DISPERSION MEASURES AND ANALYSIS FOR FACTORIAL DIRECTIONAL DATA WITH REPLICATES

Christine M. Anderson & C.F. Jeff Wu

IIQP Research Report
RR-94-08

June 1994

DISPERSION MEASURES AND ANALYSIS FOR FACTORIAL DIRECTIONAL DATA WITH REPLICATES

Christine M. Anderson
University of Western Ontario

C.F. Jeff Wu[1]
University of Waterloo and
University of Michigan

ABSTRACT

Studying the influence of linear factors on the spread of a directional response in industrial experimentation has not been considered much in the literature. Several dispersion measures are explored and their relationships described. The circular variance is a good dispersion measure that transforms the angular dispersion into a statistic measured on a linear scale. Once this transformation has been performed, established techniques for analysis can be employed for analyzing factor influences on the directional dispersion. The proposed method is used to analyze data from an actual experiment involving the balancing of automotive flywheels.

Key words: Directional data, circular data, variation reduction, dispersion analysis.

---

[1] Research supported by the Natural Sciences and Engineering Research Council of Canada, General Motors of Canada, and the Manufacturing Research Corporation of Ontario.

1 Introduction

Studying factor effects on dispersion and improving quality through variation reduction are the main ideas in robust parameter design that were popularized by Taguchi (1986). In some applications, reducing the spread of the data by selecting an optimal combination of factors is the primary goal of experimentation. In this paper we examine techniques for analysing the influence of experimental factors on the dispersion of a *directional* response located on a unit circle. The issue of control versus noise factors is also explored.

For example in the automotive industry, a number of rotating parts (such as brake rotors, flywheels, crank shafts, and tires) need to be precisely balanced to prevent excessive vibration. We can measure imperfection in the part by identifying the direction, which is disproportionately heavy (or light), and the magnitude of the imbalance. This papers considers analysing the spread of the directional component only. If a combination of factors could be found that locates all of the imbalances close together, then one of several strategies can reduce production costs. In some cases, some global corrective action could be taken to adjust the process to reduce the number of unbalanced parts. In other cases, the parts still may need to be individually corrected, but the cost of corrections can be lowered simply by having the imbalances all located in close proximity to one another on the part.

We now present some of the issues that arose from a real industrial experiment at an automotive production plant involving the balancing of engine flywheels. The response obtained from each flywheel was a location on the circumference of the part where a corrective adjustment would be required to balance the part. The process of determining the location of the imbalance is quite precise and uniquely determines a single point where the corrective action should be taken. A 2^4 full-factorial experiment was run with ten observations at each set of factor combinations. The four factors thought to influence the dispersion of the imbalance are as follows:

A The location of a buttweld to the flywheel, either Fixed (F) or Random (R). In current production, the selection of the location for joining these two critical pieces together was determined randomly.

B Flywheel radius grade, either Low (L) or High (H). Current levels used the lower grade flywheel where a larger difference in radius was tolerated.

C Flywheel thickness grade, either (L) or (H). Currently, the low grade thickness was used with a larger difference in flywheel thickness deemed acceptable.

D Size of Counter-weight attached at $0°$, either (L) or (H). The size of the counter-weight in production was presently at the low level.

In addition, a fifth factor E (ring gear imbalance) which is very difficult and expensive to control was observed and a number of each of the three levels (L, M, H) were used in each of the 16 runs. Typically, each of the runs consisted of between 1-3 Low, 5-7 Medium and 1-3 High observations. The data are provided in angular form (measured in degrees) in Table 1.

As a preliminary test of group dispersion differences, Bartlett's test for homogeneity of von Mises concentration parameters as described in Stephens (1982) can be used. Further details of the test are provided in Section 3. The test disregards the structure of the factorial experiment and considers each combination of factors as a different group. For the flywheel data, the significance level of this test is approximately 0.0001. Hence, we conclude that there are real differences between the dispersions of 16 factor combinations. A number of questions arise from this conclusion, which will be studied in the remainder of the paper.

1. Can the relative importance of the four controllable factors be assessed to determine how improvements to the process should be approached?
2. Since Factor C is actually expensive to control, can a combination of the other factors be found that is robust to the different levels of this factor?
3. How can the information about the noise factor, E, be incorporated to give greater insight into the working of the process?

Table 1: Automotive Flywheel Data

Run	A	B	C	D	Data									
1	R	L	L	L	133	175	178	178	153	190	221	177	281	190
2	R	L	L	H	139	61	109	187	74	351	309	236	69	320
3	R	L	H	L	111	122	105	49	189	188	177	151	62	329
4	R	L	H	H	170	162	19	337	171	114	341	10	266	201
5	R	H	L	L	127	215	125	188	187	175	162	172	169	82
6	R	H	L	H	150	84	113	318	84	353	301	12	82	351
7	R	H	H	L	152	164	180	187	159	149	127	148	175	201
8	R	H	H	H	184	128	177	186	163	178	196	155	150	120
9	F	L	L	L	154	200	147	133	171	318	100	108	86	73
10	F	L	L	H	198	165	31	51	314	84	267	135	318	14
11	F	L	H	L	345	43	4	295	75	138	149	141	198	175
12	F	L	H	H	153	194	207	136	144	206	151	202	104	188
13	F	H	L	L	140	134	170	62	109	127	132	116	94	183
14	F	H	L	H	340	111	128	327	81	301	3	335	215	334
15	F	H	H	L	160	152	187	158	143	91	200	143	84	191
16	F	H	H	H	171	156	171	195	159	153	188	125	107	98

Before addressing these experiment-specific questions, a number of more fundamental issues need to be addressed. Section 2 discusses three dispersion statistics: the circular variance, the projection of the data onto the maximum eigenvector diameter, and the circular standard deviation. Their relative strengths are outlined and some distributional results provided. Section 3 outlines a strategy for modeling the dispersion for a factorial experiment with replication, and discusses two models with intuitive interpretations often suggested by the strategy. Finally, Section 4 illustrates the technique by giving a complete analysis of the flywheel experiment data.

2 Measures of Dispersion

In this section we consider a number of possible dispersion statistics which might be suitable for studying the spread of the observed directional data. It is desirable that the statistic be a simple, intuitively pleasing and computationally convenient measure of the spread of the data, regardless of the shape or distribution of the original data. In addition, it would be advantageous if the measure has known and manageable distributional properties under more restrictive assumptions about the original data.

