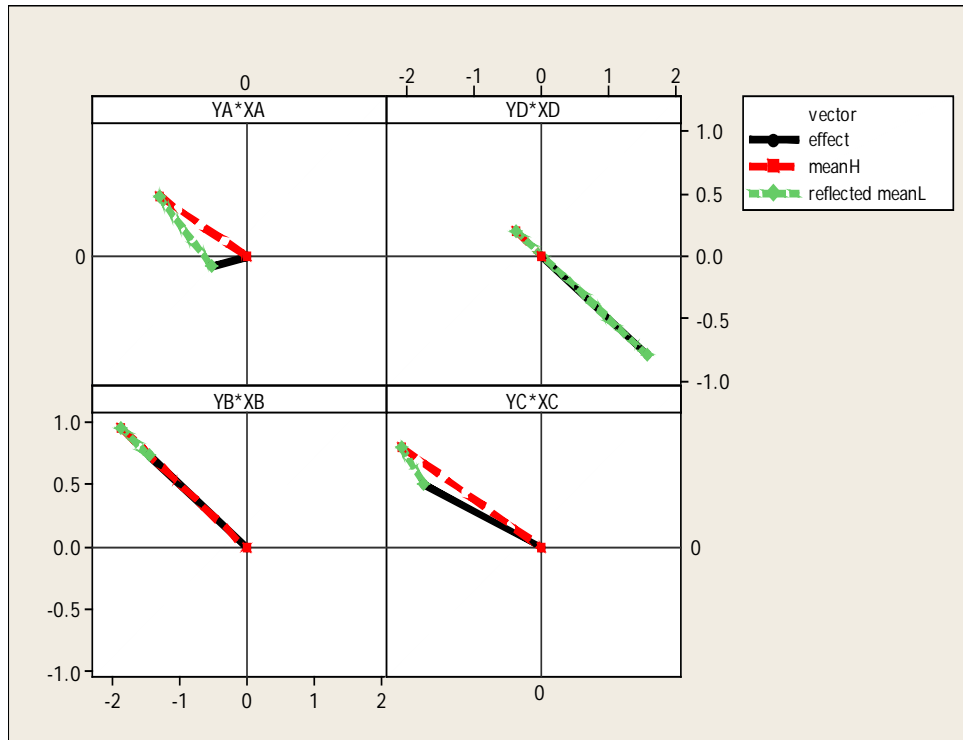


Figure 3: C-ME of A, B, C, and D in the fitted full model

C-IE plots of all the two way interactions in the  $2^4$  FD full factorial design are shown in Figure 4 for the fitted full model. The relatively most important interaction is the  $BC$  while the least important one is  $BD$ . The remaining four two-way interactions seem to be of same importance.

Table 1 contains different statistics of different models that are of interest in any initial step of a factorial analysis. The best model in terms of AIC is the model that contains up to all three factor interactions, while in terms of BIC, the best model is the model that contains only up to all two factor interactions. To compare different models in terms of the likelihood ratio principle, one can use the concept of deviance. Deviance allows one to test whether the fitted reduced model is significantly worse than the saturated model (lack of fit test). By definition the deviance is twice the negative difference between the log likelihood functions of the reduced and the saturated model, which is asymptotically chi-squared distributed with degrees of freedom equal to the difference of the number of parameters in the two models. Differences in deviances would be used to compare two nested

