1. (a) The averages needed are given in the table below.

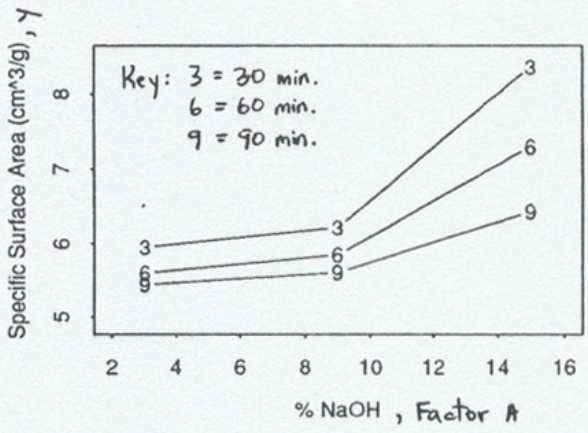
% NaOH (Factor A)	TIME (Factor B) 30	60	90	
3.0	$\bar{y}_{11} = 5.95$	$\bar{y}_{12} = 5.60$	$\bar{y}_{13} = 5.44$	$\bar{y}_{1\cdot} = 5.663$
9.0	$\bar{y}_{21} = 6.22$	$\bar{y}_{22} = 5.85$	$\bar{y}_{23} = 5.61$	$\bar{y}_{2\cdot} = 5.893$
15.0	$\bar{y}_{31} = 8.36$	$\bar{y}_{32} = 7.30$	$\bar{y}_{33} = 6.43$	$\bar{y}_{3\cdot} = 7.363$
	$\bar{y}_{\cdot 1} = 6.843$	$\bar{y}_{\cdot 2} = 6.250$	$\bar{y}_{\cdot 3} = 5.827$	$\bar{y}_{\cdot\cdot} = 6.307$

The fitted main effects are

$a_1 = \bar{y}_{1\cdot} - \bar{y}_{\cdot\cdot} = -.643$
$a_2 = \bar{y}_{2\cdot} - \bar{y}_{\cdot\cdot} = -.413$
$a_3 = \bar{y}_{3\cdot} - \bar{y}_{\cdot\cdot} = 1.057$
$b_1 = \bar{y}_{\cdot 1} - \bar{y}_{\cdot\cdot} = .537$
$b_2 = \bar{y}_{\cdot 2} - \bar{y}_{\cdot\cdot} = -.057$
$b_3 = \bar{y}_{\cdot 3} - \bar{y}_{\cdot\cdot} = -.480$

The fitted interactions are

$$ab_{11} = \bar{y}_{11} - (\bar{y}.. + a_1 + b_1) = -.250$$
$$ab_{12} = \bar{y}_{12} - (\bar{y}.. + a_1 + b_2) = -.007$$
$$ab_{13} = \bar{y}_{13} - (\bar{y}.. + a_1 + b_3) = .257$$
$$ab_{21} = \bar{y}_{21} - (\bar{y}.. + a_2 + b_1) = -.210$$
$$ab_{22} = \bar{y}_{22} - (\bar{y}.. + a_2 + b_2) = .013$$
$$ab_{23} = \bar{y}_{23} - (\bar{y}.. + a_2 + b_3) = .197$$
$$ab_{31} = \bar{y}_{31} - (\bar{y}.. + a_3 + b_1) = .460$$
$$ab_{32} = \bar{y}_{32} - (\bar{y}.. + a_3 + b_2) = -.007$$
$$ab_{33} = \bar{y}_{33} - (\bar{y}.. + a_3 + b_3) = -.453.$$



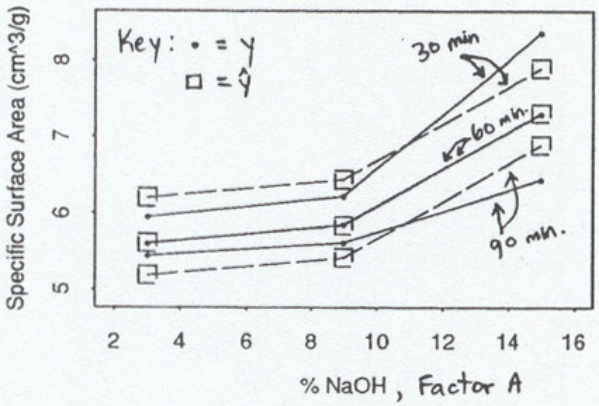
The fitted interactions ab_{31} and ab_{33} are large (relative to the fitted main effects) indicating that the effect on y of changing NaOH from 9% to 15% depends on the Time (non-parallelism in the plot). The a's are somewhat larger than the b's, indicating that Time has a slightly smaller overall effect than %NaOH. Overall, increasing Time decreases the specific surface area and increasing % NaOH increases the specific surface area. However, in each case the size of the change depends on the level of the other factor. It would not be wise to use the fitted main effects alone to summarize the data, since there may be an importantly large interaction between the two factors.