Before considering specific candidates, we begin with a brief review of notation and fundamental quantities for directional data. Consider the simplest situation of a single population of directional data responses located on the circumference of the circle. From this population we obtain a sample, $(\theta_1, \cdots, \theta_n)$. The θ_i's can also be identified as vectors of unit length starting at the origin and pointing in the direction of their angle, with the usual convention that $0°$ points horizontally to the right, with positive angles rotating counter-clockwise. The vector that corresponds to angle θ_i is called u_i, and from this vector representation, we can calculate the resultant vector by summing the vectors. The overall resultant vector, $u_.$, of the sample has length $R = (c^2 + s^2)^{\frac{1}{2}}$, where $c = \sum_i \cos \theta_i$ and $s = \sum_i \sin \theta_i$. Another quantity of interest is $\bar{R} = R/n$, the standardized length of the resultant vector. The average direction for the sample, frequently called the mean direction in the literature, is defined to be the angle of the resultant vector and can be obtained as follows

$$\theta_{.} = \begin{cases} \arctan(s/c), & \text{if c} > 0 \ , \\ \frac{\pi}{2}, & \text{if } c = 0 \text{ and } s > 0 \\ \frac{-\pi}{2}, & \text{if } c = 0 \text{ and } s < 0 \\ \pi + \arctan(s/c), & \text{otherwise.} \end{cases} \tag{1}$$

A common choice for distribution on the circle is the von Mises distribution with mean direction μ and concentration parameter k, denoted $VM(\mu, k)$. It possesses some of the desirable properties associated with the normal distribution for traditional data measured on a linear scale (Mardia, 1972, pp.55-58) and has the probability density:

$$f(\theta) = \frac{1}{2\pi I_0(k)} \exp\{k\cos(\theta - \mu)\},$$

where $I_0(k)$ is the modified Bessel function with $k \geq 0$ and $\theta \in (-\pi, \pi)$.

One possible candidate for quantifying the dispersion presented by Fisher and Lee (1992) models the von Mises concentration parameter, k, as a function of a linear combination of explanatory variables. Frequently, data sets can be reasonably assumed to come from this distribution. The maximum likelihood estimate of k can be found by solving the equation, $A(\hat{k}) = \bar{R}$, where $A(\hat{k}) = I_1(\hat{k})/I_0(\hat{k})$, the quotient of modified Bessel functions. Dobson (1978) and Best and Fisher (1981) suggest a number of improvements to this estimate to reduce, but not eliminate, the bias and instability of this estimate. However, the estimate of the concentration parameter, $\hat{k}$, is too dependent on distributional assumptions. For data that do not follow the von Mises distribution (i.e. heavy-tailed or asymmetric), this quantity does not have a sensible interpretation. In contrast to the variance for traditional linear data, $\hat{k}$ cannot be used to describe a general characteristic of the data without making the von Mises assumption. Therefore, a more robust general summary is sought. Alternate statistics are considered in the following subsections.

2.1 Circular Variance

The circular variance, $S_0 = 1 - \bar{R}$, is a common dispersion statistic used to quantify the variability of a sample of directional data. It ranges in value from zero to one. A value of zero corresponds to no variation in the data, while $S_0 = 1$ means that the data is uniformly distributed on the circumference of the circle. It follows the usual convention that a small value for the variance means that the data is concentrated near the average, but unlike the variance obtained for linear data, it has a finite maximum.

Calculation of $\bar{R}$, and hence the variance, is straightforward and well-defined for all data sets. Rivest (1982) noted that for highly concentrated samples the circular variance properly normalized has approximately the same distribution as the linear variance of the angles measured on a $(0, 2\pi)$ scale. Various authors, including Watson and Williams (1956), have studied the distributional properties of $nS_0 = n - R$ for a sample of n observations from a von Mises distribution. Mardia (1972, p. 113) summarizes results about the expectation and variance of S_0 under the von Mises assumption, while Watson and Williams (1956) show that $2k(n - R) = 2nk(1 - \bar{R}) \sim \chi^2_{n-1}$ for data from a concentrated von Mises distribution.

To improve these approximations for small concentrations and small sample sizes which are often typical of industrial data sets, the method of matching the first two moments of the circular variance is used. Using this method, we assume that $\gamma(1 - \bar{R}) \sim \chi^2_f$, for some γ, f to be determined by the moments. Solving the two equations obtained by equating means and variances,

$$\gamma \left[1 - A - \frac{1}{2nk}\right] = f \tag{2}$$

and

$$\gamma^2 \left[\frac{1}{n}(1 - A^2) - \frac{1}{nk}A - \frac{1}{4n^2k^2}\right] = 2f \tag{3}$$

gives the following solutions for the multiplicative coefficient, γ, and the estimated degrees of freedom, f,

$$\gamma = \frac{2nk(1 - A) - 1}{k(1 - A^2) - A - \frac{1}{4nk}} \tag{4}$$

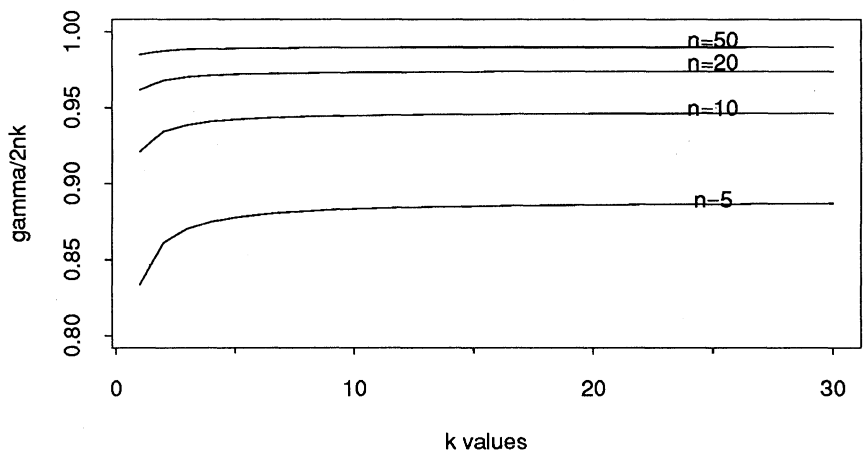

Figure 1: Ratio of Estimated Coefficient and $2nk$

and

$$f = \gamma \left[1 - A - \frac{1}{2nk}\right] \quad , \tag{5}$$

where $A = A(\hat{k})$.