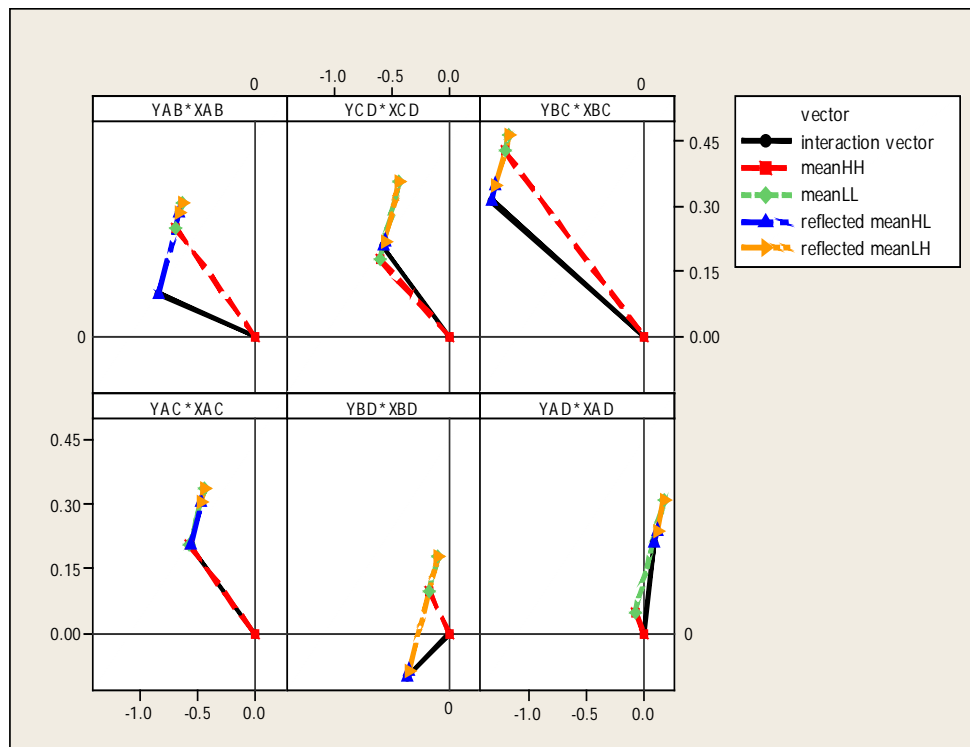


Figure 4: C-IE of all 2-way interactions in  $2^4$  FD using the fitted full model

models [Myers *et al.*, 2002, p. 114]. In terms of the deviance analysis, the model without any higher than two-factor interactions is reasonable and does not suffer from lack of fit ( $p$ -value = 0.66113). This model is the best model in terms of BIC, too. Although the model without all the two-factor interaction does not suffer from lack of fit ( $p$ -value = 0.06480), models without  $CD$  or  $BC$  do suffer from lack of fit ( $p$ -value = 0.005424 and 0.031635, respectively). Hence, keeping  $CD$  and  $BC$  is recommended, which leads to keep also main effect  $B$ ,  $C$ , and  $D$ . The deviance analysis shows also that the model without main effect  $A$  fits reasonably ( $p$ -value = 0.436049). Models that do suffer from lack of fit (given an asterisk in Table 1) indicate that dropping the chosen effect is not recommended. Accordingly, the best SPML model for this data set should contain  $B$ ,  $C$ ,  $D$ ,  $CD$  and  $BC$ . This model has  $AIC = 460.337$ ,  $BIC = 480.012$  and  $2 \log(L) = -218.168$  ( $p$ -value = 0.733592). Note that this model has the least BIC statistics among all the models considered in Table 1.

Anderson and Wu (1995, 1996) considered the same data set to study a location-only and a dispersion-only effect, respectively. Their final location-influential factors are  $A$ ,  $C$ ,  $D$ ,  $CD$  and  $AD$ , while their dispersion model contained  $B$ ,  $C$ ,  $D$ , and  $CD$ . Our SPML model contains their dispersion-factors,

Table 1: Statistics of different models of interest

Model	$-2 \log(L)$	Deviance from full	df	$p$ -value	AIC	BIC
Full	202.444				468.888	567.294
Without $ABCD$	202.553	0.109	2	0.946959	463.107	514.2958
Without $ABC$	205.029	2.585	2	0.274583	470.058	519.2478
Without $ACD$	205.669	3.225	2	0.199389	471.338	520.5278
Without $ABD$	203.260	0.816	2	0.664979	466.521	515.7098
Without $BCD$	203.418	0.974	2	0.614467	466.836	516.0258
Drop all $> 2$ interactions	210.113	7.669	10	0.66113	464.227	500.2993
Without $AB$	203.555	1.111	2	0.573785	467.11	516.2998
Without $AC$	203.204	0.76	2	0.683861	466.407	515.5968
Without $AD$	206.619	4.175	2	0.123997	473.238	522.4268
Without $BC$	209.351	6.907	2	0.031635*	478.703	527.8918
Without $BD$	203.320	0.876	2	0.645326	466.639	515.8288
Without $CD$	212.878	10.434	2	0.005424*	485.757	534.9458
Dropping all interactions	235.240	32.796	22	0.06480	498.481	527.1776
Without $A$	204.104	1.66	2	0.436049	468.207	517.3968
Without $B$	208.441	5.997	2	0.049862*	476.882	526.0718
Without $C$	210.926	8.482	2	0.014393*	481.853	531.0418
Without $D$	209.327	6.883	2	0.032017*	478.654	527.8438

