# Identifying Parametric Nonlinear Models for Computer Codes

Matthias Schonlau, Michael Hamada, and
William J. Welch
*University of Waterloo*

I.I.Q.P. Research Report
**RR-96-02**
January 1996

# Identifying Parametric Nonlinear Models for Computer Codes

**Matthias Schonlau, Michael Hamada, and William J. Welch**

Department of Statistics and Actuarial Science and

The Institute for Improvement in Quality and Productivity

University of Waterloo

Waterloo, Ontario N2L 3G1 Canada

**Abstract**

A complex mathematical model that produces output values from input values is now commonly called a computer model. The literature thus far has concentrated on obtaining fast predictors for the computer model. Because the relationship between the inputs and outputs is still mathematically complex, however, these predictors are not suitable for explanation. We show how plots of nonparametric estimates of main and interaction effects based on the predictor are useful for identifying a class of parametric nonlinear models; fitting the nonlinear model provides the desired explanation. The proposed method is illustrated using data from a computer experiment with a solar collector.

KEYWORDS: Nonparametric function fitting, Computer experiment, Computer model, Nonlinear regression, Model identification, Visualization

# 1 Introduction

Computer models or codes are now frequently used in engineering design, and in many other areas of physical science. For instance, the main example discussed in this article concerns the engineering design of a solar collector. This code computes an increase in heat transfer effectiveness, $y$, resulting from an engineering innovation. The design is characterized by six factors (engineering parameters), $x_1, \ldots, x_6$. Further details will be given in Section 3. As is often the case, the code is expensive to compute and the engineers wanted a, possibly nonlinear, parametric approximation. The surrogate would then be used to determine quantitatively the impact of the six design factors on the response and to explain the complex functional relationships embodied in their computer code.

Figure 1 shows scatter plots of the response against each $x$ variable in turn for data from an experiment on the solar collector code. They indicate some trend in $x_2$ and $x_5$. However, the scatter plots do not show, for example, the strong relationship in $x_4$, because it is masked by the effects of the other covariates. This would not matter if the effects were all linear and additive, but, as we shall see in Section 3, the effect of $x_4$ is highly nonlinear. With nonlinear effects, we need to know the form of the model to be fitted, and simple plotting of the data does not suggest a class of nonlinear parametric models here. Moreover, in our experience, nonlinearities are common in computer experiments, because the inputs often cover wide ranges.

There is already some work on the design and analysis of computer experiments. See, for example, Currin, Mitchell, Morris, and Ylvisaker (1991), Sacks, Schiller, and Welch (1989), Sacks, Welch, Mitchell, and Wynn (1989), and Welch, Buck, Sacks, Wynn, Mitchell, and Morris (1992). The methods proposed in these references take into account the deterministic nature of a code like the solar collector computer model. Given the same inputs, it always reproduces the same output(s). Typically, the code will be expensive to run, e.g., it solves a large number of differential equations which may require several hours or more of computer time.

So far this work has focused on finding a good cheap-to-compute nonparametric surrogate

(i.e., predictor) for the computer model. In the solar collector example, however, *explanation* rather than *prediction* is the overriding objective. The class of nonparametric predictors suggested in the above references is unsuitable for this task: They are computationally cheap approximations, but they are nonetheless mathematically complex.

In this article we propose identifying approximating parametric models from graphical analysis of estimated effects from a nonparametric model. Along the way we introduce some new methodology for attaching standard errors to the estimated effects. Note that here the nonparametric analysis of computer experiments is an intermediate tool rather than an end in itself. As will be shown, the visualization of effects is fairly automatic.

An overview of the article is as follows. Section 2 first outlines the nonparametric method we use for analyzing data from a computer experiment. It has several advantages, but it is by no means the only method that might accomplish this task. Section 2 then explains how parametric models can be identified graphically. Section 3 demonstrates these ideas using the solar collector code. In Section 4, the article concludes with some discussion, including comments on the choice of experimental design and alternative modeling approaches.

# 2 Identifying Parametric Nonlinear Models

Identifying a class of nonlinear models that fits the data well is easy if there is only one covariate. A simple scatter plot would reveal the functional relationship which for a computer model is exact since the relationship is deterministic. Then, the data analyst can choose a class of models suggested by the scatter plot and fit the model using standard nonlinear regression software to obtain parameter estimates. This approach was used in a case study presented in Bates and Watts (1988, Section 3.13) for physical experimental data which contained random error. While the data from a computer experiment contain no random error, the objective here remains the same, i.e., to find a good approximating model.

Scatter plots are not very useful for model identification where there is more than one covariate, however. The relationship between the response and each covariate can be masked

by the relationships between the response and the other covariates (e.g., Montgomery and Peck, 1982, Section 4.2.5). To overcome the masking problem, a plot of a function involving only the covariate of interest is needed. In other words, the effects of the other covariates need to be eliminated. Such plots will be considered shortly, after some preliminaries.

First, a brief overview of the nonparametric predictor used in this article is given because it plays a key role in the method proposed shortly. The data from a computer experiment consist of $n$ vectors of covariate values (or inputs) denoted by $\mathbf{x}_1, \ldots, \mathbf{x}_n$ for the $d$ - dimensional covariates $x_1, \ldots, x_d$ as specified by a particular experimental design. The corresponding response values (for a particular output variable) are denoted $\mathbf{y} = (y_1, \ldots, y_n)^t$. Then, following the approach of, e.g., Welch et al. (1992), the response $y$ is treated as a realization of a stochastic process:

$$Y(\mathbf{x}) = \beta + Z(\mathbf{x}), \tag{1}$$

where $E(Z(\mathbf{x})) = 0$ and $\mathrm{Cov}(Z(\mathbf{w}), Z(\mathbf{x})) = \sigma_z^2 R(\mathbf{w}, \mathbf{x})$ for two input vectors $\mathbf{w}$ and $\mathbf{x}$. The correlation function $R(\cdot, \cdot)$ can be tuned to the data, which for this article is assumed to have the form:

$$R(\mathbf{w}, \mathbf{x}) = \prod_{j=1}^{d} \exp(-\theta_j |w_j - x_j|^{p_j}), \tag{2}$$

where $\theta_j \geq 0$ and $0 < p_j \leq 2$. The $p_j$'s can be interpreted as smoothness parameters—the response surface is smoother with respect to $x_j$ as $p_j$ increases—and the $\theta_j$'s indicate how local the estimate is. If the $\theta_j$'s are large, only data at points in the immediate vicinity of a given point are highly correlated with $Y$ at that point and are thus influential in the prediction at that point. If the $\theta_j$'s are small, data at points further away are still highly correlated and still influence the prediction at that point.

The best linear unbiased predictor of $Y$ at an untried $\mathbf{x}$ can be shown to be:

$$\hat{Y}(\mathbf{x}) = \hat{\beta} + \mathbf{r}^t(\mathbf{x})\mathbf{R}^{-1}(\mathbf{y} - \mathbf{1}\hat{\beta}), \tag{3}$$

where $\mathbf{r}(\mathbf{x})$ is the $n \times 1$ vector of the correlations between $Y(\mathbf{x})$ and $\mathbf{y}$, $\hat{\beta}$ is the generalized least squares estimator of $\beta$, $\mathbf{R}$ is the $n \times n$ correlation matrix with element $i, j$ defined by

$R(\mathbf{x}_i, \mathbf{x}_j)$ in (2) and $\mathbf{1}$ is an $n \times 1$ vector of 1's. While this cheap-to-compute predictor has proven to be accurate for numerous applications, it does not reveal the relationship between $y$ and $x_1, \ldots, x_d$ in a readily interpretable way. Consequently, this predictor is unsuitable for *explaining* the functional relationship between the covariates and the response.

When the functional relationship between the covariates $x_1, \ldots, x_d$ and the response $y$ is approximately additive, i.e.,

$$y \simeq \mu_0 + \mu_1(x_1) + \mu_2(x_2) + \ldots + \mu_d(x_d),$$

the difficult problem of identifying a nonlinear function $y(x_1, \ldots, x_d)$ has turned into the much easier problem of identifying $\mu_i(x_i)$ for $i = 1, \ldots, d$. Note that while the method proposed later in this section does not assume additivity, an important point is that additivity does make the model identification much easier.

