Figure 1: Best fitted line for the response variables
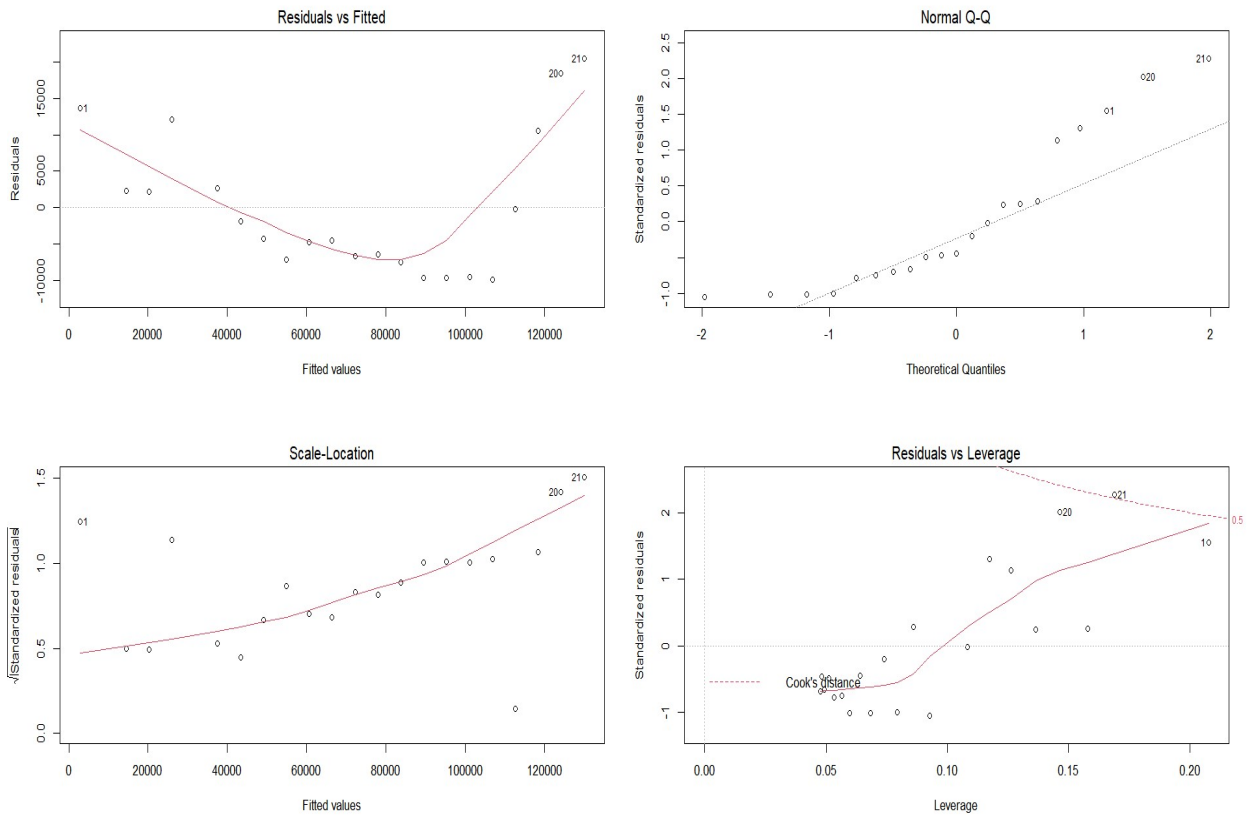

Figure 2: Diagnostic plots for model of total assessed species

24

Fitting of non linear models

Now since the diagnostic plots suggest a non-linear model, a number of non-linear models were fitted and the best possible model was taken into account.

For both the response variables certain models were fitted and the results have been summarised. The table indicates which assumptions have been violated leading to rejection of the model.

Model I:-

For modelling total species assessed:-

Model	Residuals vs Fitted values	Standardized residuals	Autocorrelation	Normality
$Y = a \cdot e^{bX}$	Yes	No	Yes	No
$Y = a \cdot b^X$	No	No	Yes	No
$Y = a \cdot X^b$	Yes	No	Yes	Yes
$Y = a + b \cdot X + c \cdot X^2$	Yes	No	Yes	No
$Y = a + b \cdot X + c \cdot X^2 + d \cdot X^3$	No	No	No	No

