Data Set 1B: *Mastocarpus* Photosynthesis *vs.* Salinity and Temperature, including higher temperatures

Statistical setting

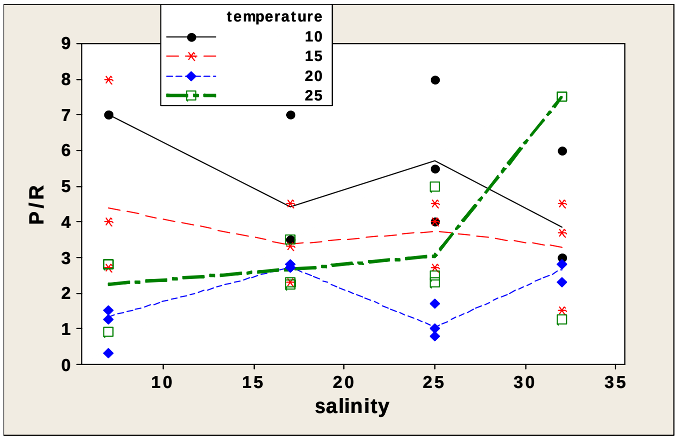
This handout uses the data presented in Data Set 1A, augmented by observations at a higher temperature. The new feature of the analysis is that the response to temperature now is clearly **nonlinear**, requiring a **quadratic** term in addition to the linear terms and interaction used in Data Set 1A.

Data

		temperature											
		10°			15°			20°			25°		
salinity	7%	7.0	7.0	7.0	2.7	4.0	8.0	0. 3	1.25	1.5	0.9	2.8	2.8
	17%	3.5	3.5	7.0	2.3	3.3	4.5	2.7	2.7	2.8	2.25	2.3	3.5
	25%	4.0	5.5	8.0	2.7	4.0	4.5	0.8	1.0	1.7	2.3	2.5	5.0
	32%	3.0	3.0	6.0	1.5	3.7	4.5	2.3	2.8	2.8	1.25	7.5	7.5