Recall that in order to identify the functional relationship between a group of covariates and the response, the effect of these covariates needs to be isolated from the others. When we want to isolate the effect of a single covariate, the true *main* effect of the covariate can be defined in the following two ways:

1. *Integrating out the other factors.* The main effects are defined as:

$$\mu_i(x_i) = \int y(\mathbf{x}) \prod_{h \neq i} dx_h$$

   (Sacks et al., 1989). They can be estimated by replacing $y(\mathbf{x})$ by $\hat{Y}(\mathbf{x})$. Standard errors for the estimated main effects are derived in Appendix A.

2. *Keeping the other variables fixed.* For example, the other variables might be fixed at their respective midranges. Standard errors for the estimated effects using this method are available directly from $\text{MSE}(\hat{Y}_{\mathbf{x}})$ as given for example in Sacks et al. (1989).

In both calculations, the unknown $y(\mathbf{x})$ needs to be replaced by $\hat{Y}(\mathbf{x})$ from Equation (3). The first approach is preferred because it is analogous to analysis of variance in that all the other covariates are averaged out. Note also that integrating $\hat{Y}(\mathbf{x})$ is numerically easy to

perform if the $\mathbf{x}$ region is cuboidal and if the correlations are in product from as in (2). In a similar fashion, the joint effect of two or more covariates can be investigated by integrating out all the other covariates or fixing the other covariates at specific values.

Main effects for each $x_i$ and joint effects of, say, two covariates for each pair $(x_i, x_j)$ can then be displayed graphically. By choosing a tentative model for each of the effect plots which displays some key feature (i.e., impacts the response), an overall model can be developed by adding up all the corresponding candidate models.

If there are no interactions (and hence, additivity holds) the $d$-dimensional problem has been reduced to $d$ one-dimensional problems. If large interactions are present, then the interacting covariates need to be considered jointly. Covariates can then be grouped so that covariates in two different groups do not interact. Provided that the groups contain no more than two variables, candidate models may still be identified from contour plots of the response. For larger sized groups, such plots will generally not be helpful. In this case, when faced with many interactions, transforming the response may help in reducing the apparent complexity. Experience with a number of computer models, however, suggests the complexity of computer models tends to arise from additive nonlinearities rather than through interactions.

Subsequently, the identified parametric model can be fit using standard nonlinear regression techniques. When there is additivity, starting values for the parameter estimates can be estimated from the main effect plots.

# 3 Application to a Solar Collector Code

In this section, the proposed method is applied to an expensive-to-compute computer model for the heat exchange effectiveness between the air and an unglazed transpired-plate solar collector with slot-like perforations (henceforth, referred to as holes). The use of equally spaced slot-like holes replaces the unrealistic assumption of infinitesimally small and infinitesimally close holes and thus, represents an engineering novelty in the design of unglazed solar

collectors. Golneshan (1994) showed that the heat exchange effectiveness for these solar collectors is a function of six covariates, (1) inverse wind velocity, (2) dimensionless slot width, (3) Reynolds number, (4) admittance, (5) dimensionless plate thickness, and (6) the radiative Nusselt number, as defined by a system of differential equations. The computer code (Cao, 1993) solves the system of differential equations for given covariate values and requires around two hours of computing time on a workstation. The response considered here is the increase in heat exchange effectiveness attributed to the heat transfer in the holes from the hole sides and is expressed as a percentage (0-100). For further details, see Cao (1993). For notational simplicity, in the following, the six covariates listed above will be referred to as $x_1, x_2, \ldots, x_6$ and the response as $y$.

The mechanical engineers who had developed the solar collector code were interested specifically in explaining the impact of the six covariates (which are design factors) on the response heat exchange effectiveness; ultimately, the explanation would help to identify better solar collector designs. Note that such understanding was not apparent from inspecting the system of differential equations. The engineers were interested in developing a surrogate *parametric* model because empirical models of this type existed in the literature for solar collectors based on older technologies; they had no preconceived idea of what form the model should take because the collectors with slot-like holes represented state-of-the-art technology. Hence, the need arose for performing an experiment on the solar collector code, i.e., a computer experiment.

The experimental design used for the computer design was one that filled the six dimensional cuboidal region, a so-called space filling design. Specifically, a Latin hypercube design (McKay, Beckman, and Conover, 1979) consisting of 100 points was chosen in which the minimum distance between points (i.e., the covariate vectors) in low-dimensional projections was maximized. The design was found using ACED (Algorithms for Constructing Experimental Designs) which was developed by Welch. All the two-dimensional projections of the Latin hypercube design can be seen in Figure 2 which shows that the design is indeed space-filling.

Because the response " heat exchange effectiveness" is a percentage, models in the logit of the response were considered. Scatter plots of the logit data (Figure 1) indicate a possible linear trend in $x_2$ and $x_5$. The remaining relationships, if any, are masked by the presence of the other covariates. In the following, the proposed method for identifying a class of surrogate nonlinear models will be applied.

The stochastic process predictor (3) for the logit response was fit using the nonparametric method outlined in the previous section, implemented in GaSP (Gaussian Stochastic Processes), also developed by Welch. One of the observations is very extreme on the logit scale. Based on predicting each $y_i$ using all the data except $y_i$, the cross validation prediction error for this observation is very large (see Figure 3). Standardizing this cross-validation residual by dividing by its standard error gives a value of 7.1, also very large.

We remove this observation for modeling purposes and refit the model. Figure 4 displays a cross validation plot of the new fit with 99 data points. The predictor appears to be reasonably accurate. Main effect plots, generated by integrating out the other covariates, are as shown in Figure 5 for covariates $x_1$ through $x_6$. The main effect for covariate $x_6$ is very flat, and all two-way interactions are close to zero everywhere. These effects were considered negligible by the engineers. The features displayed in the main effect plots suggest that the effect of $x_1$, $x_2$, and $x_3$ are approximately linear and the effect of $x_5$ is approximately quadratic.

The main effect plot for $x_4$ is rather ragged. Although the plot gives a good indication of the apparently nonlinear $x_4$ effect, it is doubtful that the true $x_4$ relationship is that bumpy. One possible explanation is that the computer code may have some numerical convergence problems in certain regions of the $\mathbf{x}$ space. This possible erratic behavior may then be erroneously attributed to $x_4$ which clearly has the most nonlinear or complex impact on the response. Engineering knowledge suggests that the increase in heat efficiency is a monotone increasing function of the admittance rate of the plate $x_4$. The head engineer commented: "The slight blip in the curve is almost certainly due to some numerical problem" (Hollands, 1995, personal communication). Therefore, we do not model the little down peak at $x_4 = 300$.

Plots of the main effects using the method of fixing the other variables at their respective midranges rather than averaging them out, result in very similar graphs. For example, Figure 6 shows the Method 2 main effect plot for $x_4$.

The nonlinear shape of the $x_4$ main effect plot which appears to asymptote can be captured by a Michaelis-Menten model (Bates and Watts, 1988, p. 329); the Michaelis-Menten model has long been used to model the behavior of a limiting chemical reaction which rises at a decreasing rate to an asymptote. It also arises in the context of a reciprocal link function in generalized linear model, where an inverse linear response function is assumed (McCullagh and Nelder, 1989, p. 291). Here, to give more flexibility, the Michaelis-Menten model was augmented by introducing an additional parameter $\beta_2$ and takes the following form:

$$y = \frac{1}{\beta_0 + \beta_1 / x_4^{\beta_2}}.$$

The overall model consisting of linear effects in $x_1$, $x_2$, $x_3$, and $x_5$, a quadratic effect in $x_5$ and the augmented Michaelis-Menten model for $x_4$ was then fit using standard nonlinear regression software which gave:

$$\text{logit}(y) = 6.601x_1 - 0.0028x_3 - 35.41x_2 + 53.61x_5 - 392.54x_5^2 + \frac{1}{-0.388 + 0.210/x_4^{0.488}}.$$

All of the parameters including the flexibility parameter $\beta_2$ were significant at the 0.0001 level. Also, adding $x_6$ reveals that $x_6$ is not significant at the .05 level. Although the data contain no random error so that significance testing has no theoretical grounds here, the results of the significance tests do indicate the importance of the various effects relative to the ability of the overall model to fit the data. Note that the model contains only eight parameters but fits the 99 data points quite well as indicated by the corresponding cross validation plot given in Figure 7. The fact that the parametric nonlinear model does not fit the data quite as well as the nonparametric model is not surprising, since the parametric model is much simpler.

