619 Project
Bayesian Approach to Nonlinear Regression Models

Beijia Hu, Chenhui Hao, Xin Chen

Abstract

In this project, we constructed one nonlinear model for Langmuir equation and a competitive model to fit the adsorption data and then used Maximum likelihood estimation (MLE) and Bayesian approach to estimate all parameters of each model. We developed several model determination tools based on conditional predictive ordinates (CPO) and predictive distribution to compare these two nonlinear models. Sampling-based methods are used to carry out these diagnostic tools. All of results we got are almost the same as the first example in our reference paper[1].

Key words: Nonlinear model, Langmuir equation, model determination

1 Introduction

Langmuir equation is commonly used in pharmacokinetic models. In this report, we took logarithm on both sides of this equation in order to transform it into a linear form to fit the adsorption data. And we also constructed another competitive model which fits the data pretty well by least square method.

We proposed a general model-determination strategy based on Bayesian methods to compare these two nonlinear models. Model determination mainly contains two components: model assessment or model checking and model choice or model selection. The first component is to check if a model is adequate or not while the second component is to choose the most appropriate model among a collection of models. Here, we used predictive interval to check model adequacy and deviance, conditional predictive ordinate and pseudo-Bayes factor to do model selection.

The reason why we utilized pseudo-Bayes factor instead of Bayes factor is that there is a problem of interpretation and calibration under a noninformative prior specification for Bayes factor. Therefore, we are restricted to the pseudo-Bayes factor approach which is based on a cross-validated predictive density.

The paper is organized as follows. In Section 2, we introduced the predictive distribution, the computation of Monte Carlo estimates of conditional predictive ordinates (CPO) and sampling/importance resampling (SIR) method, a kind of Monte Carlo method, to generate predictive samples from predictive distribution. In Section 3, we described tools for model determination including some graphical methods for model checking and selection. In Section 4, we illustrated the procedure of parameter estimates by MLE and Bayesian approach and model determination for nonlinear models via an example about adsorption data. Finally, some conclusions are presented in Section 5.

2 Predictive Distributions

Consider a $n \times 1$ data vector $\mathbf{Y}$ with the $n \times k$ matrix of explanatory variables $\mathbf{X}$. Let $f(\mathbf{Y}|\theta, \mathbf{X})$ be the joint density of the data given a vector of parameter θ and $\pi(\theta)$ be the prior density of θ. The definition of the predictive density is

$$f(\mathbf{Y}) = \int f(\mathbf{Y}|\theta, \mathbf{X})\pi(\theta)d\theta,$$

We define Y_r be the rth data point and $\mathbf{Y}_{(r)}$ be the remaining data points, i.e, $\mathbf{Y}_{(r)} = (Y_1, ..., Y_{r-1}, Y_{r+1}, ..., Y_n)^T$. Similarly, the definition of the marginal density of Y_r given the remaining is

$$f(Y_r|\mathbf{Y}_{(r)}) = \frac{f(\mathbf{Y})}{f(\mathbf{Y}_{(r)})} = \int f(Y_r|\theta, \mathbf{X}, \mathbf{Y}_{(r)})\pi(\theta|\mathbf{Y}_{(r)})d\theta,$$

This approach is called the cross-validation approach, where $f(Y_r|\mathbf{Y}_{(r)})$ is called the cross-validation predictive density.
The cross-validation predictive density is to be checked against y_r. This means if model holds, y_r may be viewed as a random observation from the cross-validation predictive density. So we consider a checking function $g(Y_r; y_r)$. The expectation of $g(Y_r; y_r)$ under $f(Y_r|\mathbf{y}_{(r)})$ can be calculated and be denoted by d_r. One possible choice of the checking function is

$$g_\epsilon(Y_r; y_r) = \frac{1}{2\epsilon} I_{C_r(\epsilon)}(Y_r),$$

where y_r is a random observation near the data point Y_r, $C_r(\epsilon) = \{Y_r : y_r - \epsilon \leq Y_r \leq y_r + \epsilon\}$ is an interval with 2ϵ length and $I_{C_r(\epsilon)}$ denotes the indicator function of this interval. So this checking function is equivalent to the indicator function of the interval divided by the interval length. By the definition of d_r,

$$d_r(\epsilon) = E[g_\epsilon(Y_r; y_r)|\mathbf{y}_{(r)}] = \frac{1}{2\epsilon} P(C_r(\epsilon)|Y_r).$$

When ϵ is close to 0, Y_r is close to y_r under the interval $C_r(\epsilon)$. Hence

$$\lim_{x \to 0} C_r(\epsilon) = y_r,$$
$$\lim_{x \to 0} d_r(\epsilon) = d_r = f(y_r|\mathbf{y}_{(r)}).$$

The quantity of the equation above is called the conditional predictive ordinate (CPO).
To calculate the estimates of CPO, we attempt three methods. Given the checking function $g(Y_r; y_r)$, by the definition of expectation,

$$d_r = E[g(Y_r; y_r)|\mathbf{y}_{(r)}] = \int \int g(Y_r; y_r) f(Y_r|\theta, \mathbf{X}, \mathbf{y}_{(r)})\pi(\theta|\mathbf{y}_{(r)})d\theta dY_r.$$

