STATISTICAL INFERENCES & EXAMPLE PROBLEMS[1]
- Statistical Inferences
- Example problems

We should determine the standard error in the *parameters* and in the *expected response variables* in order to obtain a quantitative measure of the quality of the fit and how trustworthy the estimates are.

INFERENCES ON THE PARAMETERS
The measurements are random variables, and hence the parameter estimates are also random variables. The probability distribution of the parameter estimates usually approaches the *normal* as the number of observations in increased.

- The parameter values ($\mathbf{k}^*$) are the *means* of the estimated distribution
- The covariance matrix of the distribution, Cov($\mathbf{k}^*$), is a measure of the reliability of the estimates. This matrix expresses the manner in which variations in the measurements would affect the parameter estimates. The diagonal elements of Cov($\mathbf{k}^*$) are the variances of the parameter estimates and their square roots are the *standard deviations*. Off-diagonal elements indicate interdependence of the estimates of the various parameters.

Occurrence of large variances[2]
- *The model is inappropriate.* You observe large systematic deviations between model predictions and data. The obvious remedy is to modify the model
- *The measurement precision is low.* This case is characterized by large random residuals i.e. randomly high and low deviations. If no improvement of the measurement techniques is feasible, the only remedy is to make more measurements
- *The experiments were not properly designed*. The attainable precision in estimates for a given number of experiments is maximized when the experimental conditions are properly chosen.

The covariance matrix of the parameter estimates is given by

---

[1] Englezos, P. and N. Kalogerakis, "*Applied Parameter Estimation for Chemical Engineers*", Marcel-Dekker, New York, 2001 (chapter 8)

[2] Parameters may have large variances even though the overall fit is good i.e, even though the residuals are small. An example is a situation when two parameters appear in a model as a product of functions of themselves. It is possible that an identical response can be obtained by widely varying values of the two parameters.

$$COV(\mathbf{k}*) = \sigma_\varepsilon^2 \left[\mathbf{A}*\right]^{-1} \qquad (11.1)$$

where $\mathbf{A}*$ is matrix $\mathbf{A}$ evaluated at $\mathbf{k}*$ and the required estimate of the variance σ_ε^2 is obtained from

$$\hat{\sigma}_\varepsilon^2 = \frac{S(\mathbf{k}*)}{(\text{d.f.})} = \frac{S(\mathbf{k}*)}{N\,m - p} \qquad (11.2)$$

where (d.f.)= $Nm-p$ are the degrees of freedom (the total number of measurements minus the number of unknown parameters).

The *(1-α)100% marginal confidence interval* for each parameter, k_i, i=1,2,…,p, is given by

$$k_i^* - t_{\alpha/2}^\nu \, \hat{\sigma}_{k_i} \leq k_i \leq k_i^* + t_{\alpha/2}^\nu \, \hat{\sigma}_{k_i} \qquad (11.5)$$

where $t_{\alpha/2}^\nu$ is obtained from the tables of Student's T-distribution with $\nu=(Nm-p)$ *degrees of freedom*. The standard error of parameter k_i, $\hat{\sigma}_{k_i}$, is obtained as the square root of the corresponding diagonal element of the inverse of matrix $\mathbf{A}*$ multiplied by $\hat{\sigma}_\varepsilon$, i.e.,

$$\hat{\sigma}_{k_i} = \hat{\sigma}_\varepsilon \sqrt{\left\{\left[\mathbf{A}*\right]^{-1}\right\}_{ii}} \qquad (11.6)$$

NOTE: for $\nu \geq 30$ the approximation $t_{\alpha/2}^\nu \approx z_{\alpha/2}$ can be used where $z_{\alpha/2}$ is obtained from the standard *normal distribution* tables. In that case we choose $z_{0.025}$=1.96 or 2.0 ($t_{0.025}^{30}$ =2.042) for the 95% confidence interval.

STANDARD DEVIATIONS WHEN USING SCALING OF MATRIX A

Scaling of Matrix A

When the parameters differ by more than one order of magnitude, matrix **A** may appear to be ill-conditioned even if the estimation problem is well-posed. In order to overcome this problem the *reduced sensitivity coefficients* are defined

$$G_{Rij} = \left(\frac{\partial y_i}{\partial k_j} \right) k_j \qquad (8.37)$$

and the *reduced parameter sensitivity matrix*, **G$_R$**, is related to **G**, as follows

$$\mathbf{G_R} = \mathbf{G\,K} \qquad (8.38)$$

where

$$\mathbf{K} = diag(k_1, k_2, \ldots, k_p) \qquad (8.39)$$

As a result, the normal equations $\mathbf{A}\Delta\mathbf{k}^{(j+1)} = \mathbf{b}$ become

$$\mathbf{A_R}\,\Delta\mathbf{k}_R^{(j+1)} = \mathbf{b_R} \qquad (8.40)$$

where

$$\mathbf{A_R} = \mathbf{K}\left[\sum_{i=1}^{N} \mathbf{G_i^T Q_i\, G_i} \right]\mathbf{K} = \mathbf{KAK} \qquad (8.41)$$

$$\mathbf{b_R} = \mathbf{K}\left[\sum_{i=1}^{N} \mathbf{G_i^T Q_i}\left[\hat{\mathbf{y}}_i - \mathbf{f}(\mathbf{x}_i, \mathbf{k}^{(j)}) \right] \right] = \mathbf{Kb} \qquad (8.42)$$

and

$$\Delta\mathbf{k}_R^{(j+1)} = \mathbf{K}^{-1}\Delta\mathbf{k}^{(j+1)} \qquad (8.43)$$

The estimates for the next iteration are obtained ($0 < \mu \le 1$) as follows

$$\mathbf{k}^{(j+1)} = \mathbf{k}^{(j)} + \mu\,\mathbf{K}\,\Delta\mathbf{k}_R^{(j+1)} \qquad (8.44)$$

<u>NOTE</u>: $\mathbf{A}\,\Delta\mathbf{k} = \mathbf{b}$ ⬅➡ $\mathbf{K\,A}\,\Delta\mathbf{k} = \mathbf{K\,b}$ ⬅➡ $\mathbf{K\,A\,K\,K^{-1}}\,\Delta\mathbf{k} = \mathbf{K\,b}$ ⬅➡ $\mathbf{K\,A\,K}\,\Delta\mathbf{k_R} = \mathbf{b_R}$

Recall

$$\hat{\sigma}_{k_i} = \hat{\sigma}_\varepsilon \sqrt{\left\{ [\mathbf{A}*]^{-1} \right\}_{ii}} \qquad (11.6)$$

Since $\mathbf{G_R} = \mathbf{G\,K}$ and $\mathbf{A_R} = \mathbf{KAK}$ ➡ $\mathbf{A}_R^{-1} = \mathbf{K^{-1} A^{-1} K^{-1}}$.

