

5-1-2005

# Some Guidelines For Using Nonparametric Methods For Modeling Data From Response Surface Designs

Christine M. Anderson-Cook

*Los Alamos National Laboratory Arizona State University*

Kathryn Prewitt

*Arizona State University*

Follow this and additional works at: <http://digitalcommons.wayne.edu/jmasm>



Part of the [Applied Statistics Commons](#), [Social and Behavioral Sciences Commons](#), and the [Statistical Theory Commons](#)

## Recommended Citation

Anderson-Cook, Christine M. and Prewitt, Kathryn (2005) "Some Guidelines For Using Nonparametric Methods For Modeling Data From Response Surface Designs," *Journal of Modern Applied Statistical Methods*: Vol. 4 : Iss. 1 , Article 12.

DOI: 10.22237/jmasm/1114906320

Available at: <http://digitalcommons.wayne.edu/jmasm/vol4/iss1/12>

This Regular Article is brought to you for free and open access by the Open Access Journals at DigitalCommons@WayneState. It has been accepted for inclusion in *Journal of Modern Applied Statistical Methods* by an authorized editor of DigitalCommons@WayneState.

## Some Guidelines For Using Nonparametric Methods For Modeling Data From Response Surface Designs

Christine M. Anderson-Cook  
Statistical Sciences Group  
Los Alamos National Laboratory

Kathryn Prewitt  
Mathematics and Statistics  
Arizona State University

---

Traditional response surface methodology focuses on modeling responses using parametric models with designs chosen to balance cost with adequate estimation of parameters and prediction in the design space. Using nonparametric smoothing to approximate the response surface offers both opportunities as well as problems. This article explores some conditions under which these methods can be appropriately used to increase the flexibility of surfaces modeled. The Box and Draper (1987) printing ink study is considered to illustrate the methods.

Key words: Edge-Corrections, data sparseness, bandwidth, lowess, Nadaraya-Watson

---

### Introduction

In his review of the current status and future directions in response surface methodology, Myers (1999) suggests that one of the new frontiers is to utilize nonparametric methods for response surface modeling. Explored in this article are some of the key issues influencing the success of these methods used together. Combining nonparametric smoothing approaches, which typically depend on space-filling samples of points in the desired prediction region, with response surface designs, which primarily focus on an economy of points for adequate prediction of prespecified parametric models, presents some unique challenges. Nonparametric approaches are typically used either as an exploratory data analytic tool in conjunction with a parametric method or exclusively because a parametric model didn't

provide the necessary sensitivity to curvature.

The number and location of design points impose a limitation on the order of the polynomial the parametric model can accommodate. This, in turn, imposes a limitation on the type of curvature of the fitted model. Standard response surface techniques using parametric models often assume a quadratic model. Nonparametric techniques assume a certain amount of smoothness, but do not impose a form for the curvature of the target function. Local polynomial models which fit a polynomial model within a window of the data can pick up important curvature, which a parametric fit typically cannot. Issues of what designs are suitable for utilizing nonparametric methods, appropriate choices of smoother types as well as bandwidth considerations will all be discussed. Important limitations exist for incorporating these methods into surface modeling, because ill-defined or nonsensical models can easily be generated without careful consideration of how to blend the method and design.

---

Christine Anderson-Cook is a Technical Staff Member at Los Alamos National Laboratory. Her research interests include response surface methodology and design of experiments. Email: c-and-cook@lanl.gov. Kathryn Prewitt is Associate Professor at Arizona State University. Her research interests include nonparametric function estimation methods, time series and goodness-of-fit. Email: kathryn.prewitt@asu.edu

Vining and Bohn (1998) utilized the Gasser-Mueller estimator (G-M) (see Gasser & Mueller, 1984) which is a kernel based smoothing method to estimate the process variance for a dual response system for the Box and Draper (1987) printing ink study. In that study, a full  $3^3$  factorial design was used with three replicates per combination of factors. Each

variable was considered in the range  $[-1, 1]$  for the coded variables. Dual models were developed to find an optimal location by modeling both the mean of the process, which has a desired target value of 500, and the variance of the process, which ideally would be minimized. Using a parametric model for the mean and a nonparametric model for the variance, Vining and Bohn (1998) obtained a location in the design space with a substantially improved estimate mean square error (MSE) over parametric models for both mean and variance presented by Del Castillo and Montgomery (1993) and Lin and Tu (1995). The estimated MSE was the chosen desirability function for simultaneously optimizing the mean and variances of the process.

The Box and Draper (1987) example has some interesting features that suggest consideration of nonparametric methods for modeling the variability of the data set. Because the mean response has been carefully studied and appears to be relatively straightforward to model, the focus is on the characteristics and modeling of the standard deviation. This is only half of the problem for the dual modeling approach, but the nonparametric issues here are many.

First, an overview of the characteristics of that part of the data set is provided. Figure 1 shows a plot of the 27 estimates of the standard deviation at the  $3^3$  factorial locations. Clearly, there is no easily discernible pattern in this response such as a simple function of the three factors. In addition, the range of the data should give us some concern. Within the range of the experimental design space, the standard deviation varies from a value of 0 (all three observations at each of  $(-1, -1, 0)$  and  $(0, 0, 0)$  were measured to be exactly the same) to a value of 158.2 at  $(1, 0, 1)$ .

This should alert one to a possible problem immediately as this range occurring in an actual process seems extreme. Figure 2 shows several different ranges of response standard deviations from 1 to 20. It is uncertain as to what a maximal proportional difference between minimum and maximum variance should be, however, a 1:10 or 1:20 ratio already seems excessive for most well-controlled industrial processes. Hence, one of the goals of the

modeling should likely be to moderate this range of observed variability to more closely reflect what is believed to be realistic for the actual process.

If the modeling undersmooths the data (approaching interpolating between observed points), a risk exists of basing the dual response optimization on non-reproducible idiosyncrasies of the data. If the data is oversmoothed, important curvature is flattened making it difficult to find the best location for the process. This perpetual problem of modeling is doubly important here as the results of the model are being used to determine weights for the modeling of the mean of the process as well as for the optimization of the global process through the dual modeling paradigm. Hence, as different models for the variability are considered, predicted ranges will be noted throughout the design space.

Reviewed in this article are some of the basics of nonparametric methods and their implications for the designed experiment are discussed with limited sample size and structured layout of design points. Then compared are different nonparametric approaches to the existing parametric choices and those presented in Vining and Bohn (1998) for this particular example, and conclude with some general recommendations for how to sensibly and appropriately use nonparametric methods for response surface designs

### Smoothing Methods

Smoothing methods are distinct from traditional response surface parametric modeling in that they use different subsets of the data and different weightings for the selected points at different locations in the design space. There are several popular nonparametric smoothing methods such as the Nadaraya-Watson (Nadaraya, 1964) and Watson (1964) which fits a constant to the data in a window, the Gasser-Mueller (Gasser & Mueller, 1984) which is a convolution-type estimator, spline smoothing (Eubank, 1999), and local polynomial methods (Fan & Gijbels, 1996) which fit polynomials in the local data window.