We find this formula involves a multidimensional integral and it is impossible to find an analytical expression for d_r. To solve this problem, we use Monte Carlo integration, which is our first method.
If $(\theta_s, Y_{rs}), s = 1, ..., B$ are samples from the joint conditional distribution for θ and Y_r, $f(Y_r|\theta, \mathbf{X}, \mathbf{y}_{(r)})\pi(\theta|\mathbf{y}_{(r)})$, then a Monte Carlo estimate of d_r is

$$\hat{d_r} = \frac{1}{B} \sum g(Y_{rs}; y_r).$$

However, sampling from $\pi(\theta|\mathbf{y}_{(r)})$ is not a easy task. Hence we consider the second method, importance sampling to compute CPO estimates.
If $h(\theta)$ is an importance sampling density for $\pi(\theta|\mathbf{y}_{(r)})$ and $\theta_s, s = 1, ..., B$ are drawn from $h(\theta)$, the estimate of d_r is

$$\hat{d_r} = \sum g(Y_{rs}; y_r) w_s,$$

where

$$w_s = \frac{\pi(\theta_s|\mathbf{y}_{(r)})/h(\theta_s)}{\sum \pi(\theta_s|\mathbf{y}_{(r)})/h(\theta_s)}, s = 1, ..., B.$$

The third method is to estimate the CPO value itself, i.e, calculate $f(y_r|\mathbf{y}_{(r)})$. We observe that

$$\begin{aligned} f(y_r|\mathbf{y}_{(r)}) &= \frac{f(\mathbf{y})}{f(\mathbf{y}_{(r)})} = \frac{\int f(\mathbf{y}|\theta)\pi(\theta)d\theta}{\int f(\mathbf{y}_{(r)}|\theta)\pi(\theta)d\theta} \\ &= \frac{\int \frac{f(\mathbf{y}|\theta)\pi(\theta)}{\pi(\theta|\mathbf{y})f(\mathbf{y})}\pi(\theta|\mathbf{y})d\theta}{\int \frac{f(\mathbf{y}_{(r)}|\theta)\pi(\theta)}{\pi(\theta|\mathbf{y})f(\mathbf{y})}\pi(\theta|\mathbf{y})d\theta} \\ &= \frac{\frac{\int f(\mathbf{y}|\theta)\pi(\theta)d\theta}{f(\mathbf{y})}}{\int \frac{f(\mathbf{y}_{(r)},\theta)}{f(y_r,\mathbf{y}_{(r)},\theta)}\pi(\theta|\mathbf{y})d\theta} \\ &= \frac{1}{\int \frac{1}{f(y_r|\mathbf{y}_{(r)},\theta)}\pi(\theta|\mathbf{y})d\theta}. \end{aligned}$$

If θ_s are samples from the joint posterior $\pi(\theta|\mathbf{y})$, a Monte Carlo integration of CPO is

$$\hat{f}(y_r|\mathbf{y}_{(r)}) = (\frac{1}{B}\sum \frac{1}{f(y_r|\mathbf{y}_{(r)},\theta_s)})^{-1} = B(\sum \frac{1}{f(y_r|\mathbf{y}_{(r)},\theta_s)})^{-1}$$

Note that if $\{Y_r, r = 1, ..., n\}$ are conditionally independent given θ, $f(y_r|\mathbf{y}_{(r)},\theta_s) = f(y_r|\theta_s)$. For model checking, the presence of many small CPOs criticizes the model. Finally, we present a sampling/importance resampling (SIR) method to generate Y_r from its predictive distribution $f(Y_r \mid y_{(r)})$, which could be implemented as follows:

- Generate B independent proposed samples $\theta_\mathbf{1}, \ldots, \theta_\mathbf{B}$ from an importance sampling distribution, $h(\theta) = \pi(\theta \mid y)$.
- Calculate the standardized weights w_B with the formula $w_B = \frac{(f(y_r|\theta_\mathbf{B}))^{-1}}{\sum_{j=1}^{B}(f(y_r|\theta_\mathbf{j}))^{-1}}$.
- Generate an approximate realization $\theta^\star_\mathbf{s}$ from $(\theta_\mathbf{1}, \ldots, \theta_\mathbf{B})$ with probability $w_1, \ldots, w_B$.
- Generate Y_{rs} from $f(y_r \mid \theta^\star_\mathbf{s})$.

Under mild regularity conditions, it can be shown that

$$Y_{rs} \xrightarrow{D} f(Y_r \mid y_{(r)}) \quad \text{as} \quad B \to \infty$$

3 Tools for Model Determination

In our project, we introduce four different tools which could be used to do model determination. They are predictive interval, CPO plot, Deviance plot and l_r plot, respectively.

The Predictive Interval
Based on the samples of predictive values for each observation from predictive distributions, $f(Y_r \mid \mathbf{y_{(r)}}), r = 1, \ldots, 16$, we could count the number of samples that fall within $100 \times (1-\alpha)\%$ predictive intervals for each Y_r. If too many samples are in the predictive interval with a large α (0.5), the predictive distribution might be overdispersed. Conversely, if too few observations are in the interval with a small α (0.05), then the predictive distribuiton might be underdispersed. i.e. The model is inadequate.

CPO Plot
The conditional predictive ordinate (CPO) is a kind of Bayesian diagnostic approach which detects surprising observations in models. Since the better model has the majority of its CPOs (d_rs) above those of the worse one, we could figure out all of observations with CPO (d_r) values under two different models in a single plot to determine which model is better than the other one.