To verify that these expressions approach the values predicted by Watson and Williams (1956), we substitute in the Taylor expansion of $A(k) \approx 1 - 1/2k - 1/8k^2$ for k large (Mardia 1972, p.63) to obtain

$$\gamma \approx \frac{2k\left\{(n-1) + \frac{1}{4k}\right\}}{1 - \frac{1}{2n}} \rightarrow 2nk \tag{6}$$

and

$$f \approx \left\{(n-1) + \frac{1}{4k}\right\} \frac{1 + \frac{1}{4k} + \frac{1}{n}}{1 - \frac{1}{2n}} \rightarrow (n-1) \tag{7}$$

as anticipated when $n \rightarrow \infty$ and $k \rightarrow \infty$. Figures 1 and 2 illustrate the difference in values of the coefficient and degree of freedom using the asymptotic and (γ, f) estimates. The plots show the ratio of the two estimates and highlight the differences for small concentrations and small samples. However, if k is moderate (i.e. $k \geq 2$) and the sample size, n, is at least 10, then the difference between the estimated and asymptotic values will be small. Therefore,

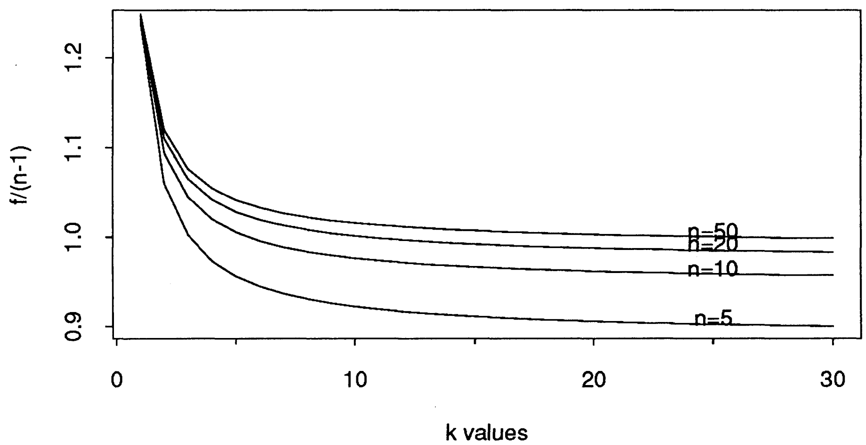

Figure 2: Ratio of Estimated Degrees of Freedom and $n - 1$

for von Mises data, the distribution of the circular variance can be well approximated by a chi-squared random variable with some adjustment to the degrees of freedom and the multiplicative constant. Empirically, quantile-quantile plots of 1000 simulated observations each from a von Mises distribution with a variety of sample sizes (n=5 to 100) and concentration parameters ($k \in (2, 16)$) show that the approximation to a chi-squared distribution is good for a wide range of these values (Anderson, 1993).

The circular variance will be a good dispersion statistic for a large number of distributions. It is a simple and intuitively pleasing measure of dispersion, and is robust to outliers and distributional shape (see Anderson, 1993). In addition it has manageable distributional properties for data originating from a von Mises distribution. Hence, the circular variance or its monotonic transformation satisfies the criteria established earlier.

2.2 Projection onto the Maximum Eigenvector Diameter

Consider data located on the circle as a bivariate data set, $(x_i, y_i) = (\cos\theta_i, \sin\theta_i)$ with the obvious restriction that $x_i^2 + y_i^2 = 1$. From this representation the 2×2 matrix $(x, y)^T(x, y)$

can be determined, along with its two eigenvalues and corresponding eigenvectors. The eigenvector which corresponds to the maximum eigenvalue defines a diameter through the circle. For example, if the eigenvector is (x^*, y^*), then the diameter is defined to run from (x^*, y^*) to $(-x^*, -y^*)$. The rationale for examining the eigenvector defined by the maximum eigenvalue is that this direction is frequently used as a measure of direction in multivariate analyses, and hence by projecting data onto it, we obtain a linear measure of spread from the centre of the data. For a unimodal symmetric directional distribution, this corresponds to the skewed cross-section along the axis of symmetry.

The projection of the original data onto this diameter can be obtained by multiplying $c_i = (x_i, y_i)^T(x^*, y^*)$, to obtain a univariate value on the interval $(-1, 1)$. A useful convention is to adjust the sign of the eigenvector to give a projection close to 1, not -1. From this projection, we define the contribution to the statistic of each individual observation to be $\delta_i = (1 - c_i)/2$, where $\delta_i \approx 0$ corresponds to an observation near the centre of the data, and $\delta_i \approx 1$ denotes an observation opposite to the centre of the data. Hence, the dispersion statistic is $\sum_i \delta_i$, the sum of the individual contributions.

An interesting connection between this new measure and the circular variance exists. If the angular average of the data and the direction defined by the maximum eigenvalue are the same, then $2\delta_i$ is equivalent to $1 - \cos(\theta_i - \theta_.)$, which is the contribution of an observation to the circular variance sum of squares. Therefore these two entirely different approaches to dispersion analysis yield the same statistic, if the same choice of the location is used. Hence an alternate interpretation of the circular variance is to think of each observation's contribution to the overall circular variance as a function of its projection onto the diameter defined by the circular average.

Simulation results for a comparison of the direction of the eigenvector for the maximum eigenvalue with the circular average direction show that there is more variation in the new estimate of the centre of the data. If the concentration of the data is small, then the circular average has much less variability. Table 2 summarizes the results from 1000 data

Table 2: Comparison of Eigenvector and Circular Average Approaches

		Variance		Variance Ratio	Correlation
n	k	Eigenvector Centre	Circular Average	Eigen to Average	
10	2	.348	.0806	4.29	.311
	4	.0369	.0282	1.31	.877
	8	.0145	.0142	1.02	.987
	16	.00541	.00539	1.00	.997
20	2	.101	.0390	2.58	.622
	4	.0167	.0140	1.19	.909
	8	.00689	.00650	1.06	.985
	16	.00330	.00329	1.00	.997

sets each with 20 observations from a von Mises distribution (generated using the algorithm suggested by Best and Fisher, 1979) with population mean zero and a variety of concentration parameters. The fifth column, considers the ratio of the variances of the eigenvector mean and the circular average. This column highlights the increased variability in the eigenvector method.

The circular variance is a preferred measure because it is simple, has less variability then the eigenvector method and has a rich literature on its distributional properties.

2.3 Circular Standard Deviation

The circular standard deviation is another alternate statistic to the circular variance. Unlike the linear case where the standard deviation is simply the square root of the variance, for directional data, the form of this new statistic is $s_0 = \{-2\log(1 - S_0)\}^{\frac{1}{2}}$, where S_0 is the circular variance. Hence it can be simplified to $\{-2\log(\bar{R})\}^{\frac{1}{2}}$.

The circular standard deviation is a non-negative statistic ranging from zero for no dispersion in the data, to infinity for the data uniformly distributed around the circle. Less is known about the distributional properties of this statistic than the circular variance when the data comes from a von Mises distribution. Mardia (1972, p. 24) commented that the circular variance is "more useful than s_0 for theoretical investigations."

