Comparison of errors within monthly weather forecasting in different sized areas of the United States to observed data

Hope Johnson

ABSTRACT –

In the method of forecasting, its use of past data can create a somewhat accurate prediction, but this project examines the errors that can appear throughout this process. This project looks specifically at a reliable climate variable, temperature, through datasets measuring the monthly maximum temperature between the years 2001 and 2020. This data is used in this process and forecasts the monthly maximum temperature for the subsequent year 2021. The error of the forecasted data against the actual measured data is analyzed at different regions of the United States with different climate behaviors and seasonality trends and is investigated for potential reasons why.

INTRODUCTION –

During the 1960's, mathematician Edward N. Lorenz wanted to establish if it was possible to make a long-term prediction of weather forecasting per a mathematical model. Lorenz arrived at a system of equations that involve 3 state variables representing the spatial average of the hydrodynamic velocity, temperature, and temperature of the gradient with constants related to the Prandtl number and Rayleigh number of the fluid, and the aspect ratio of the domain under consideration. He found that the solution to these systems of equations, or rather their approximate repetitions have a finite duration. Meaning that eventually after a certain amount of forecasting, the accuracy of the system becomes what is known as chaotic.

In support of this idea, this project examines the relationship between observed weather data and weather data found from a simple prediction model. More precisely, using monthly maximum temperature data between the years 2001 and 2020 of two overlapping areas to create a prediction model based on its linear fit. The two overlapping locations we first look at are using are the Northeast Climate Region recognized as being a climate consistent region which includes the states of Connecticut, Delaware, Maine, Maryland, Massachusetts, New Hampshire, New Jersey, New York, Pennsylvania, Rhode Island, and Vermont, and the State Region, New York (Karl and Kloss). Both of these areas contain an overlapping area, New York State, that potentially could show a relationship between the inaccuracies of a predictive model and the amount of area it is forecasting.

This project takes monthly data from the two previously mentioned areas within the United States, referred earlier as The Northeast and New York State from January 2001 to December of 2020, a range of 20 years. This data is used in the creation of a prediction model with an output of monthly maximum temperatures of the subsequent year, 2021.

This project will first discuss the findings of the prediction model used from this method. I will also investigate how this prediction model fairs with multiple areas within the continental United States that have different climate behaviors and seasonality. These areas consist of the Southeast including the states, Alabama, Florida, Georgia, North Carolina, South Carolina, Virginia, and the West which is made up of the states California and Nevada. Using these findings, I will be able to come to conclusions about this simple prediction model and its implications for models like this and what its inaccuracies may signify.

RESULTS –

I. Northeast

In this prediction model a linear fit of the historical data of the monthly maximum temperature in the Northeastern U.S. is created, its parameters were specified earlier from the dataset between the years 2001 and 2020. In this case the tool used was Excel's forecast function. This function requires constant intervals so using monthly data works to calculate/predict future values from existing values using linear regression. The input of this being the monthly maximum temperature from 2001-2020 and the output resulting in monthly data of the year 2021 seen in Figure 1.

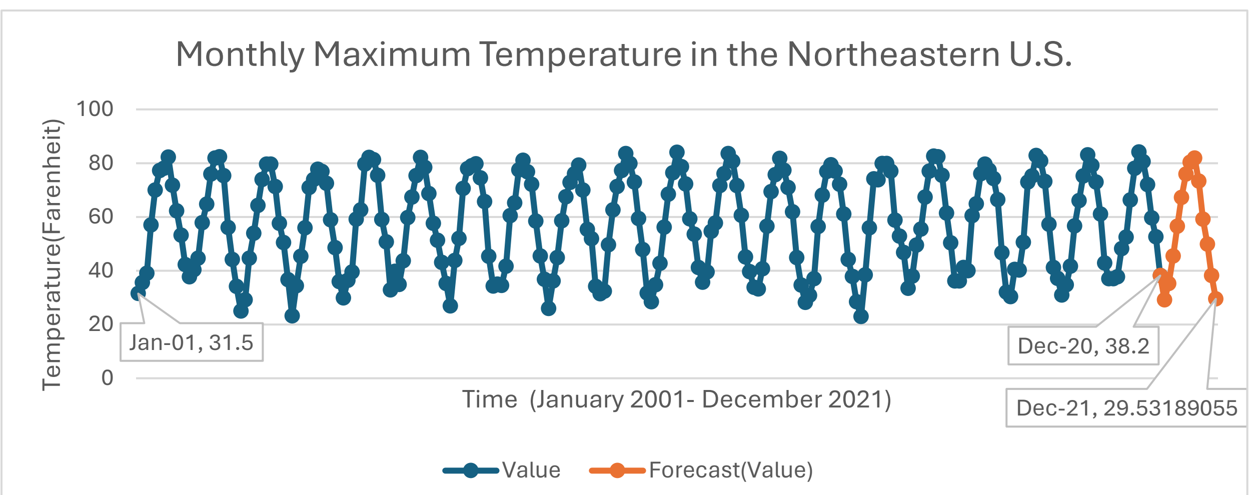


Figure 1

In Figure 1, it is shown the past seasonal data is marked in blue and months are shown by a circle dot. The lowest of these maximum temperatures usually occurs in January where temperatures typically stay around the low 20 degrees Fahrenheit, whilst the highest months reach highs in the 80-85 degree area. The predicted data follows this as it reaches similar highs and lows as seen in orange. However, you may notice that there are more orange dots than months in the year 2021. This is because with this function there were two highest temperatures predicted for December of 2021, however in the analysis of this data further on, the lower measurement of temperature in December was disregarded as this is a focus on maximum temperature.

To look more closely at the predicted data, we can see Table 1 below, showing the exact numerical results we get from both the forecasted monthly data and the observed monthly maximum temperatures measured throughout the year 2021. To visualize this data, these are also shown on a graph (Figure 2), where the forecast data and observed data are represented in red and blue respectively in a time series over every month throughout the year 2021.

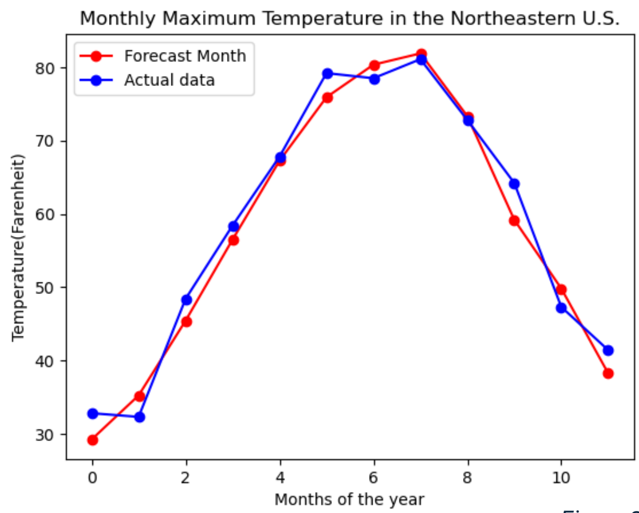


Figure 2

Month of the Year	Predicted Max Temperature	Actual Max Temperature
January	29.531891	32.8
February	35.225374	32.3
March	45.423732	48.4
April	56.486698	58.4
May	67.240100	67.8
June	75.910687	79.2
July	80.347826	78.5
August	81.887540	81.1
September	73.322890	72.8
October	59.179204	64.2
November	49.811334	47.3
December	38.227278	41.1

Table 1

The error of the forecasted modeled data in comparison to the actual observed data of each month was calculated as follows. In this instance, the magnitude of the error for each month was found from the absolute value of the difference in predicted and actual temperatures. In Table 2, the result of the calculation is shown by the corresponding months. Here, the largest absolute error occurred in October with an estimated 5.021 error whilst the smallest absolute error happened in the month of September with an error around 0.523. The standard deviation of the monthly absolute error was calculated to be at ~1.359 and had an absolute error mean of ~2.426 degrees Fahrenheit.