Deviance Plot
Given the samples generated from the predictive distribution $f(Y_r \mid y_{(r)}), r = 1, \ldots, 16$, we could compute the deviance measure $|y_r - \mu_r|$ or $|y_r - m_r|$ and spread measure $V_r = var(Y_r \mid y_{(r)})$ or $I_R = IQR(Y_r \mid y_{(r)})$. Note that μ_r and m_r represent the mean and median of predictive values, Y_{rs}, we generated for the rth observation. V_r and I_r represent the variance and interquartile range of the predictive distribution $f(Y_r \mid y_{(r)})$. The deviance plot could be either a plot of $|y_r - \mu_r|$ vs. V_r or a plot of $|y_r - m_r|$ vs. I_r to compare several models. In our project, I applied $|y_r - m_r|$ vs. I_r to figure out the deviance plot for Model 1 and Model 2.

l_r Plot
In order to compare different models, we could calculate their Bayes factor. If the Bayes factor is larger than 10, we could conclude that this model is better than its competitive model. Pettit and Yound (1990) proposed a quantity $k_r = \log_{10} B_{12} - \log_{10} B_{12}^r$ to measure the effect of observation r on the Bayes factor, where B_{12}^r represents the Bayes factor excluding the rth observation with Model 1 and Model 2. In our project, instead of using k_r, we define the similar quantity $l_r = \log_{10} PBF_{12} - \log_{10} PBF_{12}^r$ to do model determination. Note that PBF_{12}^r represents pseudo-Bayes factor excluding the rth observation, which could be used to measure the effect of observation r on the pseudo-Bayes factor.
By the definition of pseudo-Bayes factor (PBF), we have

$$PBF_{12} = \frac{\prod_{r=1}^{n} \pi(y_r \mid y_{(r)} \mid M_1)}{\prod_{r=1}^{n} \pi(y_r \mid y_{(r)} \mid M_2)}$$

where the cross validated predictive density is

$$\begin{aligned}\pi(y_r \mid y_{(r)}) &= \int \pi(y_r \mid \theta, y_{(r)})\pi(\theta \mid y_{(r)})d\theta \\ &= \frac{1}{\int \frac{1}{\pi(y_r|\theta,y_{(r)})}\pi(\theta \mid y_{(r)})d\theta}\end{aligned}$$

Since d_r is estimated by $\hat{\pi}(y_r \mid y_{(r)}) = B\left(\sum_{s=1}^{B} \frac{1}{\pi(y_r|\theta_s)}\right)^{-1}$ which is approximate to the cross validated predictive density, PBF could also be written as

$$PBF_{12} = \frac{\prod_{r=1}^{n} d_r^1}{\prod_{r=1}^{n} d_r^2}$$

and we could get $\frac{PBF_{12}}{PBF_{12}^r} = \frac{d_r^1}{d_r^2}$. i.e. $l_r = \log_{10} PBF_{12} - \log_{10} PBF_{12}^r = \log_{10} d_r^1 - \log_{10} d_r^2$. Therefore, l_r plot is just the plot for the difference of d_r values under two models in logarithm. Specifically, a negative l_r indicates less support for the model from observation r and a positive l_r suggests the observation r favors the model. Thus, the model with more positive l_rs is better than that with less positive l_rs.

4 Example with Adsorption data

4.1 Nonlinear Regression Models of the Langmuir Equation

The Langmuir equation correlates the amount of adsorbed gases y on airplane surfaces with the equilibrium aqueous concentration x by a nonlinear function,

$$y = \frac{\alpha\beta x}{1 + \alpha x}, \tag{4.1}$$

where $\alpha > 0$ is the Langmuir constant, $\beta > 0$ is the maximum adsorption capacity of the solid phase. To do a linear regression analysis, linearization of the Langmuir adsorption equation is able to simplify the task of determining α and β values. The kinetic equation of Michaelis and Menten (1913) is similar in form of the Langmuir equationm but it instead correlates the enzyme reaction rates with the amount of substrate present. We denote (4.1) as the first model M_1. The following Table 1 shows the adsorption isotherm of poly (vinyl alcohol) (PVA) onto an Si oxide as described by Schulthess and Dey (1996) and Schulthess and Tokunaga (1996).

OBS	Amount adsorbed (y)	Equilibrium aqueous concentration (x)
1	46.79	3.17
2	46.54	3.48
3	95.82	3.56
4	95.57	3.86
5	201.48	7.14
6	201.28	7.39
7	471.19	101.27
8	469.27	103.65
9	602.63	281.47
10	598.54	286.56
11	696.43	637.41
12	691.17	643.96
13	773.07	1126.94
14	744.45	1162.55
15	835.45	1725.30
16	805.88	1761.88

Table 1: **Adsorption Data**

After examining the plot of y and x in the following figure 1, we see that the data could also be well fitted by the model

$$y = \alpha + \beta \log(x),$$

We denote this model as M_2, Therefore, we consider model M_2 as a competing model with model M_1. Notice that the parameters here are different from α, β in model M_1. By using the traditional least square fitting technique, both of the models M_1, M_2 fit the data very well. For model one, we linearize the Langmuir equation to

$$\frac{x}{y} = \frac{1}{\alpha\beta} + \frac{1}{\alpha}x$$

to do the least square fit. The R^2-values for the two models are $0.9958, 0.9952$ respectively.