Figure 1: Printing Ink standard deviation raw data

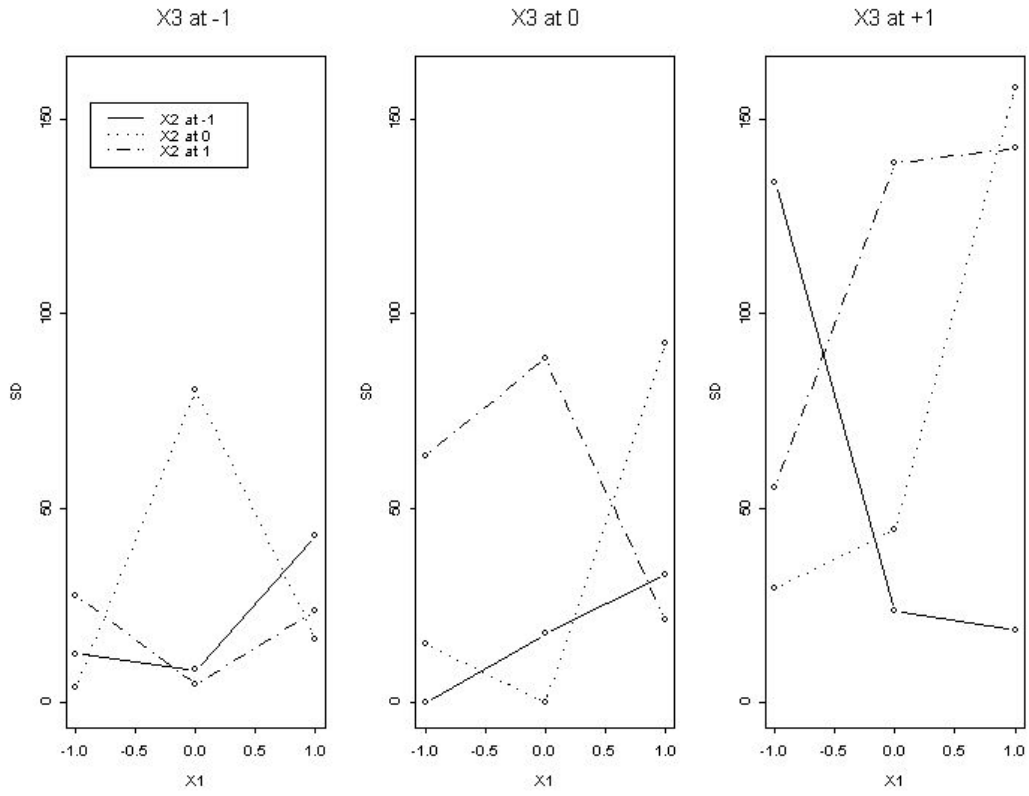
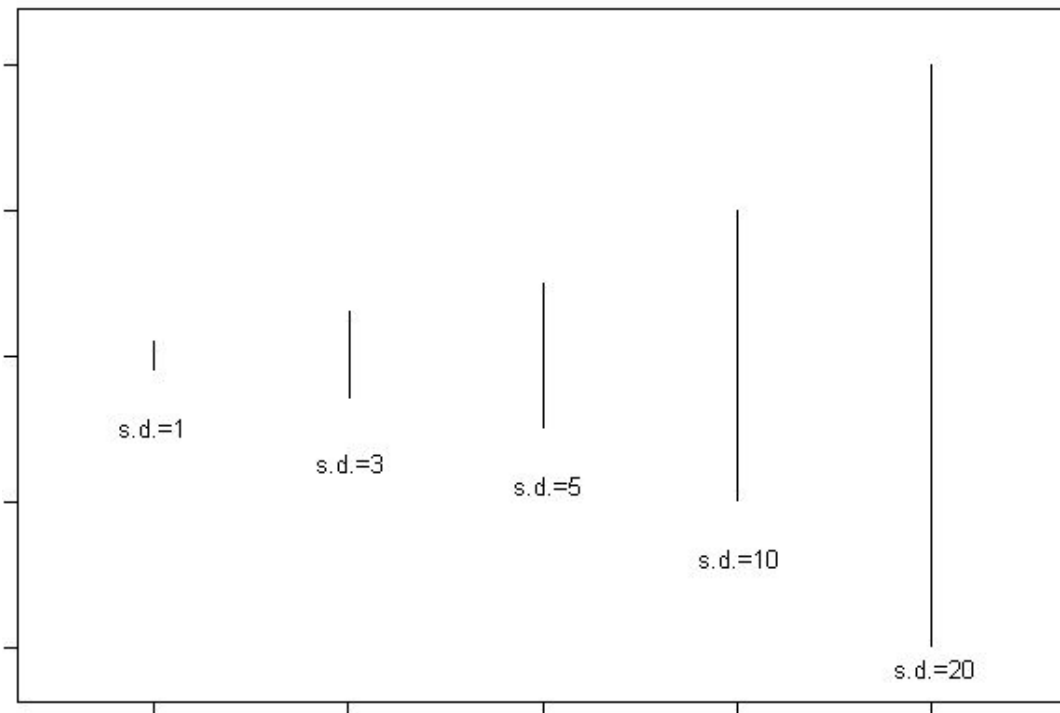


Figure 2: Range of observed responses likely with different values for a variety of standard deviations



The local polynomial methods (lowess) have several positive properties for this problem. Generally, these methods fit a polynomial within a window of data determined by a bandwidth with the kernel function providing weights for the points. The estimators are linear combinations of the responses just as the familiar parametric regression estimators. The Nadaraya-Watson estimator fits a constant in the window so it is a special case of the local polynomial method. When fitting polynomials of order greater than one, these estimators have been shown to naturally account for the bias issues on the boundary (Ruppert & Wand, 1994). In the problem, most of the data (26 of 27 locations) are on the edge, so the concern should be with the behavior of estimators on the boundary.

Typically the nonparametric literature differentiates between behavior on the boundary and behavior in the interior. The variance of this estimator conditioned on the data is unbounded. The unconditional variance of the estimator is actually infinite (Seifert & Gasser, 1996) if the number of points in the window is small (two or less in the local linear, univariate X case). Consequently, the number of points in the window should be greater than two in the univariate X case and in practice greater than the minimum necessary to calculate the estimator. The conditional unbounded variance is due to the fact that the coefficients of the  $Y_i$ 's in the estimator can be positive or negative).

Problem can be envisioned as all data points are on the edges of a cube with the exception of the point (0, 0, 0). The minimum bandwidth which would include at least 4 data points would be larger than 1 (half of the range of each coded variable) otherwise the number of points in the window would be too small to allow estimation.

Observe  $n$  independent data points  $(\mathbf{X}_i, Y_i)$  where  $\mathbf{X}_i = (X_{i1}, X_{i2}, \dots, X_{ip})$  are the locations in the design space, and  $Y_i$  is the response. The model assuming homoscedastic error for nonparametric function estimation is:

$$Y_i = m(\mathbf{X}_i) + \sigma(\mathbf{X}_i)\varepsilon_i \quad (1)$$

where  $E(\varepsilon_i) = 0$  with  $Var(\varepsilon_i) = 1$  and  $\sigma^2(x) = Var(Y | X = x)$ . The smoothing function,  $m(\cdot)$  is also called the regression function,  $E(Y | X = x)$ . It is assumed that the variance of the error term for the problem of modeling the printing ink standard deviations would be reasonably constant.