Hence,

$$\sqrt{A^{-1}_{Rii}} = \frac{\sqrt{A^{-1}_{ii}}}{k_i}$$

If we then denote by β the above quantity multiplied by 100 we have

$$100\sqrt{A^{-1}_{Rii}} = \frac{\sqrt{A^{-1}_{ii}}}{k_i}100 = \beta\,\%$$

We can now report the parameters as

$$k_i = k_i^* \pm \beta\,\% \qquad\qquad k_i = k_i^* \pm t^v_{\alpha/2}\sigma_{k_i}$$

And the variance given by equation 11.6 is related to β by the following equation

$$\frac{\sigma_{ki}}{k_i}100 = \beta\hat{\sigma}_\varepsilon \qquad\longrightarrow\qquad \sigma_{ki} = \frac{\beta k_i}{100}\hat{\sigma}_\varepsilon$$

From:

$$\hat{\sigma}_{k_i} = \hat{\sigma}_\varepsilon\sqrt{\left\{[\mathbf{A}^*]^{-1}\right\}_{ii}} = \hat{\sigma}_\varepsilon k_i\sqrt{\left\{[\mathbf{A}^*]^{-1}\right\}_{Rii}} = \hat{\sigma}_\varepsilon k_i\frac{\beta}{100} =$$

INFERENCES ON THE EXPECTED RESPONSE VARIABLES

1. <u>Algebraic Models</u>: **y=f(x,k)**.

The *100(1−α)% confidence interval of the expected mean response* of the variable y_j at $\mathbf{x}_0$ is given by

$$f_j(\mathbf{x}_o,\mathbf{k}^*) - t_{\alpha/2}^{\nu}\hat{\sigma}_{y_{j0}} \leq \mu_{y_{j0}} \leq f_j(\mathbf{x}_o,\mathbf{k}^*) + t_{\alpha/2}^{\nu}\hat{\sigma}_{y_{j0}} \qquad (11.8)$$

where $t_{\alpha/2}^{\nu}$ is obtained from the tables of Student's t-distribution with $\nu = (Nm-p)$ degrees of freedom and the standard prediction error of y_{j0}, $\hat{\sigma}_{y_{j0}}$, is

$$\hat{\sigma}_{y_{j0}} = \hat{\sigma}_{\varepsilon}\sqrt{\left(\frac{\partial f_j}{\partial \mathbf{k}}\right)^{T}[\mathbf{A}^*]^{-1}\left(\frac{\partial f_j}{\partial \mathbf{k}}\right)} \qquad (11.11)$$

Equation 11.8 represents the confidence interval for the *mean* expected response rather than a *future observation (future measurement)* of the response variable, $\hat{y}_0$. In this case, besides the uncertainty in the estimated parameters, we must include the uncertainty due to the measurement error (ε_0).

The *(1−α)100% confidence interval* of $\hat{y}_{j0}$ is

$$f_j(\mathbf{x}_o,\mathbf{k}^*) - t_{\alpha/2}^{\nu}\hat{\sigma}_{\hat{y}_{j0}} \leq \hat{y}_{j0} \leq f_j(\mathbf{x}_o,\mathbf{k}^*) + t_{\alpha/2}^{\nu}\hat{\sigma}_{\hat{y}_{j0}} \quad (11.12)$$

where the standard prediction error of $\hat{y}_{j0}$ is given by

$$\hat{\sigma}_{\hat{y}_{j0}} = \hat{\sigma}_{\varepsilon}\sqrt{1+\left(\frac{\partial f_j}{\partial \mathbf{k}}\right)^{T}[\mathbf{A}^*]^{-1}\left(\frac{\partial f_j}{\partial \mathbf{k}}\right)} \qquad (11.13)$$

2. <u>ODE Models:</u>

$$\frac{d\mathbf{x}(t)}{dt} = \mathbf{f}(\mathbf{x}(t), \mathbf{u}, \mathbf{k}) \quad ; \quad \mathbf{x}(t_0) = \mathbf{x}_0 \quad (6.1)$$

$$\mathbf{y}(t) = \mathbf{C}\mathbf{x}(t) \quad (6.2)$$

We are interested in establishing confidence intervals for each of the response variables y_j, j=1,…,*m* at any time t=t$_o$.

The *(1−α)100% confidence interval of the expected mean response* of y_j at t=t$_o$,

$$y_j(t_o, \mathbf{k}^*) - t_{\alpha/2}^{\nu}\hat{\sigma}_{y_{jo}} \le \mu_{y_{jo}} \le y_j(t_o, \mathbf{k}^*) + t_{\alpha/2}^{\nu}\hat{\sigma}_{y_{j0}} \quad (11.18)$$

where the estimated standard prediction error of $y_j(t_o)$ is obtained as follows

$$\hat{\sigma}_{y_{jo}} = \hat{\sigma}_{\varepsilon}\sqrt{\left\{\mathbf{C}\,\mathbf{G}(t_0)\left[\mathbf{A}*\right]^{-1}\mathbf{C}^T\mathbf{G}^T(t_0)\right\}_{jj}} \quad (11.17)$$

It is noted that the sensitivity coefficients matrix $\mathbf{G}(t_0)$ is evaluated at $\mathbf{k}^*$. Recall that the sensitivity coefficient matrix $\mathbf{G}$ is as follows

$$\mathbf{G}(t_i) \equiv (\partial \mathbf{x}^T/\partial \mathbf{k})^T$$

The *(1−α)100% confidence interval of the response* of y_j at t=t$_o$, also takes into account the error term (ε_o) in the calculation of the standard error

$$y_j(t_o, \mathbf{k}^*) - t_{\alpha/2}^{\nu}\hat{\sigma}_{\hat{y}_{jo}} \le \hat{y}_j(t_o) \le y_j(t_o, \mathbf{k}^*) + t_{\alpha/2}^{\nu}\hat{\sigma}_{\hat{y}_{j0}} \quad (11.19)$$

with

$$\hat{\sigma}_{\hat{y}_{jo}} = \hat{\sigma}_{\varepsilon}\sqrt{1 + \left\{\mathbf{C}\mathbf{G}(t_o)\left[\mathbf{A}*\right]^{-1}\mathbf{C}^T\mathbf{G}^T(t_o)\right\}_{jj}} \quad (11.20)$$

MODEL ADEQUACY TESTS

These tests are based on the comparison of the experimental error variance estimated by the model to that obtained experimentally or through other means.