To estimate the distribution of s_0 we exploit its relationship to the circular variance. Using Taylor series expansions, we obtain

$$s_0 \approx \sqrt{2S_0}\left\{1 + \frac{1}{4}S_0 + \frac{13}{96}(S_0)^2 + \frac{43}{384}(S_0)^3\right\}, \tag{8}$$

where $S_0 = 1 - \bar{R}$ is assumed to be small. This supports the conclusion drawn in Mardia (1972, p. 24) that for small values of S_0, the circular standard deviation reduces to a multiple of the square root of the variance. However, for data from a von Mises distribution with moderate concentration parameters (say $k \in (1, 20)$), the additional terms of the expansion will not be negligible and will influence the shape of the distribution. Quantile-quantile plots of 1000 simulated circular standard deviation values for data from von Mises distributions with sample sizes ranging from 5 to 100, and concentration parameter $k \in (2, 16)$ show that the distribution of the circular standard deviation is quite nearly normal for a variety of sample sizes and dispersions. See Anderson (1993).

Therefore, the circular standard deviation and the circular variance are two strong choices for a dispersion modeling, both satisfying the criteria established earlier in this section. It will subsequently be convenient to utilize the connection between them, namely $s_0 \approx \sqrt{2S_0}$ for concentrated data.

3 Dispersion Modeling

Analogous to the study of dispersion effects for traditional linear data as described by Nair and Pregibon (1988) and Box (1988), this section describes a method for determining the effects of factor levels on the spread of directional data. Factors can often be broken into two categories. Control factors are those which are relatively easy to adjust (or control).

Noise factors, on the other hand are expensive or impractical to control in production, but can be controlled during experimentation. Desensitizing the process to noise variation is the objective of robust design.

Two common choices of general designs are available. The first involves genuine replicates for the observations at each factor combination, while the second involves sampling across a variety of noise factor levels to determine what levels of control factors are robust to changes in the noise factors. If the data is of the first form, then our dispersion measures within each cell gives an indication of the short term variation in the system, but may not give an accurate assessment of variation over the total range of production conditions. The second approach strives to simulate a range of possible operating conditions by changing the levels of some factors which are known to be variable but are typically hard to control in production. This approach gives a more realistic assessment of long term variability in the process, and allows the experimenter to gain information about what combinations of the control factors might reduce this variability. Both approaches can be incorporated within the dispersion modeling framework that we now describe.

First, we determine if there appear to be any significant differences between the estimates of dispersion for the groups. Stephens (1982) describes Bartlett's test for the homogeneity of concentration parameters from von Mises data. For each of the groups in a 2^r factorial design, define $Q_l = mS_l$ and $q_l = m - 1$ where l ranges from 1 to 2^r for the different groups and S_l is the circular variance of group l. We also define $T = \sum_l Q_l$ and $t = \sum_l q_l$. The test statistic for testing if a difference between groups exists is Z/C, where

$$Z = t \log T - \sum_l (q_l \log Q_l) - t \log t + \sum_l (q_l \log q_l) \tag{9}$$

and

$$C = 1 + \frac{1}{3(s-1)} \left(\sum_l \frac{1}{q_l} - \frac{1}{t} \right), \tag{10}$$

where s is the number of groups.

The test statistic, Z/C, is approximately chi-squared with $(s-1)$ degrees of freedom under the null hypothesis of no difference between groups. Therefore, for a test of size $1-\alpha$, we would reject that hypothesis if

$$P(\chi^2_{(s-1)} \geq \frac{Z}{C}) \leq \alpha. \tag{11}$$

However, it is important to note that Bartlett's test is sensitive to assumptions of normality, which in this case corresponds to the data originating from a von Mises distribution. Therefore, this test should be viewed primarily as a diagnostic method for determining if large differences exist between groups. If there is no evidence against the hypothesis that the variance estimates are constant, then the remaining analysis will likely not be beneficial. However, if differences between group dispersions are noted, as in the flywheel example, we proceed with further analyses.

Using the circular variance as our starting point for choice of a dispersion measure, we obtain the resultant length for each combination of factors in the experiment and calculate the circular variance. A suitable transformation of the data is sought using the one parameter Box-Cox power transformation family of the form:

$$(1 - \bar{R})^\lambda = X\Upsilon + \varepsilon \tag{12}$$

where λ is the transformation power (if $\lambda = 0$, then the natural logarithm is used). X is the design matrix (comprised of -1's and 1's for a two-level factorial design), Υ is the vector of parameters, and $\varepsilon \sim MVN(0, \sigma^2 I_n)$ is the vector of error terms. Because we are using a dispersion statistic measured on a linear scale, the model has the same form as dispersion analyses for traditional linear data and the error term can be assumed to be normal, rather than from a directional distribution. This transformation to a linear scale is essential, because we have little intuitive feel for directional dispersion measures, and we are able to use the existing methods for a linear response.

The Box-Cox (1964) approach to data transformations strives to balance three separate goals: simplicity of structure, variance homogeneity, and normality. As in our example, some simplifying assumptions about the model may be required to have some degrees of freedom available for estimating an error term. For example, just the main effects and two-factor interaction terms can be considered for the initial choice of transformation. Once an optimal transformation has been identified, we can examine the full model and determine the relative importance of the different factors and their interactions. Because the circular variances can be reasonably approximated by a chi-squared distribution in many cases and Hawkins and Wixley (1986) showed that the optimal choice of a chi-squared variable is near $\lambda = \frac{1}{3}$, the suggested power transformation parameter will frequently lie in $\lambda \in (-1, 1)$. In addition, $\lambda = 0$ and $\lambda = \frac{1}{2}$ are often contained in the 95% confidence interval of the Box-Cox procedure.