(b) For the factor-level combination with Factor A at level i and Factor B at level j, the fitted/predicted response for a "main effects only" model is computed as

$$\hat{y} = \bar{y}.. + a_i + b_j$$

The 9 fitted/predicted responses are given below.

$$\hat{y}_{11} = \bar{y}.. + a_1 + b_1 = 6.20$$
$$\hat{y}_{12} = \bar{y}.. + a_1 + b_2 = 5.61$$
$$\hat{y}_{13} = \bar{y}.. + a_1 + b_3 = 5.18$$
$$\hat{y}_{21} = \bar{y}.. + a_2 + b_1 = 6.43$$
$$\hat{y}_{22} = \bar{y}.. + a_2 + b_2 = 5.84$$
$$\hat{y}_{23} = \bar{y}.. + a_2 + b_3 = 5.41$$
$$\hat{y}_{31} = \bar{y}.. + a_3 + b_1 = 7.90$$
$$\hat{y}_{32} = \bar{y}.. + a_3 + b_2 = 7.31$$
$$\hat{y}_{33} = \bar{y}.. + a_3 + b_3 = 6.88$$



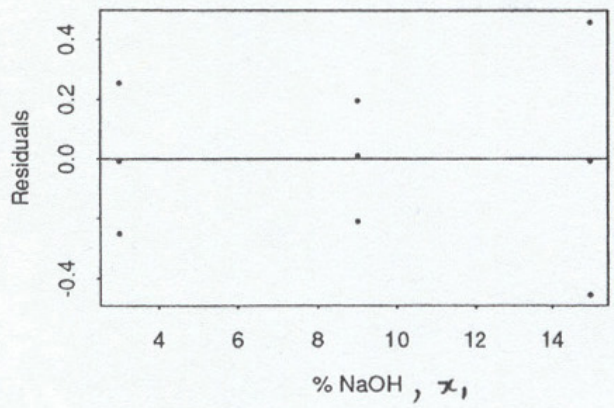
Like the plot in Chapter 4, section 2, problem 2 the fitted values for each level of B (x_2) must be parallel; no interactions are allowed. However, unlike Ch. 4, sec. 2, prob. 2, the current model allows these fitted values to be non-linear in x_1. Factorial models are generally more flexible than lines, curves, and surfaces.

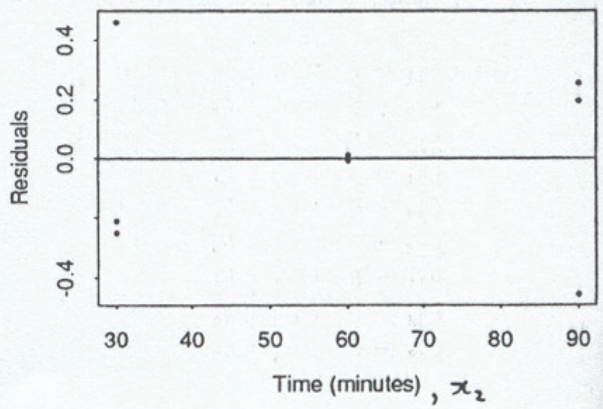
(c) The computations are given in the table below. (Note: $\bar{y}$ represents the average of all observations. It is equal to $\bar{y}_{..}$ only because all the sample sizes are equal.)