Thus, the cubic equation was overall best fit, where no assumptions were violated.
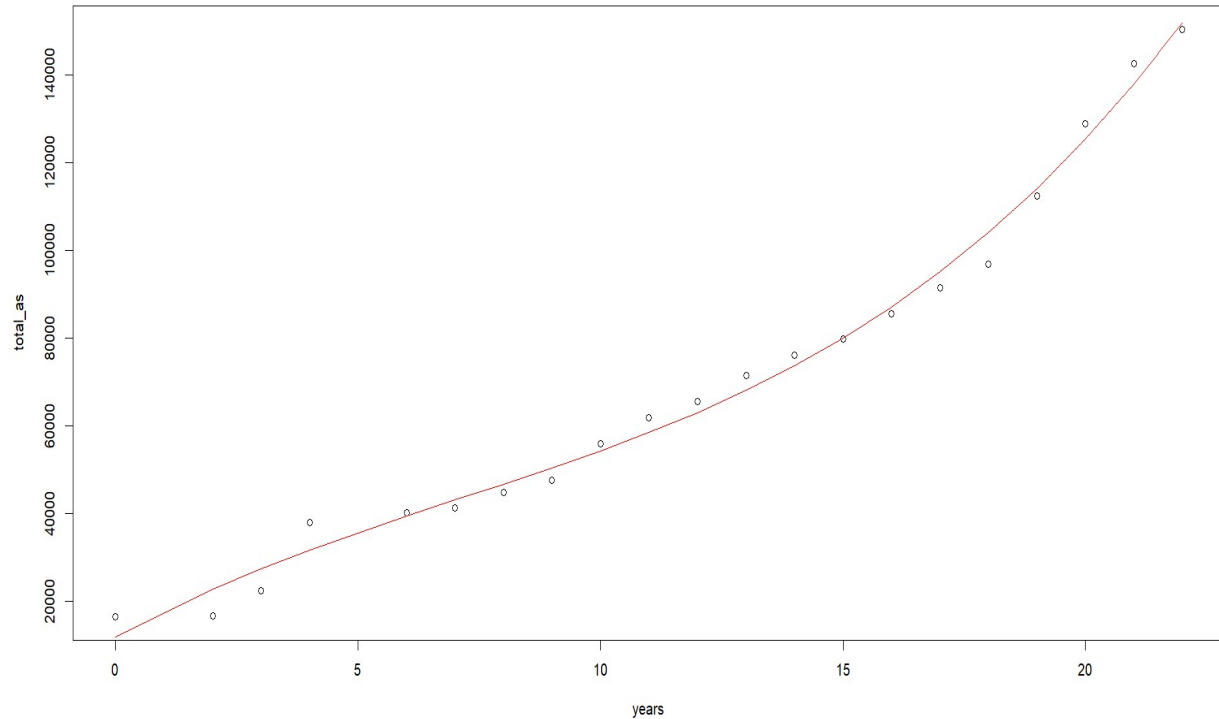

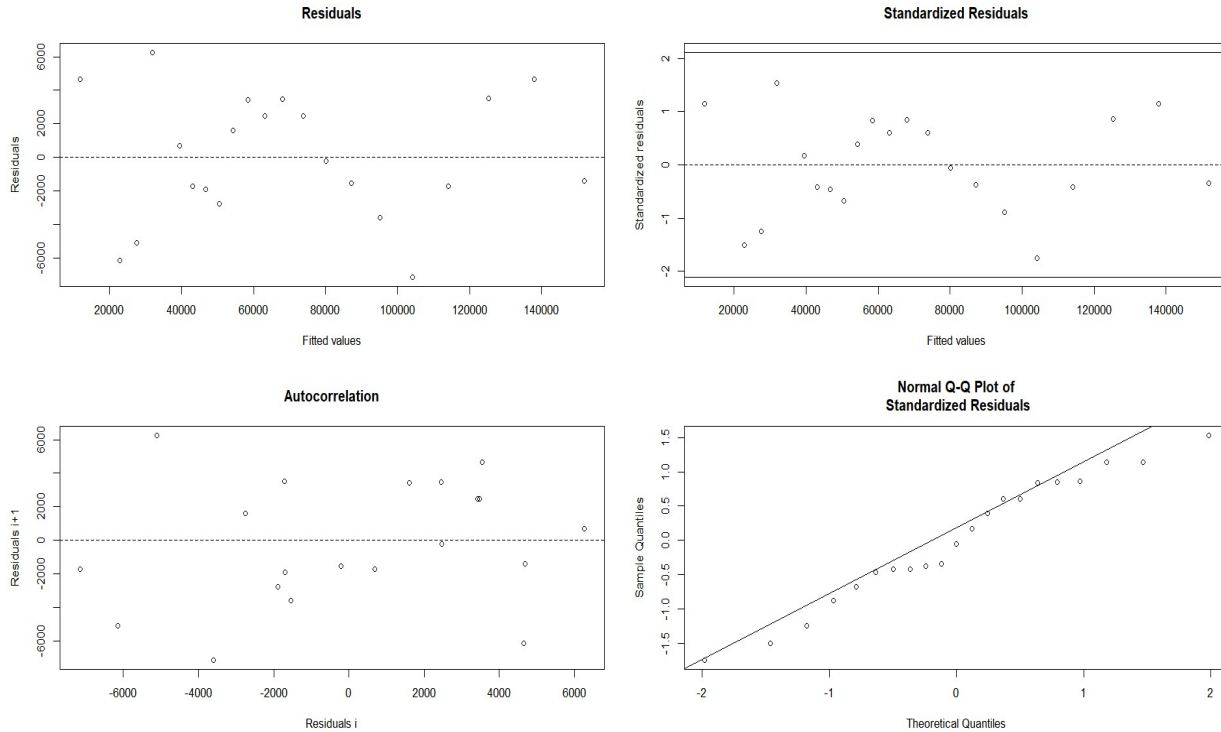Figure 3: Regression curve on actual values