We obtain a model with multiplicative effects and errors on the circular variance for situations where $\lambda = 0$ and the model takes the form $\log(1 - \bar{R}) = X\Upsilon + \varepsilon$. For example for a two-way design we obtain

$$\log(1 - \bar{R}_{ijk}) = \sigma_0 + A_i + B_j + AB_{ij} + \varepsilon_{ijk}. \tag{13}$$

This can be interpreted as the circular variance being influenced by the factors in the following way

$$\begin{aligned}(1 - \bar{R}_{ijk}) &= e^{\sigma_0}\ e^{A_i}\ e^{B_j}\ e^{AB_{ij}}\ e^{\varepsilon_{ijk}} \\ &= \sigma_0^*\ A_i^*\ B_j^*\ AB_{ij}^*\ \varepsilon_{ijk}^*\ ,\end{aligned} \tag{14}$$

where σ_0^* is the baseline measure of variability of the data, and A_i^*, B_j^*, and AB_{ij}^* are the main and interaction effects of the factors, respectively. To interpret the factor effects, if $A_i > 0$, and hence $A_i^* > 1$, then level i of factor A increases the circular variance. Conversely, if $A_i < 0$, then level i of the factor reduces the variance. Because the range of S_0 is restricted to the range $1 - \bar{R} \in [0, 1]$, we have an additional concern for this modeling which is not present for traditional linear data where the variance does not have a finite upper bound. For directional data, $\log(1 - \bar{R}) \in (-\infty, 0]$ which means that the linear combination of factor effects must also be restricted to lie in this range. For a general linear model, there is no convenient way to adjust the range of the $X\Upsilon$ to accommodate this restriction, since extrapolation into some regions of the design space could lead to an expected value of $\log(1-\bar{R})$ which might be positive and hence lie outside of the interpretable range of values. McCullagh & Nelder (1989) comment that a transformation is less desirable if the possibility exists of obtaining a value for $X\Upsilon$ outside of defined boundaries. However, the boundary of the range corresponds to an unlikely extreme of the data being uniformly distributed around the circle. For several industrial examples considered by the authors, the common range for the resultant vector length is $(.45, .95)$, which corresponds roughly to the von Mises concentrations parameter, $k \in (1, 20)$. and circular variances of $(.05, .55)$. This yields a $\log(S_0)$ range of $(-3.0, -0.6)$ which is a reasonable distance away from the problem area near zero. If there is a noticeable gap between the edge of the projected region and zero relative to the expected spread of the data, we would not expect obtaining estimates of $\log(S_0)$ outside of the acceptable range to be a major problem when prediction of variance effect is restricted to the usual interpolation between the high and low levels of the factors. In addition, if we are doing a dispersion analysis with the goal of variance reduction, then this problem area lies at the opposite extreme to our desired target region of minimal variance. As $1-\bar{R}$ decreases, $\log(1-\bar{R}) \to -\infty$ which is in a stable area away from the boundary. In this region we can expect the variance estimates for different factor combinations to be well-defined.

Alternately if $\lambda = \frac{1}{2}$, then we might choose to model the circular standard deviation, instead of the square root of the circular variance. As demonstrated in the previous section, if the spread of the data is small, the two are nearly proportional. The advantage of this choice is that it can be much more easily interpreted as it yields an additive model with an additive error,

$$s_0 = X\Upsilon + \varepsilon \,. \tag{15}$$

Again illustrating with the same two-way design, we obtain the following model:

$$s_{ijk} = \sigma_0 + A_i + B_j + AB_{ij} + \varepsilon_{ijk} \;, \tag{16}$$

where σ_0 is the baseline estimate of the circular standard deviation of the data, and A_i, B_j, and AB_{ij} are the main and interaction effects of the factors, respectively. Once again, $A_i < 0$ corresponds to level i of factor A reducing the spread of the data.

In different industrial applications, one of the models described above may agreeur more closely with the physical understanding of the process, and hence be preferable.

After a suitable transformation has been selected, a half normal plot of the factor effects can provide insights into the relative influence of different factors. Factors influencing the directional dispersion can then be identified and further examined to suggest a suitable combination of factor levels to attain the minimum variability.

Therefore, a strategy for analyzing dispersion from a factorial experiment involving directional data had been outlined. We now address a further approach to analysing the robustness of the process to changing noise factor levels. For linear data, it is well-established that in many situations the use of robust design can give an overall reduction in response variability without having to control the levels of noise factors. Two major approaches are taken to study possible exploitable relationships between control and noise factors (Shoemaker, Tsui & Wu, 1991):

1. Loss modeling involves studying a measure of the dispersion directly as a function of control and noise effects to determine an optimal setting for control factors levels.
2. Examining the response directly can provide insights into specific relationships between control and noise factors.

For directional data, no location model for exploring the factor effects of a factorial design currently exists (see Anderson, 1993 for an explanation of the difficulties associated with such a model). However, the two methods for examining the dispersion effects can be extended to the circular data situation, even in the absence of a working model.

The first method has already been described, but we now clarify how the control and noise factor are directly incorporated. If we have control factors, C, and noise factors, N, we can choose an orthogonal array with factor effects of interest for the control factors (called CA, for control array) and a similar array for the noise factors (NA). Subsequently, the noise array is run for each row of the control array, to give the product array. Ideally, it would be desirable to have replicates at each combination of factors. Once the standardized length of the resultant vector and hence the circular variance are calculated for each combination, analysis of control, noise and control-by-noise factor effects could be studied directly. However, in practice this replication may not be feasible if the cost of multiple runs is prohibitive.

If only one observation is available from each combination of the product array, then the standardized length of the resultant vector for all observations at a given control factor setting is calculated and an overall estimate of the circular variance is obtained across all noise level settings. In this way, a measure is taken of how the variation of the process behaves under a wide variety of noise conditions. Once the circular variances have been obtained for each of the combinations of the CA, we model the dispersions using the procedure outlined earlier. For this model we attempt to minimize the spread of the data across the range of the noise array.

The second approach parallels the examination of control-by-noise interactions for linear data, but because of the absence of a model a quantitative assessment of control-by-noise interactions using the response-model approach is not possible. However, qualitative comparison of different relationships between control and noise factors can give insight into these relationships and suggest dispersion reduction strategies. The key to these qualitative methods is a directional data interaction plot (see Figure 6 for an example and Anderson (1994) for details). The four radial lines mark the circular averages of the factor combinations being considered. The circumscribed arcs show the change in response values across one of the factor's levels while the other factor is held fixed. As Shoemaker, Tsui & Wu (1991) noted, the absolute magnitude of the control-by-noise interaction is not of primary interest, but rather the existence of a control factor combination which gives responses robust to changes in the noise factor levels.

Recall for the linear case, real interest lies not in the particular slopes of the interaction lines, but rather with the overall width of the band that encompassed all noise combinations for a given set of control factor levels. In this way, we can obtain a graphical qualitative analysis of the range of data over the set of noise levels considered. Because of the lack of an underlying model for the directional response from a multi-way design, the loss model approach is superior. It gives a quantitative assessment of which control factor combinations are best for minimizing the spread of the data over the range of noise factor levels. However, analyzing dispersion through a single statistic, like the circular variance or the circular standard deviation, can sometimes disguise interesting attributes in the data. The graphical presentation of control-by-noise interactions can complement the more formal quantitative methods obtained by modeling a function of the dispersion directly.

4 Automotive Example

In this section we consider a complete analysis of the flywheel data, and illustrate the previously described techniques. Because of the strong evidence obtained by applying Bartlett's test (here $Z/C \approx 44$ with significance level 0.0001, see Section 1), we proceed with the dispersion analysis. First, the standardized resultant length, $\bar{R}$, (see Table 3) for each of the factor combinations is obtained, and the circular variance calculated.

