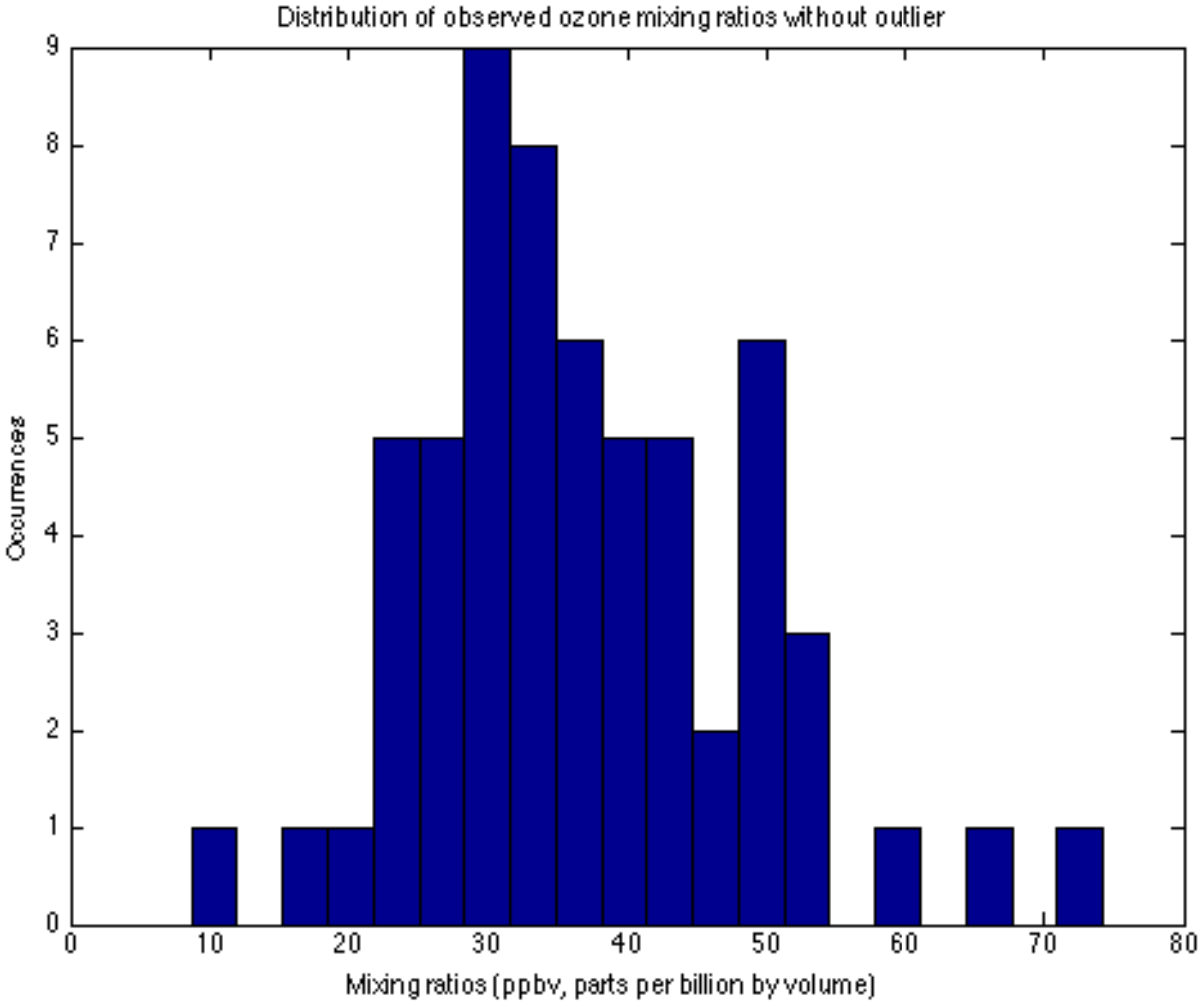