Figure 4: Diagnostic plots for the model

Concise summary of the model and its interpretation is given as:-

```
> overview(model)

------
Formula: total_as ~ a + b * years + c * years^2 + d * years^3

Parameters:
   Estimate Std. Error t value Pr(>|t|)
a 11847.815   3376.729   3.509 0.002693 **
b  6142.485   1282.478   4.790 0.000171 ***
c  -355.686    134.181  -2.651 0.016813 *
d    16.620      3.953   4.204 0.000597 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 4083 on 17 degrees of freedom

Number of iterations to convergence: 1
Achieved convergence tolerance: 2.791e-08

------
Residual sum of squares: 2.83e+08

------
t-based confidence interval:
        2.5%       97.5%
a 4723.538988 18972.09190
b 3436.692166  8848.27768
c -638.783504   -72.58897
d    8.278848    24.96081

------
Correlation matrix:
          a          b          c          d
a  1.0000000 -0.8274881  0.6831721 -0.5913462
b -0.8274881  1.0000000 -0.9638657  0.9081558
c  0.6831721 -0.9638657  1.0000000 -0.9853145
d -0.5913462  0.9081558 -0.9853145  1.0000000
```

Interpretation:-

The fitted equation is,

$$Y = 11847.815 + 6142.485X - 355.686X^2 + 16.62X^3$$

Further,the t-values indicate that all the regression coefficients are statistically significant,that is all the coefficients used in the model are important for predicting the number of species that will be assessed in the coming years.

Thus the very same procedure used for the fitting of the above model,was employed while fitting the curve for total threatened species. It was overall found out that the best fitted curve is once again a cubic one,just like the one above. The summary tables have been drawn, along with the plots for better visualisation.

26

Model II:-

For modelling total species threatened:-

Model	Residuals vs Fitted values	Standardized residuals	Autocorrelation	Normality
$Y = a \cdot e^{bX}$	Yes	No	Yes	No
$Y = a \cdot b^X$	Yes	No	Yes	No
$Y = a \cdot X^b$	Yes	No	Yes	Yes
$Y = a + b \cdot X + c \cdot X^2$	Yes	No	Yes	No
$Y = a + b \cdot X + c \cdot X^2 + d \cdot X^3$	No	No	No	No

Thus,here too the cubic equation was overall best fit, where no assumptions were violated.
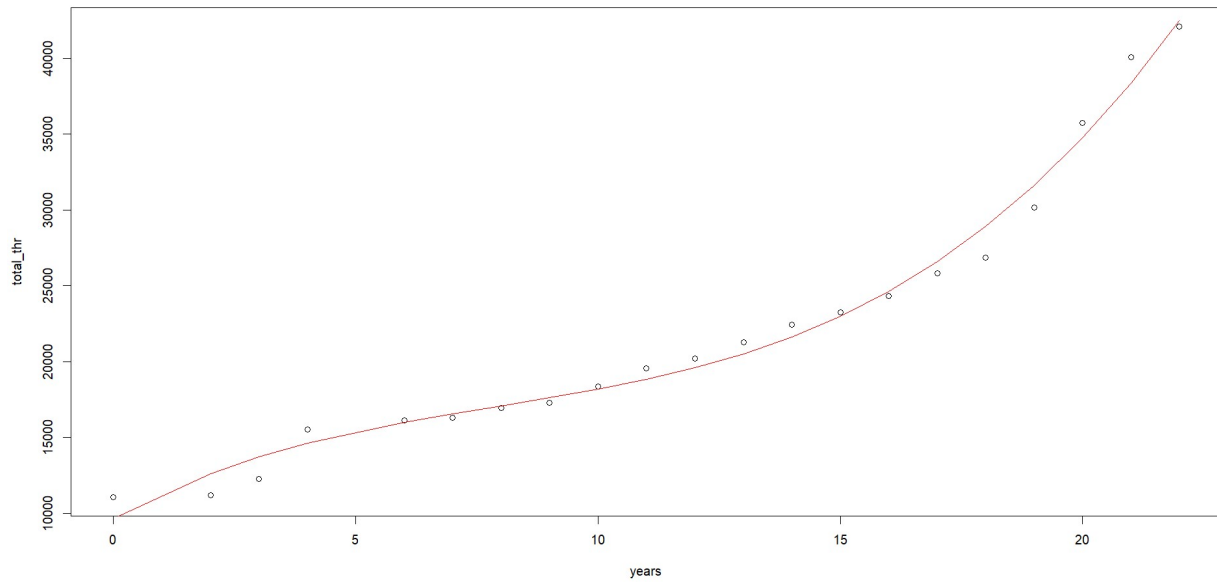