# 4 Discussion

The examples presented in nonlinear regression books typically deal with only a single covariate $x$, where the functional relationship between $x$ and the response $y$ is unknown. On the other hand, the method proposed in this article can be applied to an arbitrarily large number of covariates.

Throughout this paper, we have used model (1) for the initial nonparametric analysis. Other nonparametric methods, like Generalized Additive Models, could be used. However, the model we use has three main advantages: first, the model is truthful to the deterministic nature of the data, second, error bounds for the effects are available, and third, interactions do not need to be modeled explicitly. Breiman (1991) criticized algorithms for producing "only one picture" of the functional relationship, thus ignoring the many other "pictures" which are almost as good. The error bounds given for the effects can serve here as an assessment of the variability of the effect fit.

There are certainly other ways to identify a parametric nonlinear model. For example, clever residual analyses in the hand of a skilled data analyst may well lead to the same results. For the solar collector experiment, an added variable (partial regression) plot for $x_4$ based on a linear regression model for the remaining covariates shows the effect of $x_4$ is nonlinear, albeit with considerable scatter as displayed in Figure 8. This success is not surprising since the assumption of a linear model for the remaining variables turns out to be a good approximation. If the true model had contained several strong nonlinearities, then added variable plots on their own would not have sufficed.

Elaborate residual analyses are often not done for three reasons: (1) They are hard to do, especially when the "true" model contains more than one nonlinear effect. (2) Data analysts, especially inexperienced ones, may not always know about them. (3) They can take a lot of time to perform. The method presented here is easy and fairly automatic for detecting nonlinear effects. It is not a panacea for all "true" models, however. If the "true" model cannot be transformed to an additive model with few or no interaction effects, then identification of nonlinear relationships with several covariates will still be a challenge. For

these cases, it is doubtful whether alternate methods will work either.

The effect plots play a key role in the proposed method and their resolution depends on the experimental design used. The Latin hypercube design is a desirable choice because the design points fill the experimental region well and produce high-resolution plots.

Originally, a $4^{6-2}$ fractional factorial design was considered for the solar collector computer experiment. A fractional factorial or even full factorial design would have had several drawbacks, however. First, if only a few covariates (factors) had an impact, the design effectively collapses into a design in the active factors with replications. But, replications in a computer experiment are non-informative because of the deterministic nature of the computer code and therefore would have been a waste of resources. Second, it could have been easy to miss an unknown effect by only experimenting at a few different points for each factor. For example, the exact nature of the nonlinear $x_4$ effect would have been difficult to identify with only four levels; in fact, the dramatic nonlinear behavior of $x_4$ surprised the engineers. Third, the decision of where to place the levels becomes much more crucial for the factorial design; lower dimensional projections of Latin hypercube design typically consist of $n$ distinct and spread-out points so that their exact position is less important. Finally, a $4^{6-2}$ fractional factorial design would have required 256 runs. Contrast this with the 100-run Latin hypercube design that was used; even fewer runs might have been sufficient.

Computer experiments typically use such space filling designs so the proposed method is particularly suited to computer experiments. While physical experiments typically collect much less data than computer experiments, in principle the proposed method can be applied to physical experiments by adding a random error term to the model.

# Acknowledgments

We thank S. Cao (now at the National Research Council of Canada) and T. Hollands, Solar Thermal Lab, Department of Mechanical Engineering, University of Waterloo, for

involving us in the solar collector design application. M. Hamada's research was supported by General Motors of Canada Limited, the Manufacturing Research Corporation of Ontario, and the Natural Sciences and Engineering Research Council of Canada. M. Schonlau and W. Welch's research was supported by the Natural Sciences and Engineering Research Council of Canada. We also thank M. Abt for comments.

# A The Derivation of the Main Effects Bounds

Suppose we want to plot the estimated effect of some of the $x$ variables, denoted by $\mathbf{x}_{\text{effect}}$. The remaining $x$ variables, denoted by $\mathbf{x}_{\text{out}}$ , have to be integrated out of the predictor. The estimated effect is

$$\hat{\mu}(\mathbf{x}_{\text{effect}}) = \frac{1}{V}\int \hat{Y}(\mathbf{x})d\mathbf{x}_{\text{out}}, \tag{4}$$

where $V$ is the volume of the $\mathbf{x}_{\text{out}}$ region over which we integrate. For example, in the first plot of Figure 5, $\mathbf{x}_{\text{effect}} = x_1$, and the plotting coordinates $\hat{\mu}(x_1)$ require an integration over $\mathbf{x}_{\text{out}} = (x_2, \ldots, x_6)^t$ for each value of $x_1$ plotted. The integral in (4) is easy to approximate if the $x$-space is cuboidal, and if the correlation function is a product of correlation functions for each $x$ variable. Here we show that the mean squared error of $\hat{\mu}$ is also fairly easy to compute.

Numerically, we approximate (4) by a sum over a grid of $m$ points $\mathbf{x}_{\text{out}}^{(1)}, \ldots, \mathbf{x}_{\text{out}}^{(m)}$ representing the $\mathbf{x}_{\text{out}}$ space. Thus (4) becomes

$$\hat{\mu}(\mathbf{x}_{\text{effect}}) = \frac{1}{m}\sum_{i=1}^{m} \hat{Y}(\mathbf{x}_{\text{effect}}, \mathbf{x}_{\text{out}}^{(i)}), \tag{5}$$

where we have decomposed the vector of $x$ variables as $\mathbf{x} = (\mathbf{x}_{\text{effect}}, \mathbf{x}_{\text{out}})$. The sum in (5) estimates the corresponding sum of true function values, with mean squared error

$$\begin{aligned} & E\left(\frac{1}{m}\sum_{i=1}^{m} \hat{Y}(\mathbf{x}_{\text{effect}}, \mathbf{x}_{\text{out}}^{(i)}) - \frac{1}{m}\sum_{i=1}^{m} Y(\mathbf{x}_{\text{effect}}, \mathbf{x}_{\text{out}}^{(i)})\right)^2 \\ = \; & E\left(\frac{1}{m}\sum_{i=1}^{m} \left(\hat{Y}(\mathbf{x}_{\text{effect}}, \mathbf{x}_{\text{out}}^{(i)}) - Y(\mathbf{x}_{\text{effect}}, \mathbf{x}_{\text{out}}^{(i)})\right)\right)^2 \end{aligned}$$

$$= \frac{1}{m^2}\sum_{i,i'} E\left(\left(\hat{Y}(\mathbf{x}_{\text{effect}}, \mathbf{x}_{\text{out}}^{(i)}) - Y(\mathbf{x}_{\text{effect}}, \mathbf{x}_{\text{out}}^{(i)})\right)\left(\hat{Y}(\mathbf{x}_{\text{effect}}, \mathbf{x}_{\text{out}}^{(i')}) - Y(\mathbf{x}_{\text{effect}}, \mathbf{x}_{\text{out}}^{(i')})\right)\right)$$

$$= \frac{1}{m^2}\sum_{i,i'} \text{Cov}\left(\hat{Y}(\mathbf{x}_{\text{effect}}, \mathbf{x}_{\text{out}}^{(i)}) - Y(\mathbf{x}_{\text{effect}}, \mathbf{x}_{\text{out}}^{(i)}), \hat{Y}(\mathbf{x}_{\text{effect}}, \mathbf{x}_{\text{out}}^{(i')}) - Y(\mathbf{x}_{\text{effect}}, \mathbf{x}_{\text{out}}^{(i')})\right). \quad (6)$$

Note that the last expression in (6) follows from unbiasedness, i.e., $E(\hat{Y}(\mathbf{x}) - Y(\mathbf{x})) = 0$.

In this appendix we generalize the model (1) by replacing the intercept term, $\beta$, by a regression model:

$$Y(\mathbf{x}) = \sum_{j=1}^{k} \beta_j f_j(\mathbf{x}) + Z(\mathbf{x}), \quad (7)$$

where $f(\mathbf{x}) = [f_1(\mathbf{x}), \ldots, f_k(\mathbf{x})]^t$ are $k$ known regression functions and $\boldsymbol{\beta} = (\beta_1, \beta_2, \ldots, \beta_k)^t$ are the corresponding unknown parameters.