Single Response Models (one measured variable, m=1**)**

<u>CASE 1:</u> The *experimental error variance,* σ_ϵ^2 , is known precisely:

Using Equation 11.2 we obtain an estimate of the experimental error variance under the assumption that the model is adequate. Therefore, to test whether the model is adequate we simply need to test the hypothesis

$$H_O: \quad \sigma_{model}^2 \ = \ \sigma_\epsilon^2$$

$$H_1: \quad \sigma_{model}^2 \ > \ \sigma_\epsilon^2$$

at any desirable level of significance, e.g., $\alpha=0.05$. Here with σ_{model}^2 we denote the error variance estimated by the model equations (Eqn 11.2); namely, $\hat{\sigma}_\epsilon^2$ is an estimate of σ_{model}^2 .

$$\hat{\sigma}_\epsilon^2 = \frac{S(\mathbf{k}^*)}{(\text{d.f.})} = \frac{S(\mathbf{k}^*)}{N\,m - p} \qquad (11.2)$$

where (d.f.)= Nm−p are the degrees of freedom.

Since σ_ϵ^2 is known exactly (i.e., there is no uncertainty in its value, it is a given number) the above hypothesis test is done through a χ^2-test. Namely,

$$\text{If} \quad \chi_{\text{data}}^2 \ > \ \chi_{\nu=(Nm-p),1-\alpha}^2 \ \Rightarrow \ \text{Reject } H_O$$

where

$$\chi_{\text{data}}^2 = (N\,m - p)\,\frac{\hat{\sigma}_\epsilon^2}{\sigma_\epsilon^2} = \frac{S(\mathbf{k}^*)}{\sigma_\epsilon^2} \qquad (11.21)$$

and $\chi_{\nu=(Nm-p),1-\alpha}^2$ is obtained from the tables of the χ^2-distribution with degrees of freedom $\nu=(Nm-p)$.

NOTE: <mark>When Ho is rejected the model is inadequate.</mark>

CASE 2: The *experimental error variance*, σ_ε^2, is known approximately:

We assume that σ_ε^2 is not known exactly, however, we have performed n *repeated* measurements of the response variable. From this small sample of multiple measurements we can determine the sample *mean* and sample *variance*. If s_ε^2 is the sample estimate of σ_ε^2, estimated from the n repeated measurements it is given by

$$S_\varepsilon^2 = \frac{1}{n-1} \sum_{i=1}^{n} (y_i - \bar{y})^2 \qquad (11.22)$$

where the sample *mean* is obtained from

$$\bar{y} = \frac{1}{n} \sum_{i=1}^{n} y_i \qquad (11.23)$$

Again, we test the hypothesis at any desirable level of significance, for example $\alpha=0.05$

$$H_0: \quad \sigma_{model}^2 = \sigma_\varepsilon^2$$

$$H_1: \quad \sigma_{model}^2 > \sigma_\varepsilon^2$$

In this case, since σ_ε^2 is known only approximately, the above hypothesis is tested using an F-test, i.e.,

$$\text{If} \quad F_{data} > F_{1-\alpha}^{\nu_1=(Nm-p),\nu_2=n-1} \qquad \Rightarrow \qquad \text{Reject } H_0$$
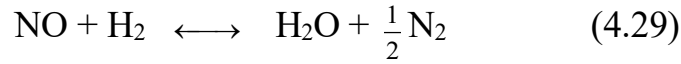
where

$$F_{data} = \frac{\hat{\sigma}_\varepsilon^2}{s_\varepsilon^2} \qquad (11.24)$$

and $F_{1-\alpha}^{\nu_1=\text{d.f.},\nu_2=n-1}$ is obtained from the tables of the F-distribution.

EXAMPLES

***Example 1*: Catalytic Reduction of Nitric Oxide by H_2**[3]

$$NO + H_2 \longleftrightarrow H_2O + \tfrac{1}{2} N_2 \qquad (4.29)$$

Data were taken at 375, 400 ℃, and 425 ℃ using N_2 as the diluent and given in Table 4.5. The reaction rate in *gmol/(min·g-catalyst)* and the total NO conversion were measured at different partial pressures for H_2 and NO.

A *Langmuir-Hinshelwood reaction rate model* for the reaction between an adsorbed NO molecule and one H_2 molecule is described by:

$$r = \frac{k K_{H_2} K_{NO} p_{H_2} p_{NO}}{\left(1 + K_{NO} p_{NO} + K_{H_2} p_{H_2}\right)^2} \qquad (4.30)$$

where r is the reaction rate in *gmol/(min·g-catalyst)*, p_{H2} is the partial pressure of hydrogen (*atm*), p_{NO} is the partial pressure of NO (*atm*), K_{NO} (*atm^{-1}*) is the adsorption equilibrium constant for NO, K_{H_2} (*atm^{-1}*) is the adsorption equilibrium constant for H_2 and k is the forward reaction rate constant for surface reaction.

Using our standard notation:

Parameter vector:	$\mathbf{k} = [k_1, k_2, k_3]^T$	where $k_1=k$, $k_2=K_{H2}$ & $k_3=K_{NO}$
Independent variables	$\mathbf{x} = [x_1, x_2]^T$	where $x_1=p_{H2}$, $x_2=p_{NO}$
Output vector	$\mathbf{y} = [y_1]$	where $y_1=r$
Model Equation	$\mathbf{f} = [f_1]$	

where

$$f_1(x_1, x_2, k_1, k_2, k_3) = \frac{k_1 k_2 k_3 x_1 x_2}{\left(1 + k_3 x_2 + k_2 x_1\right)^2} \qquad (4.31)$$

The elements of the (m × p) or (*1×3*) sensitivity coefficient matrix **G** are

$$G_{11} = \left(\frac{\partial f_1}{\partial k_1}\right) = \frac{k_2 k_3 x_1 x_2}{\left(1 + k_3 x_2 + k_2 x_1\right)^2} \qquad (4.32a)$$

$$G_{12} = \left(\frac{\partial f_1}{\partial k_2}\right) = \frac{k_1 k_3 x_1 x_2}{\left(1 + k_3 x_2 + k_2 x_1\right)^2} - \frac{2 k_1 k_2 k_3 x_1^2 x_2}{\left(1 + k_3 x_2 + k_2 x_1\right)^3} \qquad (4.32b)$$

---

[3] Ayen and Peters, *Ind Eng Chem Proc Des Dev.*, 1, 204-207, 1962

$$G_{13} = \left(\frac{\partial f_1}{\partial k_3}\right) = \frac{k_1 k_2 x_1 x_2}{\left(1 + k_3 x_2 + k_2 x_1\right)^2} - \frac{2k_1 k_2 k_3 x_1 x_2^2}{\left(1 + k_3 x_2 + k_2 x_1\right)^3} \qquad (4.32c)$$