but their set of location-factors is not matching our combined location-dispersion factors. While they identified  $A$  and  $AD$  as important factors in their location-effect analysis, both effects are not important in our SPML model, with  $p$ -values of 0.436049 and 0.123997, respectively. Actually  $AD$  was identified as the second most important effect in their analysis and  $CD$  is the first most important one, while  $A$  came at the 13th position among the 15ths effects. Fitting SPML model with their location-factor set gives the following statistics: AIC = 473.844, BIC = 493.5197 and  $2 \log(L) = -224.922$  ( $p$ -value = 0.315146). The model is still adequate but our best model beats it in terms of AIC and BIC.

Our fitted SPML model is

$$\hat{u}'_{ijk} = [-0.874332^{\cos} - 0.324906^{\cos} x_{ik}^B - 0.440622^{\cos} x_{jk}^C + 0.316346^{\cos} x_{lk}^D - 0.391979^{\cos} x_{ik}^B x_{jk}^C - 0.464214^{\cos} x_{lk}^D x_{jk}^C, 0.464643^{\sin} + 0.154155^{\sin} x_{ik}^B + 0.054378^{\sin} x_{jk}^C - 0.153721^{\sin} x_{lk}^D + 0.050118^{\sin} x_{ik}^B x_{jk}^C + 0.078266^{\sin} x_{lk}^D x_{jk}^C].$$

### 6. Conclusion

The SPML model is used in the factorial design setup to help in modeling circular data with predictors measured on the real line in a relatively easy way than the models proposed in the circular regression setup. SPML solved the

dilemma of not having a suitable model to complete a factorial analysis for factorial data with a response measured on the circle. It enables us to determine and visualize the factor effects on the response. The proposed C-ME and C-IE plots with the circular response plays the same role the effect plots with a linear response is playing in the factorial analysis. The deviance analysis is used to identify adequate model with significant effects.

The use of SPML model in the factorial design setup along with the two effect-plots and the effect interpretation, proposed in this paper, enables the researcher to perform a complete analysis of a factorial experiment for a circular response.

## Appendix 1

For the  $2^2$  factorial design with interaction defined in (3.1), the main effect vector of  $A$  is defined as  $[A] = \bar{u}_{2..} - \bar{u}_{1..}$ .

If  $[A] = \underline{0}$  (no main effect), then

$$\rightarrow s_2 \|\bar{u}_2\| - s_1 \|\bar{u}_1\| = 0 \quad \text{and} \quad c_2 \|\bar{u}_2\| - c_1 \|\bar{u}_1\| = 0.$$

For the cosine equation, we have

$$\begin{aligned} (2r\beta_0^{\cos} + 2r\beta_A^{\cos})\|\bar{u}_2\| - (2r\beta_0^{\cos} - 2r\beta_A^{\cos})\|\bar{u}_1\| &= 0, \\ \rightarrow 2r\beta_0^{\cos}(\|\bar{u}_2\| - \|\bar{u}_1\|) + 2r\beta_A^{\cos}(\|\bar{u}_2\| + \|\bar{u}_1\|) &= 0. \end{aligned}$$

This holds if one of the following three solutions satisfied:

$$(1) \beta_0^{\cos} = 0 \quad \text{and} \quad \beta_A^{\cos} = 0 \quad (2) \|\bar{u}_2\| = \|\bar{u}_1\| \quad \text{and} \quad \beta_A^{\cos} = 0 \quad (3) \|\bar{u}_2\| = \|\bar{u}_1\| = 0.$$

Similarly, for the sine equation we have

$$(1) \beta_0^{\sin} = 0 \quad \text{and} \quad \beta_A^{\sin} = 0 \quad (2) \|\bar{u}_2\| = \|\bar{u}_1\| \quad \text{and} \quad \beta_A^{\sin} = 0 \quad (3) \|\bar{u}_2\| = \|\bar{u}_1\| = 0.$$

The first and the third solutions are not acceptable (trivial solution), since it means SPML model is not significant, while the second solution means that the dispersion is fixed and sub-effects are zeros.