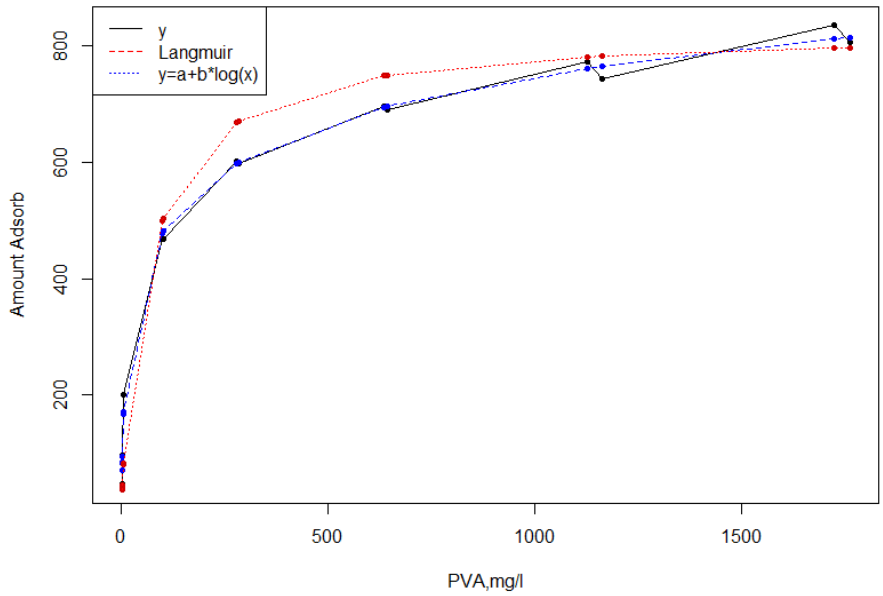

Figure 1: **Plot of adsorption data and least square fit**

From the figure and the results in table 2, it implies that the coefficient of determination approach does not provide clear evidence to distinguish between two models. The results are showing in Table 2.

OBS	y	Predicted values under M_1	Residual under M_1	Predicted values under M_2	Residual under M_2	Supporting sign [a]
1	46.79	37.68	9.11	71.12	−24.34	+
2	46.54	41.18	5.36	82.10	−35.56	+
3	95.82	42.07	53.74	84.77	11.04	−
4	95.57	45.43	50.14	94.29	1.28	−
5	201.48	80.28	121.20	166.65	34.83	−
6	201.28	82.81	118.47	170.70	30.58	−
7	471.19	499.27	−28.08	478.64	−7.45	−
8	469.27	503.85	−34.58	481.37	−12.10	−
9	602.63	668.77	−66.14	598.89	3.73	−
10	598.54	671.04	−72.50	601.00	−2.46	−
11	696.43	748.52	−52.09	695.05	1.38	−
12	691.17	749.23	−58.06	696.25	−5.09	−
13	773.07	780.48	−7.41	762.09	10.98	+
14	744.45	781.82	−37.37	765.74	−21.30	−
15	835.45	795.83	39.62	812.19	23.26	−
16	805.88	796.45	9.43	814.66	−8.78	−

'+' means in favor of model M_1 and '−' means in favor of model M_2.

Table 2: **Least square fit results for adsorption data**

Using a Maximum Likelihood Estimates technique, we first transfer Langmuir equation to,

$$\log y = \log \alpha + \log \beta + \log x - \log(1 + \alpha x) + \epsilon$$

where $\epsilon \sim N(0, \sigma^2)$. Then we reparameterize by letting $\alpha = e^{\alpha^*}$ and $\beta = e^{\beta^*}$.
Hence, model 1 becomes

$$\log y = \alpha^* + \beta^* + \log x - \log\left(1 + e^{\alpha^*} x\right) + \epsilon$$

where $\epsilon \sim N(0, \sigma^2)$.
We calculate the log-likelihood function for M1

$$\ell_{M1} = -n \log \sigma - \frac{1}{2\sigma^2} \sum_i (\log y_i - \alpha^* - \beta^* - \log x_i + \log(1 + e^{\alpha^*} x))^2$$

The log-likelihood function for M2

$$\ell_{M2} = -n \log \sigma - \frac{1}{2\sigma^2} \sum_i (y_i - \alpha^* - \beta^* \log x_i)^2$$

Also we calculate the 95% confidence interval by

$$\mathrm{SE}\left(\hat{\theta}_{\mathrm{ML}}\right) = \frac{1}{\sqrt{\mathbf{I}\left(\hat{\theta}_{\mathrm{ML}}\right)}} = \frac{1}{\sqrt{-\mathbf{H}\left(\hat{\theta}_{\mathrm{ML}}\right)}}$$

where I is the fisher information and H is the hessian matrix, with

$$\mathbf{H}(\theta) = \frac{\partial^2}{\partial\theta_i \partial\theta_j} \ell(\theta), \quad 1 \le i, j \le 3.$$

The estimates and 95% confidence interval are showing in Table 3.

4.2 Bayesian Approach

4.2.1 Metropolis Hasting for M1

Using a Bayesian model-fitting technique, we transfer Langmuir equation to,

$$\log y = \log \alpha + \log \beta + \log x - \log(1 + \alpha x) + \epsilon$$

where $\epsilon \sim N(0, \sigma^2)$, and then we reparameterize model 1 by letting $\alpha = e^{\alpha^*}$ and $\beta = e^{\beta^*}$.
Hence, model 1 becomes

$$\log y = \alpha^* + \beta^* + \log x - \log\left(1 + e^{\alpha^*} x\right) + \epsilon$$

where $\epsilon \sim N(0, \sigma^2)$.
We employ noninformative priors on model parameters.

$$\pi\left(\alpha^*, \beta^*, \sigma^2\right) \propto 1/\left(\sigma^2\right)$$

The conditional posterior distribution of σ^2 given α^*, β^* is

$$\begin{aligned}
&\pi\left(\sigma^2 | \alpha^*, \beta^*, D\right) \propto \pi\left(D | \alpha^*, \beta^*, \sigma^2\right) \pi\left(\alpha^*, \beta^*, \sigma^2\right) \\
&\propto (\sigma^2)^{\frac{-n}{2}} \exp\left\{-\frac{1}{2\sigma^2} \sum_i (\log y_i - \alpha^* - \beta^* - \log x_i + \log(1 + e^{\alpha^*} x_i))^2\right\} \cdot \frac{1}{\sigma^2} \\
&= \exp\left\{-\frac{1}{2} \sum_i (\log y_i - \alpha^* - \beta^* - \log x_i + \log(1 + e^{\alpha^*} x_i))^2 \frac{1}{\sigma^2}\right\} \cdot \left(\sigma^2\right)^{-\frac{n}{2}-1}
\end{aligned}$$

