Bayesian Optimal Designs for Fitting Fractional Polynomial Response Surface Models

Steven Gilmour

Latest Advances in the Theory and Applications
of the Design and Analysis of Experiments
Banff Workshop
6-11 August 2017

Acknowledgements

Joint work with Luzia Trinca (UNESP-Botucatu, Brazil).

Partially funded by the Royal Society, EPSRC, Fapesp and CNPq.

Outline

1. Background
2. Motivation
3. Fractional polynomial models
4. Analysis
5. Design

Response Surfaces

Given a response variable y and explanatory variables $x_1, \ldots, x_q$, a response surface is a function describing the relationship between the expectation of Y and $\mathbf{x}$,

$$E(Y|\mathbf{x}) = \eta(\mathbf{x}).$$

Note:

- Assume $Y \overset{ind.}{\sim} N(\eta(\mathbf{x}), \sigma^2)$, although distributional assumption is not essential;
- Can be generalised, so that $\eta(\mathbf{x})$ is the linear predictor in a GLM or other model;
- We are interested in response surface experimentation.

Motivation

First and second order polynomial regression models are widely, and often successfully, used as empirical response surface models in small- to moderate-sized experiments.

However:

- second order model may not fit the data and higher order models are difficult to interpret and have multiple stationary points;
- a factor may be known to have a monotonic, but not linear, effect.

Example 1

An experiment to optimise a bioreactor.

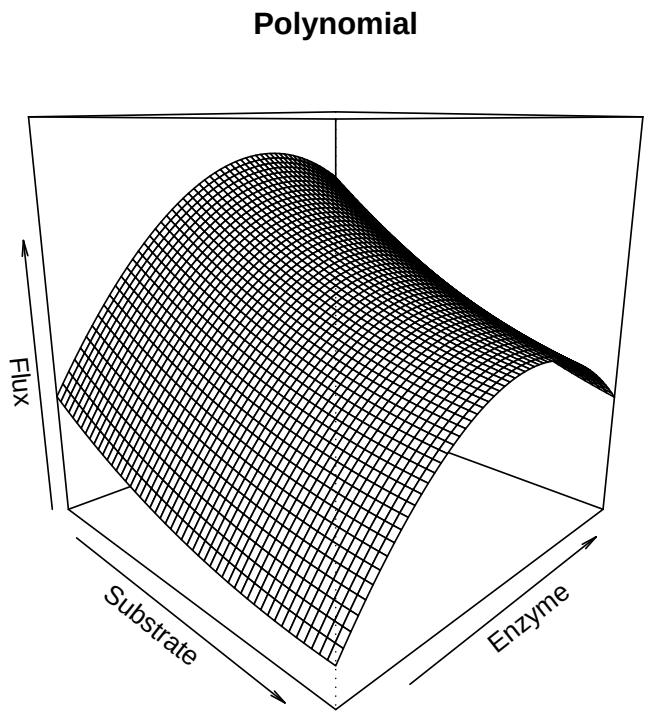
Factors: substrate concentration, enzyme concentration, pressure.

Design: three-level central composite with four centre points. One run lost, so only 17 runs.

Response: flux.

Fitted model: second order polynomial, plus Substrate$\times$Pressure^2 term.

Fitted Polynomial Model



Approximating Models

Taking a Taylor series expansion of $\eta(\mathbf{x})$ and truncating gives a low order polynomial, $f(\mathbf{x})$, as a local approximation, e.g.

$$f(\mathbf{x}) = \beta_0 + \sum_{i=1}^{q} \beta_i x_i + \sum_{i=1}^{q} \beta_{ii} x_i^2 + \sum_{i=1}^{q-1} \sum_{j=i+1}^{q} \beta_{ij} x_i x_j.$$

This argument applies equally if x_i is measured in a different metric, e.g. pressure is measured in units of $Pa = N/m^2$. It could equally well be measured in $\sqrt{N}/m$, m^2/N, $m/\sqrt{N}$, etc.