Data Exploration

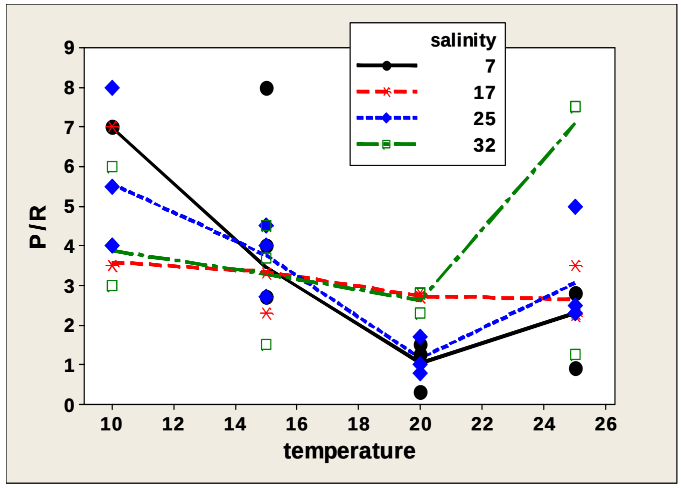
Graphical

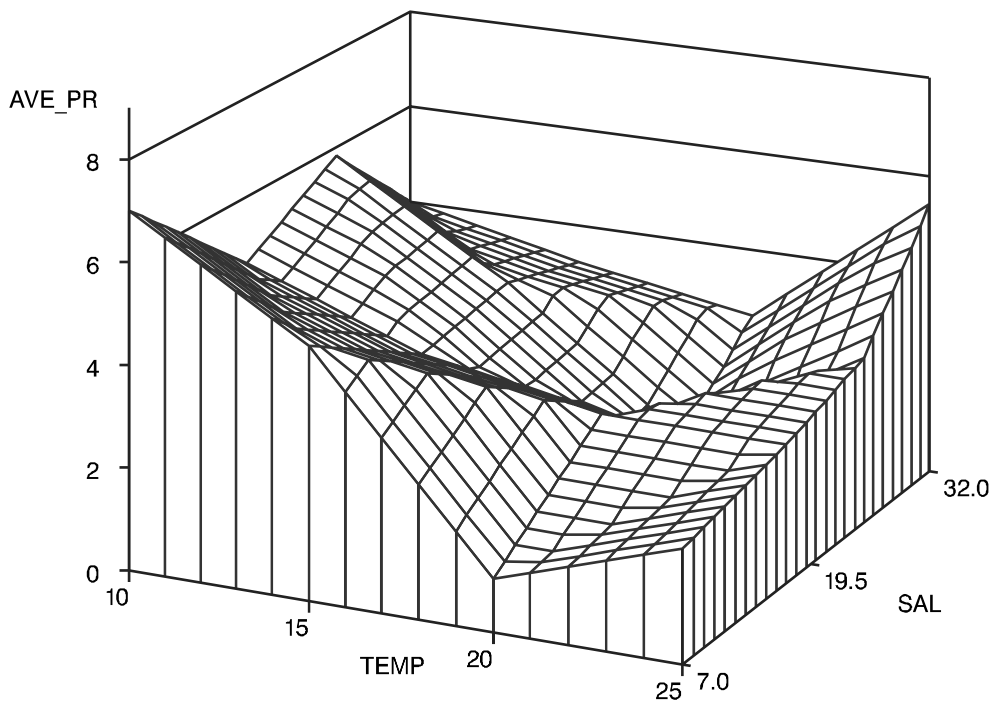


The graphs to the right and below add the data for 25° to those shown in Data Set 1A. The graph to the right, with salinity on the horizontal axis, shows that the added data continue one trend seen at the lower temperatures: from 10° to 20° the trend in P/R with increasing salinity progressively changes from negative to positive, and it is an even steeper positive trend at 25°. The other pattern seen at the lower temperatures, however, is not maintained: while the overall level of P/R decreases with increasing temperature from 10° to 20°, it is higher at 25° than 20°, and at 32% salinity P/R is far higher at 25° than any of the other temperatures.

This latter pattern is more clearly seen in the second graph, with P/R plotted against temperature. The downwards trend from 10° to 20° clearly reverses between 20° and 25° for all salinities but 17%, and at the latter, intermediate salinity there is very little effect of temperature throughout the range used.

Overall, the P/R ratio tends to be highest when salinity and temperature are either both high or both low. This outcome, as well as the more specific trends described above, are evident in a three-dimensional surface plot of the P/R means for each combination of salinity and temperature.





Summary Statistics

salinity	temp	mean	sd	min	median	max	range
7%	10°	7.00	0.000	7.0	7.00	7.0	0.0
	15°	4.90	2.760	2.7	4.00	8.0	5.3
	20°	1.02	0.633	0.3	1.25	1.5	1.2
	25°	**2.17**	**1.097**	**0.9**	**2.80**	**2.8**	**1.9**
17%	10°	4.67	2.020	3.5	3.50	7.0	3.5
	15°	3.37	1.102	2.3	3.30	4.5	2.2
	20°	2.73	0.058	2.7	2.70	2.8	0.1
	25°	**2.68**	**0.708**	**2.25**	**2.30**	**3.5**	**1.25**
25%	10°	5.83	2.020	4.0	5.50	8.0	4.0
	15°	3.73	0.929	2.7	4.00	4.5	2.8
	20°	1.17	0.473	0.8	1.00	1.7	0.9
	25°	**3.27**	**1.504**	**2.3**	**2.50**	**5.0**	**2.7**
32%	10°	4.00	1.730	3.0	3.00	6.0	3.0
	15°	3.23	1.553	1.2	3.70	4.5	3.3
	20°	2.63	0.289	2.3	2.80	2.8	0.5
	25°	**5.42**	**3.610**	**1.25**	**7.50**	**7.5**	**6.25**

These statistics quantify the trends described in the graphs, in particular the increase in mean or median P/R with the increase from 20° to 25°, as well as over the range of salinities at 25°. The statistics also show that the variability (standard deviations and ranges) in P/R shows similar trends to those of the mean and median.