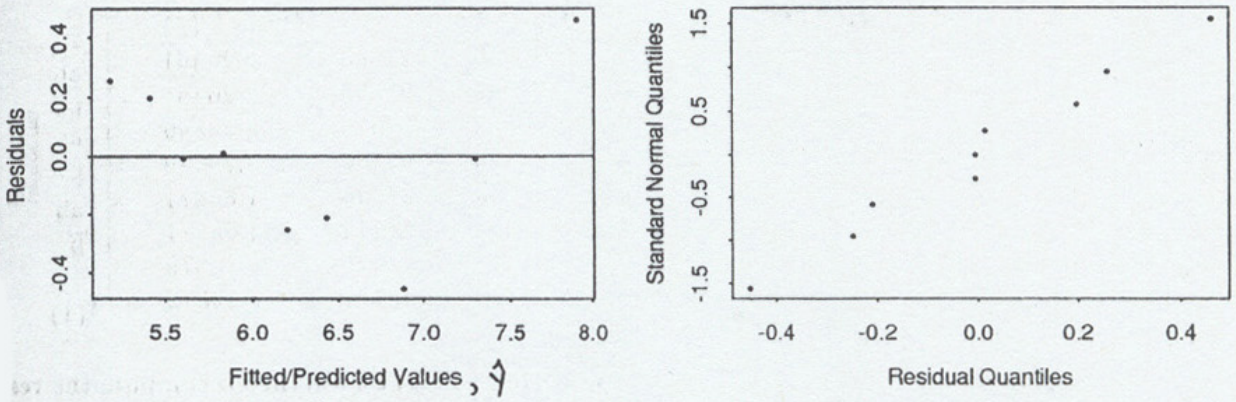
i	y_i	$\bar{y}$	$\hat{y}_i$	$(y_i - \bar{y})$	$(y_i - \bar{y})^2$	$e_i = (y_i - \hat{y}_i)$	$(y_i - \hat{y}_i)^2$
1	5.95	6.307	6.200	−.357	.127	−.250	.063
2	5.60	6.307	5.607	−.707	.499	−.007	.000
3	5.44	6.307	5.183	−.867	.751	.257	.066
4	6.22	6.307	6.430	−.087	.008	−.210	.044
5	5.85	6.307	5.837	−.457	.209	.013	.000
6	5.61	6.307	5.413	−.697	.485	.197	.039
7	8.36	6.307	7.900	2.053	4.216	.460	.212
8	7.30	6.307	7.307	.993	.987	−.007	.000
9	6.43	6.307	6.883	.123	.015	−.453	.206
					7.2972		.6285

$$R^2 = \frac{\sum(y_i - \bar{y})^2 - \sum(y_i - \hat{y}_i)^2}{\sum(y_i - \bar{y})^2} = \frac{7.2972 - .6285}{7.2972} = .914.$$

The residuals $e_i = y_i - \hat{y}_i$ are given in the table above.







The plots of residuals versus Time and residuals versus $\hat{y}_i$ both have patterns; these show that the "main effects only" model is not accounting for the apparent interaction between the two factors. Even though R^2 is higher than both of the models in 4-8, this model does not seem to be adequate.

2. (a) Using the Yates algorithm:

Combination	$\bar{y}$	Cycle 1	Cycle 2	Cycle 3	Cycle 3 ÷ 8	
(1)	21.0100	42.8333	54.843	166.337	20.7921	$= \bar{y}...$
a	21.8233	12.0100	111.493	.903	.1129	$= a_2$
b	6.0067	95.5633	.810	−110.457	−13.8071	$= b_2$
ab	6.0033	15.9300	.093	−.690	−.0863	$= ab_{22}$
c	47.7900	.8133	−30.823	56.650	7.0812	$= c_2$
ac	47.7733	−.0033	−79.633	−.717	−.0896	$= ac_{22}$
bc	7.9100	−.0167	−.817	−48.810	−6.1012	$= bc_{22}$
abc	8.0200	.1100	.127	.943	.1179	$= abc_{222}$

Other fitted effects can be obtained by appropriately changing the signs of the fitted effects in the last column. Since b_2, c_2, and bc_{22} are relatively large, the simplest possible interpretation is that Diameter, Fluid, and their interaction are the only effects on Time. Generally, the .314 diameter (B +) results in shorter times than the .188 diameter (B −), and this is reflected by the negative sign of b_2. Also, ethylene glycol (C +) results in longer times than water (C −), and this is reflected by the positive sign of c_2. The negative sign of bc_{22} indicates that the decrease in time due to changing the diameter from .188 to .314 is smaller for water than it is for ethylene glycol. All of these observations are consistent with simple graphical summaries of the sample means.