Table 4.5 Experimental Data for the Catalytic Reduction of Nitric Oxide

p_{H2} (atm)	p_{NO} (atm)	Reaction Rate, $r \times 10^5$ gmol/(min·g-catalyst)	Total NO Conversion (%)
T=375 °C, Weight of catalyst=2.39 *g*			
0.00922	0.0500	1.60	1.96
0.0136	0.0500	2.56	2.36
0.0197	0.0500	3.27	2.99
0.0280	0.0500	3.64	3.54
0.0291	0.0500	3.48	3.41
0.0389	0.0500	4.46	4.23
0.0485	0.0500	4.75	4.78
0.0500	0.00918	1.47	14.0
0.0500	0.0184	2.48	9.15
0.0500	0.0298	3.45	6.24
0.0500	0.0378	4.06	5.40
0.0500	0.0491	4.75	4.30
T=400 °C, Weight of catalyst=1.066 *g*			
0.00659	0.0500	2.52	0.59
0.0113	0.0500	4.21	1.05
0.0228	0.0500	5.41	1.44
0.0311	0.0500	6.61	1.76
0.0402	0.0500	6.86	1.91
0.0500	0.0500	8.79	2.57
0.0500	0.0100	3.64	8.83
0.0500	0.0153	4.77	6.05
0.0500	0.0270	6.61	4.06
0.0500	0.0361	7.94	3.20
0.0500	0.0432	7.82	2.70
T=425 °C, Weight of catalyst=1.066 *g*			
0.00474	0.0500	5.02	2.62
0.0136	0.0500	7.23	4.17
0.0290	0.0500	11.35	6.84
0.0400	0.0500	13.00	8.19
0.0500	0.0500	13.91	8.53
0.0500	0.0269	9.29	13.3
0.0500	0.0302	9.75	12.3
0.0500	0.0387	11.89	10.4

Catalytic Reduction of Nitric Oxide (SOLUTION)[4]

In equation 4.30 for the reaction rate $K_{NO} = A_2 exp\{-E_2/RT\}$ *atm^{-1}* is the adsorption equilibrium constant for NO, $K_{H2} = A_3 exp\{-E_3/RT\}$ *atm^{-1}* is the adsorption equilibrium constant for H_2 and $k = A_1 exp\{-E_1/RT\}$ *gmol/(min·g-catalyst)* is the forward reaction rate constant for surface reaction.

The objective of the estimation procedure is to determine the parameters k, K_{H2} and K_{NO} (if data from one isotherm are only considered) or the parameters A_1, A_2, A_3, E_1, E_2, E_3 (when all data are regressed together). The units of E_1, E_2, E_3 are in *cal/mol* and R is the universal gas constant (1.987 *cal/mol K*).

In this work, we first regressed the isothermal data. The estimated parameters from the treatment of the isothermal data are given in Table 16.6. An initial guess of (k_1=1.0, k_2=1.0, k_3=1.0) was used for all isotherms and convergence of the Gauss-Newton method without the need for Marquardt's modification was achieved in 13, 16 and 15 iterations for the data at 375, 400, and 425$°C$ respectively.

Plotting of $ln k_j$ (j=1,2,3) versus 1/T shows that only k_1 exhibits Arrhenius type of behavior. However, given the large standard deviations of the other two estimated parameters one cannot draw definite conclusions about these two parameters.

Table 16.6   Catalytic reduction of NO: Estimated Model Parameters   by the Gauss-Newton Method Using <u>Isothermal Data</u>

Temperature (°C)	$(k_1 \pm \sigma_{k1}) \times 10^4$	$k_2 \pm \sigma_{k2}$	$k_3 \pm \sigma_{k3}$
375	5.2 ± 1.2	18.5 ± 3.4	13.2 ± 3.4
400	5.5 ± 3.2	31.5 ± 13.0	35.9 ± 14.0
425	13.5 ± 8.0	25.9 ± 10.3	14.0 ± 8.9

---

[4] Ayen, R.J., and M.S. Peters, "Catalytic Reduction of Nitric Oxide", *Ind Eng Chem Proc Des Dev*, 1, 204-207 (1962).

Kittrell, J.R., R. Mezaki, and C.C. Watson, "Estimation of Parameters for Nonlinear Least Squares Analysis*", Ind. Eng. Chem*., 57(12), 18-27 (1965b).

Kittrell, J.R., W.G. Hunter, and C.C. Watson, "Nonlinear Least Squares Analysis of Catalytic Rate Models*", AIChE J*., 11(6), 1051-1057 (1965a).

Kittrell et al. (1965a) reported parameter estimates shown in <u>Tables</u> 16.7 and 16.8. Ayen and Peters (1962) have also reported values given here in <u>Table</u> 16.9.

Table 16.7   Catalytic reduction of NO: Estimated Model Parameters by Linear Least Squares Using Isothermal Data

Temperature (°C)	$(k_1 \pm \sigma_{k1}) \times 10^4$	$k_2 \pm \sigma_{k2}$	$k_3 \pm \sigma_{k3}$
375	4.9 ± 0.7	18.8 ± 4.6	14.6 ± 2.9
400	5.3 ± 8.5	38.6 ± 19.6	35.4 ± 11.3
425	8.8 ± 2.3	48.9 ± 31.3	30.9 ± 20.2

Source: Kittrell et al. (1965a).

Table 16.8   Catalytic reduction of NO: Estimated Model Parameters by Nonlinear Least Squares Using Isothermal Data

Temperature (°C)	$(k_1 \pm \sigma_{k1}) \times 10^4$	$k_2 \pm \sigma_{k2}$	$k_3 \pm \sigma_{k3}$
375	5.19 ± 0.9	18.5 ± 3.4	13.2 ± 3.4
400	5.51 ± 1.2	31.6 ± 12.9	36.0 ± 13.9
425	10.1 ± 3.0	34.5 ± 15.2	23.1 ± 11.6

Source: Kittrell et al. (1965a).

Table 16.9   Catalytic reduction of NO: Estimated Model Parameters by Nonlinear Least Squares

Temperature (°C)	$k_1 \times 10^4$	k_2	k_3
375	4.94	19.00	14.64
400	7.08	30.45	20.96
425	8.79	48.55	30.95

Source: Ayen and Peters (1962).

Kittrell et al. (1965a) also used all the data simultaneously to compute the parameter values. These parameter values are reported for each temperature and are given in Table 16.10.