Month of the Year	Calculated Error
January	3.585
February	2.925
March	2.976
April	1.913
May	0.560
June	3.289
July	1.848
August	0.788
September	0.523
October	5.021
November	2.511
December	3.173

Table 2

II. New York State

Using similar methods, New York State monthly maximum temperatures from the same years, 2001-2020, were used to create a forecast of the monthly maximum temperatures of 2021. Below the time series of New York State has most similarities to the Northeastern data (Figure 3).

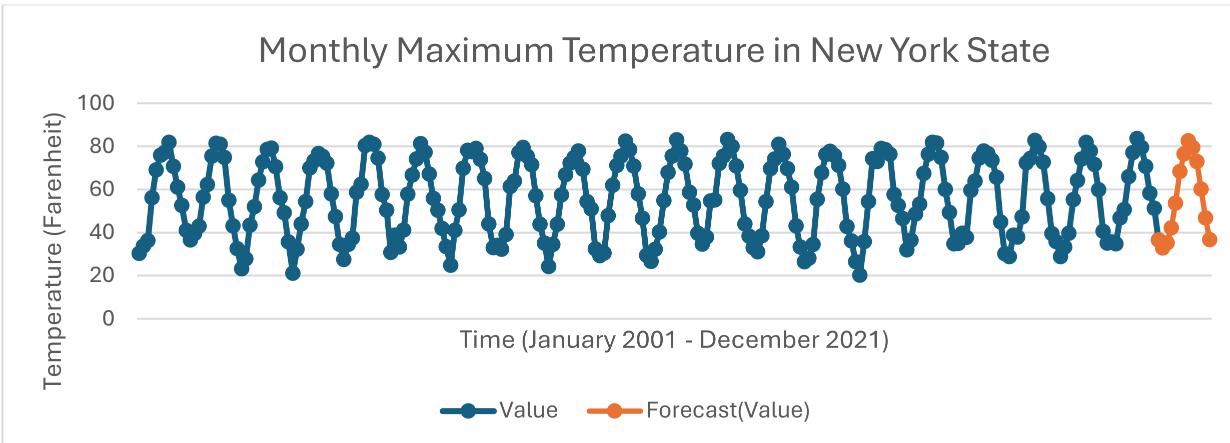


Figure 3

The seasonality of this data similarly consists of lows in January/February around 25-30 degrees and reaches its highs at around 75-80 degrees in the summer months of July/August. The predicted data follows this with a low of 32.8 degrees in January and a high of 82.6 degrees in July seen above in Figure 3 in orange.

Looking at the maximum temperature data, Table 3 shows the data we get for the observed and forecasted data. Again, this is visualized with a line graph below seen in Figure 4.

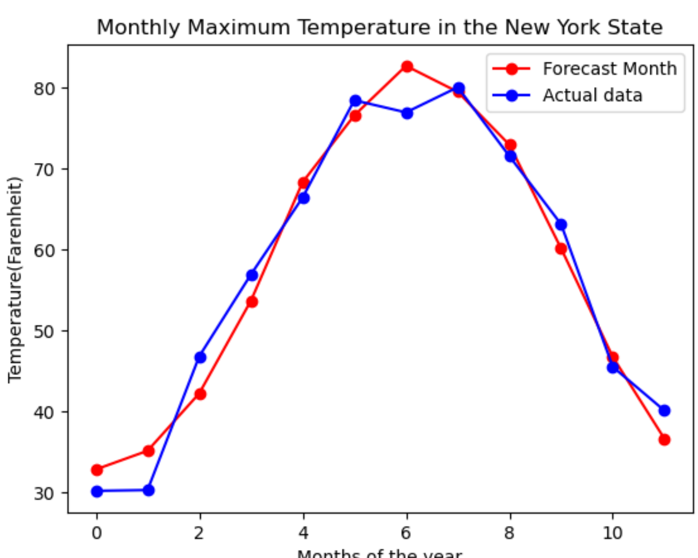


Figure 4

Month of the Year	Predicted Max Temperature	Actual Max Temperature
January	35.225374	30.2
February	45.423732	30.3
March	56.486698	46.8
April	67.240100	56.9
May	75.910687	66.4
June	80.347826	78.4
July	81.887540	76.9
August	73.322890	80.0
September	59.179204	71.5
October	49.811334	63.1
November	38.227278	45.5
December	29.531891	40.1

Table 3

The errors of this data were also recorded once again using the same method as before, corresponding to the Table 4 seen to the right, displaying the magnitude of error of the forecast model for the 2021 data. The largest absolute error occurring in July at ~5.74 degrees and the lowest occurrence in the subsequent month, August at ~0.56 degrees. The standard deviation of the absolute monthly error was calculated to be at ~1.588 and had an absolute error mean of ~2.872 degrees Fahrenheit.

Month of the Year	Calculated Error
January	2.642
February	4.843
March	4.571
April	3.237
May	1.893
June	1.869
July	5.744
August	0.565
September	1.408
October	2.990
November	1.223
December	3.484

Table 4

III. Southeast

The next set of data used for this analysis was from the Southeast region of the United States. This region includes the states of Alabama, Florida, Georgia, North Carolina, South Carolina, and Virginia. The same source of monthly data between January 2001 and December 2020 within the Southeast region was used as an input into Excel's forecast function to predict monthly maximum temperatures of the year 2021. The seasonality of the climate in this region reached different temperatures than they did in the North, with highs around 90 degrees Fahrenheit and lows around 50-55 degrees Fahrenheit.

Looking closely at the data in 2021, we can then compare the predicted data made with this forecast function against the actual data from 2021. Figure 5 and Table 5 can be used as aids to see this more in depth.

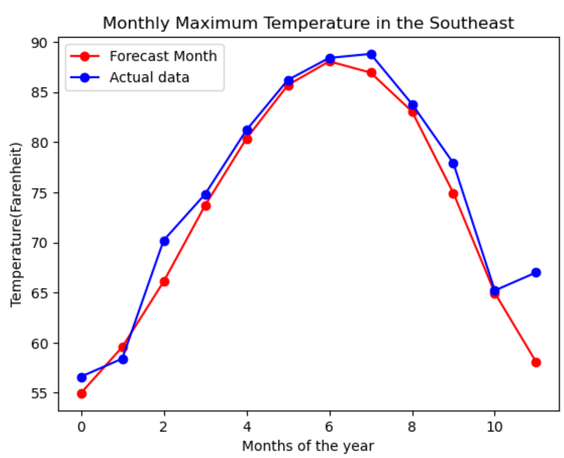


Figure 5

Month of the Year	Predicted Max Temperature	Actual Max Temperature
January	54.949960	56.6
February	59.561543	58.4
March	66.084901	70.2
April	73.674239	74.8
May	80.348563	81.2
June	85.690355	86.2
July	88.056649	88.4
August	86.935168	88.8
September	83.056345	83.8
October	74.923041	77.9
November	64.962813	65.2
December	58.061600	67.0

Table 5

The absolute errors of the monthly data were found using the same method as before yielding results that can be seen in Table 6 below. The largest absolute error occurs in December at ~8.938 degrees Fahrenheit and the lowest occurred in November at ~0.237 degrees Fahrenheit. The standard deviation of the absolute monthly errors resulted in ~2.45 and had a mean absolute monthly error that was around 2.04 degrees.