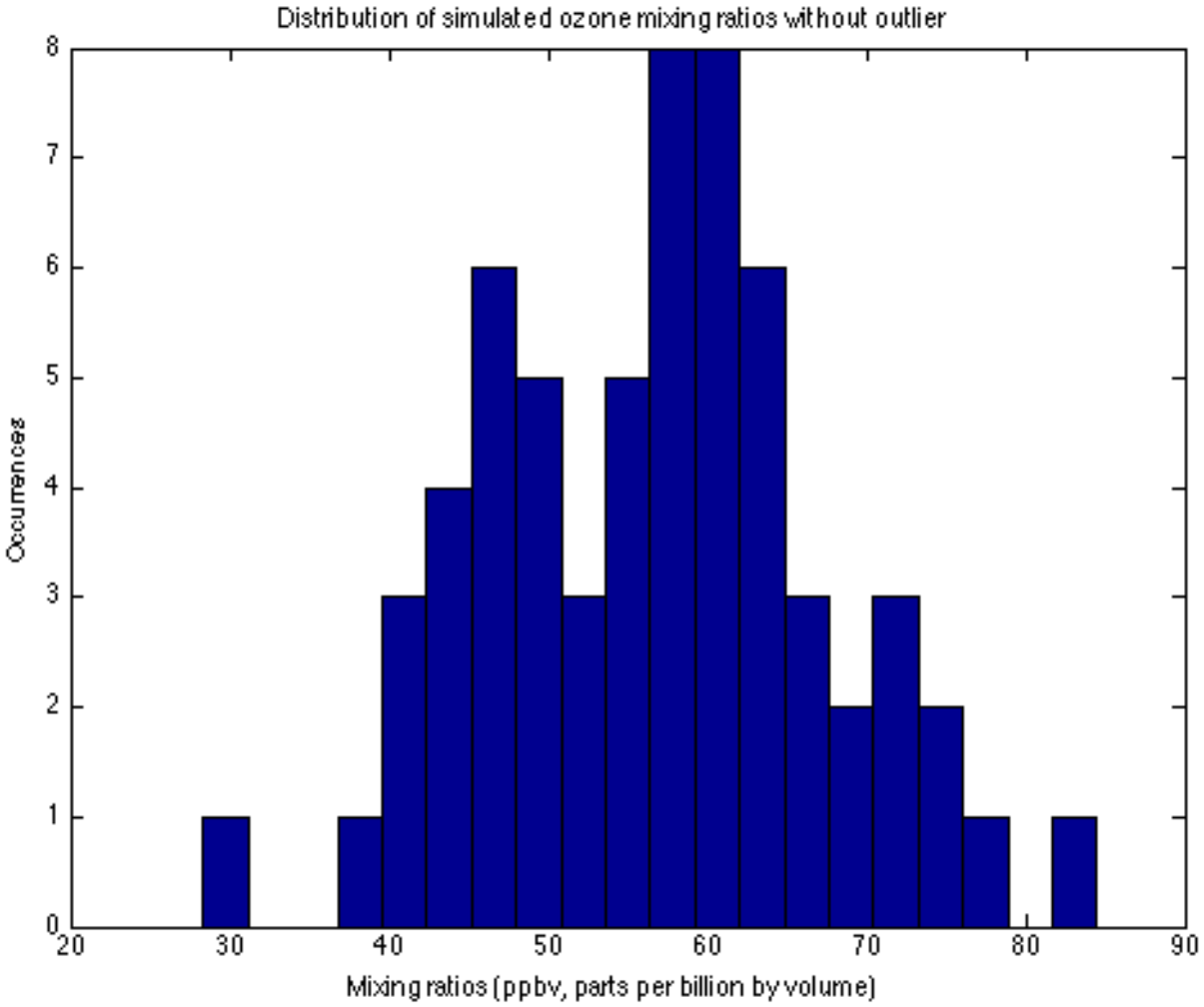