Note that fixed dispersion is satisfied if

$$\begin{aligned} \bar{u}_2 &= \bar{u}_1, \\ \rightarrow (\beta_0^{\cos} + \beta_A^{\cos})^2 + (\beta_0^{\sin} + \beta_A^{\sin})^2 &= (\beta_0^{\cos} - \beta_A^{\cos})^2 + (\beta_0^{\sin} - \beta_A^{\sin})^2 \\ \rightarrow -\beta_0^{\sin} \beta_A^{\sin} &= \beta_0^{\cos} \beta_A^{\cos} \\ \rightarrow \left( \frac{\beta_0^{\sin}}{\beta_0^{\cos}} \right) \left( \frac{\beta_A^{\sin}}{\beta_A^{\cos}} \right) &= -1 \\ \xrightarrow{SPML} \tan(\theta_{2..}) \tan(\theta_{2..} - \theta_{1..}) &= -1. \end{aligned}$$

## Appendix 2

Figure A2.1 represents two scenarios where main effect affects both dispersion and location while only one of  $[A]$  sine or cosine coordinates equals zero. Figure A2.2 shows possible scenarios in which the main effect increases the dispersion without any effect on the location while both slopes are nonzero. Possible Scenarios in which the main effect increases the location without any effect on dispersion are depicted in Figure A2.3. Note that in all of these three figures if we switched the high mean response vector (blue vector) with the low response vector (red vector), the effect on location or/and dispersion is still present but with the opposite sign. Scenario d in Figure A2.3 shows that when the high and low vectors of the factor are  $2\pi$  apart from each other,  $[A]$  will exactly lie on the high vector and with same direction and dispersion.

## Appendix 3

The SPML model of the full 2<sup>4</sup> FD for the example data

$$\begin{aligned} \mathbf{u}'_{ijltk} = & [\beta_0 \cos + x_{ik}\beta_A \cos + x_{jk}\beta_B \cos + x_{lk}\beta_C \cos + x_{tk}\beta_D \cos + x_{ik}x_{jk}\beta_{AB} \cos \\ & + x_{ik}x_{lk}\beta_{AC} \cos + x_{ik}x_{tk}\beta_{AD} \cos + x_{jk}x_{lk}\beta_{BC} \cos + x_{jk}x_{tk}\beta_{BD} \cos \\ & + x_{lk}x_{tk}\beta_{CD} \cos + x_{ik}x_{jk}x_{lk}\beta_{ABC} \cos + x_{ik}x_{jk}x_{tk}\beta_{ABD} \cos \\ & + x_{ik}x_{lk}x_{tk}\beta_{ACD} \cos + x_{jk}x_{lk}x_{tk}\beta_{BCD} \cos + x_{ik}x_{jk}x_{lk}x_{tk}\beta_{ABCD} \cos, \\ & \beta_0 \sin + x_{ik}\beta_A \sin + x_{jk}\beta_B \sin + x_{lk}\beta_C \sin + x_{tk}\beta_D \sin + x_{ik}x_{jk}\beta_{AB} \sin \\ & + x_{ik}x_{lk}\beta_{AC} \sin + x_{ik}x_{tk}\beta_{AD} \sin + x_{jk}x_{lk}\beta_{BC} \sin + x_{jk}x_{tk}\beta_{BD} \sin \\ & + x_{lk}x_{tk}\beta_{CD} \sin + x_{ik}x_{jk}x_{lk}\beta_{ABC} \sin + x_{ik}x_{jk}x_{tk}\beta_{ABD} \sin \\ & + x_{ik}x_{lk}x_{tk}\beta_{ACD} \sin + x_{jk}x_{lk}x_{tk}\beta_{BCD} \sin + x_{ik}x_{jk}x_{lk}x_{tk}\beta_{ABCD} \sin]' \\ & + \varepsilon_{ijltk}. \end{aligned}$$