Table 16.10   Catalytic reduction of NO: Estimated Model Parameters by Nonlinear Least Squares Using Nonisothermal Data

Temperature (°C)	$(k_1 \pm \sigma_{k1}) \times 10^4$	$k_2 \pm \sigma_{k2}$	$k_3 \pm \sigma_{k3}$
375	4.92 ± 3.71	15.5 ± 13.4	17.5 ± 11.5
400	6.58 ± 3.94	26.3 ± 18.3	23.8 ± 12.8
425	8.63 ± 3.92	42.9 ± 23.6	31.7 ± 14.6

Source: Kittrell et al. (1965a).

Writing **Arrhenius-type expressions**, $k_j=A_j exp(-E_j/RT)$, for the kinetic constants, the mathematical model with six unknown parameters (A_1, A_2, A_3, E_1, E_2 and E_3) becomes

$$f(\mathbf{x,k})=\frac{A_1 e^{-\frac{E_1}{RT}} A_2 e^{-\frac{E_2}{RT}} A_3 e^{-\frac{E_3}{RT}} x_1 x_2}{\left(1+A_2 e^{-\frac{E_2}{RT}} x_2 + A_3 e^{-\frac{E_3}{RT}} x_1\right)^2} \quad (16.4)$$

The elements of the (*1×6*)-dimensional sensitivity coefficient matrix **G** are obtained by evaluating the partial derivatives:

$$G_{11}=\left(\frac{\partial f}{\partial A_1}\right)=\frac{A_2 A_3 e^{-\frac{E_1+E_2+E_3}{RT}} x_1 x_2}{Y^2} \quad (16.5)$$

$$G_{12}=\left(\frac{\partial f}{\partial A_2}\right)=\frac{-2A_1 A_2 A_3 e^{-\frac{E_1+2E_2+E_3}{RT}} x_1 x_2^2}{Y^3} \quad (16.6)$$
$$+\frac{A_1 A_3 e^{-\frac{E_1+E_2+E_3}{RT}} x_1 x_2}{Y^2}$$

$$G_{13} = \left(\frac{\partial f}{\partial A_3}\right) = \frac{-2A_1 A_2 A_3 e^{-\frac{E_1+E_2+2E_3}{RT}} x_1^2 x_2}{Y^3}$$

$$+ \frac{A_1 A_2 e^{-\frac{E_1+E_2+E_3}{RT}} x_1 x_2}{Y^2} \qquad (16.7)$$

$$G_{14} = \left(\frac{\partial f}{\partial E_1}\right) = \frac{-A_1 A_2 A_3 e^{-\frac{E_1+E_2+E_3}{RT}} x_1 x_2}{RTY^2} \qquad (16.8)$$

$$G_{15} = \left(\frac{\partial f}{\partial E_2}\right) = \frac{2A_1 A_2^2 A_3 e^{-\frac{E_1+2E_2+E_3}{RT}} x_1 x_2^2}{RTY^3}$$

$$- \frac{A_1 A_2 A_3 e^{-\frac{E_1+E_2+E_3}{RT}} x_1 x_2}{RTY^2} \qquad (16.9)$$

$$G_{16} = \left(\frac{\partial f}{\partial E_3}\right) = \frac{2A_1 A_2 A_3^2 e^{-\frac{E_1+E_2+2E_3}{RT}} x_1^2 x_2}{RTY^3}$$

$$- \frac{A_1 A_2 A_3 e^{-\frac{E_1+E_2+E_3}{RT}} x_1 x_2}{RTY^2} \qquad (16.10)$$

where

$$Y = 1 + A_2 e^{-\frac{E_2}{RT}} x_2 + A_3 e^{-\frac{E_3}{RT}} x_1 \qquad (16.11)$$

Three different sets of initial guesses which were given by Kittrell et al. (1965b) were used. The first two converged to local optima and the third diverged. The <u>lowest LS objective function</u> was obtained with the first initial guess and it was 0.1464×10^{-6}. The corresponding <u>estimated parameter values</u> were: $A_1=0.8039\pm0.3352$, $A_2=1.371\times10^5\pm6.798\times10^4$, $A_3=1.768\times10^7\pm8.739\times10^6$, $E_1=9520\pm0.4\times10^{-5}$, $E_2=11500\pm0.9\times10^{-7}$ and $E_3=17{,}900\pm0.9\times10^{-7}$.

In this problem it is very difficult to obtain convergence to the global optimum as the *condition number* of matrix **A** at the above local optimum is 3×10^{18}.

Even if this was the global optimum, a small change in the data would result in widely different parameter estimates since this parameter estimation problem appears to be fairly ill-conditioned.

<mark>Is anything that could be done to reduce the ill-conditioning of the problem?</mark>

The model given by Eqn 16.4 can be rewritten as

$$f(\mathbf{x}, \mathbf{k}) = \frac{A_1^* e^{-\frac{E_1^*}{RT}} x_1 x_2}{\left(1 + A_2 e^{-\frac{E_2}{RT}} x_2 + A_3 e^{-\frac{E_3}{RT}} x_1\right)^2} \qquad (16.12)$$

where

$$A_1^* = A_1 A_2 A_3 \qquad (16.13)$$

and

$$E_1^* = E_1 + E_2 + E_3 \qquad (16.14)$$

The <u>re-parameterized model</u> has the same number of unknown parameters (A_1^*, A_2, A_3, E_1^*, E_2 and E_3) as the original problem but a simpler structure.

This is expected to offer better convergence characteristics of the iterative estimation algorithm. <u>Indeed, convergence to the global optimum was obtained after many iterations using Marquardt's modification</u>. The value of Marquardt's parameter was always kept one order of magnitude greater than the smallest eigenvalue of matrix **A**. At the optimum, a value of zero for Marquardt's parameter was used and convergence was maintained.

The <u>LS objective function</u> was found to be 0.7604×10^{-9} (almost three orders of magnitude smaller than the one found earlier at a local optimum).

The estimated parameter values were: A_1=22.672, A_2=132.4, A_3=585320, E_1=13899, E_2=2439.6 and E_3=13506 where parameters A_1 and E_1 were estimated back from A_1^* and E_1^*. The re-parameterization "reduced" the ill-conditioning of the problem since the *condition number* of matrix **A** was now 5.6×10^8.

The model-calculated reaction rates are compared to the experimental data in Table 16.11 where it can be seen that the match is quite satisfactory.