Therefore, $\sigma^2|\alpha^*, \beta^*, D \sim$ inverse-gamma(a, b), where

$$a = n/2, \quad b = \frac{1}{2}\sum_i \left(\log y_i - \alpha^* - \beta^* - \log x_i + \log\left(1 + e^{\alpha^*} x\right)\right)^2$$

The conditional posterior distribution of α^*, β^* given σ^2 is

$$\begin{aligned}
&\pi\left(\alpha^*, \beta^* | \sigma^2, D\right) \propto \pi\left(D|\alpha^*, \beta^*, \sigma^2\right) \pi\left(\alpha^*, \beta^*, \sigma^2\right) \\
&\propto (\sigma^2)^{\frac{-n}{2}} \exp\left\{-\frac{1}{2\sigma^2}\sum_i (\log y_i - \alpha^* - \beta^* - \log x_i + \log(1 + e^{\alpha^*} x_i))^2\right\} \cdot \frac{1}{\sigma^2} \\
&= \exp\left\{-\frac{1}{2\sigma^2}\sum_i (\log y_i - \alpha^* - \beta^* - \log x_i + \log(1 + e^{\alpha^*} x_i))^2\right\}
\end{aligned}$$

Therefore, we use random walk Metropolis Hasting algorithm to sample α^* and β^* from its conditional posterior distribution.
The proposal distribution is of the form bivariate normal distribution with mean $\begin{pmatrix}\alpha^* \\ \beta^*\end{pmatrix}$ and covariance matrix $V = \begin{pmatrix}0.05 & 0 \\ 0 & 0.05\end{pmatrix}$.
The iteration time is 50000, the acceptance rate is 0.25.

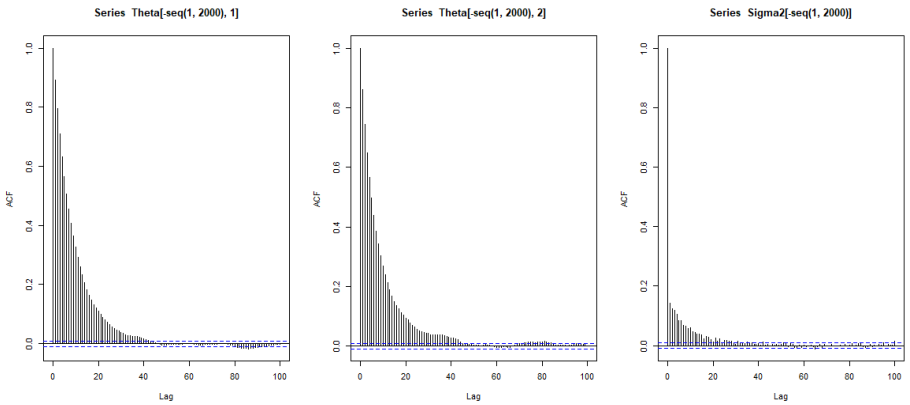

Figure 2: **ACFs for** α, β, σ

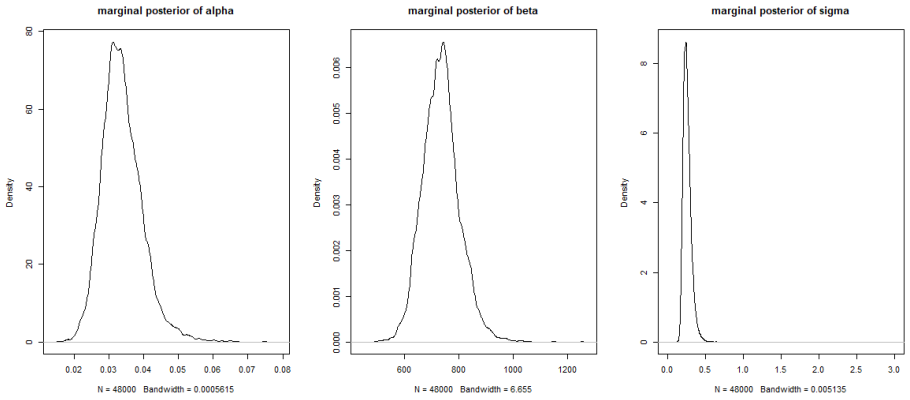

Figure 3: **Marginal posterior distributions for** α, β, σ

To check the convergence, the first three figures are auto correlation function with the first 2000 iterations being discarded. For parameter α and β, the Markov Chains converge after about 50 lags, for parameter σ, the chain converges after 30 lags. Markov

Chains for parameters are significantly independent.

4.2.2 Gibbs Sampling for M2

Using a Bayesian model-fitting technique, model 2 becomes

$$y = \alpha + \beta \log x + \epsilon,$$

where $\epsilon \sim N(0, \sigma^2)$. We employ noninformative priors on model parameters.