Kernel smoothing nonparametric methods involve the choice of a kernel function and a bandwidth as a smoothing parameter which determines the window of data to be utilized in the estimation process. The idea is to weight the data according to its closeness to the target location, hence to estimate  $m(x_0)$ , greater weight is given to the  $Y_i$  values with associated  $\mathbf{X}_i$  values close to  $x_0$ .

Spline smoothing methods are categorized as a nonparametric technique and involve a smoothing parameter but no kernel function. One of the few references to an application of nonparametric methods to response surface problems is Hardy et al. (1997) who explored the use of R-splines with a significantly larger number of design points and with the goal of selecting variables for the regression model rather than obtaining a plausible curve.

There are special considerations when using nonparametric methods for the printing data problem which are next outlined. Most of the literature regarding nonparametric methods shows application to space-filling designs and a larger number of sample points. The printing example has 27 data points which is significantly smaller than the data typically seen in the smoothing literature. Most of these points are on the boundary or edge. It is known that nonparametric estimators can exhibit so-called boundary effects. If a method such as the Gasser-Mueller (Gasser & Mueller, 1984) is used, the bias is bounded but not decreasing with increased sample size as one would want unless kernel functions called boundary kernels are used. This means that a different kernel needs to be used when a point is on the boundary.

Local polynomial methods of order greater than 1 incorporate naturally the boundary kernels necessary. These methods are easily

explained by comparing them to a weighted least squares problem where the kernel function provides the weights and the estimate is provided by solving a familiar looking matrix operation. Most of the nonparametric methods literature provides results and examples for sample sizes much larger than the printing data and also provides leading terms of the bias and variance to describe the behavior of the estimator which implies that there are negligible terms as  $n$  grows large. The problem then is much different than has been addressed before: the sample size is small, the design is not space-filling and most of the points are on the edge.

#### Bandwidth issues

One of the most important choices to make when using a nonparametric method of function estimation is the smoothing parameter. For kernel methods, the bandwidth is such a parameter. Large bandwidths provide very smooth estimates and smaller bandwidths produce a more noisy summary of the underlying relationship. The reason for this behavior can be seen in the leading terms of the bias and variance for a point in the interior in the univariate explanatory variable case:

$$\text{Bias}(\hat{m}(x)) \approx \frac{1}{2} m''(x) b^2 \int K(u) u^2 du$$

and

$$\text{Var}(\hat{m}(x)) \approx \frac{\sigma^2 \int K(u)^2 du}{n b f(x)} \quad (2)$$

where  $f(x)$  is the density of the  $\mathbf{X}$  explanatory variable,  $K(\cdot)$  the kernel function, and  $b$  the bandwidth. The effect of the bandwidth can be observed: large values of the bandwidth increase the bias and reduce the variance of the predicted function; small values decrease the bias and increase the variance. This difficulty is called the bias-variance tradeoff. Bandwidth selection methods can be local (potentially changing at each point at which the function is to be estimated) or global (where a single bandwidth is used for the entire curve). Typically, the bandwidth is often chosen to minimize an optimality criterion, such as an estimate of the leading terms of the MSE or cross-validation (see Eubank, 1988, Fan

& Gijbels, 1995; Prewitt & Lohr, 2002). The optimality quantities are more accurate when data sets are larger, i.e., the leading terms of the MSE leave out negligible terms which are often not negligible when  $n$  is small. In simulation studies with bivariate data, sample sizes of  $n < 50$  are not seen. The current problem, on the other hand, involves multivariate data with three explanatory variables and one response with a total of only 27 data points. Minimizing a quantity such as  $SSE$  where

$$SSE = \sum_{i=1}^n (Y_i - \hat{m}(\mathbf{X}_i))^2 \quad (3)$$

cannot be used for the purpose of goodness of fit because without a parametric form for  $m(\cdot)$ ,  $SSE$  is minimized with  $Y_i = \hat{m}(\mathbf{X}_i)$ , i.e. the curve estimate which minimizes this quantity is obtained by connecting the points. The purpose of the bandwidth selection method is essentially to solve the bias-variance tradeoff difficulty described previously. The second derivative in the bias term suggests that these estimators typically underestimate peaks and overestimate valleys which is sometimes an argument for using a local bandwidth choice since the expectation would be to use a smaller bandwidth in regions where there are more curvature.

Because the number of points in the problem is small, it would be more sensible to use a global bandwidth, one bandwidth for the entire curve. There are not enough points to justify accurate estimation of different local bandwidths. This is not to say that in the future it may be discovered that in fact different bandwidths should be used to estimate different portions of the surface, but existing methods (Fan & Gijbels, 1995; Prewitt, 2003) will not work. Methods for local bandwidth selection have relied on the fact that each candidate bandwidth for a particular point  $x_0$  will incorporate additional data points as the bandwidth candidates become larger which may not be the case for the problem.

## Methodology

## Nonparametric Methods for the Sparse Response Surface Designs

Two methods were considered that take into account the special circumstances of the problem as previously outlined. The fitted constant (C) version of the local polynomial (Nadaraya, 1964; Watson, 1964) and the local linear version (LL) (see Fan & Gijbels, 1996) were used. The benefit of fitting these models is that curvature can be achieved without fitting higher order polynomials as is necessary when a completely parametric model is fit. There is the potential to capture different kinds of curvature consistent with what might be reasonable given the nature of the design implemented. The Epanechnikov kernel was used

$$K(u) = 0.75(1 - u^2)I(|u| \leq 1) \quad (4)$$

which is simple and has optimal properties (Mueller, 1988).

At the point  $\mathbf{x} = (x_1, x_2, x_3)$  the weighted least squares estimate with a kernel function as the weight. The two methods can be described as follows where  $\hat{m}_C(\mathbf{x})$  is the local polynomial with fitted constant: Let the weight function be defined as:

$$K_b(x, X_1) = \frac{1}{b^3} \prod_{j=1}^3 K\left(\frac{x_j - X_{ij}}{b}\right).$$

This is called a product kernel because it is the product of three univariate kernel functions. The kernel function equals zero when data points are outside the window defined by the bandwidth and has a nonzero weight when  $\mathbf{X}_i$  is inside the window. It is appropriate to use the same bandwidth in each of the three directions because the scaling of the coded variables in the set-up of the response surface design makes units comparable in all directions. The definition below resembles a weighted least squares estimator when a constant is fit.

$$\begin{aligned} \hat{m}_{NW}(\mathbf{x}) &= \arg \min_{\beta_0} \sum_{i=1}^n (Y_i - \beta_0)^2 K_b(\mathbf{x}, \mathbf{X}_i) \\ &= \sum_{i=1}^n \frac{K_b(\mathbf{x}, \mathbf{X}_i) Y_i}{\sum_{i=1}^n K_b(\mathbf{x}, \mathbf{X}_i)} \end{aligned} \quad (5)$$