Let $\mathbf{y} = (Y_1, Y_2, ..., Y_n)^t$ denote the vector of observations at the design points, and denote the best linear predictor of $Y$ at $\mathbf{x}$ by $\hat{Y}(\mathbf{x}) = \mathbf{c}_\mathbf{x}^t\mathbf{y}$. The coefficients $\mathbf{c}_\mathbf{x}$ can be determined through minimization of $E(\mathbf{c}_\mathbf{x}^t\mathbf{y} - Y(\mathbf{x}))^2$ subject to the unbiasedness constraint $\mathbf{F}^t\mathbf{c}_\mathbf{x} = \mathbf{f}(\mathbf{x})$, where $\mathbf{F}$ is the expanded design matrix (see for example Sacks, Schiller, and Welch, 1989):

$$\mathbf{c}_\mathbf{x} = \begin{pmatrix} \mathbf{0} \\ \mathbf{I} \end{pmatrix}^t \begin{pmatrix} \mathbf{0} & \mathbf{F}^t \\ \mathbf{F} & \sigma_z^2\mathbf{R} \end{pmatrix}^{-1} \begin{pmatrix} \mathbf{f}(\mathbf{x}) \\ \sigma_z^2\mathbf{r}(\mathbf{x}). \end{pmatrix} \quad (8)$$

The quantities $\mathbf{R}$ and $\mathbf{r}(\mathbf{x})$ were defined in (3).

We now calculate the covariance between $\hat{Y}(\mathbf{x}_1) - Y(\mathbf{x}_1)$ and $\hat{Y}(\mathbf{x}_2) - Y(\mathbf{x}_2)$ at any points $\mathbf{x}_1$ and $\mathbf{x}_2$ . Because $E(\hat{Y}(\mathbf{x}) - Y(\mathbf{x})) = 0$, we obtain

$$\begin{aligned} \text{Cov}(\hat{Y}(\mathbf{x}_1) - Y(\mathbf{x}_1), \hat{Y}(\mathbf{x}_2) - Y(\mathbf{x}_2)) &= \text{Cov}(\mathbf{c}_{\mathbf{x}_1}^t\mathbf{y} - Y(\mathbf{x}_1), \mathbf{c}_{\mathbf{x}_2}^t\mathbf{y} - Y(\mathbf{x}_2)) \\ &= \sigma_z^2\left[\mathbf{c}_{\mathbf{x}_1}^t\mathbf{R}\,\mathbf{c}_{\mathbf{x}_2} + R(\mathbf{x}_1, \mathbf{x}_2) - \mathbf{c}_{\mathbf{x}_1}^t\mathbf{r}(\mathbf{x}_2) - \mathbf{c}_{\mathbf{x}_2}^t\mathbf{r}(\mathbf{x}_1)\right], \end{aligned}$$

where $R(\mathbf{x}_1, \mathbf{x}_2)$ is the correlation between $Y(\mathbf{x}_1)$ and $Y(\mathbf{x}_2)$. Substituting $\mathbf{c}_{\mathbf{x}_1}$ and $\mathbf{c}_{\mathbf{x}_2}$ from (8) yields

$$\begin{aligned} &\text{Cov}(\hat{Y}(\mathbf{x}_1) - Y(\mathbf{x}_1), \hat{Y}(\mathbf{x}_2) - Y(\mathbf{x}_2)) = \\ &\quad \sigma_z^2(R(\mathbf{x}_1, \mathbf{x}_2) - \mathbf{r}(\mathbf{x}_1)^t\mathbf{R}^{-1}\mathbf{r}(\mathbf{x}_2) + \mathbf{f}(\mathbf{x}_1)^t\mathbf{K}^{-1}\mathbf{f}(\mathbf{x}_2) + \mathbf{r}(\mathbf{x}_1)^t\mathbf{R}^{-1}\mathbf{F}\mathbf{K}^{-1}\mathbf{F}^t\mathbf{R}^{-1}\mathbf{r}(\mathbf{x}_2) \\ &\quad -\mathbf{f}(\mathbf{x}_1)^t\mathbf{K}^{-1}\mathbf{F}^t\mathbf{R}^{-1}\mathbf{r}(\mathbf{x}_2) - \mathbf{f}(\mathbf{x}_2)^t\mathbf{K}^{-1}\mathbf{F}^t\mathbf{R}^{-1}\mathbf{r}(\mathbf{x}_1)), \end{aligned} \quad (9)$$

where $\mathbf{K} = \mathbf{F}^t\mathbf{R}^{-1}\mathbf{F}$.

It turns out equation (9) can also be written in product form:

$$\mathrm{Cov}(\hat{Y}(\mathbf{x}_1) - Y(\mathbf{x}_1), \hat{Y}(\mathbf{x}_2) - Y(\mathbf{x}_2)) = \sigma_z^2 \left[ R(\mathbf{x}_1, \mathbf{x}_2) - \begin{pmatrix} \mathbf{f}(\mathbf{x}_1) \\ \mathbf{r}(\mathbf{x}_1) \end{pmatrix}^t \begin{pmatrix} \mathbf{0} & \mathbf{F}^t \\ \mathbf{F} & \mathbf{R} \end{pmatrix}^{-1} \begin{pmatrix} \mathbf{f}(\mathbf{x}_2) \\ \mathbf{r}(\mathbf{x}_2) \end{pmatrix} \right]. \tag{10}$$

Substituting expression (9) into (6) gives

$$\begin{aligned} \mathrm{MSE}(\frac{1}{m}\sum_{i=1}^{m} \hat{Y}(\mathbf{x}_{\text{effect}}, \mathbf{x}_{\text{out}}^{(i)})) = \sigma_z^2 \Bigg[ & \frac{1}{m^2} \sum_{i,i'} R\left((\mathbf{x}_{\text{effect}}, \mathbf{x}_{\text{out}}^{(i)}), (\mathbf{x}_{\text{effect}}, \mathbf{x}_{\text{out}}^{(i')})\right) \\ & - \bar{\mathbf{r}}(\mathbf{x}_{\text{effect}})^t \mathbf{R}^{-1} \bar{\mathbf{r}}(\mathbf{x}_{\text{effect}}) + \bar{\mathbf{f}}(\mathbf{x}_{\text{effect}})^t \mathbf{K}^{-1} \bar{\mathbf{f}}(\mathbf{x}_{\text{effect}}) + \bar{\mathbf{r}}(\mathbf{x}_{\text{effect}})^t \mathbf{R}^{-1} \mathbf{F} \mathbf{K}^{-1} \mathbf{F}^t \mathbf{R}^{-1} \bar{\mathbf{r}}(\mathbf{x}_{\text{effect}}) \\ & - 2\bar{\mathbf{f}}(\mathbf{x}_{\text{effect}})^t \mathbf{K}^{-1} \mathbf{F}^t \mathbf{R}^{-1} \bar{\mathbf{r}}(\mathbf{x}_{\text{effect}}) \Bigg] \end{aligned} \tag{11}$$

where $\bar{\mathbf{r}}(\mathbf{x}_{\text{effect}}) = \frac{1}{m}\sum_i \mathbf{r}(\mathbf{x}_{\text{effect}}, \mathbf{x}_{\text{out}}^{(i)})$, and $\bar{\mathbf{f}}(\mathbf{x}_{\text{effect}}) = \frac{1}{m}\sum_j \mathbf{f}(\mathbf{x}_{\text{effect}}, \mathbf{x}_{\text{out}}^{(j)})$.

The expressions in equation (11) are easy to evaluate if

1. the correlation function $R(\mathbf{x}_1, \mathbf{x}_2)$ is a product of correlations in each x variable,

$$R(\mathbf{x}_1, \mathbf{x}_2) = \prod_j R_j(x_j^{(1)}, x_j^{(2)}),$$

where $x_j^{(1)}$ denotes the value of the $j^{th}$ $x$-variable for point $\mathbf{x}_1$, and

2. the points representing the space of the variables integrated out, $\mathbf{x}_{\text{out}}^{(1)}, \ldots, \mathbf{x}_{\text{out}}^{(m)}$, are a grid. Without loss of generality, suppose $\mathbf{x}_{\text{out}}$ is the first $q$ of the $x$ variables, $x_1, \ldots, x_q$, and the grid of $\mathbf{x}_{\text{out}}$ values is

$$\{x_1^{(1)}, \ldots, x_1^{(m_1)}\} \otimes \ldots \otimes \{x_q^{(1)}, \ldots, x_q^{(m_q)}\},$$

where $x_j^{(i)}$ is the $i^{th}$ grid value for variable $x_j$, and $\prod_{j=1}^{d} m_j = m$.