Inference

Model

The preceding exploration of the data clearly shows nonlinearity of a sort that could not be eliminated by transformation: the trend in P/R changes sign over the range of temperatures. Analysis of these data therefore requires a model containing at least a quadratic term in temperature, and possibly in salinity as well. To reduce collinearity in this second-order model I centered the variables: the variables used were $s = S - 20.25$, and $t = T - 17.5$. The full, initial model thus was

$$Y_i = \beta_0 + \beta_1 s_i + \beta_2 t_i + \beta_{11} s_i^2 + \beta_{22} t_i^2 + \beta_{12} s_i t_i + \varepsilon_i$$

In the output below the variables are labelled as follows: s = ‘`salin_c`’; t = ‘`temp_c`’; $s \cdot t$ = ‘`inter_c`’; s^2 = ‘`salin_sq`’; and t^2 = ‘`temp_sq`’.

Hypothesis Tests — Full Model

ANOVA table for model:

Source	DF	SS	MS	F	P
Regression	5	96.289	19.258	7.58	**0.000**
Residual Error	42	106.660	2.540		
Total	47	202.949			

R-Sq = 47.4%

The overall model thus is highly significant and explains almost half of the variance in P/R.

Tests for individual parameters:

t tests:

Predictor	Coef	SE	T	P
Constant	2.4222	0.4669	5.19	0.000
salin_c	0.00673	0.02543	0.26	0.793
temp_c	-0.15792	0.04115	-3.84	**0.000**
inter_c	0.014944	0.004419	3.38	**0.002**
salin_sq	0.002703	0.003312	0.82	0.419
temp_sq	0.030625	0.009201	3.33	**0.002**

F tests (ANOVA table):

Source	DF	Seq. SS	Adj. SS	Adj. MS	F	P
salin_c	1	0.014	0.178	0.178	0.07	0.793
temp_c	1	37.407	37.407	37.407	14.73	**0.000**
inter_c	1	29.040	29.040	29.040	11.44	**0.002**
salin_sq	1	1.692	1.692	1.692	0.67	0.419
temp_sq	1	28.137	28.137	28.137	11.08	**0.002**
Error	42	106.660	106.660	2.540		
Total	47	202.949				

The quadratic and linear terms for temperature both are highly significant, as is the interaction. Neither term for salinity is significant in these added-last tests.

Note that since the variables were centered before the higher order terms were computed, so that there is very little collinearity, there are minimal differences between the added-in-order ("Seq SS") and added-last ("Adj SS") sums of squares.

Tests for second-order terms:

The significance of all the second-order terms (i.e. the quadratics and the interaction) can be tested by a multiple-partial-*F* test, using the added-in-order SSs from the preceding ANOVA table. The added SS for β_3, β_4, and β_5 is the sum of the last three sequential SSs, = 58.869, giving

$$F^* = \frac{SS/df}{MSE} = \frac{58.869/3}{2.53952} = 7.727 \qquad df = 3, 42 \qquad P = 0.0003$$

We therefore conclude that at least one of the second-order coefficients is not 0.

However, the added-last *t*-tests shown with the parameter estimates above clearly indicate that the quadratic term for salinity (β_{11}) is not significant. I therefore re-ran the analysis without this term.

Hypothesis Tests — Reduced Model

ANOVA table for model:

Source	DF	SS	MS	F	P
Regression	4	94.597	23.649	9.39	0.000
Residual Error	43	108.352	2.520		
Total	47	202.949			

R-Sq = 46.6%

The model still is highly significant, with a considerably larger *F* statistic than for the full model including the salinity-squared term, and explains nearly as much of the variance in P/R as did the full model. The MSE is very slightly smaller than for the full model (2.52 *vs*. 2.54).