Month of the Year	Calculated Error
January	1.650
February	1.162
March	4.115
April	1.126
May	0.851
June	0.509
July	0.343
August	1.865
September	0.744
October	2.977
November	0.237
December	8.938

Table 6

IV. West

The final set of data used for this analysis was from the West region of the United States which includes the states of California and Nevada. The same source of monthly data between January 2001 and December 2020 within the Southeast region was used as an input into Excel's forecast function to predict monthly maximum temperatures of the year 2021. The seasonality of the climate in this region reached different temperatures than they did in the Northeastern U.S., with highs around 90 degrees Fahrenheit and lows around 50-55 degrees Fahrenheit, more like the Southeast.

Looking closely at the data in 2021, we can then compare the predicted data made with this forecast function against the actual data from 2021. Figure 6 and Table 7 can be used as aids to see this more in depth.

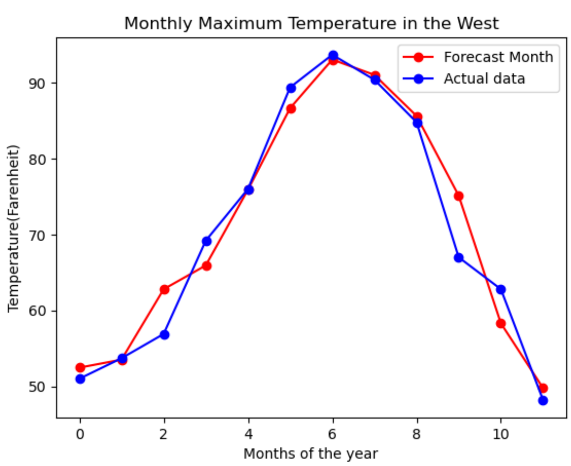


Figure 6

Month of the Year	Predicted Max Temperature	Actual Max Temperature
January	52.461752	51.0
February	53.515765	53.7
March	62.795230	56.9
April	65.936310	69.2
May	75.911802	76.0
June	86.666410	89.4
July	93.064975	93.7
August	91.041021	90.4
September	85.627878	84.8
October	75.124608	67.0
November	58.306025	62.8
December	49.771352	48.2

Table 7

The absolute errors of the monthly data were found using the same method as before yielding results that can be seen in Table 8 below. The largest absolute error occurs in October at ~8.125 degrees Fahrenheit and the lowest occurred in May at ~0.088 degrees Fahrenheit. The standard deviation of the absolute monthly errors resulted in ~2.53 and had a mean absolute monthly error that was around 2.49 degrees.

Month of the Year	Calculated Error
January	1.462
February	0.184
March	5.895
April	3.264
May	0.088
June	2.733
July	0.635
August	0.641
September	0.828
October	8.125
November	4.494
December	1.571

Table 8

DISCUSSION –

Looking at the monthly errors between all regions described as Northeast, New York State, Southeast, and Western U.S. We can compare the errors to each other month by month. In Figure 7 below, all 4 regions are depicted below showing the errors the forecasting model had made. Looking at this figure closely, we can see that there are spikes in errors typically in March, around June, and October. However, besides this observation, there is not much correlation between the regions. This, however, is expected since these regions, besides the Northeast and New York State, have vastly different climate behaviors and seasonality as discussed earlier.

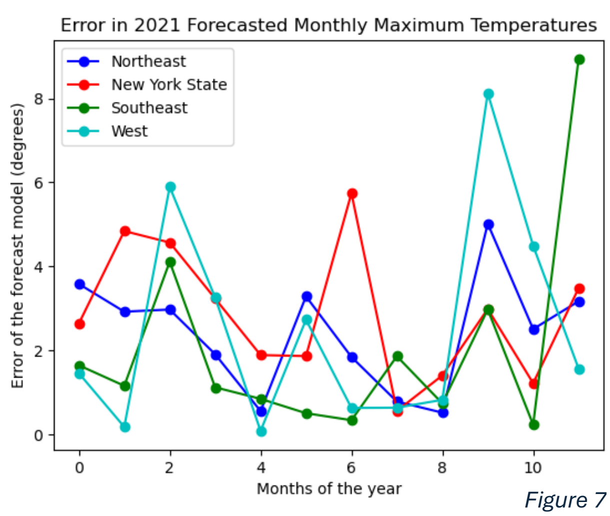


Figure 7

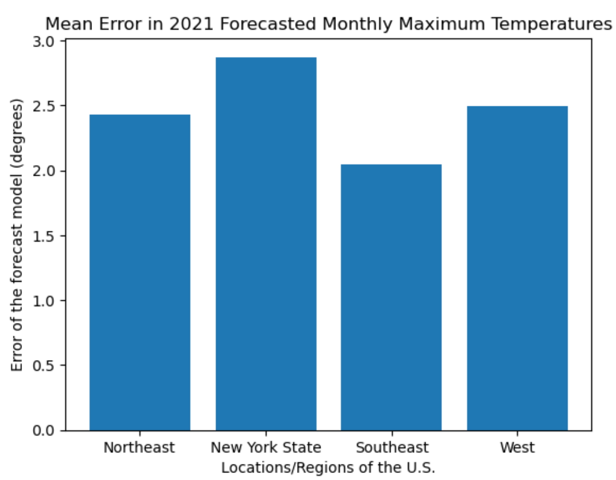


Figure 8

Looking at this pattern, and the large spikes that existed, especially in the West and Southeast regions datasets, I was interested to see what areas consisted of the most errors to see which regions forecast might have been the most reliable. In Figure 8, the mean error of the forecast of each region is shown. Here, we can see that New York State had the most errors out of all the regions. Compared to the Northeast region, New

York's mean absolute error, 2.87 is about .45 larger. We can draw a conclusion that, based on this data, larger regions have more accurate forecasts than smaller, state regions.

When looking back at Figure 7, I saw that there were spikes and wanted to investigate whether the large spikes in data certain regions have influenced the mean errors of the forecast. Thus, I was interested in where the data sets errors typically sat and how consistent they were. Figure 9 shows the standard deviations of the errors in the Forecasted Monthly maximum temperature of each region. Continuing to focus on the Northeast and New York State regions, we can see that the mean error of New York state wasn't swayed by a spiked in the error levels. This reassures my confidence in my reasoning that state-sized areas are less reliable than larger climate regions.

Now to look at differences between the larger regions, the Northeast, the Southeast, and the West, we can see in Figure 7, the West and Southeast have substantial spikes in their errors that were much larger than the rest of the months. This is backed up by Figure 9, showing the standard deviation of those regions being significantly higher than the Northeast region, meaning that the absolute errors are less consistent throughout the 2021 forecast.