Sub-effects for cosine equation:

$$\begin{aligned} [A]_{\cos} &= \{\bar{c}_{2\dots} - \bar{c}_{1\dots}\}, \\ [AB]_{\cos} &= \frac{1}{2} \{[A(b_2)]_{\cos} - [A(b_1)]_{\cos}\} = \frac{1}{2} \{(\bar{c}_{22\dots} - \bar{c}_{12\dots}) - (\bar{c}_{21\dots} - \bar{c}_{11\dots})\}, \\ [ABE]_{\cos} &= \frac{1}{2} \{[AB(e_2)]_{\cos} - [AB(e_1)]_{\cos}\} \\ &= \frac{1}{2^2} \{[A(b_2e_2)]_{\cos} - [A(b_1e_2)]_{\cos} - [A(b_2e_1)]_{\cos} + [A(b_1e_1)]_{\cos}\}, \end{aligned}$$

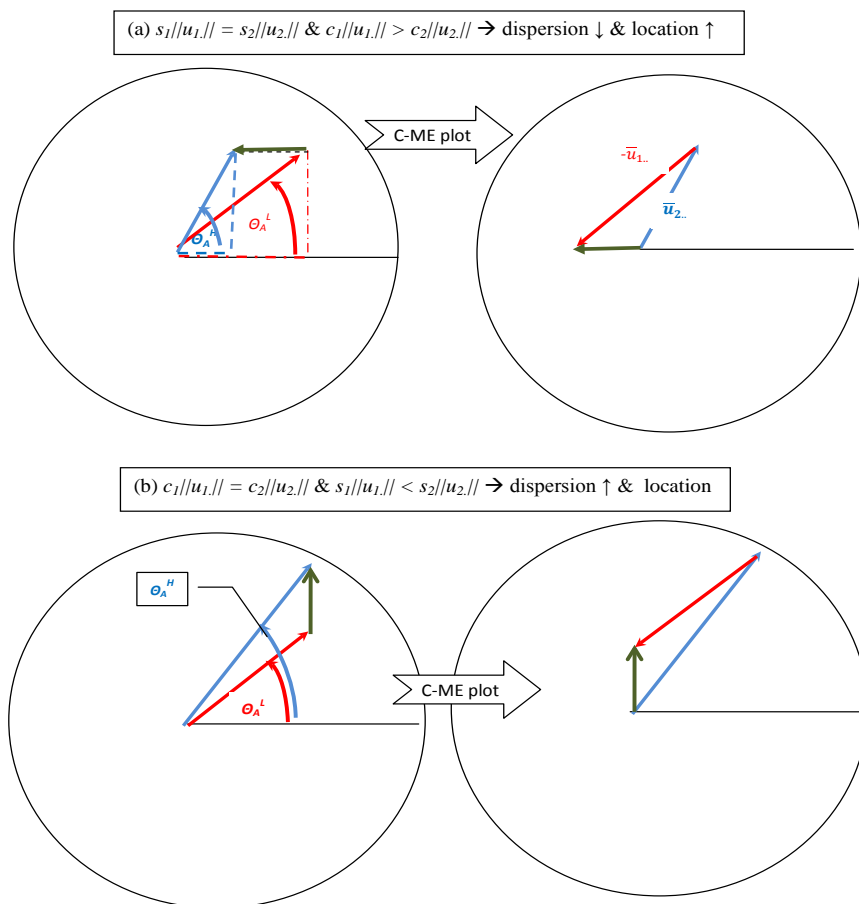


Figure A2.1: Possible Scenarios where main effect affects both dispersion and location while only one of the resultant main effect coordinates equals zero