Hence, it is equally justifiable to use a polynomial in $\sqrt{x_i}$, $\frac{1}{x_i}$, $\frac{1}{\sqrt{x_i}}$, etc.

Box-Tidwell Transformations

Box and Tidwell (1962) suggested using power transformations,

$$x_i^{(\alpha_i)} = \begin{cases} x_i^{\alpha_i}, & \alpha_i \neq 0; \\ \log x_i, & \alpha_i = 0, \end{cases}$$

and then using polynomials in $x_i^{(\alpha_i)}$.

Fractional Polynomials

One extra parameter for each factor allows:

- models with asymptotes;
- nonlinear monotonic relationships;
- asymmetrical relationships with a stationary point.

It is often useful to restrict powers to a few rational numbers, e.g.

$$\left\{-3, -2, -1, -\frac{1}{2}, -\frac{1}{3}, 0, \frac{1}{3}, \frac{1}{2}, 1, 2, 3\right\}.$$

Royston and Altman (1994) called these fractional polynomials.

Models

We can use polynomials in $x_i^{(\alpha_i)}, i = 1, \ldots, q$ as the default class of models, e.g. the second order fractional polynomial,

$$f(\mathbf{x}) = \beta_0 + \sum_{i=1}^{q} \beta_i x_i^{(\alpha_i)} + \sum_{i=1}^{q} \beta_{ii} \left\{ x_i^{(\alpha_i)} \right\}^2 + \sum_{i=1}^{q-1} \sum_{j=i+1}^{q} \beta_{ij} x_i^{(\alpha_i)} x_j^{(\alpha_j)}.$$

Prior knowledge may dictate that we should restrict the set further, e.g. if it is known that the response to some factor will be monotonic or asymptotic.

Estimation

We have had few problems with convergence using nonlinear least squares with a partial linear algorithm (e.g. in R or GenStat) exploiting the fact that only the power parameters appear nonlinearly.

We round the powers by trying the rational numbers within our set closest to the values obtained by nonlinear least squares (NLLS).

If estimates of the power parameters are outside the range $[-3, 3]$, we can use the algorithm given by Royston and Altman: set $\alpha_2, \ldots, \alpha_q$ to initial values, find the value of α_1 in our set that minimises Residual MS; cycle through the other parameters; repeat until convergence.

Estimation

A modification of this is to estimate α_i at each stage by (one-dimensional) NLLS. Another is to fix one parameter and evaluate the others by NLLS, repeating at a grid of points.

Performing model simplification before rounding (e.g. dropping some second order terms) often reduces the problem of unstable parameter estimates.

Inference

There are technical problems in performing inferences with the rounded powers - in particular, the χ^2 approximation to the deviance is invalid.

In response surface experiments, we recommend using the unrounded powers to perform inferences and using the rounded powers just for interpretation of the fitted model, which is often more important anyway.

Example 1 (ctd.)

An experiment to optimise a bioreactor.

Factors: substrate concentration, enzyme concentration, pressure.

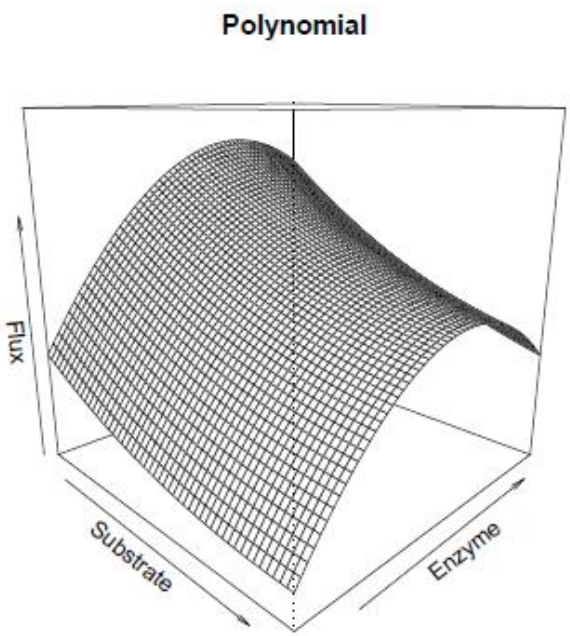
Design: three-level central composite with four centre points. One run lost, so only 17 runs.