For example, if these conditions hold, the first term in (11) becomes

$$\begin{aligned} & \frac{1}{m^2} \sum_{i,i'} R\left((\mathbf{x}_{\text{effect}}, \mathbf{x}_{\text{out}}^{(i)}), (\mathbf{x}_{\text{effect}}, \mathbf{x}_{\text{out}}^{(i')})\right) = R(\mathbf{x}_{\text{effect}}, \mathbf{x}_{\text{effect}}) \frac{1}{m^2} \sum_{i,i'} R(\mathbf{x}_{\text{out}}^{(i)}, \mathbf{x}_{\text{out}}^{(i')}) \\ = \; & \frac{1}{m^2} \sum_{i,i'} R(\mathbf{x}_{\text{out}}^{(i)}, \mathbf{x}_{\text{out}}^{(i')}) = \frac{1}{m^2} \left( \sum_{i=1}^{m_1} \sum_{i'=1}^{m_1} R(x_1^{(i)}, x_1^{(i')}) \right) \ldots \left( \sum_{i=1}^{m_q} \sum_{i'=1}^{m_q} R(x_q^{(i)}, x_q^{(i')}) \right). \end{aligned} \tag{12}$$

All other terms in equation (11) are similar in that the leading and the closing factor of a product is either $\bar{\mathbf{f}}$ or $\bar{\mathbf{r}}$. The computation of $\bar{\mathbf{r}}$ is similar to (12). Their interior factors are constant with respect to the summing. Because we restrict the terms in $f(x)$ to be polynomials $x_1^{\alpha_1} \ldots x_d^{\alpha_d}$, i.e., a product in each of the $x$ variables, the averages $\bar{\mathbf{f}}$ are also simple to compute.

The mean squared error $\mathrm{MSE}(\hat{Y}(\mathbf{x}))$ for the error bounds for Method 2, where variables not of interest are fixed at their midranges, is given in Sacks, Schiller, and Welch (1989) and can also be obtained from equation (11) as a special case for $m = 1$ and $i = i'$.

# References


[1] Bates, D.M., and Watts, D.G. (1988), *Nonlinear Regression Analysis and its Applications*, New York: John Wiley & Sons.

[2] Breiman, L. (1991), Discussion of "Multivariate Adaptive Splines" by J.H. Friedman, *The Annals of Statistics*, 19, 1-141.

[3] Cao, S. (1993), *Numerical Investigation on Unglazed Transpired Plate Solar Collector*, unpublished M.Eng. thesis, Department of Mechanical Engineering, University of Waterloo.

[4] Currin, C., Mitchell, T., Morris, M., and Ylvisaker, D. (1991), "Bayesian Prediction of Deterministic Functions, With Applications to the Design and Analysis of Computer Experiments," *JASA*, 86, 953-963.

[5] Golneshan, A. (1994), *Forced Convection Heat Transfer from Low Porosity Slotted Transpired Plates*, unpublished Ph.D. thesis, Department of Mechanical Engineering, University of Waterloo.

[6] McCullagh, P. and Nelder, J.A. (1989), *Generalized Linear Models* (2nd ed.), London: Chapman and Hall.

[7] McKay, M.D., Beckman, R.J., and Conover, W.J. (1979), "A Comparison of Three Methods for Selecting Values of Input Variables in the Analysis of Output from a Computer Code," *Technometrics*, 21, 239–245.

[8] Montgomery, D.C., and Peck, E.A. (1982) *Introduction to Linear Regression Analysis*, New York: John Wiley & Sons.

[9] Sacks, J., Schiller, S.B., and Welch, W.J. (1989), "Designs for Computer Experiments," *Technometrics*, 31, 41–47.

[10] Sacks, J., Welch, W.J., Mitchell, T.J., and Wynn, H.P. (1989), "Design and Analysis of Computer Experiments," *Statistical Science*, 4, 409–435.

[11] Welch, W.J., Buck, R.J., Sacks, J., Wynn, H.P. , Mitchell, T.J., and Morris, M.D. (1992), "Screening, Predicting, and Computer Experiments," *Technometrics*, 34, 15–25.

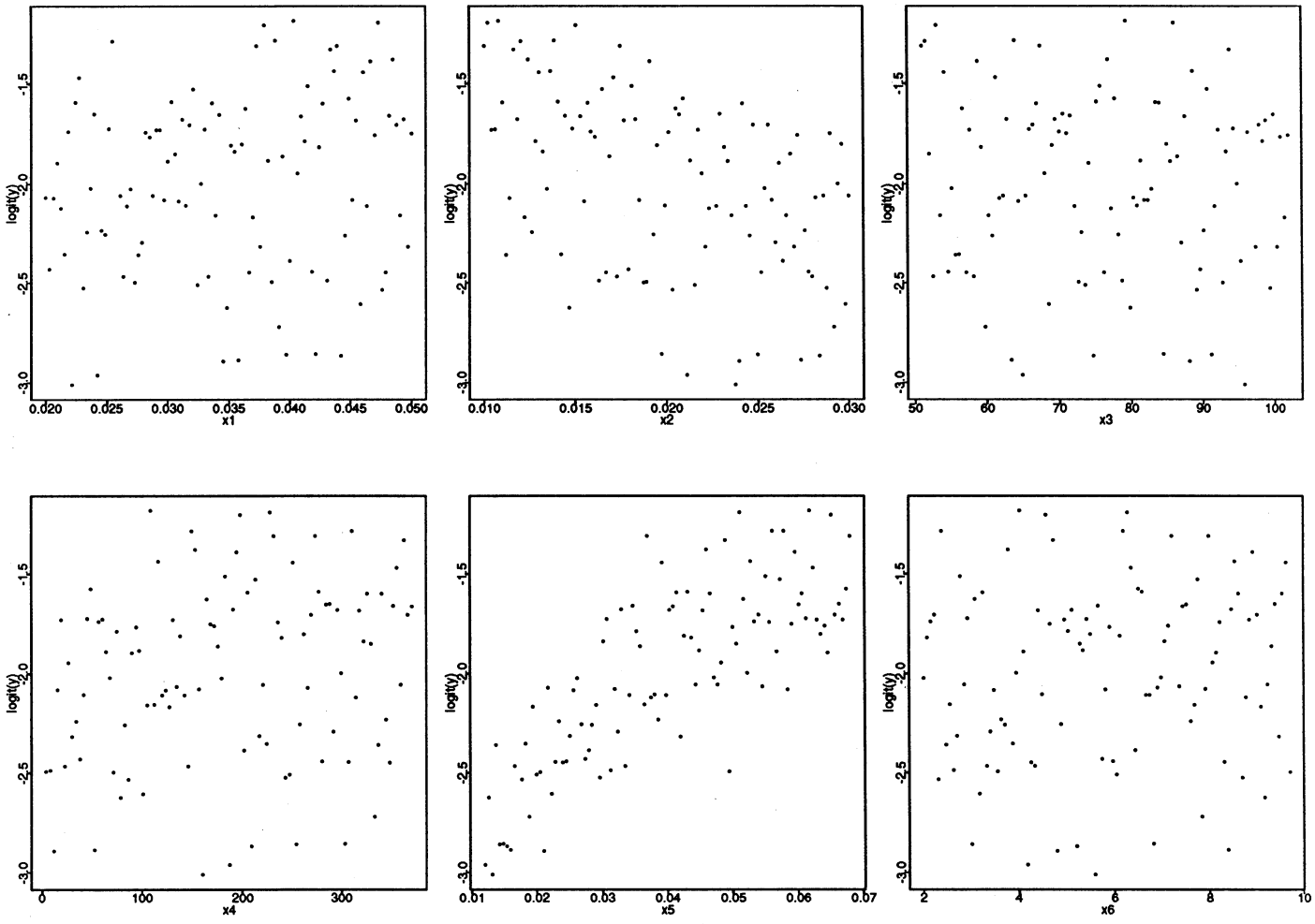


Figure 1: Scatter plots of logit(y) versus $x_i$.