The second method considered is defined below and resembles a weighted least squares estimator where a plane is fit with the data centered at  $\mathbf{x}$  so that the desired estimator is  $\hat{\beta}_0$  and the "LL" stands for local linear with no higher order terms.

$$\begin{aligned} \hat{m}_{LL}(\mathbf{x}) &= \arg \min_{\beta_0} \sum_{i=1}^n (Y_i - \beta_0 - \beta_1(x_1 - \mathbf{X}_{i1}) \\ &\quad - \beta_2(x_2 - \mathbf{X}_{i2}) - \beta_3(x_3 - \mathbf{X}_{i3}))^2 K_b(\mathbf{x}, \mathbf{X}_i) \end{aligned} \quad (6)$$

One can also think of the above estimators as motivated by a desire to estimate  $m(\mathbf{x})$  by using the first few terms of its Taylor expansion,  $m_{NW}(\mathbf{x})$  is constructed by considering an interval around  $\mathbf{x}$  and estimating the first term of the Taylor expansion around  $\mathbf{x}$  where  $m_{LL}(\mathbf{x})$  uses estimates of first order terms of the Taylor expansion as an estimate of  $m(\mathbf{x})$ .

## Printing Example Smoothing

It has already been noted that some particular issues concerning the application of nonparametric smoothing to a sparse small set of data with the vast majority of design locations on the edges. A related issue to consider is what type of surface is possible or likely. If the variability of the process can change very quickly and dramatically within the range of the design space, then the  $3^3$  factorial design is an inadequate choice and should be replaced by a much larger space filling design.

However, if the surface should change moderately slowly throughout the region, then the  $3^3$  design may be adequate. As well, if the surface is likely to be relatively smooth and undergoes changes slowly, then a nonparametric method should be selected and bandwidth that

uses information from several nearby points to estimate the surface locally. Examined now are some of the implications of choosing different bandwidths for this  $3^3$  factorial design.

The  $3^3$  factorial design is comprised of 27 locations on the cube: 8 corner points, 12 edge points, 6 face center points and one center point. Notably, all but one of the points are on the edge of the design space. This is standard practice for parametric estimation, because D- and G-efficiency both benefit from maximal spread of points to the edges of the design space.

However, this set-up coupled with the extreme small sample size is highly unusual for nonparametric approaches. One of the advantages of the structured locations selected for a response surface design is that allows the investigation of the characteristics of estimation for different nonparametric bandwidth choices. For example, using the Epanechnikov kernel weighting function, the number of design points can be specified that will be used for estimation at each of the four categories of design points.

Table 1 shows the effect of bandwidth on different locations as well as the range of the non-zero weights for particular bandwidths used for the local estimation. Bandwidths less than 0.5 of the total range of each variable use only the observation at that location, while a bandwidth of 1 uses all observations. The weights associated with each design location change for different weights. As the bandwidth increases, not only do more locations get used, but also their relative contributions to the estimate become more comparable. For example, for a bandwidth of 0.6 at one of the design points, the observation at the location to be estimated is weighted approximately 35 times more ( $1.25 / 0.036$ ) than the most distant non-zero weighted observations. As well, for a design point and a bandwidth of 1, this ratio drops to 2.4 ( $0.75 / 0.316$ ) and the points used are also further away.

Various authors considered different models for the standard deviation for this data set. Parametric models considered include a linear model in all three factors on  $\log(\text{standard deviation} + 1)$ , shown in Figure 3(a) with an  $R^2$  of 29.4 %. The transformation of the standard deviation was done to improve fit, and to avoid

negative predicted values. The range of predicted standard deviation values back on the original scale for this model range from 5.0 to 113.5, which gives a ratio of maximum to minimum standard deviation of 22.7. A full quadratic model for  $\log(\text{standard deviation} + 1)$  yields an  $R^2$  of 40.6 % and is shown in Figure 3(b). Here, the ratio of maximum to minimum standard deviation is 25.5 ( $145.1/5.7$ ).

Fitting the constant (C) and local first-order polynomial (LL) methods for a variety of bandwidths to the data were also considered. Figures 4 (a), and (b), show predicted surfaces for the untransformed standard deviation with the constant C method and bandwidths of 0.8 and 1.0, respectively. Figures 5(a), (b) and (c), show the LL method for the same response and bandwidths of 0.6, 0.8 and 1.0. For each of the parts of the figures, three slices of the design space are shown, with the third factor, C, at the low, middle and high value. Figures 6(a), (b) and 7(a), (b) show the predicted surfaces when modeling using the  $\log(\text{standard deviation} + 1)$  response and bandwidths of 0.8 and 1.0. As the bandwidth increases, the surface becomes smoother, reflecting the idiosyncrasies of the data less.

Tables 2 and 3 summarize the ranges of predicted values throughout the design space observed for the different methods for both the untransformed and  $\log(\text{standard deviation} + 1)$  responses. The C method tends to moderate the range of the predicted values considerably more than either the parametric or the lowess models. This is due the relative lack of influence of edge effects with extreme values. The transformation to the log-scale does not have a consistent effect on the range of prediction for the different approaches, with it moderating the range of predicted values for only some of the bandwidths. The LL method is susceptible to prediction of larger values near the edges of the design space, with a seeming sensitivity to edge effects. Notably missing from this comparison is the best Vining and Bohn (1998) smoother (Gasser-Mueller), which uses a bandwidth of 0.3 of the total range. As noted in Table 1, this small bandwidth is essentially an interpolator with most regions having only a single observation used for the estimation.



Table 1: Number of points contributing to local estimation for 3<sup>3</sup> factorial, with Epanechnikov kernel.

| Bandwidth          | # of Points Used |         |    | Min. & Max. Weights |               |               |
|--------------------|------------------|---------|----|---------------------|---------------|---------------|
|                    | (.25,.5)         | (.5, 1) | 1  | .6                  | .8            | 1             |
| <b>Location</b>    |                  |         |    |                     |               |               |
| <b>Corner</b>      | 1                | 8       | 27 | (0.036,1.250)       | (0.212,0.938) | (0.316,0.750) |
| <b>Edge</b>        | 1                | 12      | 27 |                     |               |               |
| <b>Face Center</b> | 1                | 18      | 27 |                     |               |               |
| <b>Center</b>      | 1                | 27      | 27 |                     |               |               |
| <b>(0,0,.5)</b>    | 2                | 18      | 27 | (0.096,1.033)       | (0.042,0.846) | (0.185,0.703) |
| <b>(0,.5,1)</b>    | 2                | 12      | 27 | (0.096,1.033)       | (0.042,0.846) | (0.185,0.703) |
| <b>(.5,.5,.5)</b>  | 8                | 8       | 27 | (0.705,0.705)       | (0.002,0.689) | (0.063,0.618) |

Table 2: Summary of Prediction Values for Lowess and Local Average on Untransformed Standard Deviations.

| Bandwidth | Lowess  |         |       | Local Average |         |       |
|-----------|---------|---------|-------|---------------|---------|-------|
|           | Minimum | Maximum | Ratio | Minimum       | Maximum | Ratio |
| .6        | 3.1     | 149.9   | 48.4  | 9.0           | 117.3   | 13.0  |
| .8        | 4.8     | 145.5   | 30.3  | 14.1          | 100.8   | 7.1   |
| 1         | 7.2     | 147.3   | 20.5  | 15.9          | 94.8    | 6.0   |

Table 3: Summary of Prediction Values for Lowess and Local Average on Transformed Log (Standard Deviations+1).

| Bandwidth | Lowess  |         |       | Local Average |         |       |
|-----------|---------|---------|-------|---------------|---------|-------|
|           | Minimum | Maximum | Ratio | Minimum       | Maximum | Ratio |
| .8        | 5.5     | 192.0   | 34.9  | 7.4           | 70.3    | 9.5   |
| 1         | 6.5     | 227.2   | 35.0  | 8.0           | 60.9    | 7.6   |

Figure 3: Contour plots for best Linear and Quadratic parametric models based on the Box and Draper (1987) data for log (standard deviations + 1).