Tests for individual parameters.

t tests:

Predictor	Coef	SE	T	P
Constant	2.6565	0.3668	7.24	0.000
salin_c	0.00181	0.02461	0.07	0.942
temp_c	-0.15792	0.04099	-3.85	**0.000**
inter_c	0.014944	0.004402	3.39	**0.001**
temp_sq	0.030625	0.009165	3.34	**0.002**

F tests (ANOVA table):

Source	DF	Seq. SS	Adj. SS	Adj. MS	F	P
salin_c	1	0.014	0.014	0.014	0.01	0.942
temp_c	1	37.407	37.407	37.407	14.84	**0.000**
inter_c	1	29.040	29.040	29.040	11.52	**0.001**
temp_sq	1	28.137	28.137	28.137	11.17	**0.002**
Error	43	108.352	108.352	2.520		
Total	47	202.949				

All terms involving temperature, including both second-order terms in the model, are highly significant in these added-last tests. The linear term for salinity still is not significant, but since salinity is contained in the interaction it should remain in the model. No further model reduction is appropriate.

Note that now without the salinity quadratic term all added-in-order SSes are identical to the added-last SSes: there is no collinearity.

Which Model?

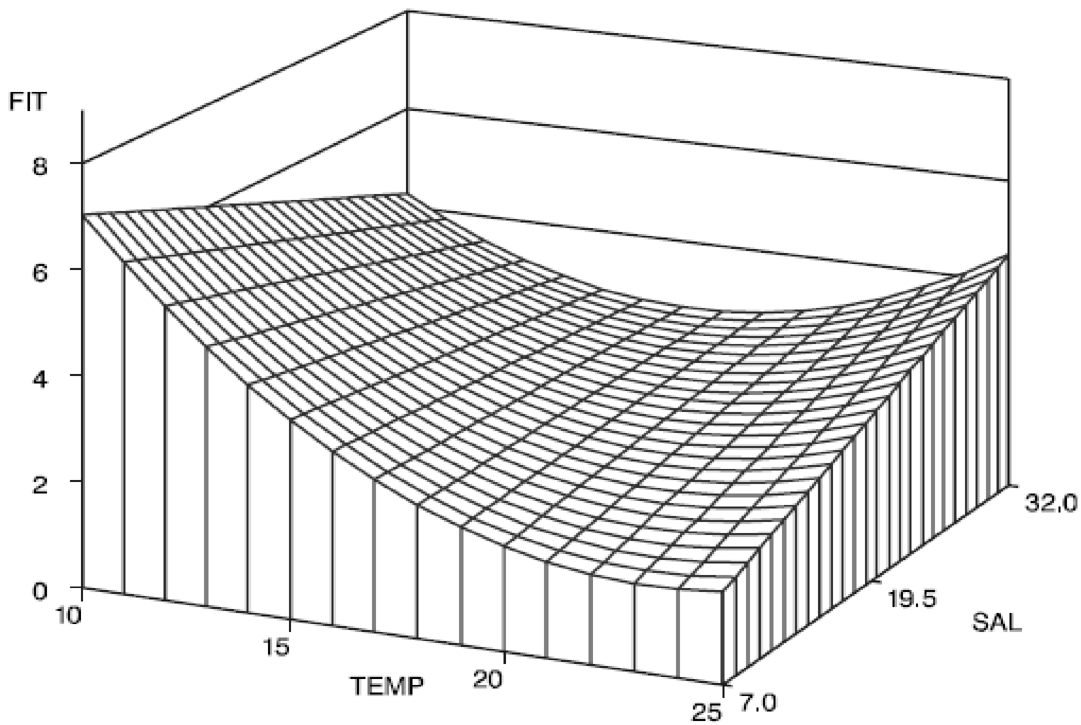
The full and reduced models fit the data nearly identically well. Furthermore, the coefficients for the temperature terms as well as the interaction are identical between the two models. The choice between them thus scarcely matters. My choice is to use the reduced model, preferring its slight advantage of simplicity and corresponding slightly smaller MSE.