Table 16.11   Catalytic reduction of NO: Experimental Measurements and Model Calculated Values (with **k** from Table 16.12)

P_{H2} (atm)	P_{NO} (atm)	Measured Rate $r \times 10^5$ gmol/(min·g-catalyst)	Calculated Rate $r \times 10^5$ gmol/(min·g-catalyst)
T=375°C, Weight of catalyst=2.39 g			
0.00922	0.0500	1.60	1.514
0.0136	0.0500	2.56	2.091
0.0197	0.0500	3.27	2.775
0.0280	0.0500	3.64	3.521
0.0291	0.0500	3.48	3.607
0.0389	0.0500	4.46	4.253
0.0485	0.0500	4.75	4.724
0.0500	0.00918	1.47	1.739
0.0500	0.0184	2.48	2.924
0.0500	0.0298	3.45	3.885
0.0500	0.0378	4.06	4.336
0.0500	0.0491	4.75	4.761
T=400°C, Weight of catalyst=1.066 g			
0.00659	0.0500	2.52	2.376
0.0113	0.0500	4.21	3.690
0.0228	0.0500	5.41	5.954
0.0311	0.0500	6.61	7.011
0.0402	0.0500	6.86	7.803
0.0500	0.0500	8.79	8.357
0.0500	0.0100	3.64	3.062
0.0500	0.0153	4.77	4.275
0.0500	0.0270	6.61	6.249
0.0500	0.0361	7.94	7.298
0.0500	0.0432	7.82	7.907
T=425°C, Weight of catalyst=1.066 g			
0.00474	0.0500	5.02	3.550
0.0136	0.0500	7.23	7.938
0.0290	0.0500	11.35	11.68
0.0400	0.0500	13.00	12.82
0.0500	0.0500	13.91	13.29
0.0500	0.0269	9.29	9.586
0.0500	0.0302	9.75	10.29
0.0500	0.0387	11.89	11.81

Based on the six estimated parameter values, the kinetic constants (k_1, k_2 and k_3) were computed at each temperature and they are shown in <u>Table</u> 16.12.

Table 16.12   Catalytic reduction of NO: Estimated Model Parameters by Nonlinear Least Squares Using All the Data.

Temperature (°C)	$k_1 \times 10^4$	k_2	k_3
375	4.65	19.91	16.29
400	6.94	21.36	24.05
425	10.07	22.80	34.53

NOTE: Having found the optimum, we returned back to the original structure of the problem and used an initial guess fairly close to the global optimum.

In this case the parameters converged very close to the optimum where the LS <u>objective function</u> was 0.774×10^{-9}.

The *condition number* of matrix **A** was found to be 1.7×10^{13} which is about 5 orders of magnitude higher that the one for the re-parameterized formulation calculated at the same point.

<u>Thus, re-parameterization should be seriously considered for hard to converge problems</u>.

<u>Model</u>: $\quad \dfrac{dC_{HPA}}{dt} = -[r_1 + r_2]C_k - [r_3 + r_4 - r_{-3}]$ (6.59a)

$$\frac{dC_{PD}}{dt} = [r_1 - r_2]C_k \qquad (6.59b)$$

$$\frac{dC_{Ac}}{dt} = r_3 - r_4 - r_{-3} \qquad (6.59c)$$

where C_k is the concentration of the catalyst (10 *g/L*) and the reaction rates are

$$r_1 = \frac{k_1 P C_{HPA}}{H\left[1 + \left(\dfrac{K_1 P}{H}\right)^{0.5} + K_2 C_{HPA}\right]^3} \qquad (6.60a)$$

$$r_2 = \frac{k_2 C_{PD} C_{HPA}}{1 + \left(\dfrac{K_1 P}{H}\right)^{0.5} + K_2 C_{HPA}} \qquad (6.60b)$$

$$r_3 = k_3 C_{HPA} \qquad (6.60c)$$

$$r_{-3} = k_{-3} C_{Ac} \qquad (6.60d)$$

$$r_4 = k_4 C_{Ac} C_{HPA} \qquad (6.60e)$$

where
$\quad$ k_j (j=1, 2, 3, -3, 4) are rate constants *(L/(mol min g)*
$\quad$ K_1 and K_2 are the adsorption equilibrium constants *(L/mol)* for H_2 and
$\qquad$ HPA respectively
$\quad$ P is the hydrogen pressure *(MPa)* in the reactor
$\quad$ H is the Henry's law constant = 1379 *(L bar/mol)* at 298 *K*.

These are the seven model parameters (k_1, k_2, k_3, k_{-3}, k_4, K_1 and K_2) to be determined from the measured concentrations of HPA and PD.

---

[5] Zhu, X. D., G. Valerius, and H. Hofmann, , *Ind. Eng. Chem. Res*, 36, 3897-2902 (1997).

Using our standard notation we have:

$$\mathbf{x} = [x_1, x_2, x_3]^T = [C_{HPA}, C_{PD}, C_{Ac}]^T$$

$$\mathbf{k} = [k_1, k_2, k_3, k_4, k_5, k_6, k_7]^T = [k_1, k_2, k_3, k_{-3}, k_4, K_1, K_2]^T$$

$$\mathbf{y} = [y_1, y_2]^T = [C_{HPA}, C_{PD}]^T$$

Hence, the differential equation model takes the form,

$$\frac{dx_1}{dt} = -u_1(r_1 + r_2) - (k_3 x_1 + k_5 x_3 x_1 - k_4 x_3) \qquad (6.63a)$$

$$\frac{dx_2}{dt} = u_1(r_1 - r_2) \qquad (6.63b)$$

$$\frac{dx_3}{dt} = (k_3 x_1 - k_5 x_3 x_1 - k_4 x_3) \qquad (6.63c)$$

where

$$r_1 = \frac{k_1 u_2 x_1}{H\left[1 + \left(\dfrac{k_6 u_2}{H}\right)^{0.5} + k_7 x_1\right]^3} \qquad (6.64a)$$

$$r_2 = \frac{k_2 x_2 x_1}{1 + \left(\dfrac{k_6 u_2}{H}\right)^{0.5} + k_7 x_1} \qquad (6.64b)$$

$$r_3 = k_3 x_1 \qquad (6.64c)$$

$$r_{-3} = k_4 x_3 \qquad (6.64d)$$

$$r_4 = k_5 x_3 x_1 \qquad (6.64e)$$

u_1 is the concentration of catalyst present in the reactor (C_k)
u_2 the hydrogen pressure (P).