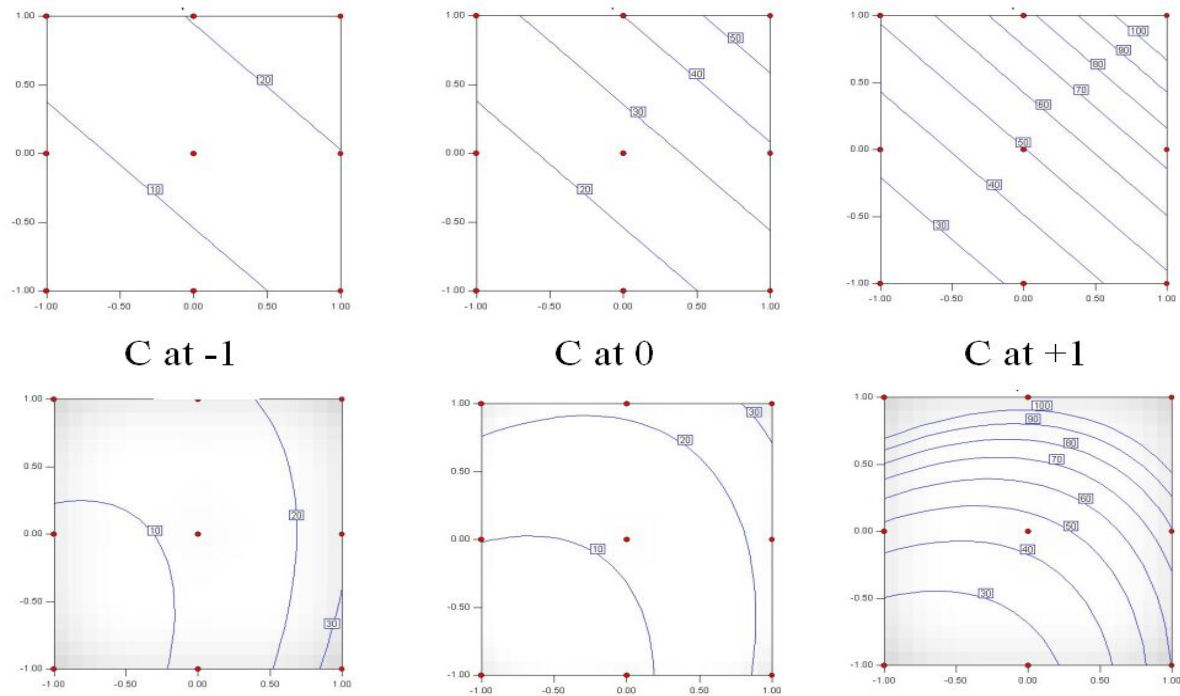
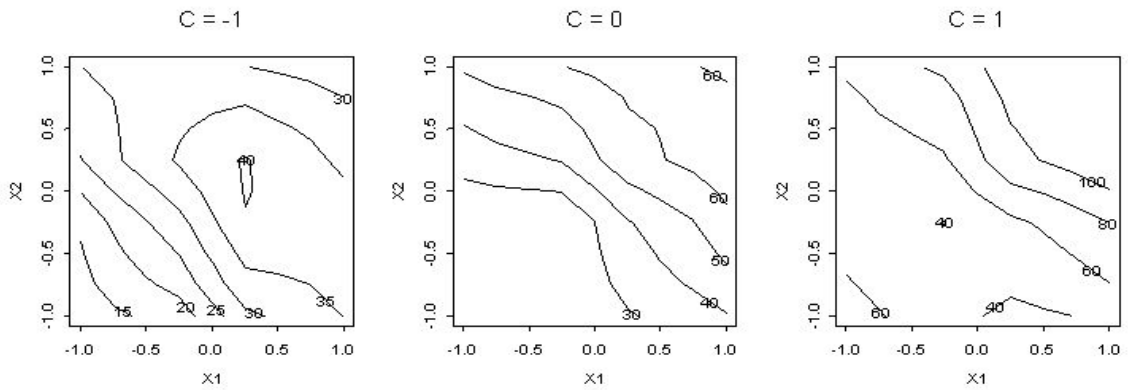


Figure 4: Contour plots of local average models for untransformed response with bandwidths 0.8 and 1.

(a)



(b)