Figure 5: Regression curve on actual values

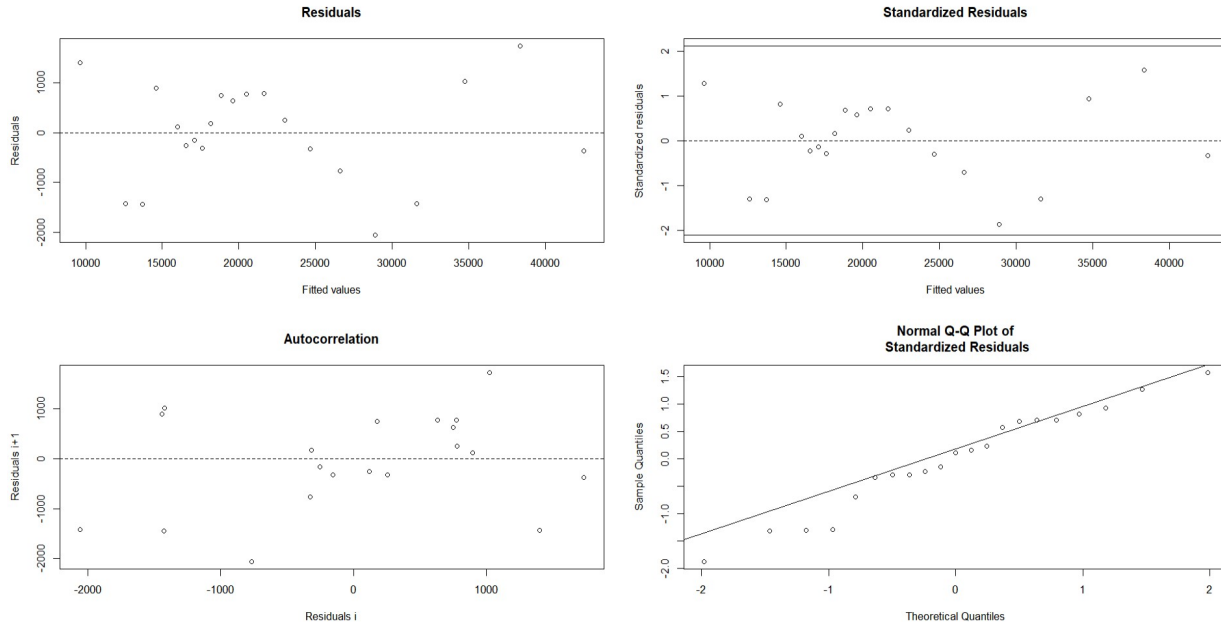Thus the diagnostic plots for the model, which validate the model are:-

Figure 6: Diagnostic plots for model

The final model and the interpretation are as follows

```
> overview(model)

------
Formula: total_thr ~ a + b * years + c * years^2 + d * years^3

Parameters:
  Estimate Std. Error t value Pr(>|t|)
a 9645.266    907.670  10.626 6.31e-09 ***
b 1761.458    344.732   5.110 8.72e-05 ***
c -156.398     36.068  -4.336 0.000449 ***
d    6.553      1.063   6.167 1.03e-05 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 1098 on 17 degrees of freedom

Number of iterations to convergence: 1
Achieved convergence tolerance: 2.475e-08

------
Residual sum of squares: 20500000

------
t-based confidence interval:
        2.5%         97.5%
a 7730.25035 11560.282640
b 1034.13699  2488.779243
c -232.49493   -80.300997
d    4.31118     8.795317

------
Correlation matrix:
           a          b          c          d
a  1.0000000 -0.8274881  0.6831721 -0.5913462
b -0.8274881  1.0000000 -0.9638657  0.9081558
c  0.6831721 -0.9638657  1.0000000 -0.9853145
d -0.5913462  0.9081558 -0.9853145  1.0000000
```

Interpretation:-
The fitted equation is,

$$Y = 965.266 + 1761.458X - 156.398X^2 + 6.553X^3$$

Further, the t-values indicate that all the regression coefficients are statistically significant,that is all the coefficients used in the model are important for predicting the number of species that will be labeled threatened out of the total species assessed in the coming years.