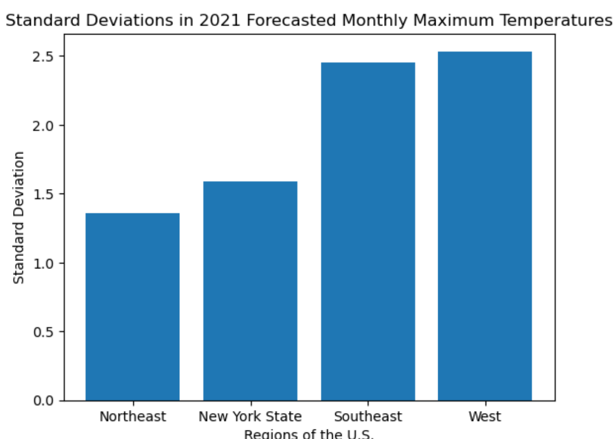


Figure 9

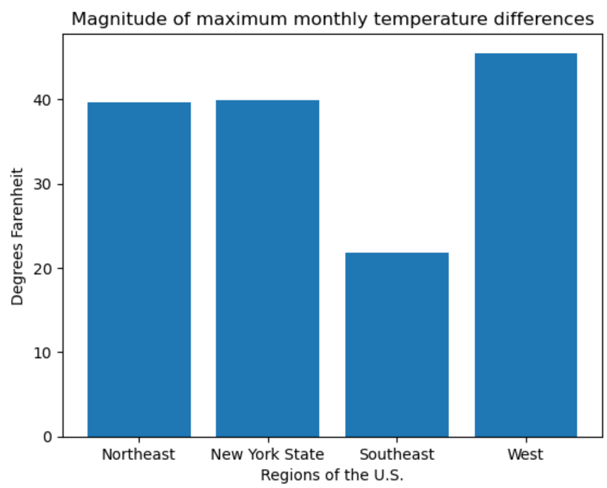


Figure 10

Since these regions are in vastly different places with vastly different climate behaviors, I also wanted to investigate whether the lowest and highest maximum temperatures of these regions affected which had higher errors and potentially explain the spikes in monthly absolute errors in the Southeast and West regions had. To do this I calculated the highest monthly maximum temperature and subtracted the lowest monthly temperature. Shown in Figure 10, we can see that the Southeast region had the least variability, and the West had the greatest. Whilst this did provide answers as to why the Southeast had the smaller mean error, the spikes in errors month-to-month were still unexplained.

Thus, based on the forecasted data's spikes in errors month-to-month with support from the analysis of the standard deviations from each region, one could infer that, based on where these regions lie within the continental United States, monthly forecasts in the Northeastern U.S. are more consistent than those in the Southeast and West regions.

SUMMARY –

With the intention to investigate and analyze the errors within monthly weather forecasting in different regions of the United States to observed data, I first had to create a forecast. I chose to use monthly maximum temperature data from the years 2001-2020 to create a forecast of the monthly maximum temperature of the regions for the year 2021. From this I found the absolute errors of the forecasts from the actual measured data from 2021. Unexpectedly, there weren't many patterns regarding the seasonality of the data that I could find an explanation for. More substantial results were found when comparing region sizes and region locations to each other. Here I observed that larger regions with similar climates (Northeast, Southeast, West) had less errors in the forecast than smaller state-sized regions. So, if I could do a further analysis, I would focus on those larger sized regions and forego smaller state-sized ones. I also learned that regions with large spikes in their errors month-to-month had less temperature variability. To analyze this further I would use a different measurement of temperature, potentially average monthly temperature instead that might yield more helpful deductions.

This analysis consisting of the temperatures within different climate regions of the United States supports Lorenz, and that forecasting can become increasingly inaccurate. While this analysis did use Excel's forecast function, which may not be optimized for climate data, the results support Lorenz and his examination about weather forecasting. Further analysis could be more accurate if done using a better forecasting model that included other climate variables besides temperature that have major effects on a region.

REFERENCES –

Kaper, Hans , Engler, Hans. Mathematics & Climate. Philadelphia (NY): Society for Applied and Industrial Mathematics; 2013.

Thomas R. Karl and Walter James Koss, 1984: *"Regional and National Monthly, Seasonal, and Annual Temperature Weighted by Area, 1895-1983."* **Historical Climatology Series 4-3**, National Climatic Data Center, Asheville, NC, 38 pp.

Karl, T. R. and Koss, W.J., 1984. *Historical Climatology Series 4-3*: Regional and National Monthly, Seasonal and Annual Temperature Weighted by Area, 1895-1983.

Microsoft. Microsoft Support. (n.d.). https://support.microsoft.com/en-us/office/forecast-and-forecast-linear-functions-50ca49c9-7b40-4892-94e4-7ad38bbeda99

Schneider, D. P., C. Deser, J. Fasullo, and K. E. Trenberth, 2013: Climate Data Guide Spurs Discovery and Understanding. Eos Trans. AGU, 94, 121–122, https://doi.org/10.1002/2013eo130001

Zhang, H.-M., B. Huang, J. Lawrimore, M. Menne, Thomas M. Smith, NOAA Global Surface Temperature Dataset (NOAAGlobalTemp), Version 5 [indicate subset used]. NOAA National Centers for Environmental Information. doi:10.25921/9qth-2p70 [indicate subset used]. doi:10.7289/V5FN144H. Accessed at NOAA/PSL [April 9 2024].

In [2]:
```
import matplotlib.pyplot as plt
import pandas as pd
import numpy as np
import statistics as st
```

Northeast Data

In [3]:
```
filename1 = r"C:\Users\hopey\OneDrive - University at Buffalo\Forecast NE.csv"   #LOADING DATA
NEdata2001 = pd.read_csv(filename1)

filename2 = r"C:\Users\hopey\OneDrive - University at Buffalo\2021 MAX TEMP NE.csv"
NEdata2021 = pd.read_csv(filename2)
```

In [4]:
```
Temp2001_NE = NEdata2001['Forecast(Value)'][240:-1]   #last 12 pieces of forecasted data
Date2001_NE = NEdata2001['Date']                      #month and year mmm-yy of data

Temp2021_NE = NEdata2021['Value']                    #actual temperature data of 2021
Date2021 = NEdata2021['Date']                        #corresponding month of 2021
```

In [5]:
```
Temp2021_NE                                          # Observed Maximum Temperatures of 2021 i
```

Out[5]:
```
0     32.8
1     32.3
2     48.4
3     58.4
4     67.8
5     79.2
6     78.5
7     81.1
8     72.8
9     64.2
10    47.3
11    41.4
Name: Value, dtype: float64
```

In [6]:
```
Temp2001_NE                                          # Forecasted Maximum Temperatures of 2021
```

Out[6]:
```
240    29.214766
241    35.225374
242    45.423732
243    56.486698
244    67.240100
245    75.910687
246    80.347826
247    81.887540
248    73.322890
249    59.179204
250    49.811334
251    38.227278
Name: Forecast(Value), dtype: float64
```

In [7]:
```
ForecastNE = []                              #create list to contain the forecasted temperatu
ActualNE = []                                #create list to contain the actual observed temp
errorNE = []                                 #create list to contain the error of the foresca
                                             #to the actual
for x,y in zip(Temp2001_NE,Temp2021_NE):
    ForecastNE.append(x)                     #adds each months forecasted temp to Forecast li
    ActualNE.append(y)                       #adds each months actual temp to error list
    errorNE.append(abs(x-y))                 #adds each months magnitude of error in degrees
                                             #list
```

In [8]:

```
errorNE                                            #error list containing each months error (unit =
```

Out[8]:

```
[3.5852339299999976,
 2.925374310000023,
 2.9762678499999993,
 1.9133023400000013,
 0.5598995699999989,
 3.289313129999965,
 1.8478262299999955,
 0.787539670000001,
 0.5228902099999999,
 5.020796340000004,
 2.511333620000002,
 3.1727217799999963]
```

In [9]:

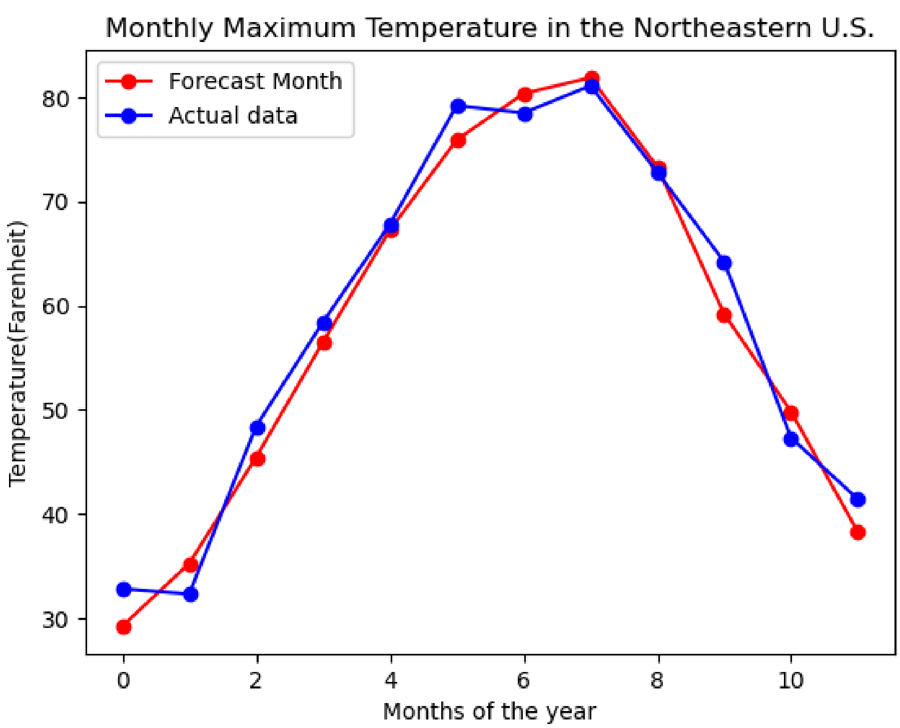
```
st_dev = []                                   #creates a standard deviation list for each region
st_dev.append(st.stdev(errorNE))              #adds the Northeastern error standard devation to
                                              
means_lst = []                                #creates a mean list for each region
means_lst.append(st.mean(errorNE))            #adds the Northeastern error mean to the list

Data_Sources = []                            #creates a Data Sources list of each region
Data_Sources.append('Northeast')           #adds the region to the list
```

In [10]:

```
plt.plot(ForecastNE, 'or-', ActualNE, 'ob-')
plt.xlabel('Months of the year')
plt.ylabel('Temperature(Farenheit)')
plt.title('Monthly Maximum Temperature in the Northeastern U.S.')
plt.legend(['Forecast Month','Actual data'])
```

Out[10]: <matplotlib.legend.Legend at 0x1c48393c8e0>



New York State Data

In [11]:
```
filename3 = r"C:\Users\hopey\OneDrive - University at Buffalo\2001-2020 MONTHLY MAX TEMP FOREC
NYSdata2001 = pd.read_csv(filename3)

filename4 = r"C:\Users\hopey\OneDrive - University at Buffalo\2021 MONTHLY MAX TEMP NYS (2).csv
NYSdata2021 = pd.read_csv(filename4)
```

In [12]:
```
Temp2001_NYS = NYSdata2001['Forecast(Value)'][240:]     #New York State Monthly Max Temp Forec
Date2001_NYS = NYSdata2001['DATE']                      # mmm-dd corresponding to the monthly m

Temp2021_NYS = NYSdata2021['Value']                     #New York State Monnthly Max Temp Obser
Date2021_NYS = NYSdata2021['DATE']                      # corresponding date
```

In [13]:
```
Temp2021_NYS  #actual temps of 2021
```

Out[13]:
```
0     30.2
1     30.3
2     46.8
3     56.9
4     66.4
5     78.4
6     76.9
7     80.0
8     71.5
9     63.1
10    45.5
11    40.1
Name: Value, dtype: float64
```

In [14]:
```
Temp2001_NYS #predicted temps of 2021
```

Out[14]:
```
240    32.842910
241    35.143588
242    42.229284
243    53.663331
244    68.292652
245    76.531114
246    82.643962
247    79.435305
248    72.908032
249    60.110483
250    46.722843
251    36.616338
Name: Forecast(Value), dtype: float64
```

In [15]:
```
ForecastNYS = []                                 #create list to contain the forecasted tempera
ActualNYS = []                                   #create list to contain the actual observed te
errorNYS = []                                    #create list to contain the error of the fore
                                                 #compared to the actual
for x,y in zip(Temp2001_NYS,Temp2021_NYS):
    ForecastNYS.append(x)                        #adds each months forecasted temp to Forecast
    ActualNYS.append(y)                          #adds each months actual temp to error list
    errorNYS.append(abs(x-y))                    #adds each months magnitude of error in degre
                                                 #error list
```

In [16]:
```
errorNYS
```

Out[16]:
```
[2.6429098300000042,
 4.843587930000002,
 4.5707156699999985,
 3.23666858,
 1.8926515099999932,
 1.8688861100000054,
 5.743961999999996,
 0.5646950000000004,
 1.4080315199999944,
 2.9895166499999988,
 1.2228429999999975,
 3.483662250000002]
```

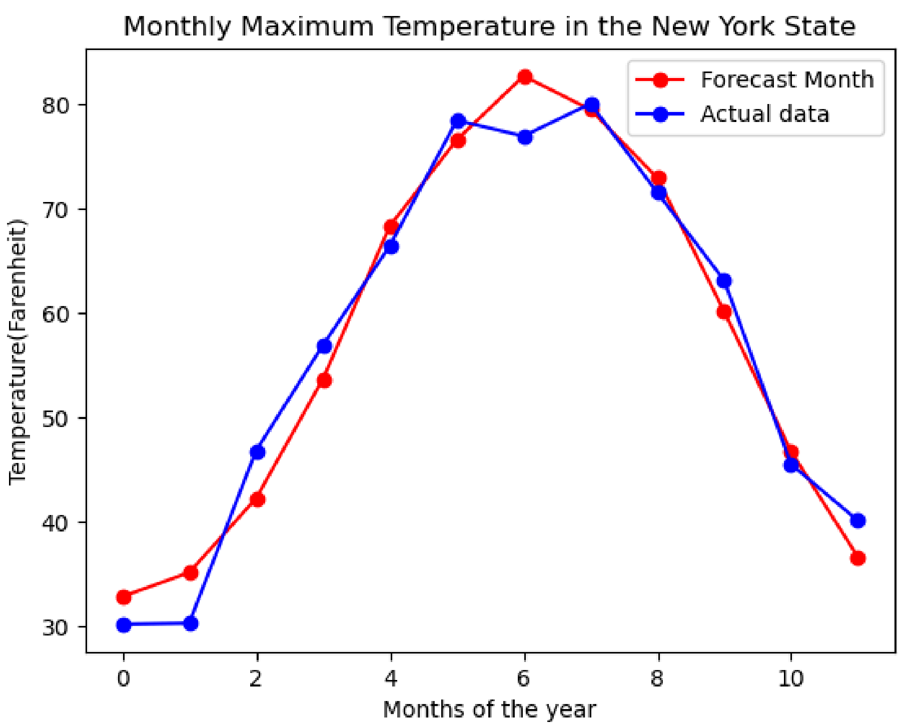
In [17]:
```
means_lst.append(st.mean(errorNYS))       #adds mean of New York State 2021 error to error mea

st_dev.append(st.stdev(errorNYS))          #adds the NYS error standard devation to error st. 

Data_Sources.append('New York State')   #adds the region to the list
```

In [18]:
```
plt.plot(ForecastNYS, 'or-', ActualNYS, 'ob-')
plt.xlabel('Months of the year')
plt.ylabel('Temperature(Farenheit)')
plt.title('Monthly Maximum Temperature in the New York State')
plt.legend(['Forecast Month','Actual data'])
```