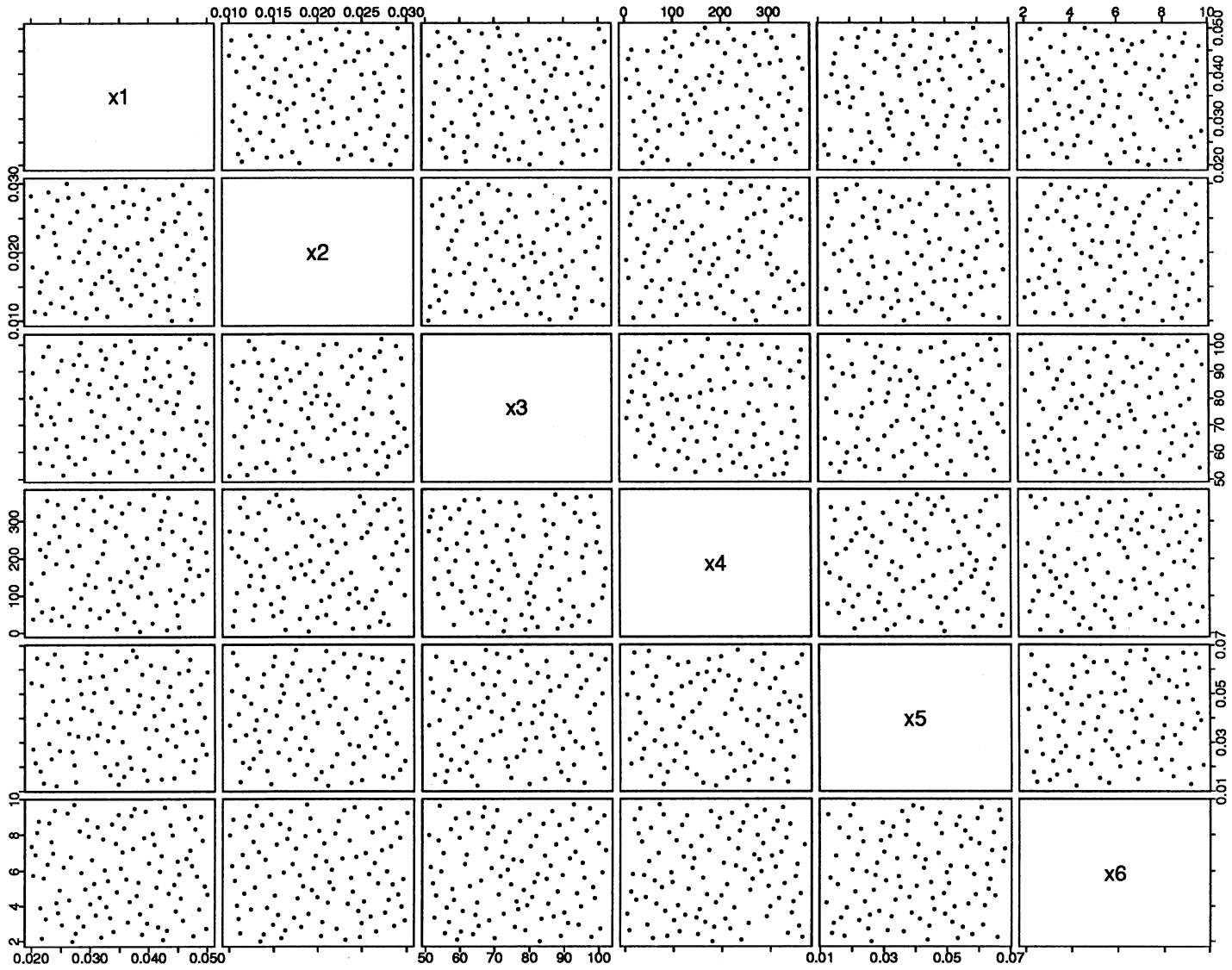


Figure 2: Two dimensional projections of the Latin hypercube design.

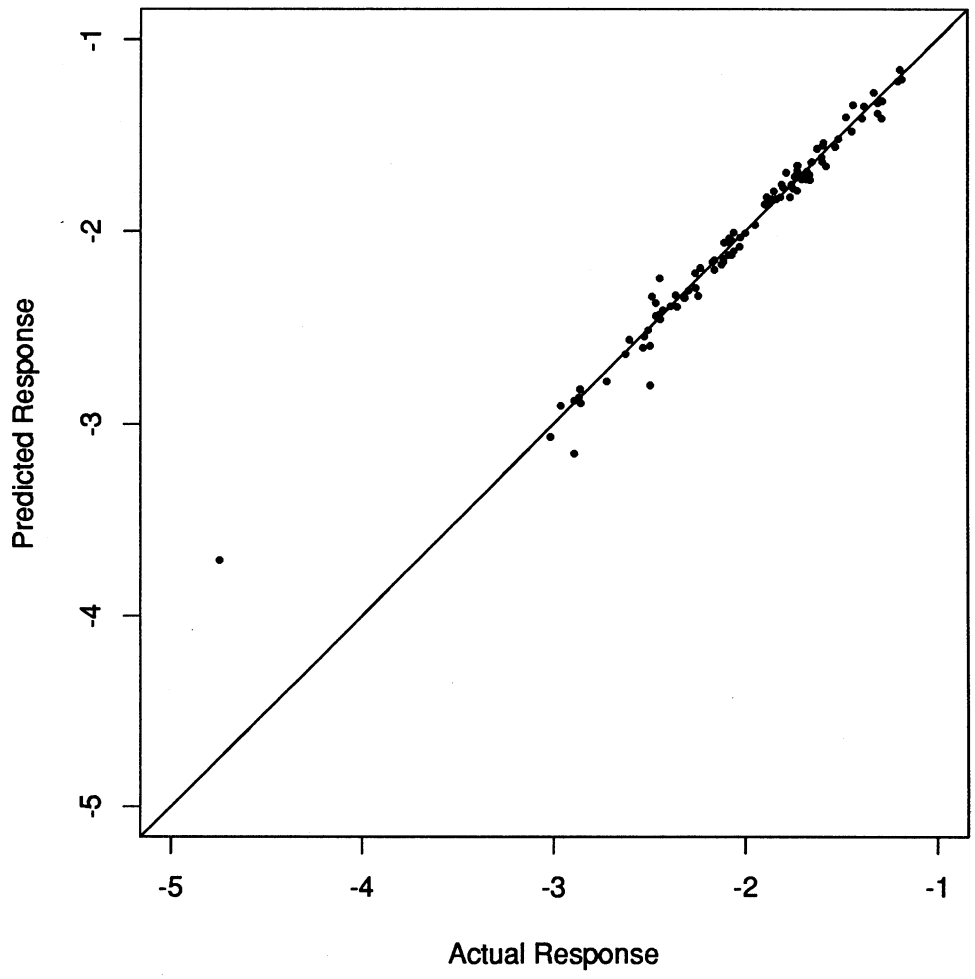


Figure 3: Cross validation predictions from the nonparametric model ($n = 100$ data points). The line Predicted Response = Actual Response is shown.