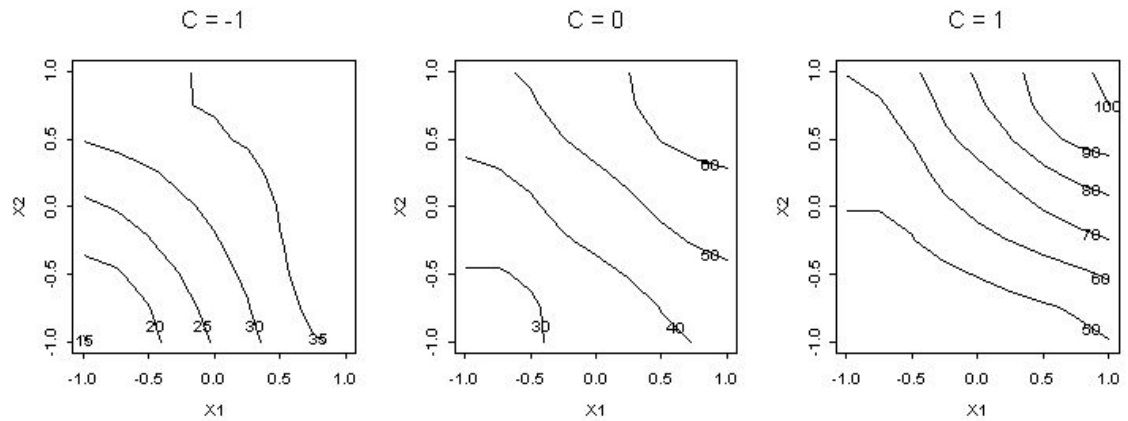
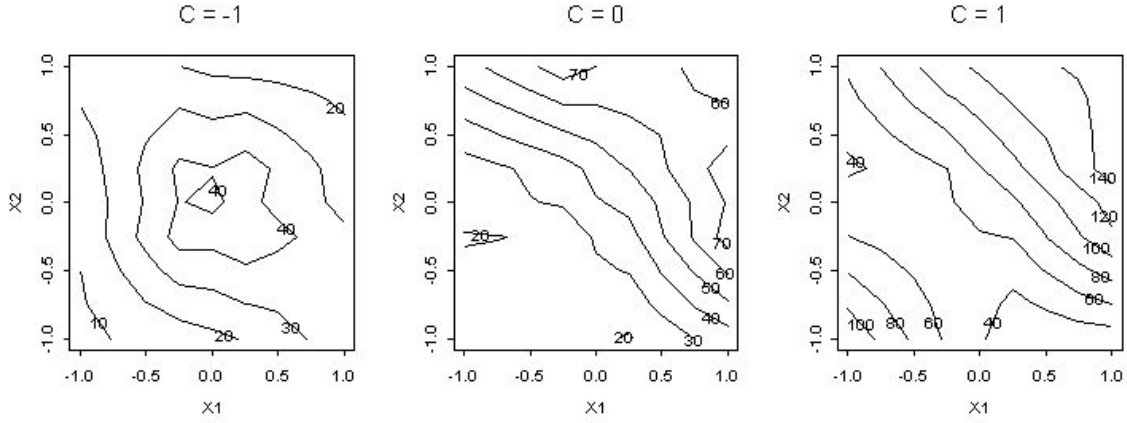
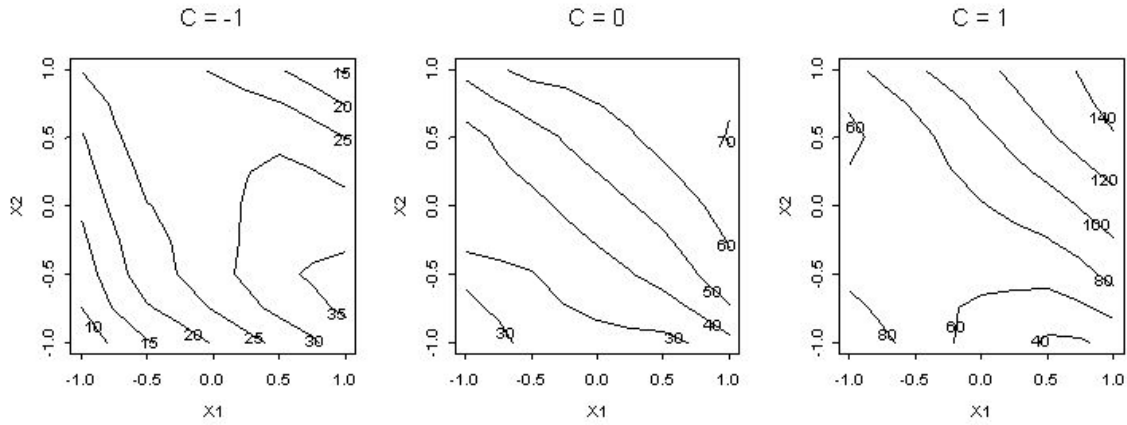


Figure 5: Contour plots of lowess models for untransformed response with bandwidths 0.6, 0.8 and 1.

(a)



(b)



(c)

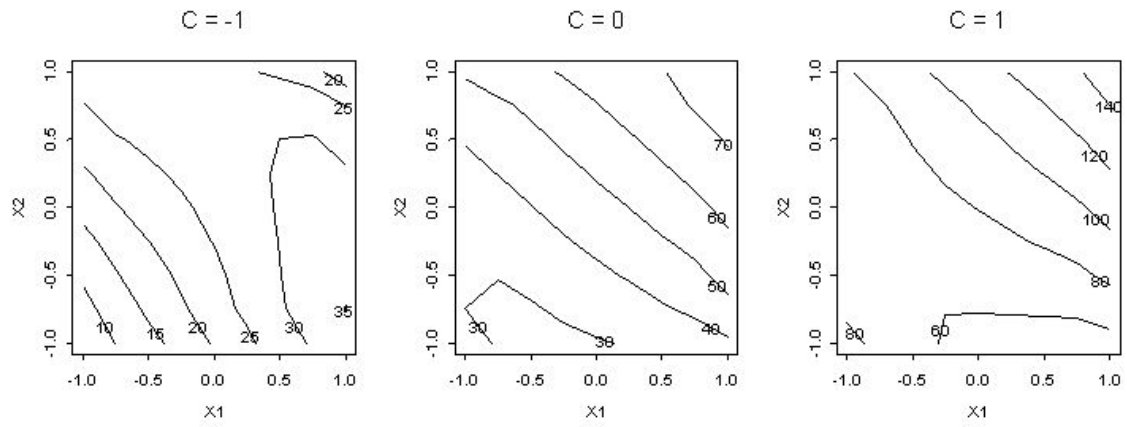


Figure 6: Contour plots of local average models for logarithm transformed response with bandwidths 0.8 and 1.