Out[18]: <matplotlib.legend.Legend at 0x1c484053640>



Southeast region

In [19]:
```
filename5 = r"C:\Users\hopey\OneDrive - University at Buffalo\2001-2020 MONTHLY MAX TEMP SE FO
SEdata2001 = pd.read_csv(filename5)

filename6 = r"C:\Users\hopey\OneDrive - University at Buffalo\2021 MONTHLY MAX TEMP SE.csv"
SEdata2021 = pd.read_csv(filename6)
```

In [20]:
```
Temp2001_SE = SEdata2001['Forecast(Value)'][240:]      #Southeast Monthly Max Temp Forecasted l
Date2001_SE = SEdata2001['DATE']                       # mmm-dd corresponding to the monthly m

Temp2021_SE = SEdata2021['Value']                     #Southeast region Monnthly Max Temp Obse
Date2021_SE = SEdata2021['DATE']                      # corresponding date
```

In [21]:
```
Temp2001_SE #PREDICTED TEMP
```

Out[21]:
```
240    54.949960
241    59.561543
242    66.084901
243    73.674239
244    80.348563
245    85.690355
246    88.056649
247    86.935168
248    83.056345
249    74.923041
250    64.962813
251    58.061600
Name: Forecast(Value), dtype: float64
```

In [22]:
```
Temp2021_SE #ACTUAL TEMP
```

Out[22]:
```
0     56.6
1     58.4
2     70.2
3     74.8
4     81.2
5     86.2
6     88.4
7     88.8
8     83.8
9     77.9
10    65.2
11    67.0
Name: Value, dtype: float64
```

In [23]:
```
 ForecastSE = []                           #create list to contain the forecasted temperature
ActualSE = []                              #create list to contain the actual observed temper
errorSE = []                               #create list to contain the error of the forescast
                                           #compared to the actual
for x,y in zip(Temp2001_SE,Temp2021_SE):
    ForecastSE.append(x)                   #adds each months forecasted temp to Forecast list
    ActualSE.append(y)                     #adds each months actual temp to error list
    errorSE.append(abs(x-y))               #adds each months magnitude of error in degrees fa
```

In [24]:
```
errorSE
```

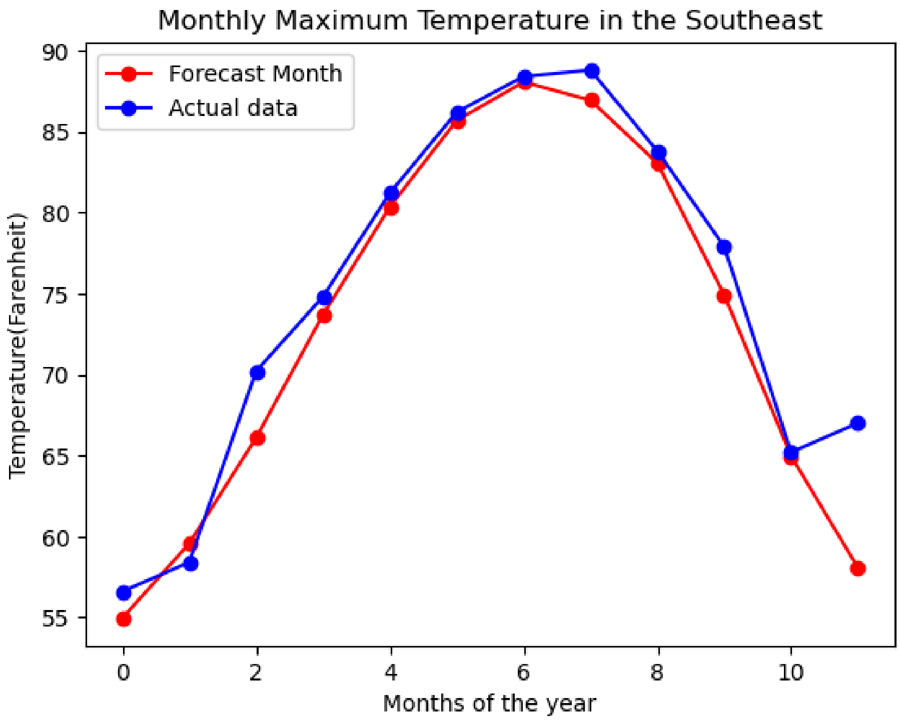
Out[24]:
```
[1.6500397100000015,
 1.1615432700000028,
 4.115099440000009,
 1.1257613599999985,
 0.8514365500000025,
 0.5096453299999979,
 0.34335143000001267,
 1.8648322299999904,
 0.7436545899999913,
 2.9769592800000027,
 0.23718669000000148,
 8.938400379999997]
```

In [25]:
```
st_dev.append(st.stdev(errorSE))        #adds the Southeast Region error standard devation to
means_lst.append(st.mean(errorSE))      #adds mean of Southeast 2021 error to error means list
Data_Sources.append('Southeast')        #adds the region to the list
```

In [26]:

```
plt.plot(ForecastSE, 'or-', ActualSE, 'ob-')
plt.xlabel('Months of the year')
plt.ylabel('Temperature(Farenheit)')
plt.title('Monthly Maximum Temperature in the Southeast')
plt.legend(['Forecast Month','Actual data'])
```

Out[26]: <matplotlib.legend.Legend at 0x1c4840d7e80>



The West (California and Nevada)

In [27]:

```
filename7 = r"C:\Users\hopey\OneDrive - University at Buffalo\2001-2020 MONTHLY MAX TEMP WEST |
Wdata2001 = pd.read_csv(filename7)

filename8 = r"C:\Users\hopey\OneDrive - University at Buffalo\2021 MONTHLY MAX TEMP WEST.csv"
Wdata2021 = pd.read_csv(filename8)
```

In [28]:

```
Temp2001_W = Wdata2001['Forecast(Value)'][240:]       #West Monthly Max Temp Forecasted Data
Date2001_W = Wdata2001['DATE']                        # mmm-dd corresponding to the monthly ma:

Temp2021_W = Wdata2021['Value']                       #The West region Monthly Max Temp Observe
Date2021_W = Wdata2021['DATE']                        # corresponding date
```

In [29]:

```
Temp2001_W #FORECAST
```

Out[29]:

```
240    52.461752
241    53.515765
242    62.795230
243    65.936310
244    75.911802
245    86.666410
246    93.064975
247    91.041021
248    85.627878
249    75.124608
250    58.306025
251    49.771352
Name: Forecast(Value), dtype: float64
```

In [30]:
```
Temp2021_W    #ACTUAL
```
Out[30]:
```
0     51.0
1     53.7
2     56.9
3     69.2
4     76.0
5     89.4
6     93.7
7     90.4
8     84.8
9     67.0
10    62.8
11    48.2
Name: Value, dtype: float64
```

In [31]:
```
ForecastW = []                                  #create list to contain the forecasted temperatures
ActualW = []                                    #create list to contain the actual observed tempera
errorW = []                                     #create list to contain the error of the forescaste
                                                #compared to the actual
for x,y in zip(Temp2001_W,Temp2021_W):
    ForecastW.append(x)                         #adds each months forecasted temp to Forecast list
    ActualW.append(y)                           #adds each months actual temp to error list
    errorW.append(abs(x-y))                     #adds each months magnitude of error in degrees far
```