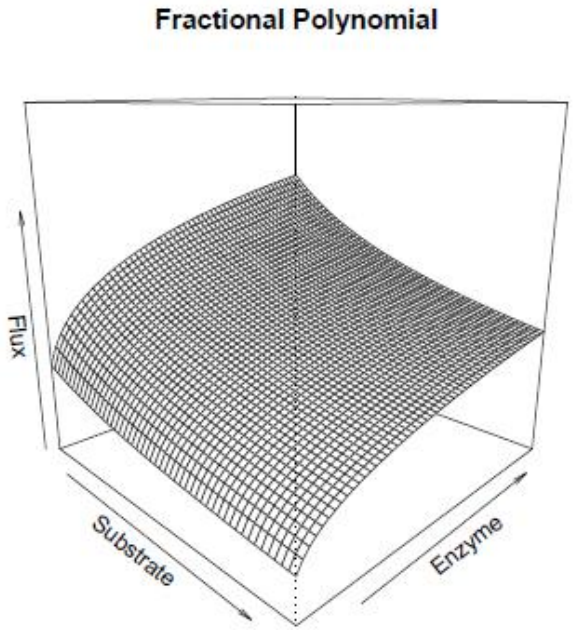
Response: flux.

Fitted polynomial model: second order polynomial, plus Substrate$\times$Pressure^2 term (9 parameters).

Fitted fractional polynomial model: first order in substrate concentration and enzyme concentration, plus Substrate$\times$Enzyme term (6 parameters).

Fitted Models





Example 2

Data from an experiment on spacing and density effects on turnip yields reported by Mead (1988) - 4×5 treatments in 3 complete blocks.

Models fitted:

1. Full model:

$$f(\mathbf{x}) = \beta_0 + \delta_i + \tau_j; i = 1, 2, 3; j = 1, \ldots, 20,$$

   where $\sum_{i=1}^{3} \delta_i = 0$ and $\sum_{j=1}^{20} \tau_j = 0$.

2. Second order polynomial:

$$f(\mathbf{x}) = \beta_0 + \delta_i + \beta_1 x_{1ij} + \beta_2 x_{2ij} + \beta_{11} x_{1ij}^2 + \beta_{22} x_{2ij}^2 + \beta_{12} x_{1ij} x_{2ij}.$$

3. Highest possible order polynomial, with non-significant terms dropped:

$$\begin{aligned} f(\mathbf{x}) \quad = \quad & \beta_0 + \delta_i + \beta_1 x_{1ij} + \beta_2 x_{2ij} + \beta_{11} x_{1ij}^2 + \beta_{22} x_{2ij}^2 + \beta_{12} x_{1ij} x_{2ij} \\ & + \beta_{111} x_{1ij}^3 + \beta_{222} x_{2ij}^3 + \beta_{122} x_{1ij} x_{2ij}^2 + \beta_{2222} x_{2ij}^4. \end{aligned}$$

Example 2

4. Second order fractional polynomial:

$$\begin{aligned} f(\mathbf{x}) &= \beta_0 + \delta_i + \beta_1 x_{1ij}^{(\alpha_1)} + \beta_2 x_{2ij}^{(\alpha_2)} \\ &\quad + \beta_{11}\left\{x_{1ij}^{(\alpha_1)}\right\}^2 + \beta_{22}\left\{x_{2ij}^{(\alpha_2)}\right\}^2 + \beta_{12} x_{1ij}^{(\alpha_1)} x_{2ij}^{(\alpha_2)}. \end{aligned}$$

