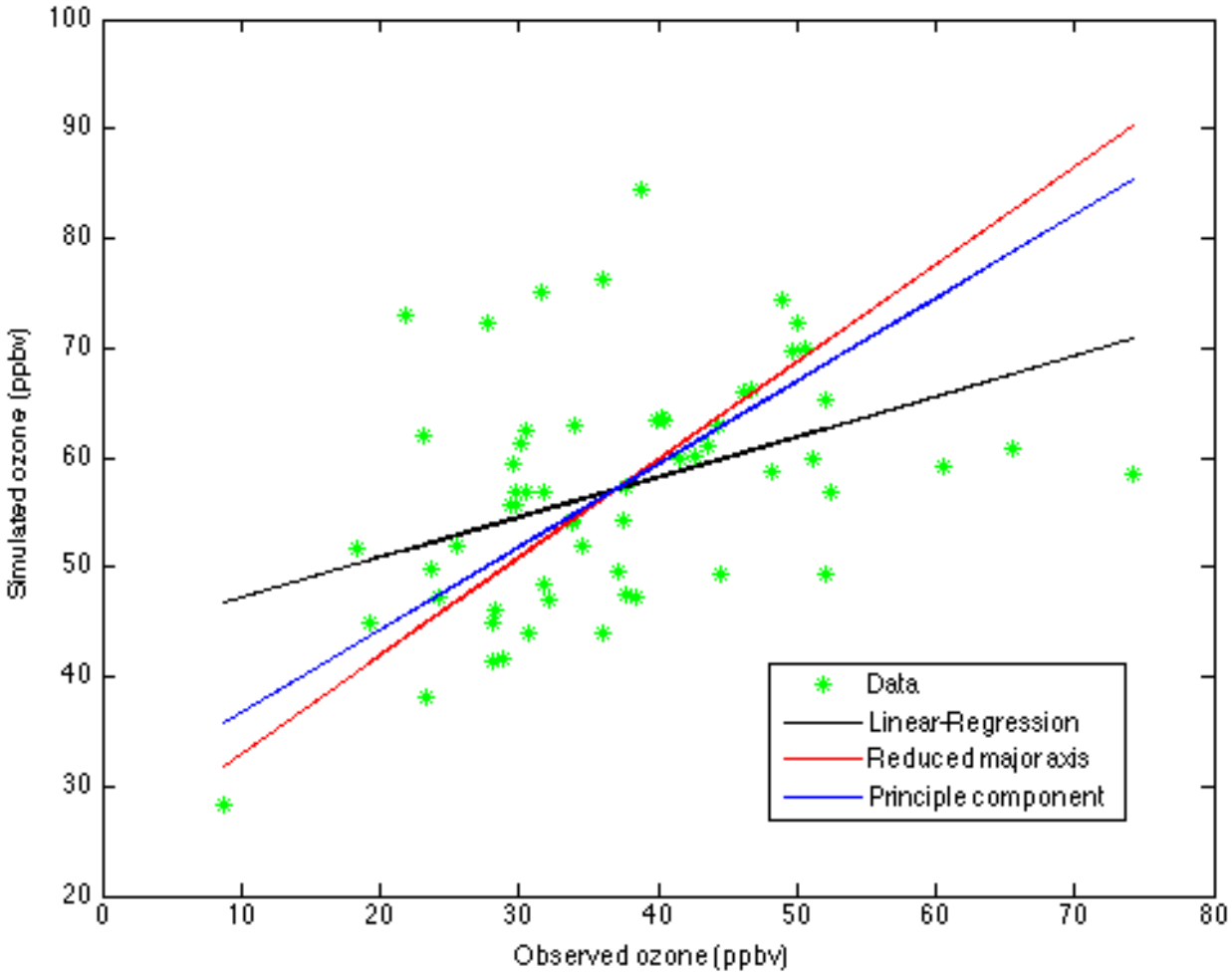
Namrata Kolla
Homework 6
Bivariate methods II