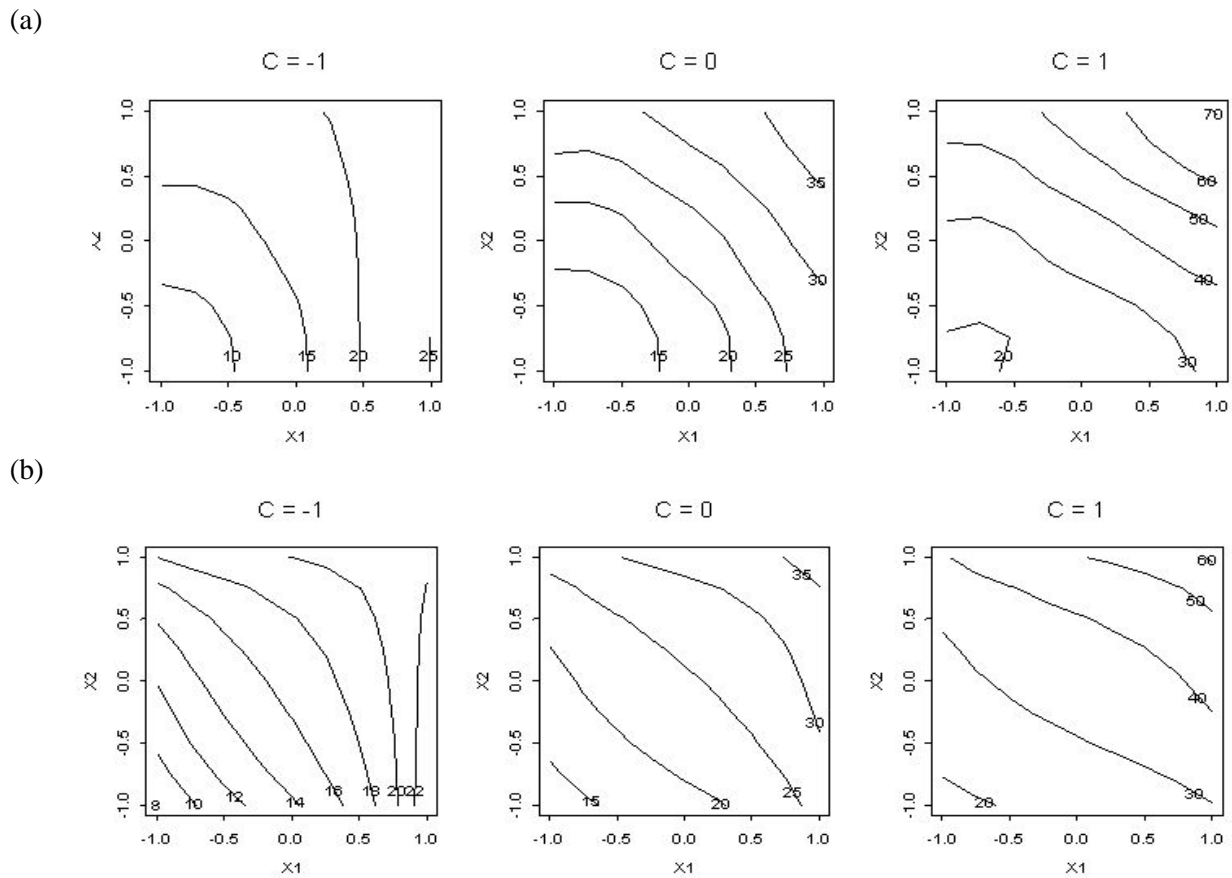
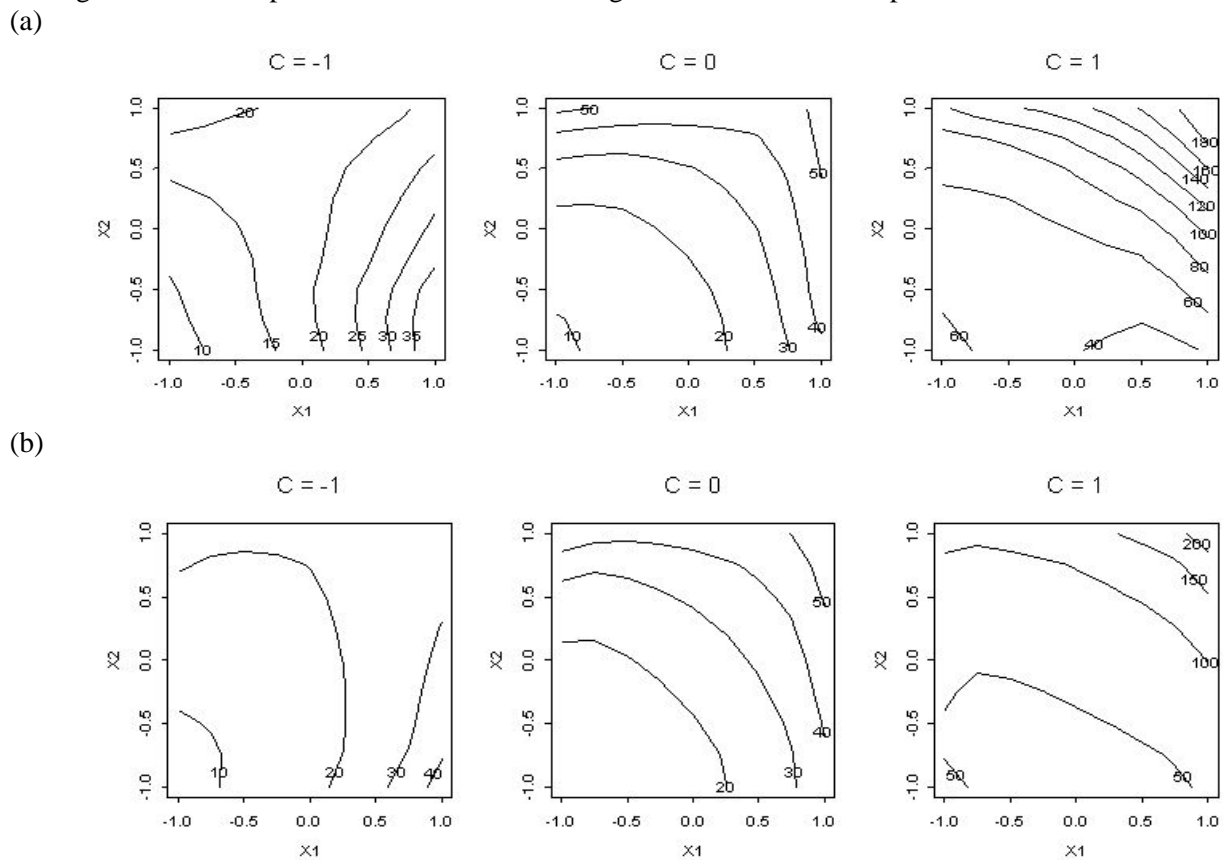


Figure 7: Contour plots of lowess models for logarithm transformed response with bandwidths 0.8 and 1.



The fit was next considered by comparing the  $R^2$  values for the different methods in Table 4. Unlike parametric models, where minimizing the  $R^2$  is desirable, here the goal is to obtain a good fit without merely interpolating between points. This is a particularly appropriate strategy given the extreme ranges of values for the standard deviation observed. Also reported are the cross-validation  $R^2$  values which were obtained by removing a single observation, refitting the model with 26 points, and then calculating the difference between the predicted value and what was observed.

Typically in other regression settings, this is a way of measuring the robustness of the model for future prediction. However, in this case, with such a small sparse data set, removing a single point (almost all of which are on the edge of the design space) has the result of leading us to do extensive extrapolation to obtain the new predicted values. As a result, the values obtained were very discouraging. For a number of the cases, including the quadratic parametric model and the small bandwidth LL method, negative  $R^2$  values were obtained, which imply that the model has predicted less well than just using a constant for the entire surface. Again, the structure of the data and the extreme amount of extrapolation involved in this calculation should be considered in interpreting these values. The 0.3 bandwidth Vining and Bohn (1998) smoother cannot be considered in this comparison, because an empty region in the design space was obtained for all of the points, which does not allow the cross-validation  $R^2$  value to be calculated.

However, there are a few general conclusions that can be reached. First, one should be quite cautious with any of these models. Due to the sparsity of the data, they can be influenced considerably by a single value. Secondly, larger bandwidths give lower  $R^2$  values, but generally perform better under the challenges of the cross-validation assessment. Finally, the LL method appears to outperform the C method for the  $R^2$  values, but consistently underperform C for the cross-validation  $R^2$ . This reflects the sensitivity to edge effects of this

method, which either yields good responsiveness if using the values near the edge, or wide extrapolation when this point is removed. This seems to imply some superiority for the C method, which outperforms both the parametric models, and appears to retain some useful predictive ability even when used for extrapolation.