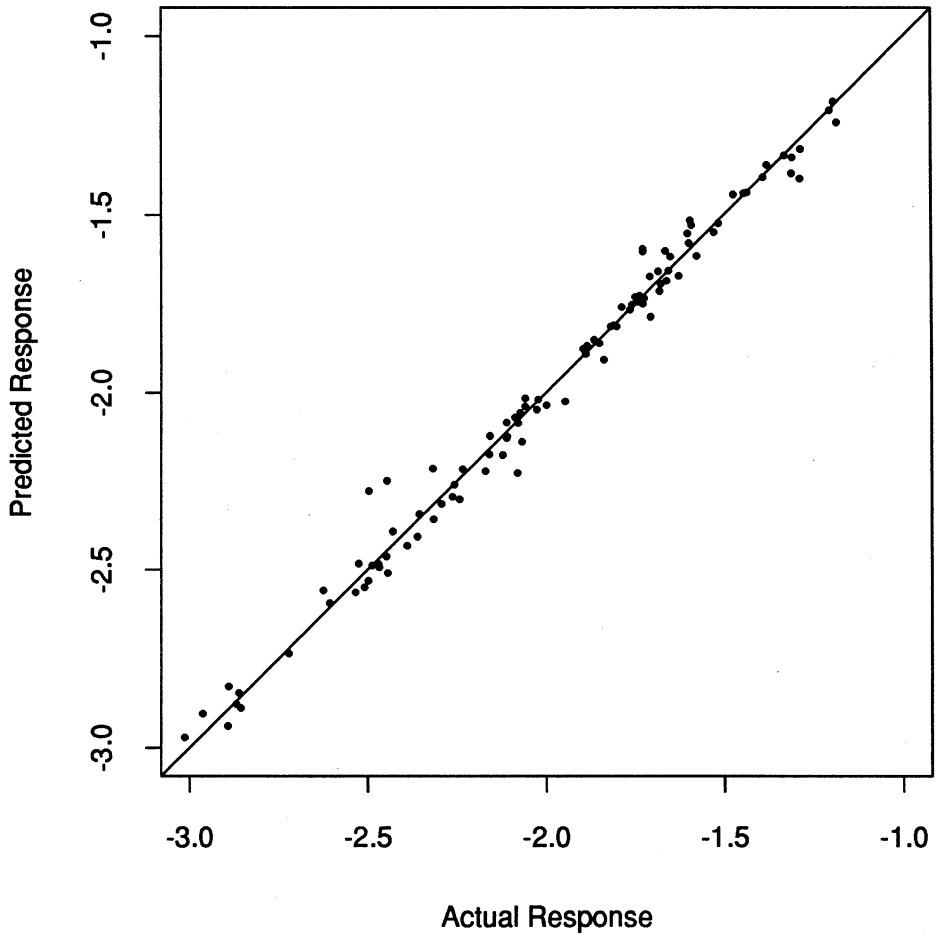


Figure 4: Cross validation predictions from the nonparametric model ($n = 99$ data points).

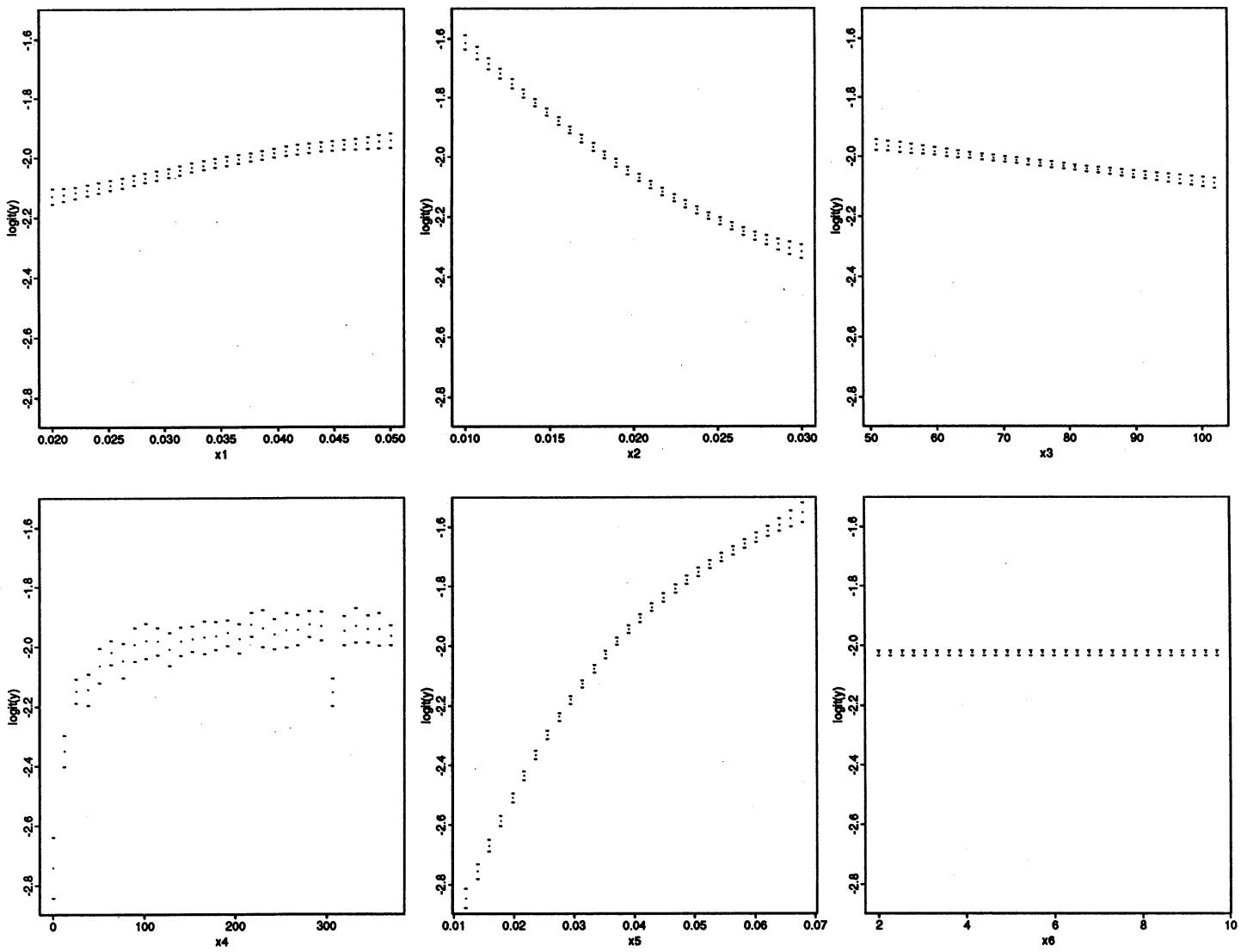


Figure 5: Main effect plots. The middle line is the estimated effect, the upper and lower lines are approximate 95% pointwise confidence limits based on the standard error derived in Appendix A.