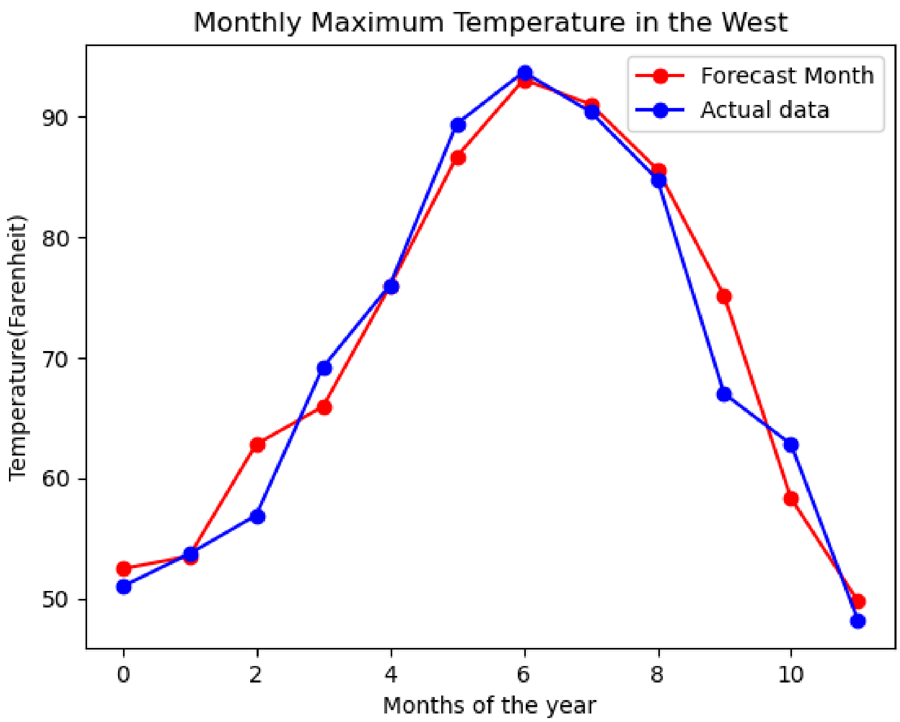
In [32]:
```
errorW
```
Out[32]:
```
[1.4617518000000018,
 0.1842349100000007,
 5.895229960000002,
 3.263690260000004,
 0.08819751000000053,
 2.7335895900000082,
 0.6350253700000081,
 0.6410213899999917,
 0.8278783300000043,
 8.124607909999995,
 4.493975469999995,
 1.5713521599999964]
```

In [33]:
```
st_dev.append(st.stdev(errorW))         #adds the West error standard devation to error st. de
means_lst.append(st.mean(errorW))       #adds mean of West 2021 error to error means list
Data_Sources.append('West')             #adds the region to the list
```

In [34]:

```
plt.plot(ForecastW, 'or-', ActualW, 'ob-')
plt.xlabel('Months of the year')
plt.ylabel('Temperature(Farenheit)')
plt.title('Monthly Maximum Temperature in the West')
plt.legend(['Forecast Month','Actual data'])
```

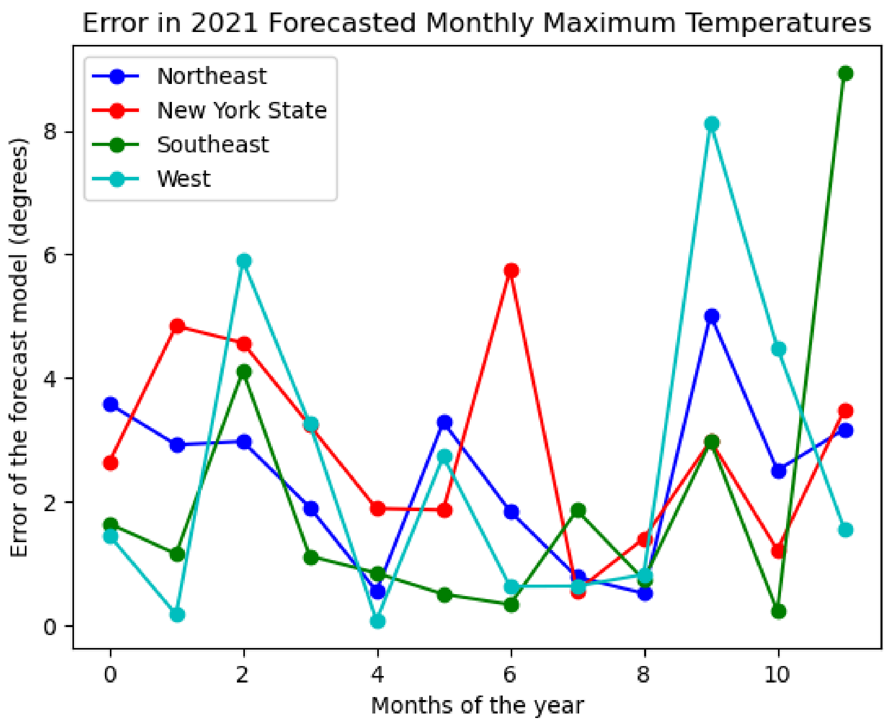
Out[34]: <matplotlib.legend.Legend at 0x1c485231df0>



Analysis between regions

In [35]:
```
plt.plot(errorNE,'-bo',errorNYS,'-ro', errorSE,'-go',errorW,'-co')
plt.xlabel('Months of the year')
plt.ylabel('Error of the forecast model (degrees)')
plt.legend(['Northeast','New York State','Southeast','West'])
plt.title('Error in 2021 Forecasted Monthly Maximum Temperatures')
```

Out[35]: Text(0.5, 1.0, 'Error in 2021 Forecasted Monthly Maximum Temperatures')

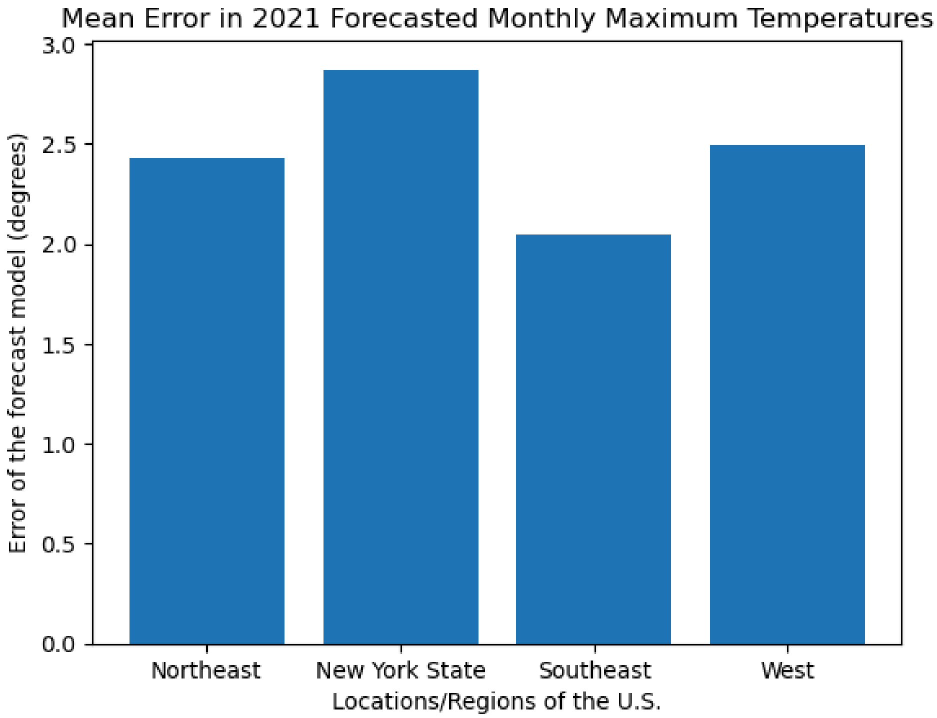


In [36]:
```
means_lst #means of the regions
```

Out[36]: [2.426041581666666, 2.872344170833333, 2.043159188333334, 2.4933795550000006]

