Team Control Number

For office use only
T1 ________________
T2 ________________
T3 ________________
T4 ________________

92609

Problem Chosen

C

For office use only
F1 ________________
F2 ________________
F3 ________________
F4 ________________

2018

MCM/ICM
Summary Sheet
(Your team's summary should be included as the first page of your electronic submission.)
Type a summary of your results on this page. Do not include the name of your school, advisor, or team members on this page.

Overall the energy consumption for the states California (CA), Arizona (AZ), New Mexico (NM), and Texas (TX) increases over the time period from 1960 to 2009. We narrow down a list of 605 variables affecting the energy profiles of these states to 56 variables. We use two different methods to create models for the energy profiles.

We then calculate the renewable energy potential for each state by taking the renewable energy produced in 2009 and dividing it by the total energy consumed in 2009. The renewable energy potential of California is 7.93%, the renewable energy potential of Arizona is 6.09%, the energy efficiency of New Mexico is 5.04%, and the energy efficiency of Texas is 2.69%. The renewable energy total consumption, total energy consumption, renewable energy consumption, and total energy consumption increase over the time period.

Variables such as geography, industry, population, and climate affect the energy consumption and production of each state. The general climate of these states is arid and hot with some mountain and desert regions. The population of each of the states has increased exponentially from 1960 to 2009. Technological advancements in energy production such as in solar panels or windmills affects the renewable energy consumption and production. More industrialization also increases the consumption and production of energy. The state with the best energy profile appears to be California, with an efficiency of 7.93%.

We use two models to create a piecewise continuous function to better re-sample the data. This allows us to create a more representative time series, which we can then fit a time series process to. Finally, we can evaluate these processes to forecast how the data will look in the future. From these predictive values, we determine that California has the best future outlook as it relates to renewable energy production, as well as energy exported.

Summary

Overall the energy consumption for the states California (CA), Arizona (AZ), New Mexico (NM), and Texas (TX) increases over the time period from 1960 to 2009. We narrow down a list of 605 variables affecting the energy profiles of these states to 56 variables. We use two different methods to create models for the energy profiles.

We then calculate the renewable energy potential for each state by taking the renewable energy produced in 2009 and dividing it by the total energy consumed in 2009. The renewable energy potential of California is 7.93%, the renewable energy potential of Arizona is 6.09%, the renewable energy potential of New Mexico is 5.04%, and the renewable energy potential of Texas is 2.69%. The renewable energy total consumption, total energy consumption, renewable energy consumption, and total energy consumption increase over the time period.

Variables such as geography, industry, population, and climate affect the energy consumption and production of each state. The general climate of these states is arid and hot with some mountain and desert regions. The population of each of the states has increased exponentially from 1960 to 2009. Technological advancements in energy production such as in solar panels or windmills affects the renewable energy consumption and production. More industrialization also increases the consumption and production of energy. The state with the best energy profile appears to be California, with a potential of 7.93%.

We use two models to create a piecewise continuous function to better re-sample the data. This allows us to create a more representative time series, which we can then fit a time series process to. Finally, we can evaluate these processes to forecast how the data will look in the future. From these predictive values, we determine that California has the best future outlook as it relates to renewable energy production, as well as energy exported.

Contents

1. Introduction

1.1. Restatement of Problem

Energy production and consumption are important to any economy. In the United States, energy policies are decentralized to the state level. An interstate compact is an agreement between two or more states on a policy. Varying geographies, industries, populations, and climates affect energy consumption and production. The problem at hand is to create an energy profile for four US states, California (CA), Arizona (AZ), New Mexico (NM), and Texas (TX) based on given date on 605 variables on each of these four states' energy consumption and production.

1.2. Overview

We present a non-standard approach towards the modeling of discrete data sets. Through the use of *Taylor Polynomials* we define a system of equations built with coefficients stored in a *Directed Graph*. From here we re-sample the system and use these data points to generate *Time Series* for the interpretation and prediction of data.

1.3. Definitions

- Renewable Energy - energy which is constantly replenished on a human timescale.
- Nonrenewable Energy - energy generated from finite sources, or sources that gradually decrease in time.
- Time Series - A series of values of a quantity usually taken at regular time intervals.

2. Assumptions

1. No policy changes resulting in markedly decreased or increased data points.
2. No new technology surfaces making any of these energy sources obsolete.
3. The ability to fit data to a distribution.
4. Linearity of data.