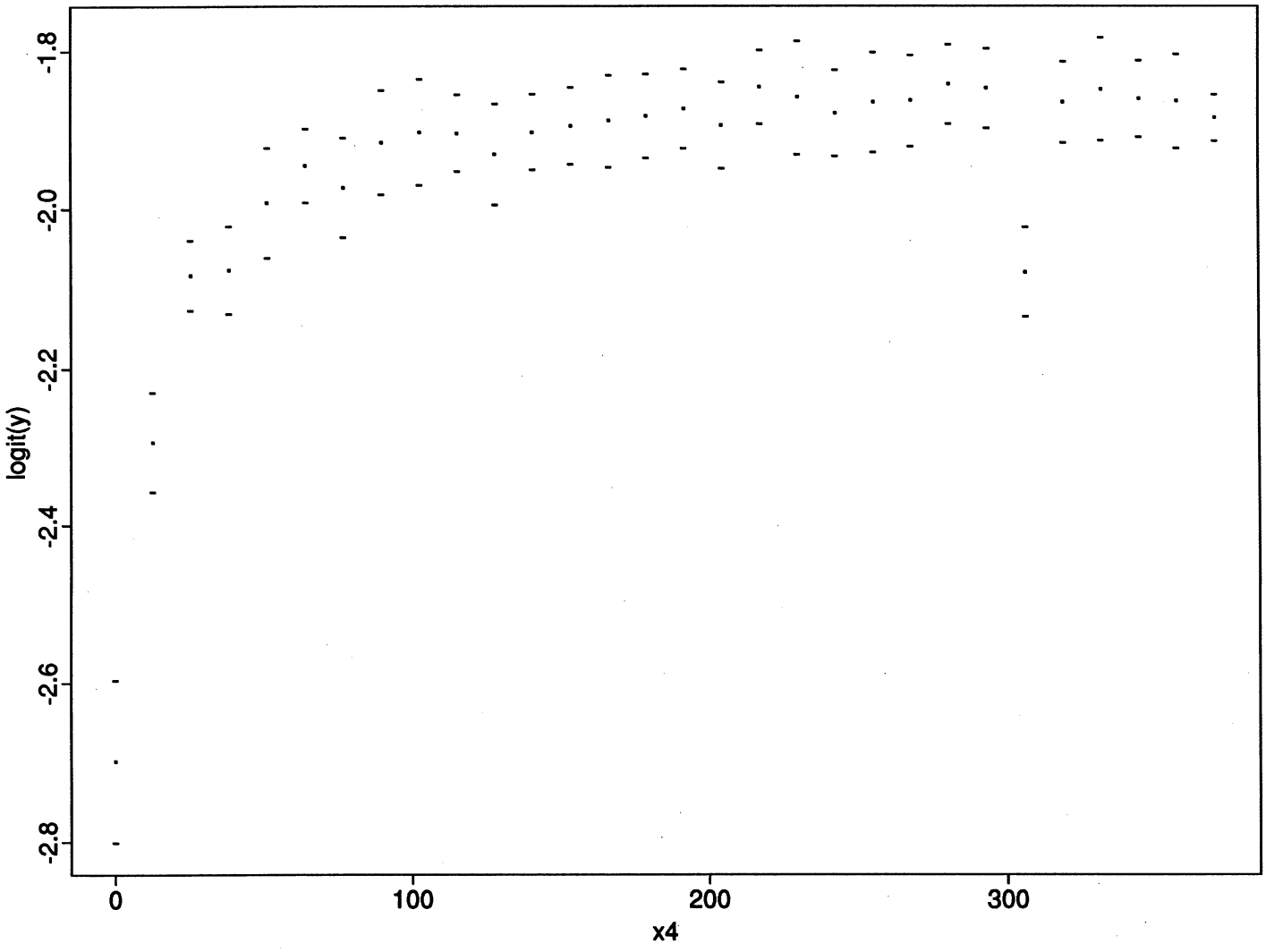


Figure 6: Predicted $\hat{Y}(\mathbf{x})$ versus $x_4$, keeping all other $x$ variables fixed at their midranges (Method 2). The middle line is the estimated effect, the upper and lower lines are approximate 95% pointwise confidence limits.

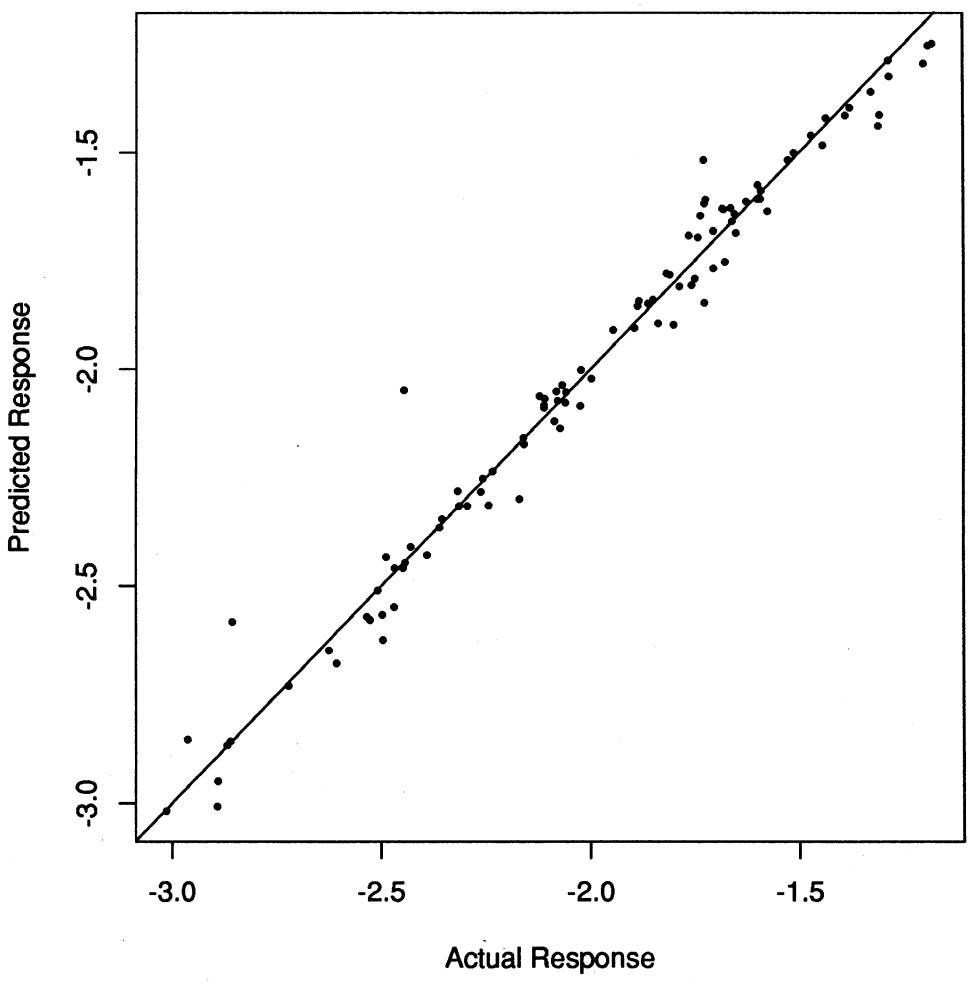


Figure 7: Cross validation predictions from the parametric nonlinear model.

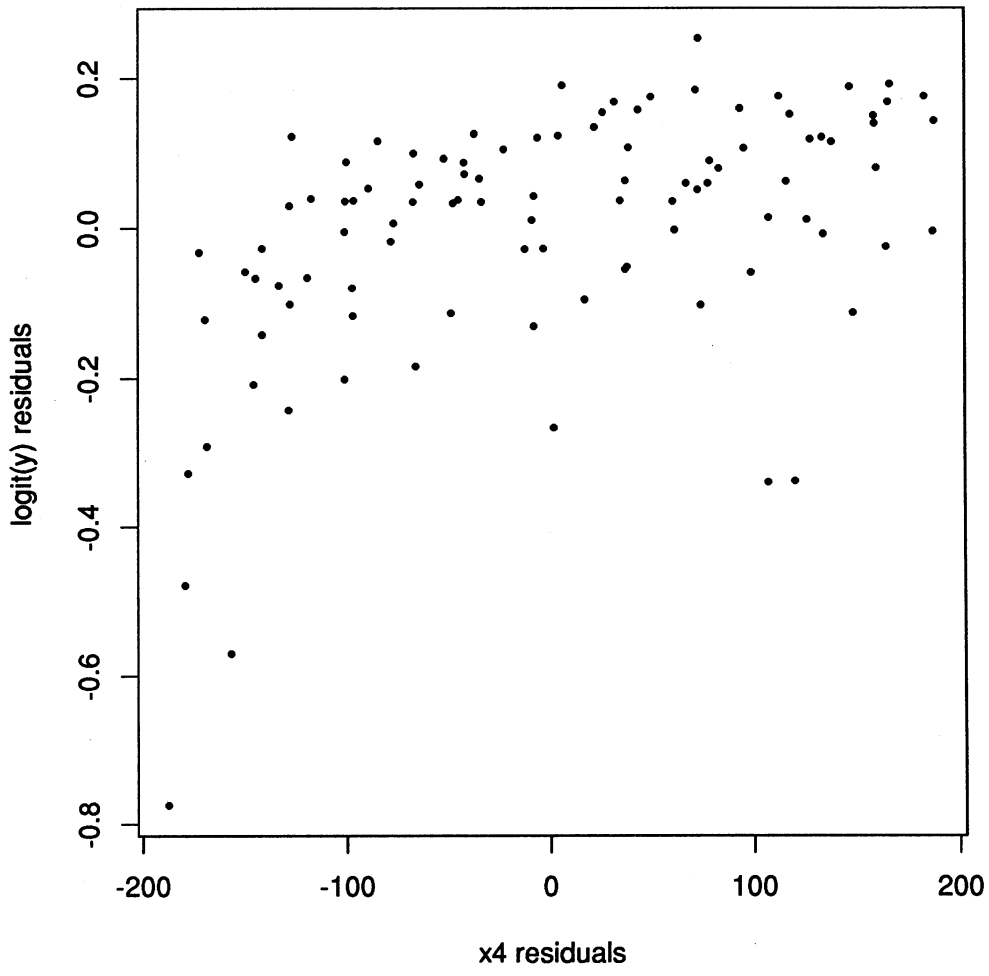


Figure 8: Added variable plot for $x_4$.