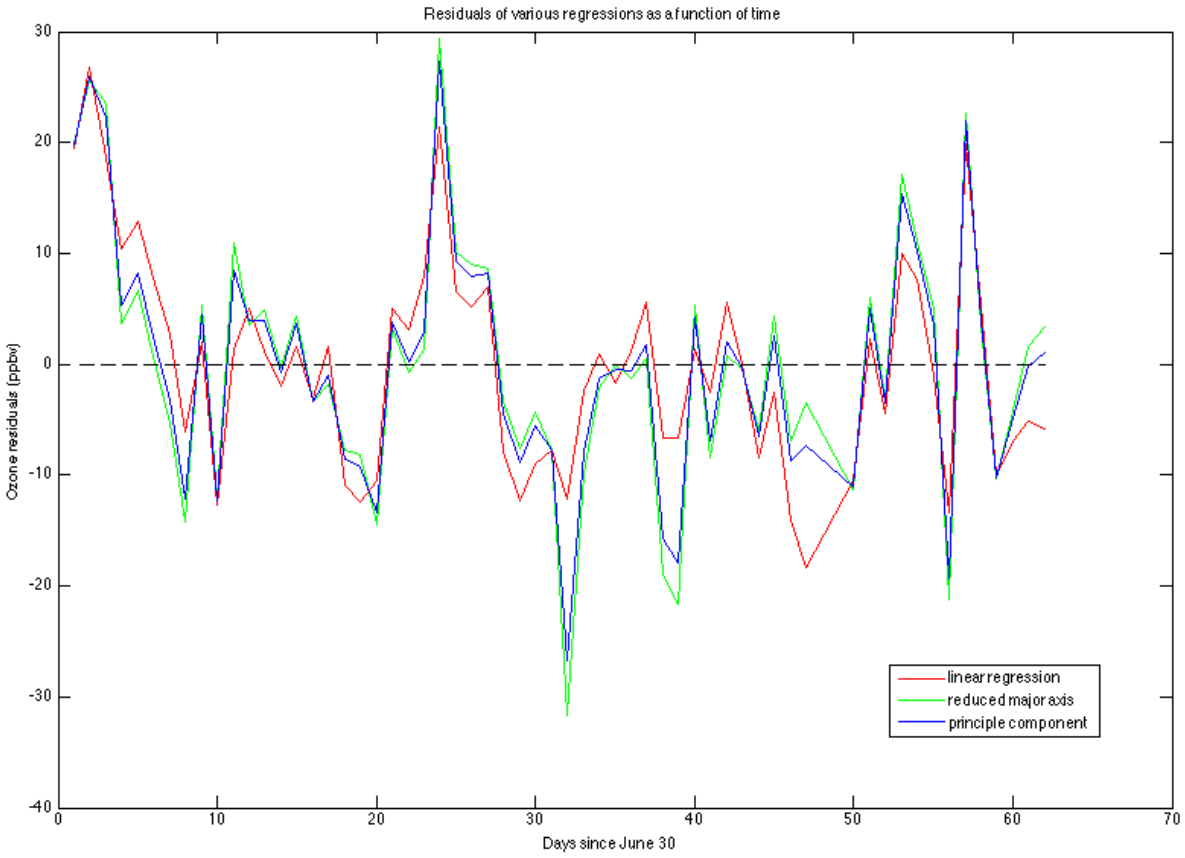
A. Regression Methods



Linear regression:
Slope = 0.367; y-intercept = 43.56
Reduced major axis regression:
Slope = 0.891; y-intercept = 44.17
Principle component regression:
Slope = 0.757; y-intercept = 29.11

Analysis:
The reduced major axis regression is the most appropriate regression method for data. One reason is that the slope for this regression is the closest to 1, which means the modeled ozone values and the observed ozone values are close to each other. The other reason it is the reduced major axis regression is the most robust against outliers. After introducing an outlier value of 30 ppbv and 70 ppbv for simulated and observed ozone respectively, the linear regression's slope and the principle component regression slope dropped while the reduced major axis slope stayed almost exactly the same.