Recall that the sensitivity matrix, $\mathbf{G}$(t), is a *(3×7)-dimensional* matrix with elements given below.

$$\mathbf{G}(t) = \begin{bmatrix} G_{11}(t) & . & . & . & G_{17}(t) \\ G_{21}(t) & . & . & . & G_{27}(t) \\ G_{31}(t) & . & . & . & G_{37}(t) \end{bmatrix} = \begin{bmatrix} \left(\dfrac{\partial x_1}{\partial k_1}\right) & . & . & . & \left(\dfrac{\partial x_1}{\partial k_7}\right) \\ \left(\dfrac{\partial x_2}{\partial k_1}\right) & . & . & . & \left(\dfrac{\partial x_2}{\partial k_7}\right) \\ \left(\dfrac{\partial x_3}{\partial k_1}\right) & . & . & . & \left(\dfrac{\partial x_3}{\partial k_7}\right) \end{bmatrix}$$ (6.65b)

The three *state equations* (Eqn 6.63) must be solved together with the 21 *sensitivity equations* (Eqn 6.68) to yield $\mathbf{x}(t)$ and $\mathbf{G}(t)$ which are used in setting up matrix $\mathbf{A}$ and vector $\mathbf{b}$ at each iteration of the Gauss-Newton method.

SENSITIVITY COEFFICIENT Equations:

$$\frac{dG_{11}}{dt} = \left(\frac{\partial f_1}{\partial x_1}\right)G_{11} + \left(\frac{\partial f_1}{\partial x_2}\right)G_{21} + \left(\frac{\partial f_1}{\partial x_3}\right)G_{31} + \frac{\partial f_1}{\partial k_1} \; ; \quad G_{11}(0) = 0$$

$$\frac{dG_{21}}{dt} = \left(\frac{\partial f_2}{\partial x_1}\right)G_{11} + \left(\frac{\partial f_2}{\partial x_2}\right)G_{21} + \left(\frac{\partial f_2}{\partial x_3}\right)G_{31} + \frac{\partial f_2}{\partial k_1} \; ; \quad G_{21}(0) = 0$$

$$\vdots$$

$$\frac{dG_{17}}{dt} = \left(\frac{\partial f_1}{\partial x_1}\right)G_{17} + \left(\frac{\partial f_1}{\partial x_2}\right)G_{27} + \left(\frac{\partial f_1}{\partial x_3}\right)G_{37} + \frac{\partial f_1}{\partial k_7} \; ; \quad G_{17}(0) = 0$$

$$\frac{dG_{27}}{dt} = \left(\frac{\partial f_2}{\partial x_1}\right)G_{17} + \left(\frac{\partial f_2}{\partial x_2}\right)G_{27} + \left(\frac{\partial f_2}{\partial x_3}\right)G_{37} + \frac{\partial f_2}{\partial k_7} \; ; \quad G_{27}(0) = 0$$

$$\vdots$$

$$\frac{dG_{37}}{dt} = \left(\frac{\partial f_3}{\partial x_1}\right)G_{17} + \left(\frac{\partial f_3}{\partial x_2}\right)G_{27} + \left(\frac{\partial f_3}{\partial x_3}\right)G_{37} + \frac{\partial f_3}{\partial k_7} \; ; \quad G_{37}(0) = 0$$

(6.68)

SOLUTION

The seven parameters (k_1, k_2, k_3, k_{-3}, k_4, K_1 and K_2) are to be determined from the measured concentrations of HPA and PD versus time. In this example, we shall consider only the data gathered at one isotherm (318 *K*) and three pressures 2.6, 4.0 and 5.15 *MPa*. The experimental data are given in <u>Table</u> 16.23.

Table 16.23HPA Hydrogenation: Experimental Data Collected at 318 K and Pressure 2.6, 4.0 and 5.15 MPa

Pressure (*MPa*)	Time (*min*)	C_{HPA} (*mol/L*)	C_{PD} (*mol/L*)
2.6	10	1.37395	0.0
	20	1.25821	0.0197109
	30	1.18707	0.0642576
	40	1.13292	0.136399
	50	1.03556	0.238633
	60	0.961339	0.304599
	80	0.734436	0.492378
	100	0.564551	0.732326
	120	0.374385	0.887254
	140	0.214799	1.04284
	160	0.100976	1.17306
	180	0.0364192	1.25769
	200	0.00530892	1.26032
4.0	10	1.3295	0.00262812
	20	1.31157	0.0525624
	30	1.22828	0.120736
	40	1.087	0.241393
	50	0.994539	0.384888
	60	0.811825	0.4682
	80	0.600962	0.773193
	100	0.386302	0.990802
	120	0.204222	1.14954
	140	0.0782304	1.28
	160	0.0277708	1.29
	180	0.00316296	1.30
	200	0.00210864	1.30
5.15	10	1.36324	0.00262812
	20	1.25882	0.0700394
	30	1.17918	0.184363
	40	0.972102	0.354008
	50	0.825203	0.469777
	60	0.697109	0.607359
	80	0.421451	0.852431
	100	0.232296	1.03535
	120	0.128095	1.16413
	140	0.0289817	1.30053
	160	0.00962368	1.31971

Zhu et al. (1997) minimized an un-weighted sum of squares of deviations of calculated and experimental concentrations of HPA and PD. They used Marquardt's modification of the Gauss-Newton method and reported the parameter estimates shown in Table 16.24.

We employed the G-N method and starting with the values given by Zhu et al. (1997) as an initial guess, the LS objective function was computed using the identity matrix as a weighting matrix. The LS objective function was 0.26325 and the corresponding condition number of matrix **A** was 0.345×10^{17}. We used the formulation with the scaled matrix **A**. Hence, the magnitude of the computed *condition number* of matrix **A** is solely due to ill-conditioning of the problem.

Indeed, the ill-conditioning of this problem is quite severe. Using as initial guess the parameter values reported by Zhu et al. (1997) we were unable to converge. A tight test of convergence was used (NSIG=5) because otherwise it may appear that the algorithm has converged.

In problems like this one which are very difficult to converge, we use Marquardt's modification first to reduce the objective function as much as possible. Then we approach closer to the global minimum in a *sequential way by varying only one or two parameters at a time*.

The results are summarized in Table 16.24. Figures 16.3a, 16.3b and 16.3c show the data and the model calculations based on the values reported by Zhu et al. (1997) and from our work ($\mathbf{k}^* = [13.502, 0.236 \times 10^{-8}, 0.3922 \times 10^{-3}, 0.126 \times 10^{-5}, 0.0273, 4.3531, 191.30]^T$). As seen, the difference between the two model calculations is very small and all the gains realized in the LS objective function (from 0.26325 to 0.21610) produce a slightly better match of the HPA and PD transients at 5.15 *MPa*.