Estimation

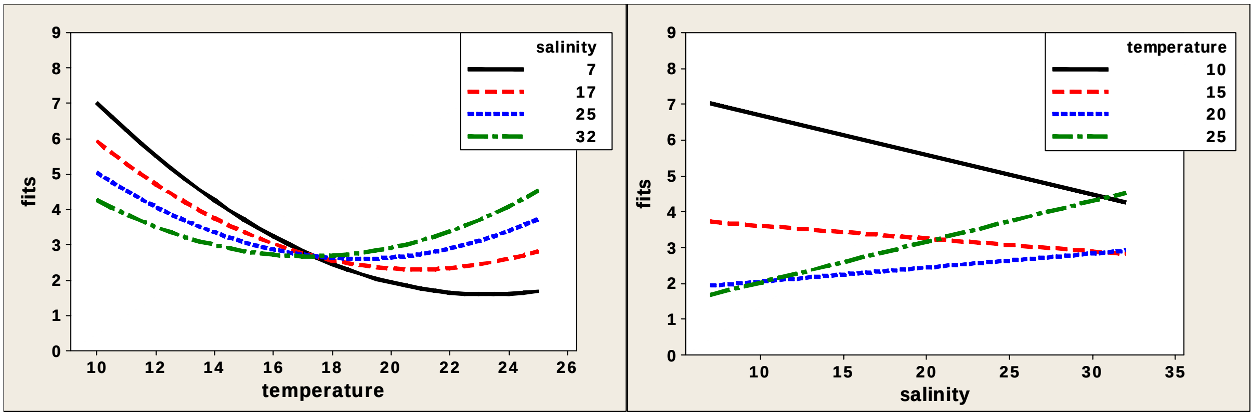
<u>The model:</u>

$$\hat{Y}_i = 2.6565 + 0.0018s_i - 0.1579t_i + 0.0306t_i^2 + 0.0149s_it_i$$

The following plots show the twisted and curved (in the temperature direction) surface produced by this estimated model. Comparison of this plot with the surface plot of the sample means (above) shows that this model describes the data reasonably well; it of course does not include the presumably random ups and downs in the data.



As in Handout 1A, this model can also be illustrated in plots of the fits *vs*. one of the independent variables, with separate lines for each level of the other independent variable. Of these plots (shown on next page), the one with temperature on the horizontal axis and separate lines for the four salinities is clearest: the partial parabola describing the relationship of (estimated) mean P/R to temperature shifts leftward and up as the salinity increases, such that the curves rotate around a common point at a temperature of about 17.5° and fits of about 2.6.