B. Fitting residuals



After performing chi-squared test, these were the results of the tests for normality of the residuals

(Null hypothesis: distribution IS normally distributed)

<u>Linear regression</u>
Critical chi value = 16.919
Calculated chi value = 14.1933
Conclusion: can reject null (normal!)

<u>Reduced major axis regression</u>
Critical chi value = 16.919
Calculated chi value = 10.3697
Conclusion: cannot reject null (normal!)

<u>Principle component regression</u>
Critical chi value = 16.919
Calculated chi value = 12.5015
Conclusion: can reject null (normal!)

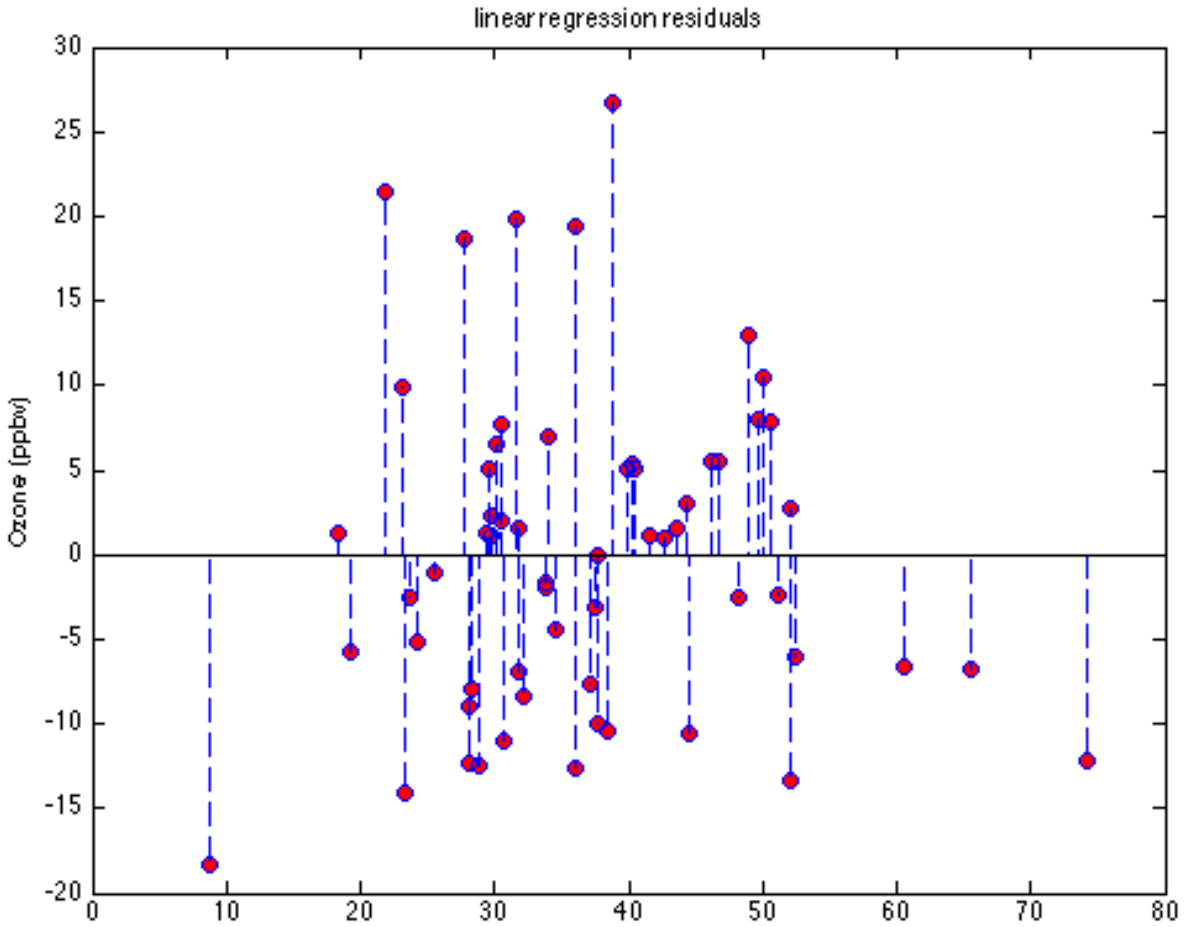
(Analysis on next page →)