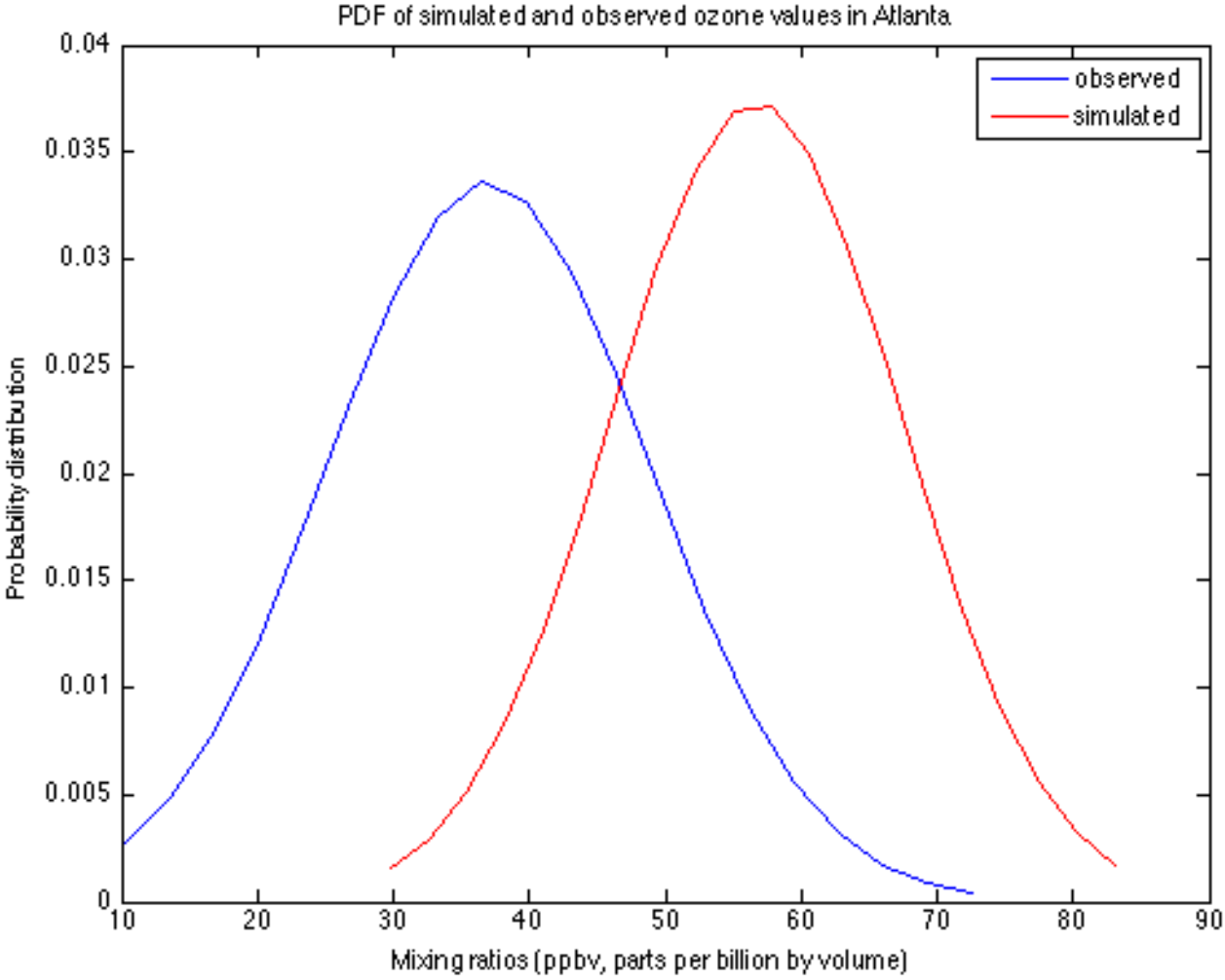
Namrata Kolla – Homework 4

