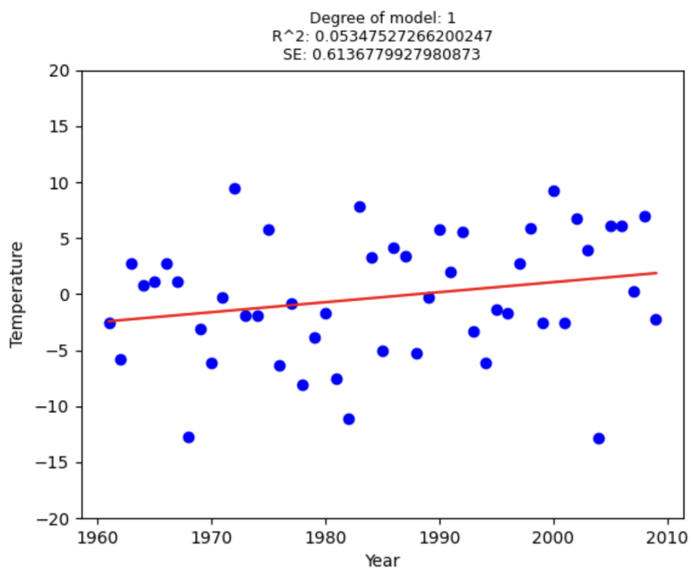
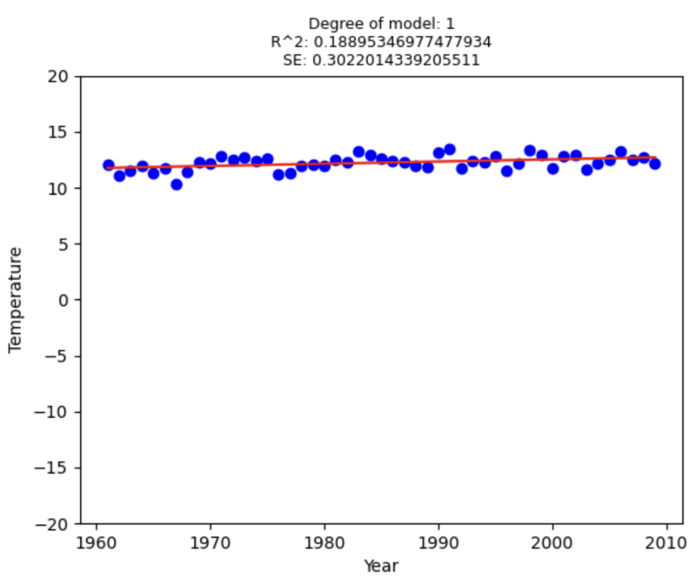
6.0002 Assignment 5

Pureumae Lee

May 23, 2024

Problem 4



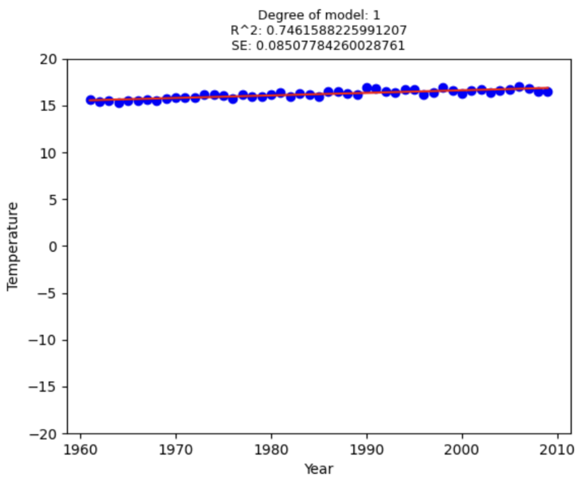


1. What difference does choosing a specific day to plot the data for versus calculating the yearly average have on our graphs (i.e., in terms of the R2 values and the fit of the resulting curves)? Interpret the results.
A: The former has less R2 value and high SE, which means the linear model doesn't fit well. Also it has high variance from the graph. On the other hand, the latter, has more consistent temperature and fits more on the linear line.

2. Why do you think these graphs are so noisy? Which one is more noisy?
A: A graph of specific day is more noisy because the complexity of climate. It cannot guarantee the consistent temperature as last year.

3. How do these graphs support or contradict the claim that global warming is leading to an increase in temperature? The slope and the standard error-to-slope ratio could be helpful in thinking about this.
A: Considering linear model, the average temperature is more convincing since it has higher R2 and lower SE value. Though the average temperature has been increasing slightly, it barely off from 10-15 degree. Thus, it could be an example that global warming isn't really leading to an increase in temperature.