$$\begin{aligned}
 [ABED]_{\cos} &= \frac{1}{2} \{ [ABE(d_2)]_{\cos} - [ABE(d_2)]_{\cos} \} \\
 &= \frac{1}{2^3} \{ [A(b_2e_2d_2)]_{\cos} - [A(b_1e_2d_2)]_{\cos} - [A(b_2e_1d_2)]_{\cos} + [A(b_1e_1d_2)]_{\cos} \\
 &\quad - [A(b_2e_2d_1)]_{\cos} + [A(b_1e_2d_1)]_{\cos} + [A(b_2e_1d_1)]_{\cos} - [A(b_1e_1d_1)]_{\cos} \},
 \end{aligned}$$

where

$$\begin{aligned}
 [A(b_j)]_{\cos} &= (\bar{c}_{2j\dots} - \bar{c}_{1j\dots}), \\
 [AB(e_i)]_{\cos} &= \frac{1}{2} \{ [\bar{c}_{22l\dots} - \bar{c}_{12l\dots}] - [\bar{c}_{21l\dots} - \bar{c}_{11l\dots}] \}, \\
 [ABE(d_t)]_{\cos} &= \frac{1}{4} \{ [\bar{c}_{222t\dots} - \bar{c}_{122t\dots}] - [\bar{c}_{212t\dots} - \bar{c}_{112t\dots}] - [\bar{c}_{221t\dots} - \bar{c}_{121t\dots}] \\
 &\quad + [\bar{c}_{211t\dots} - \bar{c}_{111t\dots}] \}.
 \end{aligned}$$