The full model to be fitted to the data follows equation (12) where X is the design matrix with a column for the overall mean plus 15 orthogonal columns, one for each main and interaction effects. To implement the Box-Cox procedure, we must make a few simplifying assumptions about the model. If we assume the full model with the effects, there are no degrees of freedom available for an error estimate and the model $(1-\bar{R})^{\lambda} = X\Upsilon + \varepsilon$ for any value of λ will fit the data perfectly. If we eliminate only the four-way interaction, that gives only one degree of freedom for error, and we have most likely overfit the data with too complicated a model. Therefore, to carry out the method we consider two possible design matrices: (i) only the four main effects, and (ii) the four main effects and their 6 two-way

Table 3: Flywheel Dispersion Summaries by Group

Group	$\bar{R}$	S_0	$\log(S_0)$	s_0
1	0.812	0.188	-1.672	0.645
2	0.203	0.797	-0.228	1.783
3	0.510	0.490	-0.713	1.161
4	0.054	0.946	-0.055	2.418
5	0.815	0.185	-1.687	0.640
6	0.434	0.566	-0.568	1.293
7	0.936	0.064	-2.752	0.363
8	0.915	0.085	-2.460	0.423
9	0.604	0.396	-0.925	1.005
10	0.154	0.846	-0.168	1.933
11	0.234	0.766	-0.267	1.703
12	0.836	0.164	-1.806	0.599
13	0.845	0.155	-1.863	0.581
14	0.349	0.651	-0.429	1.452
15	0.809	0.191	-1.653	0.652
16	0.861	0.139	-1.972	0.547

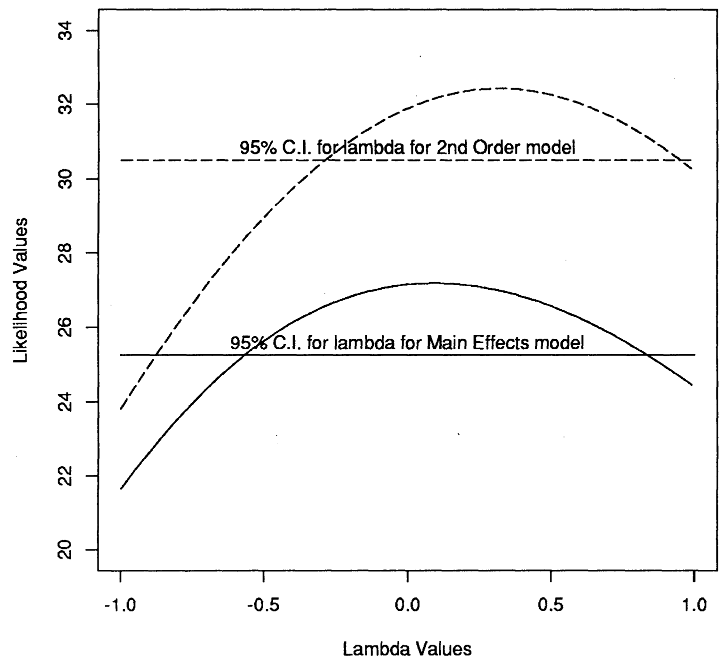

Figure 3: Box-Cox Maximum Likelihood Estimate of Lambda

interactions terms.

Figure 3 shows the plot of two Box-Cox transformation analyses for the flywheel data assuming the models (i) and (ii). For (i), the maximum value for the likelihood occurs for $\lambda = 0.09$ with the 95% confidence intervals covering the range $(-0.56, 0.83)$. For (ii), the maximum value occurs at $\lambda = 0.32$ and the 95% confidence intervals includes $(-0.28, 0.95)$. Hence the choice of either the circular standard deviation model or the log transform remain good options as both lie well within the confidence intervals for acceptable values for λ and have good interpretations. The difference between the two curves for any given λ gives the improvement in the likelihood function by extending the model to include the two-way interactions. The final two columns of Table 3 contains the log circular variance and circular standard deviation values. The current levels of production correspond to group 1, with a resultant length for the group of size 0.812. This corresponds to a group with the seventh smallest dispersion of the 16 groups, so there is some promise for improvement by selecting a better set of factor combinations.

With the set of transformations selected, we revert to the full model with all main effects and interactions included. When the log of the circular variance ($\text{Log}(S_0)$) and the standard deviation (s_0) are modeled separately against the 15 effects we obtain the analysis of variance tables summarized in Table 4. From Figures 4 and 5, which give the half-normal plots for the factor effects for the respective analyses, we see that there is considerable overlap in the dominant factors identified according to the two models. For both transformations, factor "B" is the most influential effect, with the two-way interaction "CD" also contributing significantly. Assuming that a hierarchical model is suitable, we would include the main effects "B", "C" and "D", with the two-way interactions "CD" and "BC". For the log variance, these are the five largest effects, while for the circular standard deviation model, they comprise the four largest effects, with "C" added to maintain the hierarchical structure. Since the multiplicative model for the variance is more easily interpreted for this particular data set, and the factors identified are consistent for the two models, the determination of the best factor combinations here is based on study of the log transformation. Hence the final model is

$$\log(1 - \bar{R}) = \sigma_0 + B_i + C_j + D_k + (BC)_{ij} + (CD)_{jk} + \varepsilon_{ijkl}. \tag{17}$$

The optimal choice is high-high-high for "BCD", which yields a mean for log circular variance of -2.218. If these levels of the factors are selected, we would expect the circular variance to be near 0.1, from a resultant length of 0.891. Alternately, if the present production levels are used the resultant length is 0.812. By changing from the current production levels to the new set of factor combinations, we would be able to reduce variation from 0.188 to 0.109, a 42% reduction. The estimates of variability obtained here are based on knowledge that noise factor "E" has been allowed to vary across its usual range of values within each group. Hence this assessment of the dispersion will likely be more indicative of the true variability than if "E" had not been incorporated. Therefore, a substantial savings can be realized by

Table 4: Dispersion Analysis

	$\text{Log}(S_0)$		s_0	
Source	Sum of Squares	Rank	Sum of Squares	Rank
A	0.069	(12)	0.004	(14)
B	3.563	(1)	1.909	(1)
C	1.071	(4)	0.337	(8)
D	0.924	(5)	0.706	(3)
AB	0.263	(10)	0.114	(10)
AC	0.015	(15)	0.057	(12)
AD	0.632	(7)	0.380	(5)
BC	1.123	(3)	0.524	(4)
BD	0.091	(11)	0.001	(15)
CD	2.004	(2)	1.029	(2)
ABC	0.872	(6)	0.355	(6)
ABD	0.418	(8)	0.343	(7)
ACD	0.371	(9)	0.299	(9)
BCD	0.016	(14)	0.011	(13)
ABCD	0.021	(13)	0.080	(11)