5. Second order fractional polynomial, with non-significant term dropped:

$$f(\mathbf{x}) = \beta_0 + \delta_i + \beta_1 x_{1ij}^{(\alpha_1)} + \beta_2 x_{2ij}^{(\alpha_2)} + \beta_{11}\left\{x_{1ij}^{(\alpha_1)}\right\}^2 + \beta_{12} x_{1ij}^{(\alpha_1)} x_{2ij}^{(\alpha_2)}.$$

6. Model 5 with powers rounded ($\hat{\alpha}_1 = -\frac{1}{2}$, $\hat{\alpha}_2 = -\frac{1}{3}$).

Comparison of Models

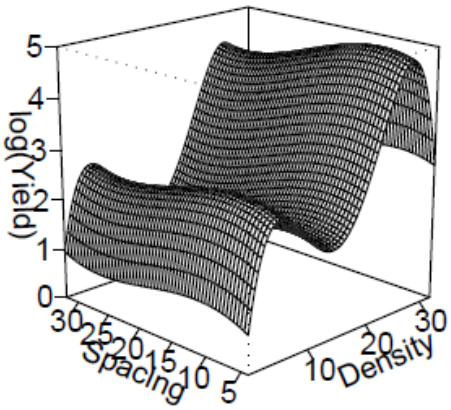
Model	SS_{Res}	df	MS_{Res}
1	1.4884	38	0.0392
2	7.5639	52	0.1455
3	1.8252	48	0.0380
4	1.7545	50	0.0351
5	1.7848	51	0.0350
6	1.7958	51	0.0352

Model 2 shows overwhelming evidence of lack of fit.

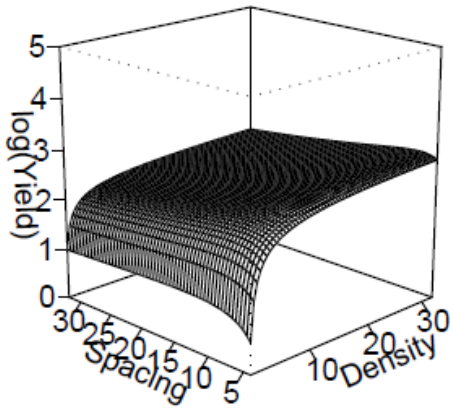
Models 3 and 6 look equally good until we consider interpretability.

Fitted Models

Polynomial



Fractional Polynomial



Comments

Fractional polynomials can and should be used routinely in experiments instead of polynomials for modelling response surfaces, . . .

. . . if more than a very approximate idea of the shape of the response surface is required.

They are not likely to be useful in situations where we would not contemplate fitting polynomials.

A problem is that standard response surface designs are very poor for fitting fractional polynomials.

Design

For computational and presentational simplicity, we present results for one factor for the second order model,

$$f(x) = \beta_0 + \beta_1 x^{(\alpha)} + \beta_{11} \left\{ x^{(\alpha)} \right\}^2.$$

Fractional polynomial models are nonlinear, so the optimal design depends on the unknown parameter values.

Unlike the first order model, the second order model is not partially nonlinear $\Rightarrow$ the optimal design depends on β_1 and β_{11}, as well as α.

Designs are very sensitive to the prior value of α, so pseudo-Bayesian optimal design seems natural.

Reparameterization

Eliciting priors and finding optimal designs are much easier if priors for different parameters are independent.

In fractional polynomial models, the interpretation of the β parameters depends on the value of α.

We reparameterize to more reasonably assume prior independence.

Define parameters which measure:

- the difference in response between x_{min} and x_{max} ($= 1$ after scaling); and
- the difference between the average response at the extremes and in the (transformed) centre.