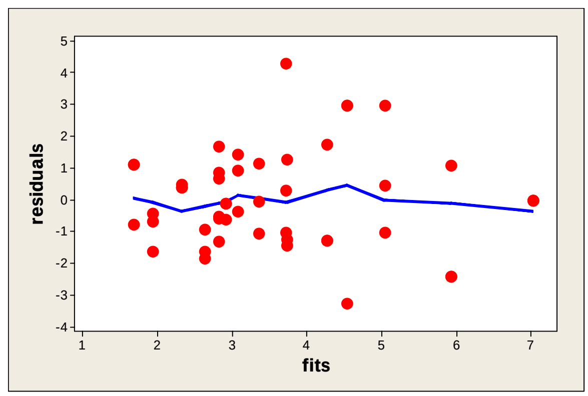


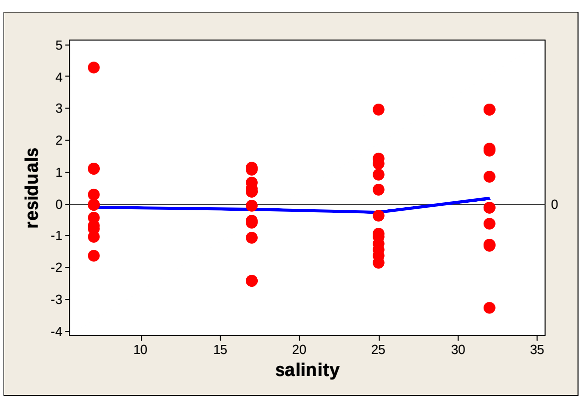
Diagnostics

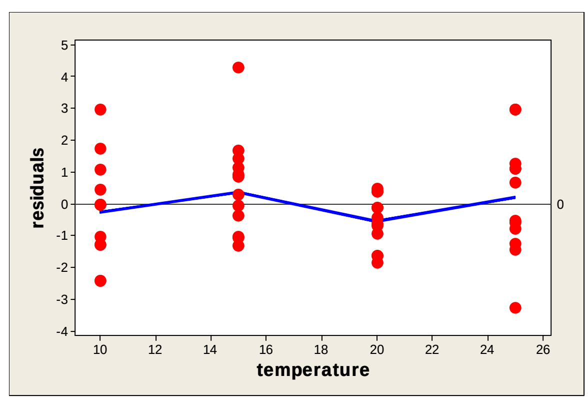
Residual Plots

<u>Residuals *vs.* variables:</u>

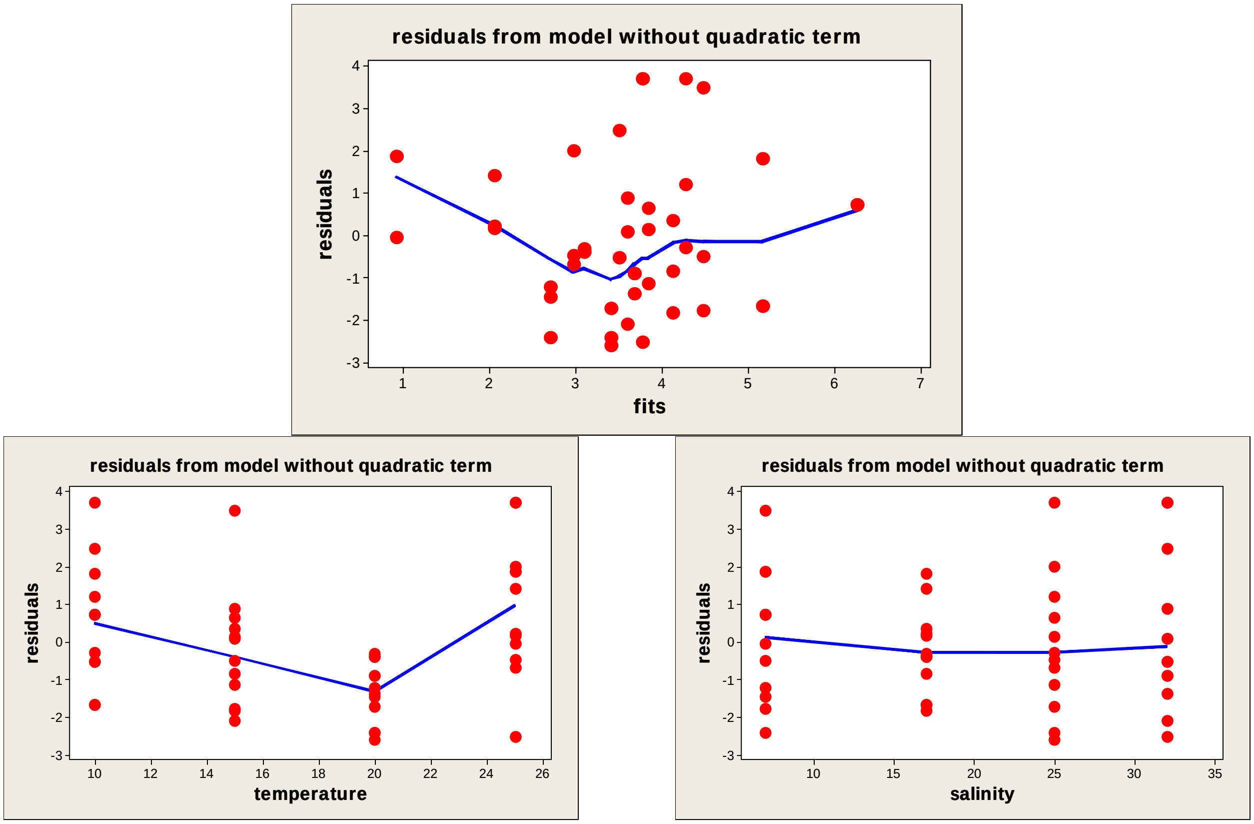
All three of the residual plots are acceptably flat (*i.e.* centered roughly around 0). It appears, however, that the variability is greater at the more extreme values of salinity and temperature; perhaps this is to be expected, and I don't think it is great enough heterogeneity to be a problem.