**Part A**

- Used the chi-squared test to determine if each dataset was normally distributed
  - Null hypothesis: the sample data comes from a normal distribution
  - Alternative hypothesis: the sample data comes from an abnormal distribution
  - For observed ozone measurements
    - Measured chi-squared value is 24.6576
    - Critical value is 27.5871
    - Because measured value < critical value, we cannot reject the null hypothesis
    - Thus, the data set ***is*** normally distributed
  - For simulated ozone measurements
    - Measured chi-squared value is 10.9941
    - Critical value is 27.5871
    - Because measured value < critical value, we cannot reject the null hypothesis
    - Thus, the data set ***is*** normally distributed




Distribution of simulated ozone mixing ratios without outlier
Occurrences
0
1
2
3
4
5
6
7
8
20
30
40
50
60
70
80
90
Mixing ratios (ppbv, parts per billion by volume)


PDF of simulated and observed ozone values in Atlanta
observed
simulated
Probability distribution
0.04
0.035
0.03
0.025
0.02
0.015
0.01
0.005
0
10
20
30
40
50
60
70
80
90
Mixing ratios (ppbv, parts per billion by volume)

**Part B**

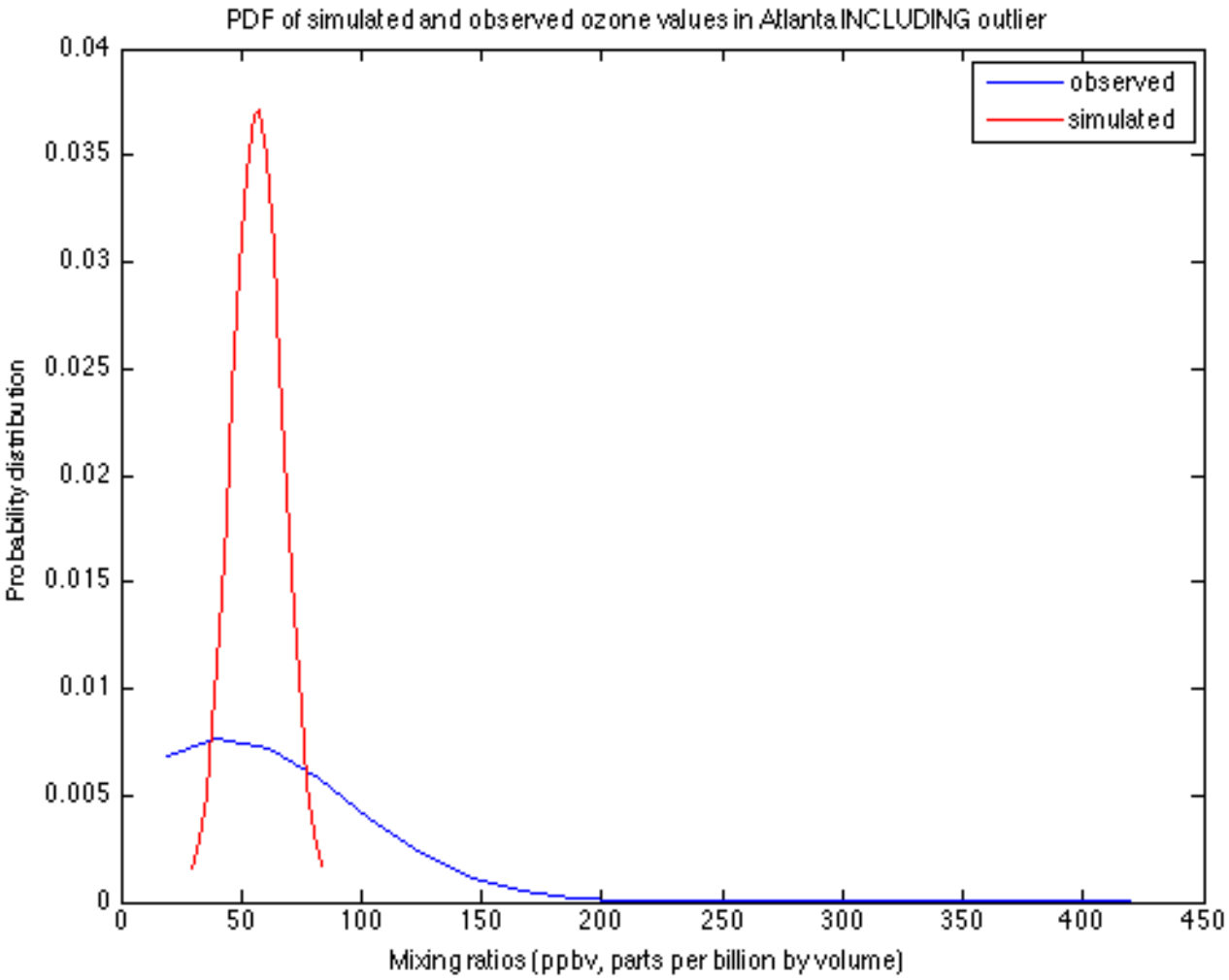
- Used the f-test to determine if the measured and simulated datasets have equivalent variances
  - Null hypothesis: the datasets have equivalent variances
  - Alternative hypothesis: the datasets have unequal variances
  - Measured f-value: 1.2378
  - Critical f-value: 1.5331
  - Because measured value < critical value, we cannot reject the null hypothesis
  - Thus, the variances ***are*** equivalent
- Used the student's t-test to determine if the measured and simulated datasets have equivalent means
  - Null hypothesis: the means of both distributions are equal
  - Alternative hypothesis: the means of both distributions are not equal
  - Result (using ttest2 function): **reject null hypothesis**
    - h (logical operator) = 1
    - p = 1.1382e-16 (significance)
    - -23.7533 to -15.6696 (confidence interval)
  - Thus, the means ***are not*** equal