Part B



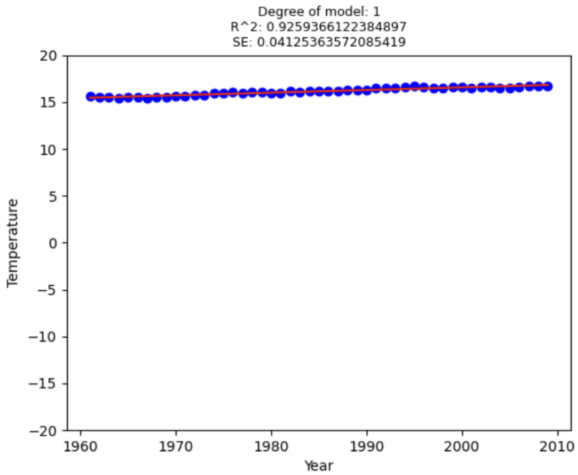
1. How does this graph compare to the graphs from part A (i.e., in terms of the R2 values, t he fit of the resulting curves, and whether the graph supports/contradicts our claim about global warming)? Interpret the results.
A: The result is more convincing then a single city's temperature. It has stronger R2 value and lower SE. The average temperature of about 20 cities has only increased 1 or 2 degrees and it doesn't seem that the global warming is leading to the extreme increase in temperature.

2. Why do you think this is the case?
A: The data itself is inclined to increasing and the model is showing decent correlation (R2, SE).

3. How would we expect the results to differ if we used 3 different cities? What about 100 different cities?
A: If we use 3 different cities, then the graph will fluctuate more and if we use 100 different cities, then it will be more stable.

4. How would the results have changed if all 21 cities were in the same region of the United States (for ex., New England)?
A: It will become more like single city since it is geologically close.

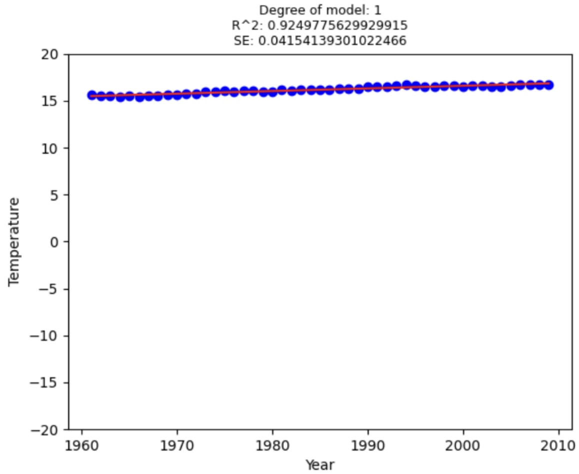
Part C

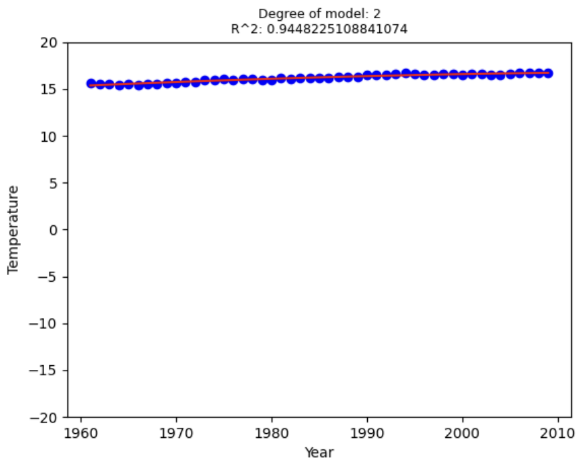


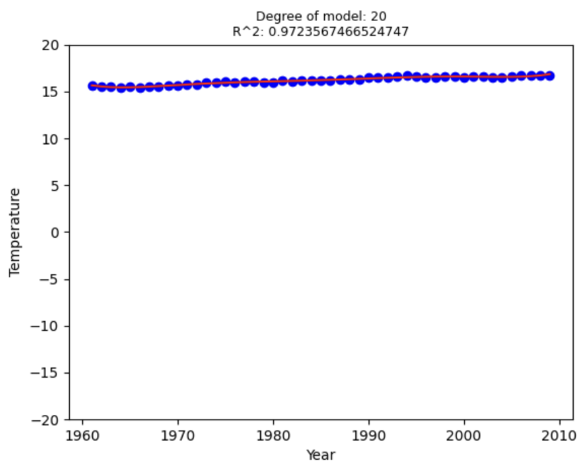
1. How does this graph compare to the graphs from part A and B (i.e., in terms of the R2 values, the fit of the resulting curves, and whether the graph supports/contradicts our claim about global warming)? Interpret the results.
A: The value of R2 and SE has improved better than average temperature of cities. Still the trend of temperature isn't that dynamic and not enough to say that global warming is really bringing the increase of temperature.

2. Why do you think this is the case?
A: It's basically averaging several times so the values are becoming same. For example, comparing t_3 and t_4, they both have t_2 and t_3 component so the value should be similar.