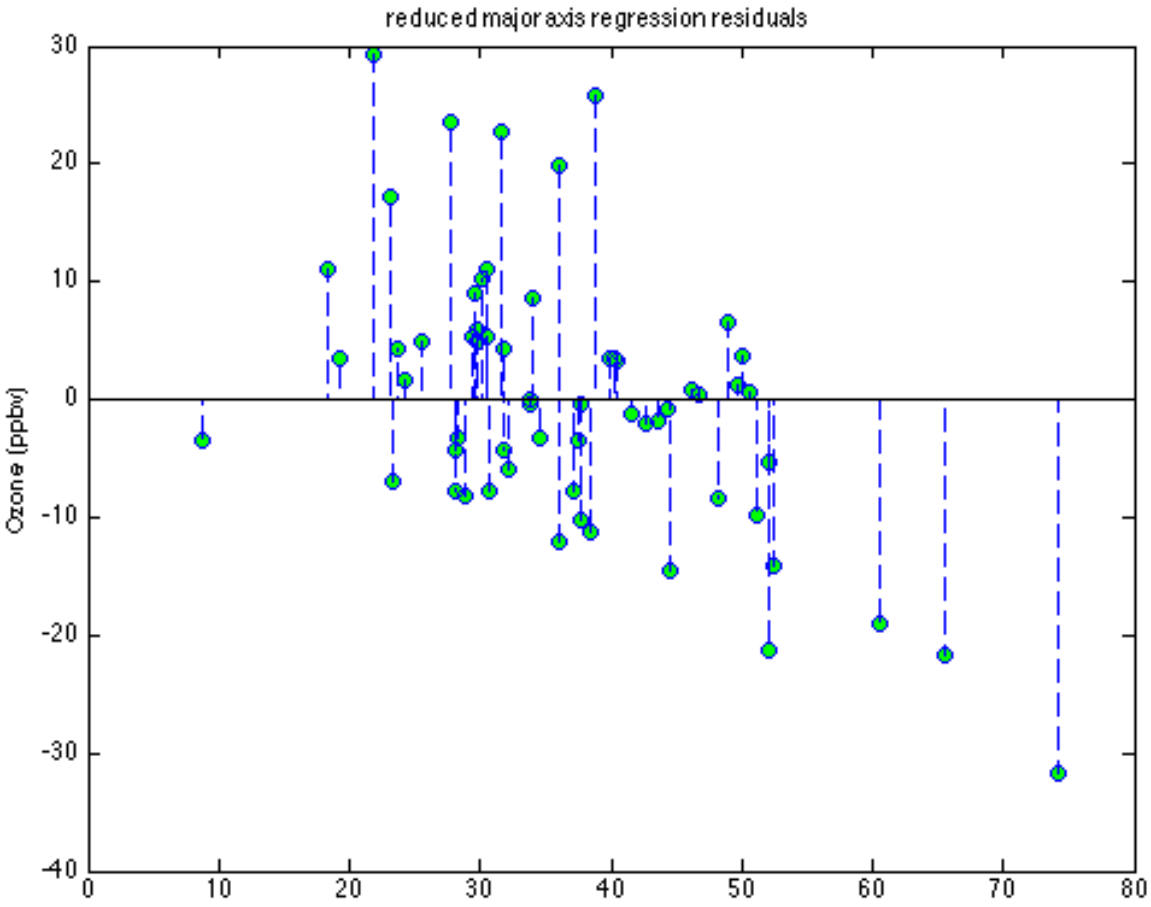
The residuals for a good statistical fit to a data set should be normally distributed because that would suggest that the regression model is erring in predicting the dependent values in a random fashion (ie. values higher and lower than the actual value with equal probability).