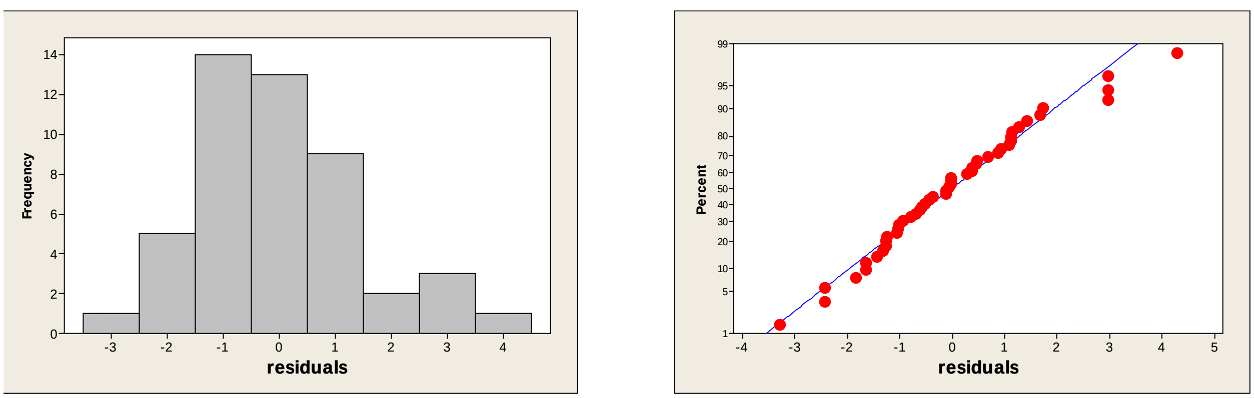




For comparison, the equivalent plots of residuals from fitting a model without the quadratic term in temperature (*i.e.* the same model as in Handout 1A) are shown on the next page. The plot of residuals against salinity is acceptable, but the plots of residuals *vs*. temperature and *vs*. the fits very clearly show the curvature that this model could not describe.



Distribution of residuals:



The residuals are mildly long-tailed; with this sample size this very slight non-normality is not a problem.

Lack-of-fit Test

Source	DF	SS	MS	F	P
Regression	4	94.597	23.649	9.39	0.000
Residual Error	43	108.352	2.520		
Lack of Fit	11	26.380	2.398	0.94	0.520
Pure Error	32	81.972	2.562		

The lack-of-fit test, being clearly non-significant, indicates that the chosen model is adequate, and that no higher-order terms are needed to describe further curvature or interaction.

Conclusion from Diagnostics

There are no substantial problems shown in these diagnostics, so I conclude that the analysis is valid.

Conclusion

The data produced in this experiment are well described by the model including an interaction between salinity and temperature (as in Handout 1A) as well as a quadratic term in temperature; there is little or no evidence of curvature in the salinity direction. Interestingly, this model (and the data it was derived from) indicate that photosynthesis was greater, relative to respiration, under the more extreme conditions: the *P/R* ratio was lowest for conditions ranging from high temperature but low salinity to moderate temperature and high salinity, and conversely the ratio was highest when both factors were at low levels or both were at high levels.

Since the entire model, and each term in it except the salinity effect (which was contained in the significant interaction), all were highly significant statistically ($P < 0.002$), and graphical diagnostics showed no major problems, the preceding conclusions can be taken as quite reliable.