Based on an overall assessment of all characteristics of the methods considered, the Nadaraya-Watson local averaging (C) method with bandwidth of either 0.8 or 1.0 emerge as leading choices. The bandwidth of 1.0 uses all of the data, with diminishing weights for more distant points. The 0.8 bandwidth excludes points on the opposite side of the design space for corner, edge and face-center points. Both of these models allow for greater flexibility than either of the parametric models, by allowing greater adaptability of the shape of the surface, while also utilizing a significant proportion of the data for estimation. They provide enough smoothing to produce a surface that likely is consistent with underlying assumptions of how the standard deviation of the process might vary across the range of the design space

### Conclusion

Based on sparseness of the data sets typical for many response surface designs, it should be evident that the use of nonparametric methods must be used with care to avoid nonsensical results. However, the printing ink example has demonstrated that nonparametric models have real potential for helping with modeling responses, when the restrictions of a parametric model are too limiting. The ability to adapt the shape of the surface locally is desirable, and can be done even when there are only a small number of values observed across the range of each variable. It is particularly important to consider a priori what the surface, range and ratio of maximum to minimum predicted values reasonably might be. The chosen method should balance optimizing fit, while still maintaining characteristics of the appropriate shape.

Table 4: Fit of models to Log (Standard Deviation +1) response.

| Model            | $R^2$ (in %) |           |          | Cross-validation $R^2$ (in %) |           |          |
|------------------|--------------|-----------|----------|-------------------------------|-----------|----------|
|                  | Linear       | 29.4      |          |                               | 8.2       |          |
| Quadratic        | 40.6         |           |          | -47.0                         |           |          |
| <i>Bandwidth</i> | <b>.6</b>    | <b>.8</b> | <b>1</b> | <b>.6</b>                     | <b>.8</b> | <b>1</b> |
| Lowess           | 83.6         | 67.6      | 61.8     | -11.0                         | -1.6      | 1.0      |
| Local Ave        | 65.7         | 42.5      | 35.7     | 12.7                          | 13.4      | 12.8     |

Table 5: Number of points contributing to local estimation for different Central Composite Designs and widths.

| Design                                    | Bandwidth         | Corner | Axial | Center Run<br>( $n_c = 1$ ) |
|---|-------------------|--------|-------|-----------------------------|
| CCD<br>$k=2, \alpha = \sqrt{2}$<br>$n=9$  | $<0.353$          | 1      | 1     | 1                           |
|   | $0.353 < b < 0.5$ | 4      | 3     | 5                           |
|   | $0.5 < b < 0.707$ | 4      | 6     | 9                           |
|   | $0.707 < b < 1.0$ | 7      | 6     | 9                           |
| CCD<br>$k=3, \alpha = \sqrt{3}$<br>$n=15$ | $<0.289$          | 1      | 1     | 1                           |
|   | $0.289 < b < 0.5$ | 5      | 5     | 9                           |
|   | $0.5 < b < 0.578$ | 5      | 10    | 15                          |
|   | $0.578 < b < 1.0$ | 12     | 10    | 15                          |
| CCD<br>$k=4, \alpha = 2$<br>$n=25$        | $< 0.25$          | 1      | 1     | 1                           |
|   | $0.25 < b < 0.5$  | 6      | 9     | 25                          |
|   | $0.5 < b < 1$     | 21     | 16    | 25                          |

Due to a large number of points on the edge of the design space, which is highly desirable for D- and G-efficiency when using a parametric model, a smoother which is insensitive to edge effects is recommended. The local averaging smoother (C) performed quite well although the (LL) supposedly has superior boundary capability in the bias term both in order and boundary kernel adjustment. The reason for this apparent contradiction may be again that the sample size is small and the boundary order results depend on larger sample sizes or as pointed out in Ruppert and Wand (1994) the boundary variance of the (LL) may

be larger than the boundary variance of the (C) estimator. Consequently the local averaging (C) estimator is recommended for this problem.

The local first-order polynomial works well in many standard applications, where the proportion of edge points is small, but does not seem like a suggested choice for most response surface designs.

To avoid near-interpolation, a moderate to large bandwidth needs to be used. Table 5 considers perhaps the most popular class of response surface designs, the Central Composite Design. It gives the number of points used for estimation for the different types of points for a

number of different bandwidths. A bandwidth of size less than 0.5, or half the range of the coded variables, yield estimates for some of the points using only a small number of observations.

By coupling moderate to large bandwidths with the Epanechnikov kernel, it is possible to downweight but not eliminate the contribution of more distant points, and hence a balance between local adaptivity and moderating extreme values is retained.

Symmetric designs, such as  $3^k$  factorials and Central Composite, are likely to perform better than non-symmetric designs, like Box-Behnken or fractional factorial designs (with some corners of the design space unexplored). While the non-symmetric design performs well for parametric models, the surface will be disproportionately poorly estimated in some regions.

Given the inherent different structure of response surface designs compared to more standard regression studies typically considered in the nonparametric smoothing literature, considerably more research is possible to determine not only reasonable, but optimal smoothing strategies in this context.

#### References

- Box, G. E. P., & Draper, N. R. (1987). *Empirical model-building and response Surfaces*. Wiley.
- Del Castillo, E., & Montgomery, D. C. (1993). A nonlinear programming solution to the dual response problem. *Journal of Quality Technology*, 25, 199-204.
- Eubank, R. L. (1988). *Spline smoothing and nonparametric regression*. Marcel Dekker.
- Eubank, R. L. (1999). *Nonparametric regression and spline smoothing*. Marcel Dekker.
- Fan, J., & Gijbels, I. (1995). Data-driven bandwidth selection in local polynomial fitting: Variable bandwidth and spatial adaptation. *Journal of the Royal Statistical Society, Series B, Methodological*, 57, 371-394.
- Fan, J., & Gijbels, I. (1996). *Local polynomial modelling and its applications* (ISBN 031298321). Chapman & Hall.
- Gasser, T., & Mueller, H.-G. (1984). Estimating regression functions and their derivatives by the kernel method. *Scandinavian Journal of Statistics*, 11, 171-185.
- Hardy, S. W., Nychka, D. W., Haaland, P. D., & O'Connell, M. (1997). Process modeling with nonparametric response surface methods. In *ASA Proceedings of the Section on Physical and Engineering Sciences*, pages 163-172. American Statistical Association (Alexandria, VA).
- Lin, D. K. J., & Tu, W. (1995). Dual response surface optimization. *Journal of Quality Technology*, 27, 34-39.
- Myers, R. H. (1999). Response surface methodology - Current status and future directions (pkg: P1-74). *Journal of Quality Technology*, 31, 30-44.
- Nadaraya, E. (1964). Some new estimates for distribution functions. *Theory of Probability and its Applications (Transl of Teorija Verojatnostei i ee Primenenija)*, 9, 497-500.
- Prewitt, K. (2003). Efficient bandwidth selection in non-parametric regression. *Scandinavian Journal of Statistics*. 30, 75-92.
- Prewitt, K., & Lohr, S. (2002). Condition indices and bandwidth choice in local polynomial regression. Under Review.
- Ruppert, D., & Wand, M. P. (1994). Multivariate locally weighted least squares regression. *The Annals of Statistics*, 22, 1346-1370.
- Seifert, B., & Gasser, T. (1996). Finite-sample variance of local polynomials: Analysis and solutions. *Journal of the American Statistical Association*, 91, 267-275.
- Vining, G. G., & Bohn, L. L. (1998). Response surfaces for the mean and variance using a nonparametric approach. *Journal of Quality Technology*, 30, 282-291.
- Watson, G. (1964). Smooth regression analysis. *Sankhya A26*, 359-372.