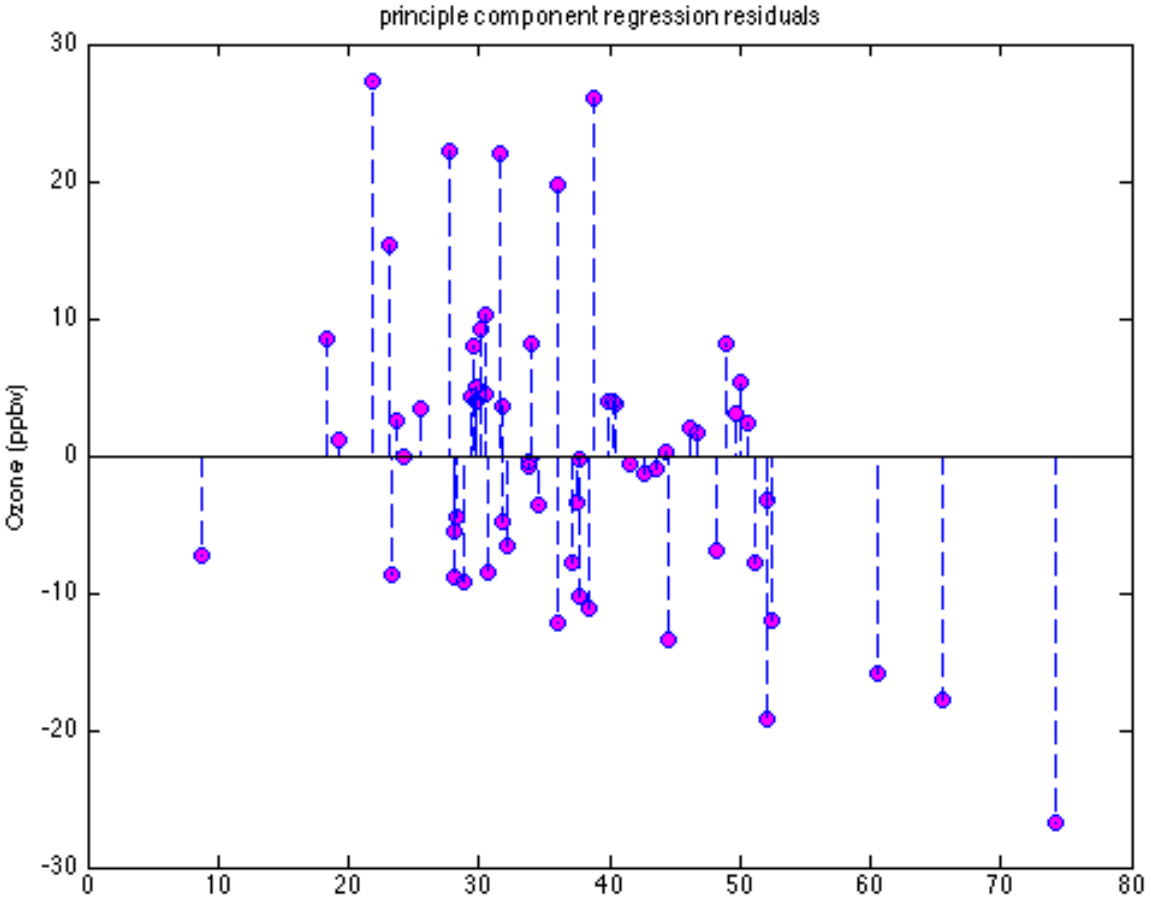
Based on the chi-squared calculation, the distributions for the residuals for all three regressions are normal (determined by comparing the critical chi-squared value to the calculated chi-squared value), so all three regressions could be used to model the data. However, it is worth noting that the reduced major axis (RMA) regression had the smallest calculated chi-squared value compared to all of the other regressions. This suggests that distribution of residuals for RMA is the most normal, so **RMA is the most appropriate aggression to use for this dataset.**

The results of the chi-square test support Part A, which also suggested using reduced major axis regression because the slope of the model was closest to 1.

Extra figures:




reduced major axis regression residuals
Ozone (ppbv)
30
20
10
0
-10
-20
-30
-40
0
10
20
30
40
50
60
70
80


principle component regression residuals
Ozone (ppbv)
30
20
10
0
-10
-20
-30
0
10
20
30
40
50
60
70
80