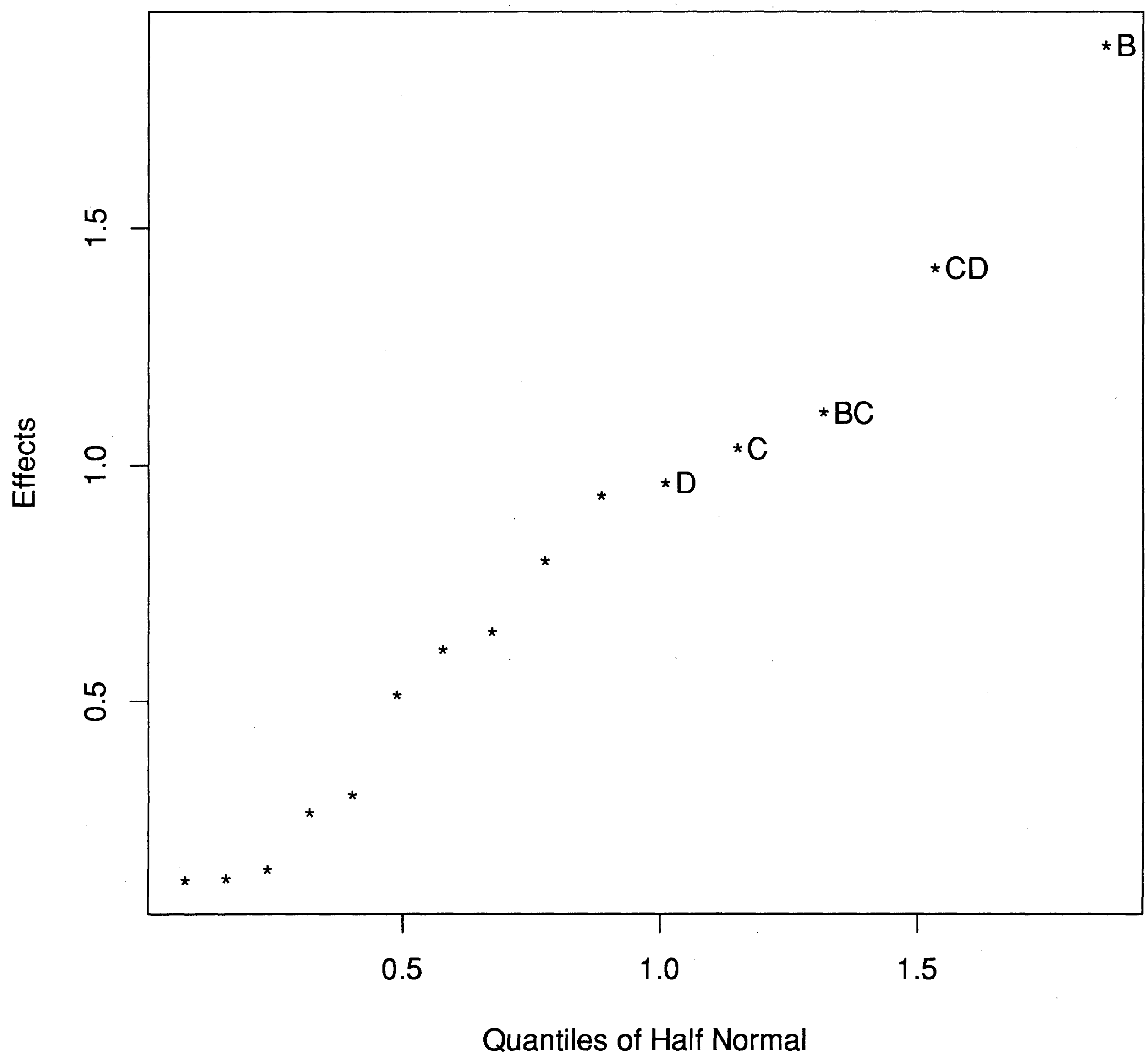

Figure 4: Half-Normal Plot of Effects from Log(S_0) Model

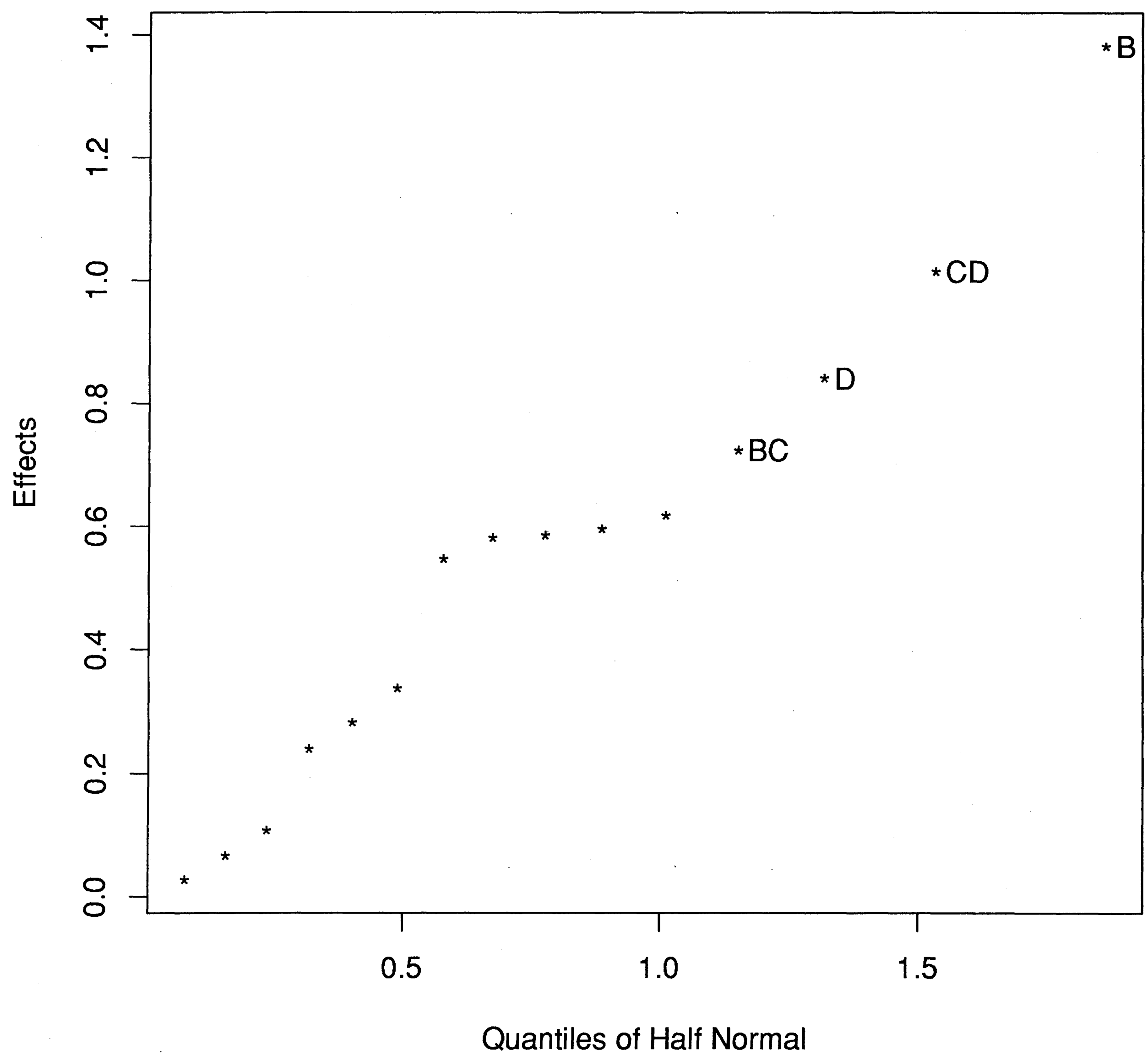

Figure 5: Half-Normal Plot of Effects from s_0 Model

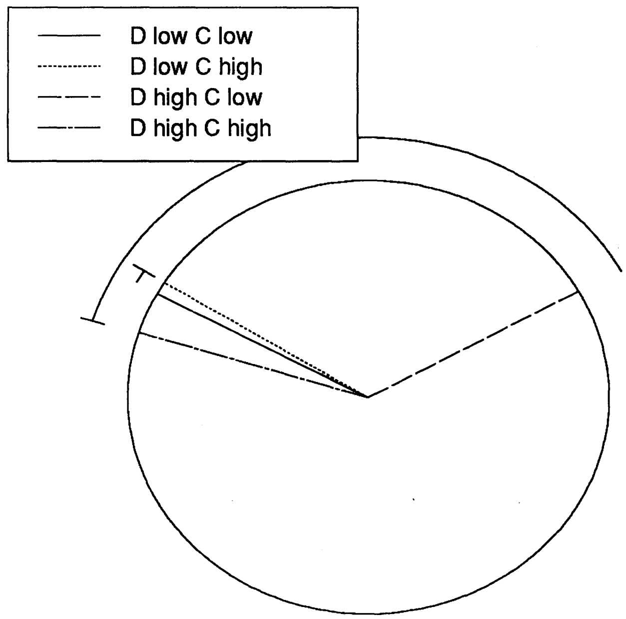

Figure 6: Circular Interaction Plot of "CD"

studying the dispersion effects from this factorial experiment.

We return now to the questions raised in Section 1. By using the dispersion modeling approach outlined here, the relative influence of the factors can be assessed. Factor "B" is the most influential in affecting the dispersion of the response.

Given that factor "C" is expensive to control, we can also treat it as a noise factor and plot some of the control-by-noise interactions and see if it may be possible to exploit one of them. The "CD" interaction is shown in Figure 6. Recall, the goal of examining this plot is to identify if one level of the control factor (here "D") gives a smaller range of values across the noise factor (here "C"). Clearly, the range of the low level of factor "D" gives a much smaller range of responses and hence would be preferable if it were too expensive to control the level of "C" in production.

This illustrates why both loss modelling and the control-by-noise interaction plots are usefully applied to the same data, since one may provide insights not revealed by the other. In this case, different results are obtained if "C" will be controlled in production or not. Hence, while the overall optimal combination of factor for "BCD" is high-high-high, if factor "C" is not controlled in production, then "D" at the low level is a superior choice. As can be seen from Table 3, all four circular variance estimates for "BD" at the high-low combination (rows 5, 7, 13 and 15) are consistently small.

This example demonstrate the methods described and also gives some practical illustrations of the insights that may be gained about the process through this type of analysis.