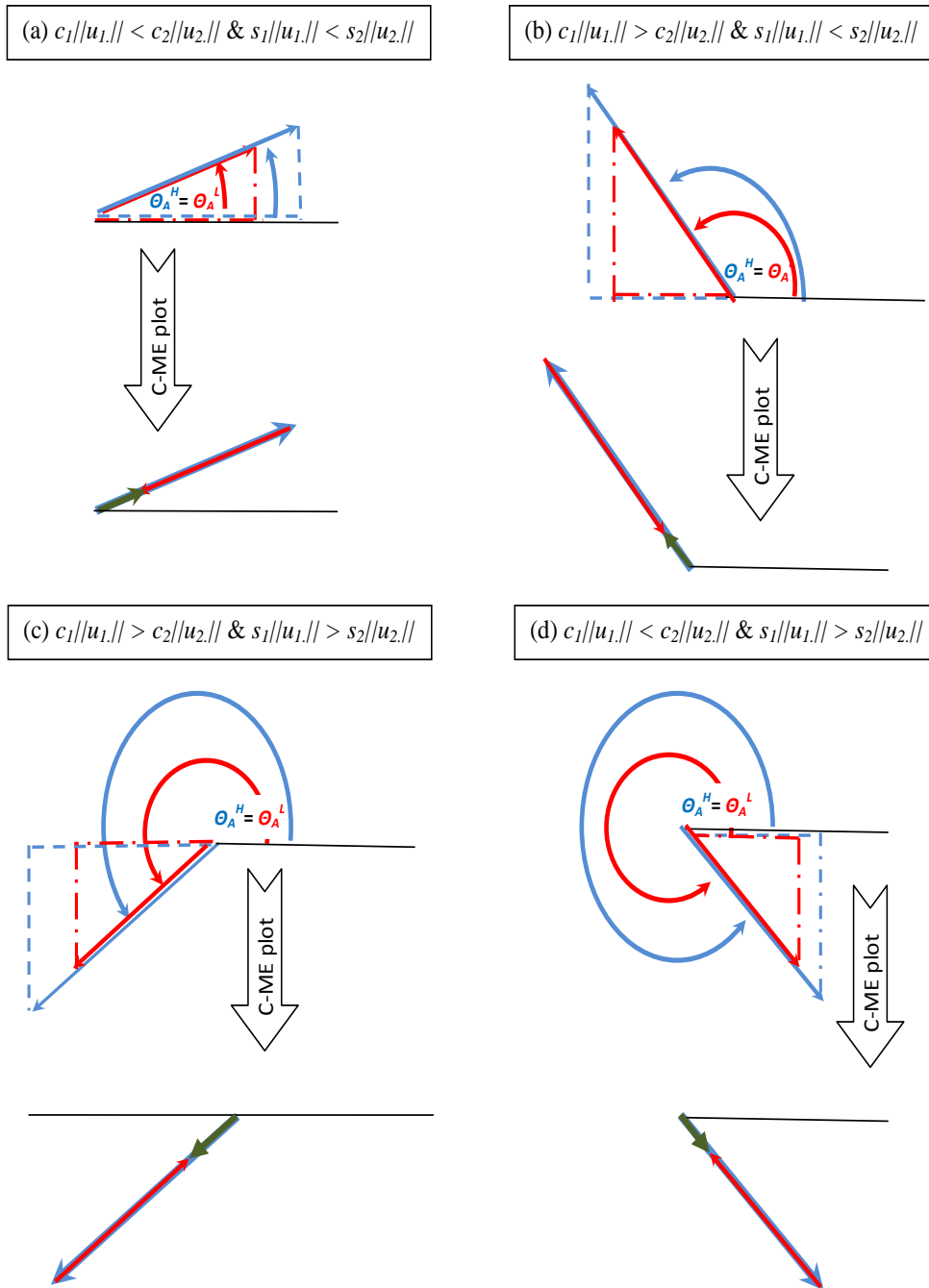


Figure A2.2: Possible Scenarios in which the main effect increases the dispersion without any effect on the location with both slopes are nonzero