$$\pi(\alpha, \beta, \sigma^2) \propto \frac{1}{\sigma^2}.$$

The conditional posterior distribution of α given β, σ^2 is

$$\begin{aligned}
\pi(\alpha|\beta, \sigma^2, x, y) &\propto \pi(y|\alpha, \beta, \sigma^2, x) \times \pi(\alpha, \beta, \sigma^2) \\
&\propto \exp[-\frac{1}{2\sigma^2}\sum(y - \alpha - \beta \log x)^2] \times \frac{1}{\sigma^2} \\
&\propto \exp\{-\frac{n}{2\sigma^2}[\alpha^2 - 2\frac{\sum(y - \beta \log x)}{n}\alpha]\},
\end{aligned}$$

i.e, $\alpha|\beta, \sigma^2, x, y \sim N(\frac{\sum(y-\beta \log x)}{n}, \frac{\sigma^2}{n})$, which is a normal distribution.
Similarly, the conditional posterior distribution of β given α, σ^2 is

$$\begin{aligned}
\pi(\beta|\alpha, \sigma^2, x, y) &\propto \pi(y|\alpha, \beta, \sigma^2, x) \times \pi(\alpha, \beta, \sigma^2) \\
&\propto \exp[-\frac{1}{2\sigma^2}\sum(y - \alpha - \beta \log x)^2] \times \frac{1}{\sigma^2} \\
&\propto \exp\{-\frac{\sum(\log x)^2}{2\sigma^2}[\beta^2 - 2\frac{\sum(y - \alpha)\log x}{\sum(\log x)^2}\beta]\},
\end{aligned}$$

i.e, $\beta|\alpha, \sigma^2, x, y \sim N(\frac{\sum(y-\alpha)\log x}{\sum(\log x)^2}, \frac{\sigma^2}{\sum(\log x)^2})$, which is also a normal distribution.
The conditional posterior distribution of σ^2 given α, β is

$$\begin{aligned}
\pi(\sigma^2|\alpha, \beta, x, y) &\propto \pi(y|\alpha, \beta, \sigma^2, x) \times \pi(\alpha, \beta, \sigma^2) \\
&\propto (\frac{1}{\sigma^2})^{\frac{n}{2}} exp[-\frac{1}{2\sigma^2}\sum(y - \alpha - \beta \log x)^2] \times \frac{1}{\sigma^2} \\
&\propto (\frac{1}{\sigma^2})^{\frac{n}{2}+1} exp[-\frac{1}{\sigma^2}\frac{\sum(y - \alpha - \beta \log x)^2}{2}],
\end{aligned}$$

i.e, $\sigma^2|\alpha, \beta, x, y \sim inverse - gamma(\frac{n}{2}, \frac{\sum(y-\alpha-\beta \log x)^2}{2})$, which is an inverse gamma distribution. Therefore, implementation of the Gibbs sampler under this model is straightforward.
To check the convergence of the Gibbs sampler, we obtain the auto correlation functions of parameters α, β, σ, which are plotted in Figure 4. By viewing the auto correlation functions, we can see the Markov Chains converge after 10 lags for α and β and converges after 8 legs for σ. Compared to Figure 2, we find the convergence for M1 occurs much earlier than M2.

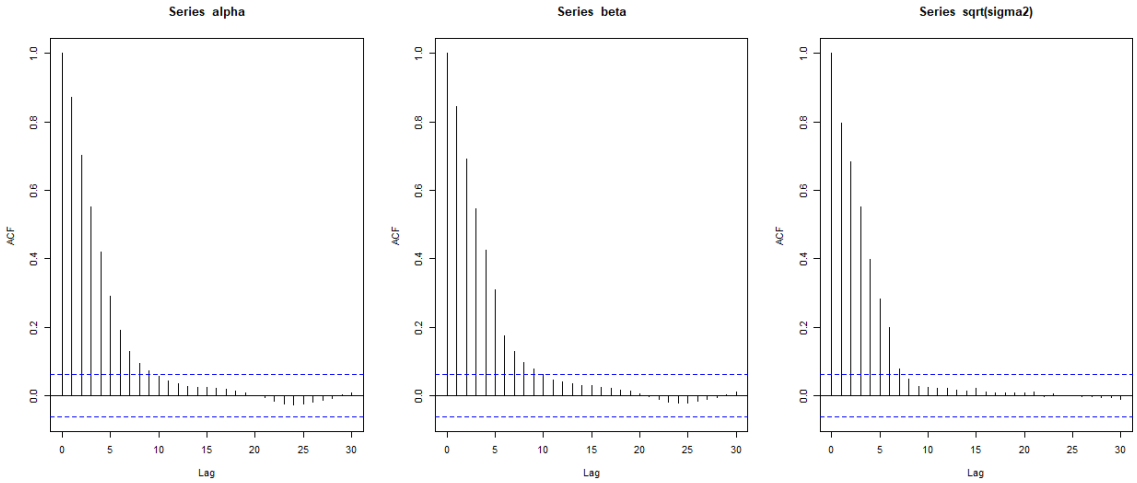

Figure 4: **ACFs for** α, β, σ

Using the Bayesian fitting technique and the third method we mentioned in Section 2, the esitimated CPO values for both models are listed in Table 3. We observe that all the CPO estimates in M2 are larger than that in M1 and all the supporting signs are negative, which means M2 is better than M1 for every observation. Table 4 provides the inference summaries for all parameters of each model, including MLEs, posterior means and their 95% confidence intervals.