References

[1] ANDERSON, C.M. (1993), "Location and Dispersion Analyses for Factorial Experiments with Directional Data," Unpublished thesis, University of Waterloo.

[2] ANDERSON, C.M. (1994), "Graphical Methods for Circular and Cylindrical Data," (submitted).

[3] BEST, D.J. & FISHER, N.I. (1979) "Efficient Simulation of the von Mises Distribution," *Appl. Statist.* 28, 152-157.

[4] BEST, D.J. & FISHER, N.I. (1981) "The Bias of the Maximum likelihood Estimators of the von Mises-Fisher Concentration Parameters," *Commun. Statist.-Simula. Computa.* 5, 493-502.

[5] BOX, G.E.P. (1988), "Signal-to-Noise Ratios, Performance Criteria, and Transformations" (with discussion), *Technometrics* 30, 1-40.

[6] BOX, G.E.P. & COX, D.R. (1964), "An Analysis of Transformations" (with discussion), *J.R.S.S. B* 26 211-243.

[7] DOBSON, A.J. (1978), "Simple Approximations for the von Mises Concentration Statistic," *Appl. Statist.* 27, 345-347.

[8] FISHER, N.I. & LEE, A.J. (1992), "Regression Models for an Angular Response," *Biometrics* 48, 665-677.

[9] HAWKINS, D.M. & Wixley, R.A.J. (1986), "A Note on the Transformation of Chi-Squared Variables to Normality," *Amer. Statist.* 40, 296-298.

[10] MARDIA, K.V. (1972) *Statistics of Directional Data.* London: Academic Press.

[11] NAIR, V.N. & PREGIBON, D. (1988) "Analyzing Dispersion Effects from Replicated Factorial Experiments," *Technometrics* 30, 247-257.

[12] RIVEST, L.-P. (1982) "Some Statistical Methods for Bivariate Circular Data," *J.R.S.S. B* 44, 81-90.

[13] SHOEMAKER, A.C., Tsui, K.-L. & WU, C.F.J. (1991) "Economical Experimentation Methods for Robust Design," *Technometrics* 33, 415-427.

[14] STEPHENS, M.A. (1982) "Use of the von Mises Distribution to Analyse Continuous Proportions," *Biometrika* 69, 197-203.

[15] TAGUCHI, G. (1986), *Introduction to Quality Engineering: Designing Quality Into Products and Processes,* Tokyo, Japan: Asian Productivity Organization.