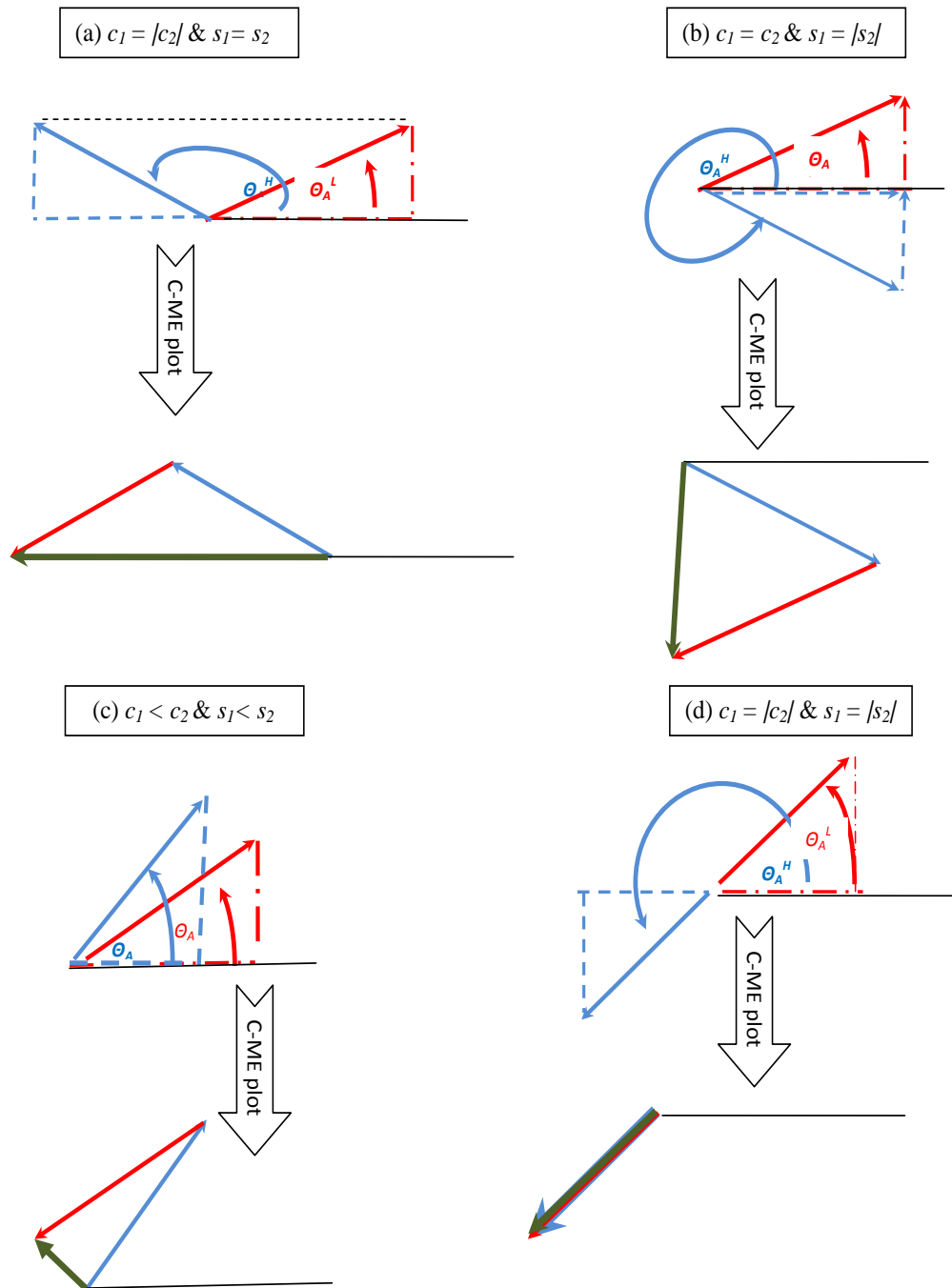


Figure A2.3: Possible Scenarios in which the main effect increases the location without any effect on dispersion  $\{||u_2.|| = ||u_1.||\}$



Other main effects, 2-way interactions and 3-way interactions are defined like the above ones with its suitable subscript. In addition, the definitions of the sine sub-effects are similar to those of the cosine counterparts where the sine function is used instead of the cosine one.

## References

- Anderson, C. M. and Wu, C. F. J. (1995). Measuring location effects from factorial experiments with a directional response. *International Statistical Review* **63**, 345-363.
- Anderson, C. M. and Wu, C. F. J. (1996). Dispersion measures and analysis for factorial directional data with replicates. *Applied Statistics* **45**, 47-61.
- Fisher, N. I. and Lee, A. J. (1992). Regression models for an angular response. *Biometrics* **48**, 665-677.
- Fisher, R. A. (1960). *The Design of Experiments*, 7th edition. Oliver and Boyd, Edinburgh.
- Gould, A. L. (1969). A regression technique for angular variates. *Biometrics* **25**, 683-700.
- Hinkelmann, K. (2004). *Evaluating and Interpreting Interactions*. Technical Report Number 04-5, Department of Statistics, Virginia Tech, Blacksburg, Virginia.
- Hinkelmann, K. and Kempthorne, O. (1994). *Design and Analysis of Experiments, Volume 1: Introduction to Experimental Design*, 2nd edition. Wiley, New York.
- Johnson, R. A. and Wehrly, T. E. (1978). Some angular linear distributions and related regression models. *Journal of the American Statistical Association* **73**, 602-606.
- Kuehl, R. O. (1994). *Statistical Principles of Research Design and Analysis*, 1st edition. Duxbury Press, California.
- Mardia, K. V. and Jupp, P. E. (1999). *Directional Statistics*, 2nd edition. Wiley, Chichester.
- Montgomery, D. C. (2009). *Design and Analysis of Experiments*, 7th edition. Wiley, New York.

- Myers, R. H. and Montgomery, D. C. (2002). *Response Surface Methodology: Process and Product Optimization Using Designed Experiments*, 2nd edition. Wiley, New York.
- Myers, R. H., Montgomery, D. C. and Vining, G. G. (2002). *Generalized Linear Models: With Applications in Engineering and the Sciences*, 1st edition. Wiley, New York.
- Presnell, B., Morrison, S. P. and Littell, R. C. (1998). Projected multivariate linear models for directional data. *Journal of the American Statistical Association* **93**, 1068-1077.
- Underwood, A. J. and Chapman, M. G. (1985). Multifactorial analyses of directions of movement of animals. *Journal of Experimental Marine Biology and Ecology* **91**, 17-43.

Received October 21, 2012; accepted January 17, 2013.

Alyaa Roshdy Zahran  
Statistics Department  
Faculty of Economics and Political Sciences  
Cairo University  
P.O. Box 12613, Nahdet Misr Street, Giza, Cairo, Egypt  
azahran@vt.edu, azahran@qu.edu.qa