OBS	d_r for M_2	d_r for M_2	$\log_{10}$ difference	supporting sign[a]
1	0.0007040442	0.006303215	-0.9519622	-
2	0.0007040764	0.001994990	-0.4523209	-
3	0.0006966145	0.013720290	-1.2943708	-
4	0.0006966808	0.015464678	-1.3463071	-
5	0.0006631989	0.003228009	-0.6872910	-
6	0.0006632955	0.004941961	-0.8721922	-
7	0.0004938062	0.015646921	-1.5008723	-
8	0.0004953031	0.014068964	-1.4533911	-
9	0.0003885033	0.015561272	-1.6026504	-
10	0.0003918202	0.015766067	-1.6046365	-
11	0.0003136740	0.014315393	-1.6593247	-
12	0.0003177801	0.014126810	-1.6479174	-
13	0.0002558821	0.010912530	-1.6298855	-
14	0.0002769818	0.008194787	-1.4710865	-
15	0.0002123626	0.005841685	-1.4394600	-
16	0.0002325449	0.010610843	-1.6592430	-

Table 3: **The values of** d_r **for adsorption data**

	M_1		M_2	
Para-meters	MLE ($95\%CI^a$) .	Posterior mean ($95\%CI^b$).	MLE ($95\%CI^a$) .	Posterior mean ($95\%CI^b$).
α	0.033	0.034	-64.591	-64.935
	(0.024,0.044)	(0.0238,0.0472)	(-86.851,-42.323)	(-87.498,-39.867)
β	731.382	734.470	117.640	117.658
	(625.146,854.895)	(612,876)	(113.264,122.007)	(113.104 ,122.005)
σ	1.257	0.259	18.660	22.335
	(1.161,1.359)	(0.179 0.386)	(12.194,25.125)	(14.385, 33.281)

Table 4: **Inference summaries for adsorption data**

4.3 Model Determination

For the predictive interval, in order to compare two modelswe constructed before, Table 5 and Table 6 shows the predictive intervals based on samples from the predictive distribution for each observation under Model 1 and Model 2, respectively. The two last columns in the tables are two indicator variables, Ind1 and Ind2. Ind1=1 indicates that the actual observation falls within 50% PI while Ind2=1 indicates that the actual observation falls within 95% PI. In summary, 11 of 16 actual observations fall within 50% predictive interval and all of them fall within 95% predictive interval under Model 1. On the other hand, for Model 2, 10 of 16 actual observations fall within 50% PI all of them are in the 95% predictive interval. Since there are neither too few observations falling within the 50% PI nor too many observations being in the 95% PI, we could conclude that both models are adequate for our adsorption data.

OBS	y	2.5%	25%	75%	97.5%	Ind1	Ind2
1	46.79	39.25	55.88	80.09	113.97	0	1
2	46.54	41.46	58.79	87.49	124.60	0	1
3	95.82	49.29	68.08	96.64	133.18	1	1
4	95.57	52.51	72.15	102.20	143.25	1	1
5	201.48	87.29	123.04	178.32	251.98	0	1
6	201.28	90.31	127.84	182.54	255.70	0	1
7	471.19	339.16	466.71	664.37	921.45	1	1
8	469.27	334.07	469.94	670.05	930.69	0	1
9	602.63	399.35	553.16	781.38	1091.46	1	1
10	598.54	399.37	553.00	783.67	1074.01	1	1
11	696.43	428.17	592.47	838.01	1162.60	1	1
12	691.17	426.03	597.97	837.43	1169.90	1	1
13	773.07	441.98	612.27	860.35	1176.02	1	1
14	744.45	438.46	606.98	859.48	1187.04	1	1
15	835.45	440.98	614.28	867.07	1207.18	1	1
16	805.88	446.52	620.38	871.75	1196.37	1	1
						$\Sigma = 11$	$\Sigma = 16$

Table 5: **Predictive intervals of** M_1

OBS	y	2.5%	25%	75%	97.5%	Ind1	Ind2
1	46.79	24.31	52.60	81.96	108.99	0	1
2	46.54	31.46	60.88	91.85	121.82	0	1
3	95.82	43.89	71.96	100.71	127.96	1	1
4	95.57	53.14	80.18	107.78	134.73	1	1
5	201.48	125.85	154.75	185.92	214.77	0	1
6	201.28	131.10	159.23	188.76	218.29	0	1
7	471.19	437.96	463.94	492.49	519.07	1	1
8	469.27	440.50	466.45	494.48	520.06	1	1
9	602.63	559.40	585.27	613.38	640.37	1	1
10	598.54	559.48	586.74	613.93	642.78	1	1
11	696.43	654.55	681.41	709.53	735.29	1	1
12	691.17	655.37	681.69	709.63	735.15	1	1
13	773.07	722.77	749.30	777.50	804.07	1	1
14	744.45	721.60	748.77	777.37	803.86	0	1
15	835.45	772.88	800.76	829.18	858.28	0	1
16	805.88	772.04	799.00	827.69	853.04	1	1
						$\Sigma = 10$	$\Sigma = 16$

Table 6: **Predictive intervals of** M_2

According to CPO plot, by the Bayesian fitting technique, the estimated CPO values for both models are displayed in Figure 5. It indicates that Model 2 can fit the adsorption data better than Model 1 as all of CPO values in Model 2 are larger that those in Model 1.

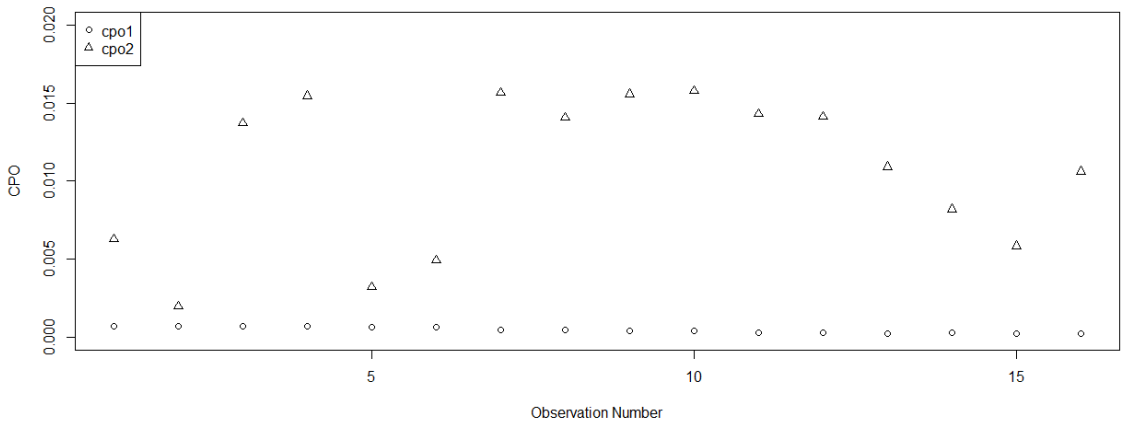

Figure 5: **CPO plot for adsorption data**

The deviance $\mid y_r - m_r \mid$ vs. the interquartile ranges (IQR) for each observation are plotted in Figure 6. We have strongly evidence that Model 2 is better than Model 1 since most of observations in Model 2 have smaller deviance values and interquartile ranges.