In [37]:
```
plt.bar(Data_Sources,means_lst)
plt.xlabel('Locations/Regions of the U.S.')
plt.ylabel('Error of the forecast model (degrees)')
plt.title('Mean Error in 2021 Forecasted Monthly Maximum Temperatures')
```

Out[37]: Text(0.5, 1.0, 'Mean Error in 2021 Forecasted Monthly Maximum Temperatures')



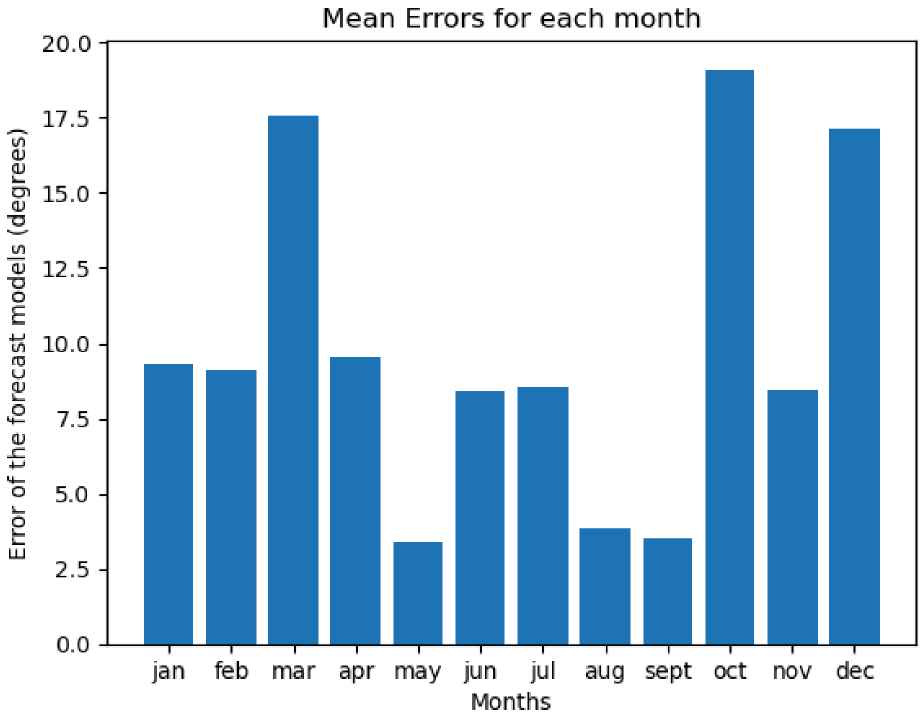
In [38]:
```
cmb_error_lst = []  #not meaningful
for i,j,k,l in zip(errorNE,errorNYS,errorSE,errorW):
    cmb_error_lst.append(i+j+k+l)
cmb_error_lst

```

Out[38]:
```
[9.339935270000005,
 9.114740420000008,
 17.557312920000008,
 9.539422540000004,
 3.392185139999995,
 8.401434160000008,
 8.570165030000013,
 3.8580882899999835,
 3.50245464999999,
 19.11188018,
 8.465338779999996,
 17.166136569999992]
```

In [39]:
```
plt.bar(['jan','feb','mar','apr','may','jun','jul','aug','sept','oct','nov','dec'],cmb_error_l
plt.xlabel('Months')
plt.ylabel('Error of the forecast models (degrees)')
plt.title('Mean Errors for each month')
```

Out[39]: Text(0.5, 1.0, 'Mean Errors for each month')



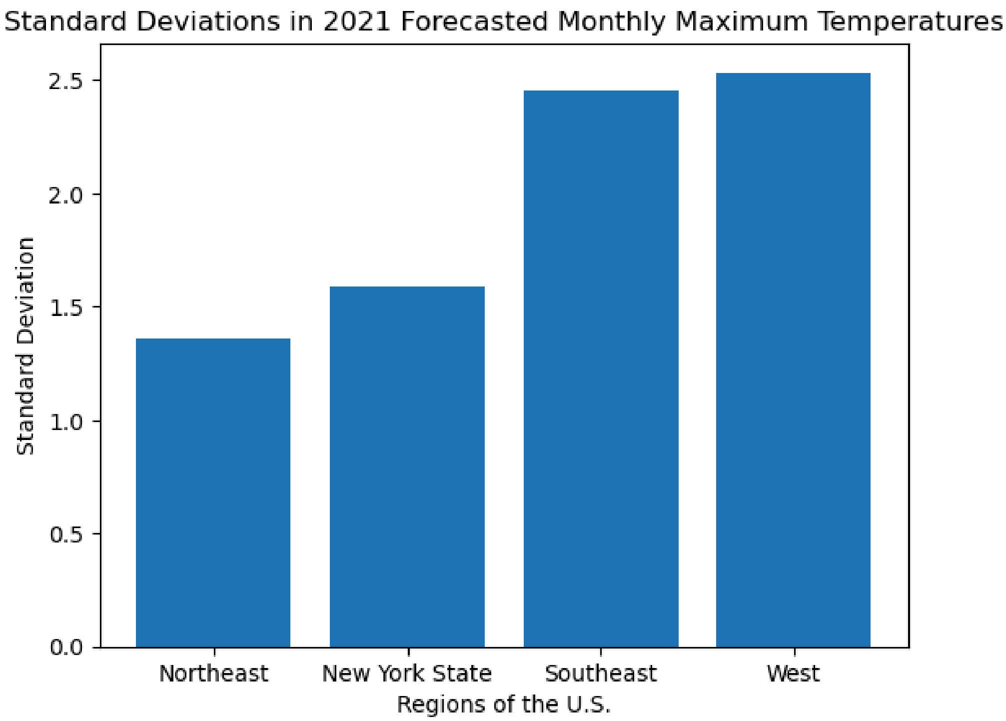
In [40]:
```
st_dev #standard deviations of the regions
```

Out[40]: [1.3593112989301914, 1.5882037989457358, 2.4518256986407168, 2.533894376151006]

In [41]:

```
plt.bar(Data_Sources,st_dev)
plt.xlabel('Regions of the U.S.')
plt.ylabel('Standard Deviation')
plt.title('Standard Deviations in 2021 Forecasted Monthly Maximum Temperatures')
```

Out[41]: Text(0.5, 1.0, 'Standard Deviations in 2021 Forecasted Monthly Maximum Temperatures')



In [42]:

```
maxi = 0
mini = 0
for i in Temp2021_W:
    if i < maxi:
        mini = i
    if i > mini:
        maxi = i
diff_W = maxi - mini

maxi = 0
mini = 0
for i in Temp2021_NYS:
    if i < maxi:
        mini = i
    if i > mini:
        maxi = i
diff_NYS = maxi - mini

maxi = 0
mini = 0
for i in Temp2021_SE:
    if i < maxi:
        mini = i
    if i > mini:
        maxi = i
diff_SE = maxi - mini

maxi = 0
mini = 0
for i in Temp2021_NE:
    if i < maxi:
        mini = i
    if i > mini:
        maxi = i
diff_NE = maxi - mini

```

In [43]:

```
 diffs = [diff_NE,diff_NYS,diff_SE, diff_W]
```

In [44]:

```
diffs
```

Out[44]:

```
[39.699999999999996, 39.9, 21.799999999999997, 45.5]
```

In [45]:

```
plt.bar(Data_Sources,diffs)
plt.xlabel('Regions of the U.S.')
plt.ylabel('Degrees Farenheit')
plt.title('Magnitude of maximum monthly temperature differences')
```

Out[45]: Text(0.5, 1.0, 'Magnitude of maximum monthly temperature differences')