Reparameterized Model

We get

$$\gamma_1 = \beta_1 \left\{1 - x_{min}^{(\alpha)}\right\} + \beta_{11} \left[1 - \left\{x_{min}^{(\alpha)}\right\}^2\right]$$

and

$$\gamma_{11} = \beta_{11} \frac{\left\{1 - x_{min}^{(\alpha)}\right\}^2}{4}.$$

The model becomes

$$f(x) = \beta_0 + \gamma_1 \frac{x^{(\alpha)}}{1 - x_{min}^{(\alpha)}} + 4\gamma_{11} \frac{\left\{x^{(\alpha)}\right\}^2 - \left\{1 + x_{min}^{(\alpha)}\right\} x^{(\alpha)}}{\left\{1 - x_{min}^{(\alpha)}\right\}^2}.$$

Another complication

Normal priors for γ_1 and γ_{11} seem natural, but we have to be aware of a singularity.

If $\gamma_1 = \gamma_{11} = 0$, any value of α gives an identical fit and $V(\hat{\alpha}) \to \infty$.

We get round this by truncating the priors for γ_1 and γ_{11} on both sides of zero. (This makes it difficult to use quadrature methods.)

Comment: in pseudo-Bayesian design, "priors" should not represent strength of belief in different parameter values, but strength of interest in being optimal for different parameter values.

Optimality Criteria

As usual, we obtain the asymptotic variance matrix, $\sigma^2 \mathbf{M}(\mathbf{X}, \boldsymbol{\theta})^{-1}$, where $\mathbf{M}(\mathbf{X}, \boldsymbol{\theta})^{-1} =$ messy, but computable.

We illustrate using weighted-A-optimality, with appropriate weights on each parameter.

Exact optimal design

Continuous optimal designs can be found and rounded.

These work well when $n >> p$, e.g. most cases with one or two factors.

Otherwise, we need exact designs.

We use a modified Fedorov exchange algorithm, with sampling from the truncated priors and a fine candidate grid.

Results

Design space: $x \in [0.1, 1]$, $n = 20$.

Priors: $\gamma_1 \sim N(1, 0.25)$, $\gamma_{11} \sim N(-2.5, 2.25)$,
$p(\alpha) = (0.15,\ 0.25,\ 0.25,\ 0.15,\ 0.10,\ 0.07,\ 0.03)$ for
$\alpha = (-2,\ -1,\ -1/2,\ 0,\ 1/2,\ 1,\ 2)$.

Design type	*Design*						Efficiency
True Prior	x_i	0.1	0.1400	0.1742	0.5100	0.6200	1	100
	n_i	3	1	5	6	1	4	
Point Prior $\alpha = 0$	x_i	0.1	0.1507	0.6073	1			86.25
	n_i	2	7	7	4			
Point Prior $\alpha = -1/2$	x_i	0.1	0.1532	0.4492	1			60.86
	n_i	6	4	4	6			

Results

Design type		*Design*					Efficiency
5 levels	x_i	0.1	01778	0.3162	0.5623	1	78.57
$\alpha = 0$	n_i	4	4	4	4	4	
5 levels	x_i	0.1	0.1455	0.2309	0.4213	1	69.57
$\alpha = -1/2$	n_i	4	4	4	4	4	
4 levels	x_i	0.1	0.2154	0.4642	1		65.32
$\alpha = 0$	n_i	5	5	5	5		
4 levels	x_i	0.1	0.1678	0.3377	1		49.71
$\alpha = -1/2$	n_i	5	5	5	5		

Conclusions

- Bayesian exact design is feasible, at least for a few factors, and necessary.
- For many factors we require further improvements (or simplification, e.g. point priors for γ parameters).
- Fractional polynomial response surfaces are very promising for many practical problems.

References

Box, G. E. P. and Tidwell, P. W. (1962) Transformation of the independent variables. *Technometrics*, **4**, 531-550.

Gilmour, S. G. and Trinca, L. A. (2005) Fractional polynomial response surface models. *Journal of Agricultural, Biological and Environmental Statistics*, **10**, 50-60.

Royston, P. and Altman, D. G. (1997) Approximating statistical functions by using fractional polynomial regression. *The Statistician*, **46**, 411-422.