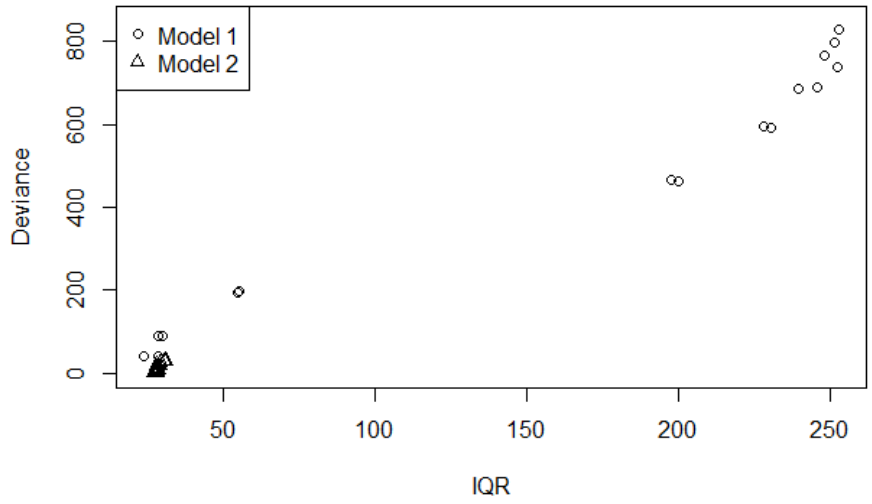

Figure 6: **Deviance vs. IQR plot for adsorption data**

As for l_r plot in Figure 7, l_rs are all less than 0, which indicates that all of observations criticized Model 1 and favor Model 2.
In conclusion, we prefer Model 2 to fit the adsorption data since it is adequate and better than Model 1 in terms of results we got for four model determination tools.

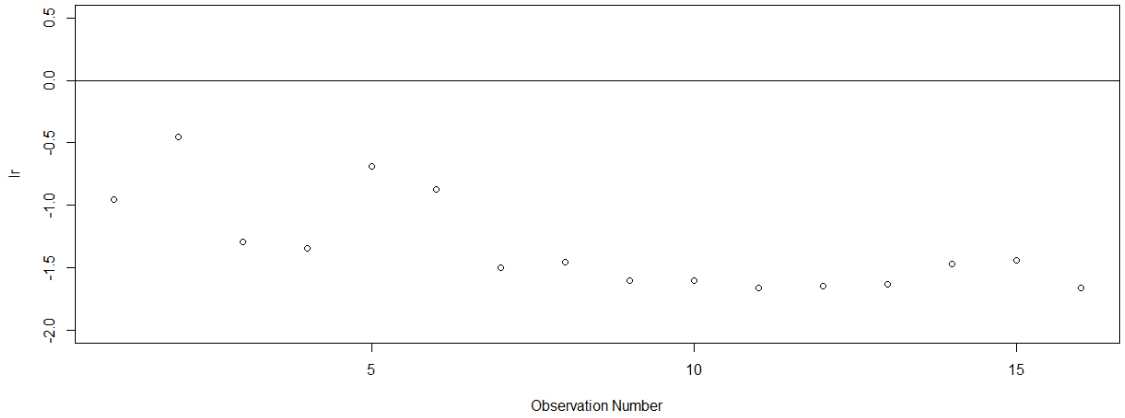

Figure 7: l_r **plot for adsorption data**

5 Conclusion

In our project, we programmed in R and reproduced most of the results shown in the reference paper[1]. We compared the parameter estimates by Bayesian approach with MLE for nonlinear models and further found out that the results of two methods are quite similar. Then, we considered how to select models within the general context of Bayesian inference. Diagnostic tools used for model determination indicate that both models are adequate but Model 2 is better than Model 1 to fit the adsorption data. Our purpose in this project is to learn how to estimate parameters by Bayesian approach in nonlinear cases and to address how sampling-based methods can make Bayesian diagnostics for model determination.

References

[1] Dipak K. Dey. (Dec. 1997). Bayesian Approach for Nonlinear Random Effects Models. *Biometrics.* Vol. 53, No.4.

[2] Aitkin, M. (1991). Posterior Bayes factors (with discussion). *Journal of the Royal Statistical Society, Series B 53, 111-142.*

[3] Devroye, L. (1986). *Non-uniform Random Variate Generation.* New York: Springer-Verlag.

[4] Gelfand, A. E., Dey, D. D., and Chang, H. (1992). Model determination using predictive distributions with implementation via sampling-based methods. In *Bayesian Statistics*, Volume 4, J. Bernardo et al. (eds), 147-167. Oxford: Oxford University Press.

[5] Schulthess, C. P. and Dey, D. K. (1996). Estimation of Langmuir constants using linear and nonlinear lest square regression analysis. *Soil Science Society of America Journal* **60,** 433-442.