(b)

Combination	$\bar{y}$	Cycle 1	Cycle 2	Cycle 3	Cycle 3 ÷ 8	
(1)	3.04497	6.12795	9.7130	21.5936	2.69920	$= \bar{y}...$
a	3.08298	3.58506	11.8806	.0507	.00634	$= a_2$
b	1.79282	7.73100	.0374	−6.1243	−.76554	$= b_2$
ab	1.79224	4.14960	.0133	−.0251	−.00314	$= ab_{22}$
c	3.86554	.03801	−2.5429	2.1676	.27095	$= c_2$
ac	3.86547	−.00058	−3.5814	−.0241	−.00301	$= ac_{22}$
bc	2.06811	−.00007	−.0386	−1.0385	−.12981	$= bc_{22}$
abc	2.08149	.01338	.0134	.0520	.00650	$= abc_{222}$

Yes, but the Diameter×Fluid interaction still seems to be important.

(c) Using the reverse Yates algorithm:

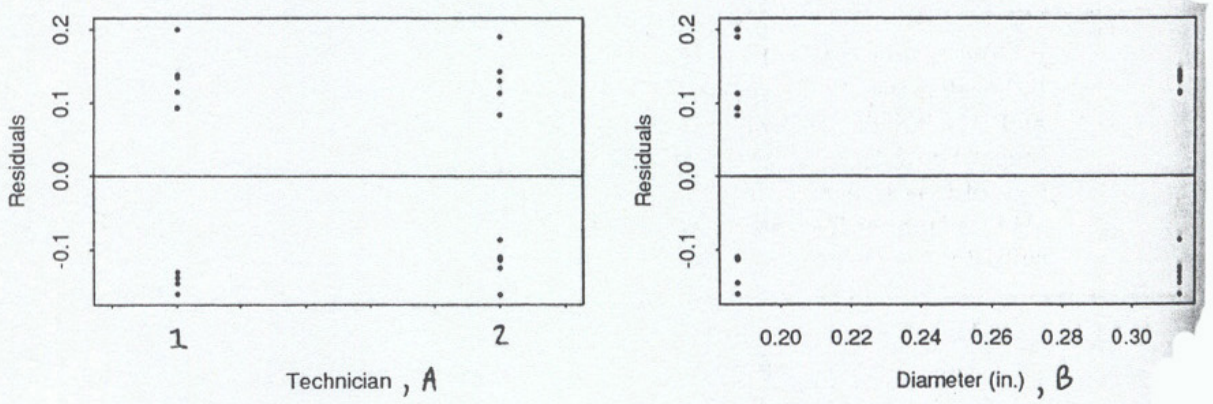
Fitted Effect	Value	Cycle 1	Cycle 2	Cycle 3 ($\hat{y}$)	
abc_{222}	0	0	.27095	2.20461	$= \hat{y}_{abc}$
bc_{22}	0	.27095	1.93366	2.20461	$= \hat{y}_{bc}$
ac_{22}	0	−.76554	.27095	3.73569	$= \hat{y}_{ac}$
c_2	.27095	2.69920	1.93366	3.73569	$= \hat{y}_{c}$
ab_{22}	0	0	.27095	1.66271	$= \hat{y}_{ab}$
b_2	−.76554	.27095	3.46474	1.66271	$= \hat{y}_{b}$
a_2	0	−.76554	.27095	3.19379	$= \hat{y}_{a}$
$\bar{y}...$	2.69920	2.69920	3.46474	3.19379	$= \hat{y}_{(1)}$

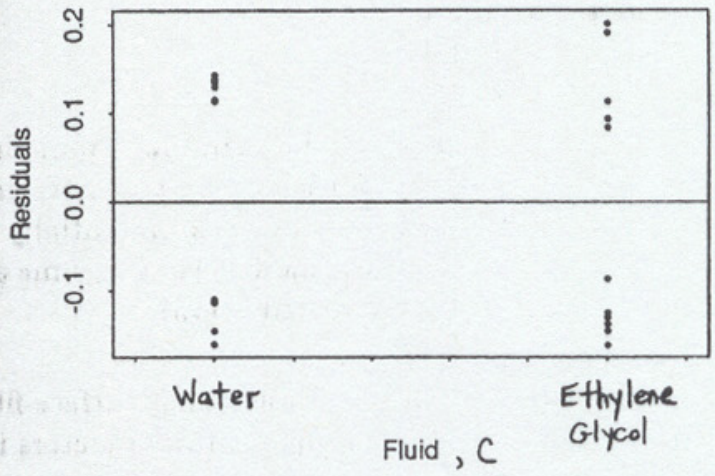
There will be a total of 24 residuals, one for each observation. To compute the residuals, take each (transformed) observation and subtract the $\hat{y}$ that corresponds to the factor-level combination from which the observation came. For example, $\hat{y}_{(1)} = 3.19379$ should be subtracted from the natural logs of each of the 3 observations from combination (1), 21.12, 21.11, and 20.80, producing the 3 residuals −.143569, −.110963, and .138995.