Table 16.24 HPA Hydrogenation: Systematic Estimation of Parameter Values Using the Data Collected at 318 K

	k_1	k_2	k_3	k_3	k_4	K_1	K_2	LS Object. Function	cond(A)
Initial Guess	10^{-3}	10^{-3}	10^{-3}	10^{-3}	10^{-3}	10^{-3}	10^{-3}	40.654	
G-N Marquardt, $\gamma=10^{-4}$	2.6866	0.108×10^{-6}	0.672×10^{-3}	0.68×10^{-5}	0.0273	35.56	2.57	0.30452	
G-N Marquardt, $\gamma=10^{-10}$	2.6866 ±22.2%	0.236×10^{-8} ±0.37×10^{8}%	0.672×10^{-3} ±352%	0.126×10^{-5} ±0.3×10^{7}%	0.0273 ±1820%	35.56 ±30.3%	2.57 ±14.5%	0.30447	
Optimizing K_2	*	*	*	*	*	*	2.5322	0.27807	
Optimizing k_1	2.7397	*	*	*	*	*	*	0.26938	
Optimizing K_2 & k_1	3.0436	*	*	*	*	*	2.7119	0.26753	
Optimizing K_2, k_1 & K_1	13.354 ±3.1%	*	*	*	*	4.5435 ±3.6%	172.25 ±9.3%	0.24357	
Optimizing K_2, k_1, K_1 & k_3	13.502 ±8.8%	*	0.3922×10^{-3} ±135%	*	*	4.3531 ±9.8%	191.30 ±8.5%	0.21610	0.62×10^{5}
Optimizing all except k_2,k_3	13.502 ±9.5%	*	0.3922×10^{-3} ±226%	*	0.0273 ±1452%	4.3531 ±10.9%	191.30 ±8.8%	0.21610	0.11×10^{13}
Optimizing all except k_2	13.502 ±21.1%	*	0.3922×10^{-3} ±445%	0.126×10^{-5} ±0.6×10^{6}%	0.0273 ±2535%	4.3531 ±12.7%	191.30 ±18.0%	0.21610	0.12×10^{17}
Optimizing all 7 parameters	13.502 ±21.1%	0.236×10^{-8} ±0.6×10^{8}%	0.3922×10^{-3} ±547%	0.126×10^{-5} ±0.4×10^{7}%	0.0273 ±2857%	4.3531 ±12.7%	191.30 ±19.4%	0.21610	0.12×10^{17}
Parameters reported by Zhu et al. (1997)	6.533 ±0.045	3.048×10^{-4} ±1.07×10^{-4}	6.233×10^{-6}	7.219×10^{-4}	3.902×10^{-6}	95.00 ±1.28	3.227 ±0.033	0.26325	0.34×10^{17}

* Parameters with a star are assumed to be known and have a constant value equal to the one determined in the previous run.
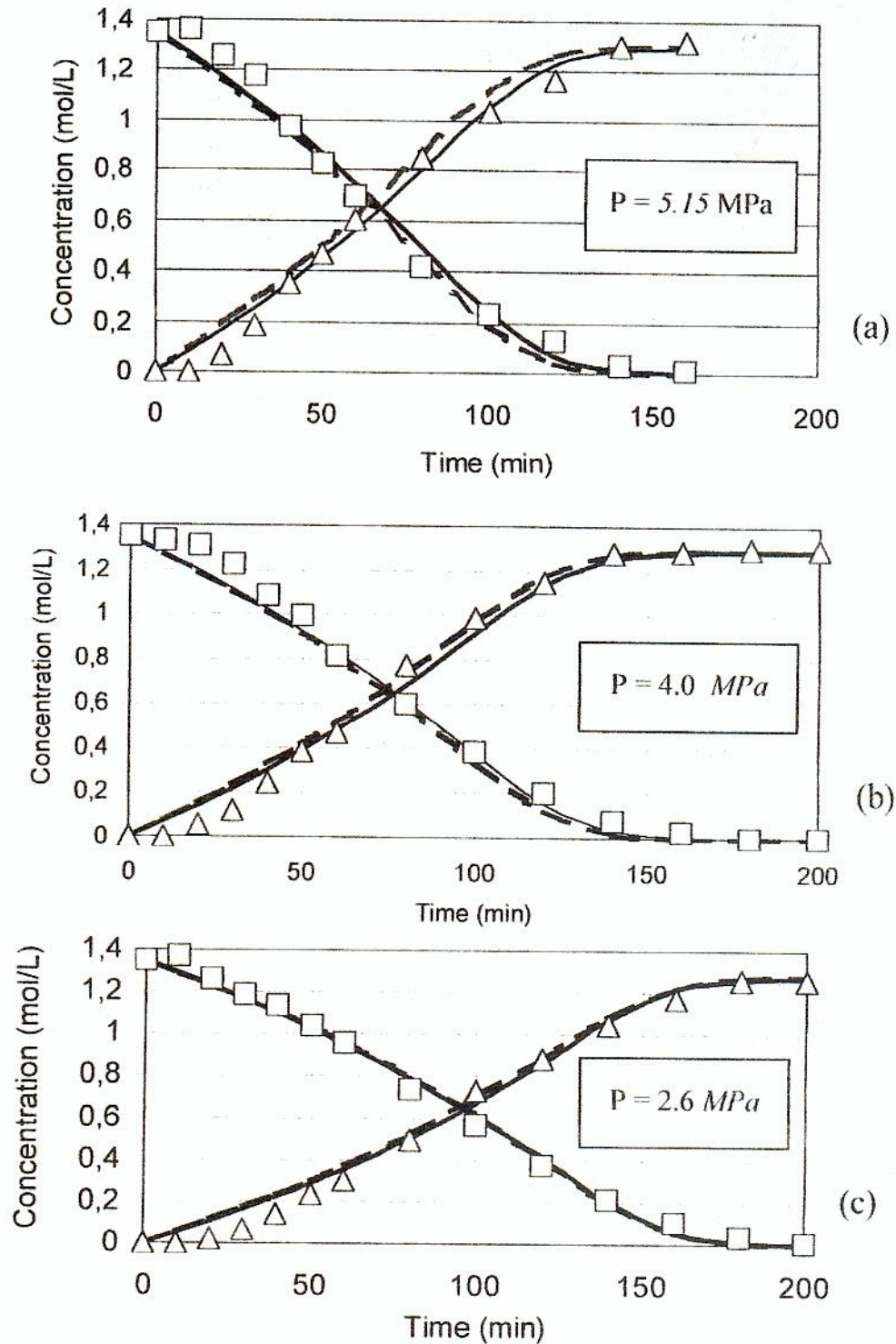
*Figure 16.3 HPA Hydrogenation: Experimental data at 318 K of HPA ( )
and PD (Δ) concentrations compared to model calculated values
(- - - - parameters by Zhu et al., 1997; ——— parameters from
this work) at (a) 5.15 MPa, (b) 4.0 MPa and (c) 2.6 MPa.*