Problem 2.I







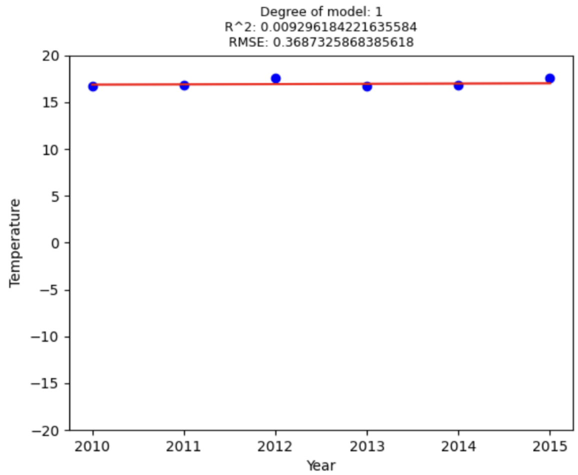
1. How do these models compare to each other?
A: Based on the R2 value, it gets better when the degree increases.

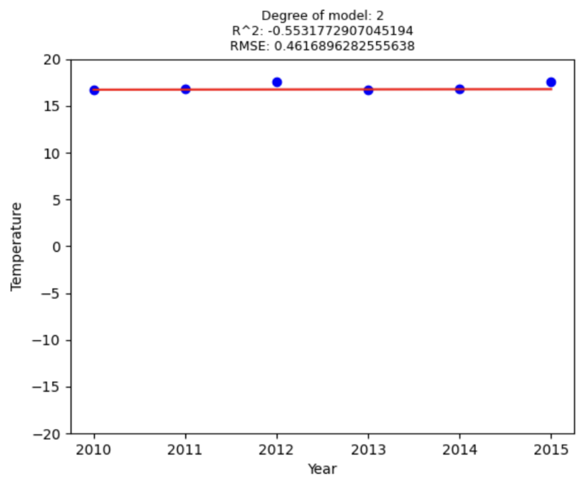
2. Which one has the best R2 ? Why?
A: The one with degree 20 because it can have 19 inflection points and able to fit each points smoothly.

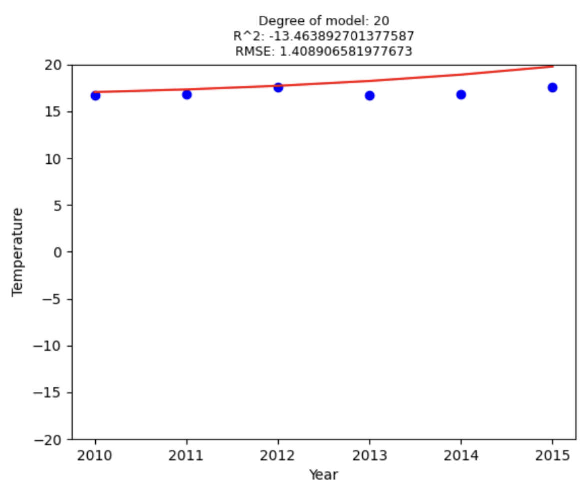
3. Which model best fits the data? Why?
A: It is also the one with degree 20. But it seems to be overfitted and it might not work well in the

test set.

Problem 2.II







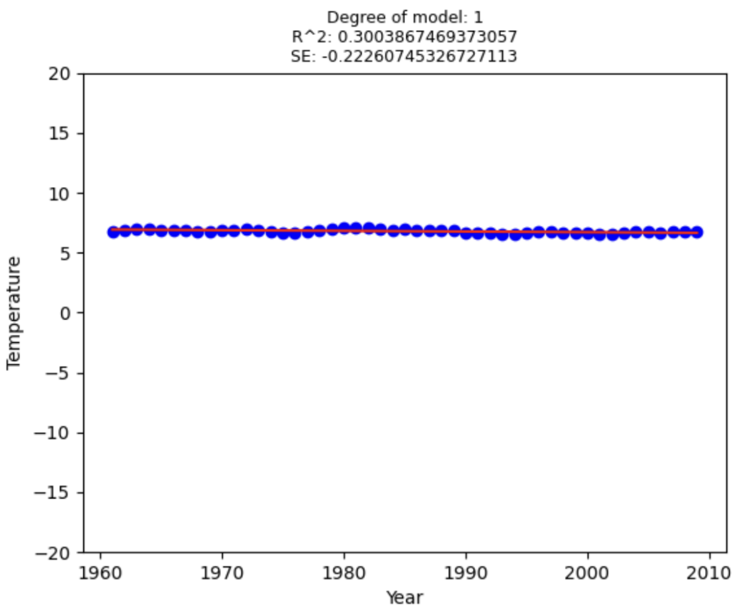
1. How did the different models perform? How did their RMSEs compare?
A: It gets bigger as the degree increases.

2. Which model performed the best? Which model performed the worst? Are they the same as

those in part D.2.I? Why?
A: It's the opposite of part D.2.I. Because higher degree on training set means that the model is trying to fit the training data, not the general data.

3. If we had generated the models using the A.4.II data (i.e. average annual temperature of New York City) instead of the 5-year moving average over 22 cities, how would the prediction results 2010-2015 have changed?
A: The model with degree 1 might not change that much. On the other hand, the degree 2 and 20 models would might perform worse than D.2.I since it will be more concaved.

Part E



1. Does the result match our claim (i.e., temperature variation is getting larger over these years)?
A: The variation stays the same as the time passes.

2. Can you think of ways to improve our analysis?
Maybe we can group some regions based on geographic features.