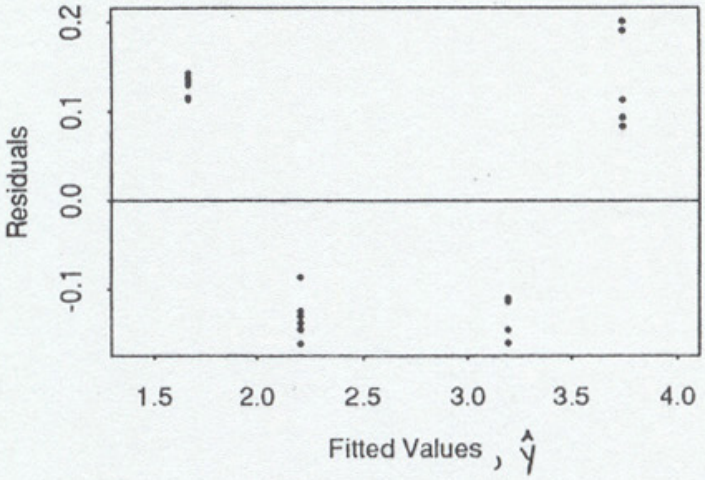
$\bar{y}$ = the average of all 24 observations = 2.69920. (This is equal to $\bar{y}...$ in this case because the data are balanced—all sample sizes are equal.) Use this and the 24 residuals e_i to compute R^2:

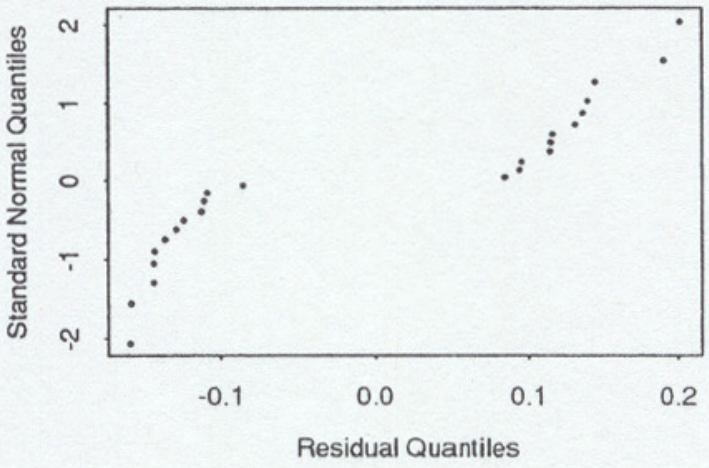
$$\begin{aligned} R^2 &= \frac{\sum(y_i - \bar{y})^2 - \sum(y_i - \hat{y}_i)^2}{\sum(y_i - \bar{y})^2} \\ &= \frac{\sum(y_i - \bar{y})^2 - \sum e_i^2}{\sum(y_i - \bar{y})^2} \\ &= \frac{16.25126 - .424197}{16.25126} = .974 \end{aligned}$$

A model with all factorial effects is a "saturated" model. The fitted/predicted values for this model will exactly match the $\bar{y}$'s for each factor-level combination. The residuals for this model are just the differences between each observation and the sample mean from its combination. The resulting R^2 is .995.









The plots of Residuals versus Technician, Diameter, and Fluid all show that there is a gap in the residuals; there are no residuals near zero. The plot of residuals versus $\hat{y}$ shows a positive-negative-positive pattern. All of these plots show that the current model is inadequate (even though its R^2 is high). It does not account for the apparent interaction between Diameter and Fluid.

(d) If you believe that there are no interactions, there is an approximate $b_1 - b_2 = 1.532$ ln(sec) decrease in log drain time. The change in raw drain time is then a multiplicative change. You would need to divide the .188 raw drain time by $e^{1.532}$ to get the .314 raw drain time. This suggests that (.188 drain time/.314 drain time) $= e^{1.532} = 4.63$; the theory predicts this ratio to be

$$\frac{\frac{1}{(\frac{.188}{2})^4}}{\frac{1}{(\frac{.314}{2})^4}} = 7.78.$$