**Part C – including the outlier**

- Used the chi-squared test to determine if the observed dataset is still normally distributed after including the outlier
  - Null hypothesis: the sample data comes from a normal distribution
  - Alternative hypothesis: the sample data comes from an abnormal distribution
  - For new observed ozone measurements
    - Measured chi-squared value is 1.6339e10
    - Critical value is 27.5871
    - Because measured value > critical value, we can reject the null hypothesis
    - Thus, the data set ***is not*** normally distributed
- Used the f-test to determine if the measured and simulated datasets have equivalent variances
  - Null hypothesis: the datasets have equivalent variances
  - Alternative hypothesis: the datasets have unequal variances
  - Measured f-value: 23.7838
  - Critical f-value: 1.5331
  - Because measured value > critical value, we can reject the null hypothesis
  - Thus, the variances ***are not*** equivalent
- Used the student's t-test to determine if the measured and simulated datasets have equivalent means
  - Null hypothesis: the means of both distributions are equal
  - Alternative hypothesis: the means of both distributions are not equal
  - Result (using ttest2 function): **cannot reject null hypothesis**
    - h (logical operator) = 0
    - p = 0.0589 (significance)
    - -26.2265 to 0.4926 (confidence interval)
  - Thus, the means ***are*** equal

**Summary of impact of outlier**

- Observed data set: normal distribution → not normal distribution
- Observed vs simulated: Equal variances → Unequal variances
- Observed vs simulated: Unequal means → Equal means
- You can conduct the chi-square test (note that it's not robust)
  - However, because the chi-square test says that observed data with the outlier is NOT normal, the results of the t-test and f-test are invalid


PDF of simulated and observed ozone values in Atlanta INCLUDING outlier
observed
simulated
Probability distribution
0.04
0.035
0.03
0.025
0.02
0.015
0.01
0.005
0
0
50
100
150
200
250
300
350
400
450
Mixing ratios (ppbv, parts per billion by volume)