3. Our Proposal

As seen above in 1 the total amount of renewable energy consumption generally increases over time for California, Arizona, New Mexico, and Texas. A possible reason for the increase in renewable energy consumption in these states is rising temperatures [1] [2] [3] [4]. Much of these states is dry and warm. Rises in temperature over time disrupts participation. California in particular is known for its wildfires and drought. Increases in temperature causes an increase in the usage of air conditioning. The population in these

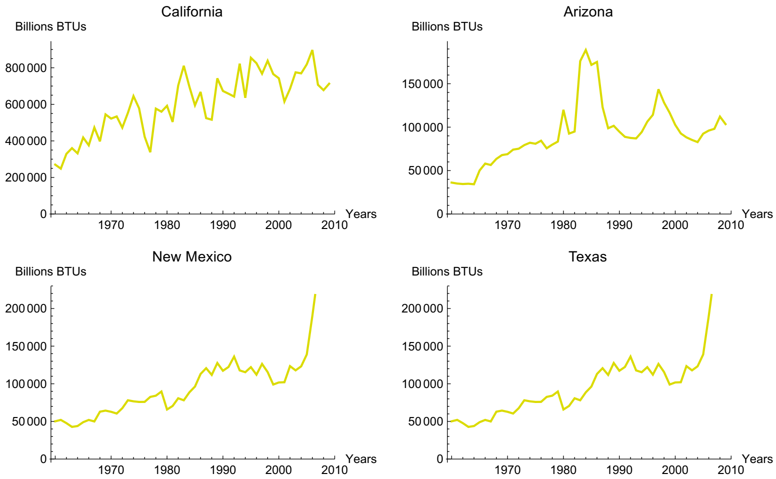

Figure 1: Renewable Energy Total Consumption

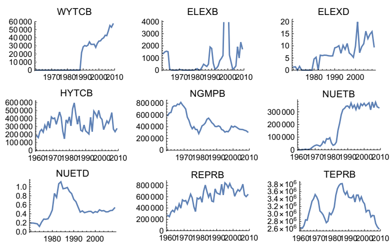

Figure 2: California produced energy

states has also been rising exponentially from 1960 to 2009 [5] [6] [7] [8]. Increases in population correlate with increases in energy consumption. Solar panels are becoming more popular, cheaper, and effective. Wind power is also becoming more wide-spread. Much of Southern California is windy and California's wind capacity has increased by 350% since 2001 as we can see in 2. As a result, there is more renewable energy consumed and produced.

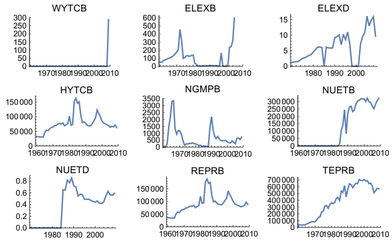

Figure 3: Arizona produced energy

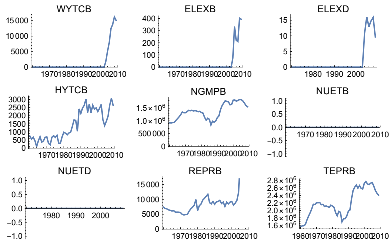

Figure 4: New Mexico produced energy

4. Modeling

4.1. Primary model

In our model, we create functions of each relevant variable in each state with respect to time. To do this we approximate the average rate of change of each variable for each year by subtracting the value of an arbitrary monomial $f(x) = x^r$, evaluated at each year $f(c)$ of the following year $f(c+1)$.Once we have these rates of change, we can estimate the second order derivative by repeating the process. We discuss the process further in the next section, and leave it to the reader to verify this process terminates at the n^{th} iteration. For each $f^{(n)}$ we take the last element of the subtraction process (indicating our approximation is centered at 2009). This allows us to evaluate $f^{(n)}(c)$ and construct a *Taylor Polynomial* [9]:

$$\sum_{n=0}^{m} f^{(n)}(m-n) \times (t-m)^n \times \frac{1}{\Gamma(n+1)} \tag{1}$$

where m is the approximated highest order derivative. We've implemented a modified version of this process, which allows for modification of which $f^{(n)}(c)$ we choose, as well as where the polynomial is centered.

4.1.1. Calculation of Slope

The polynomial generated gives an approximation of the variable over time. Through the repeated subtraction process is illustrated below with a sample data set. With the

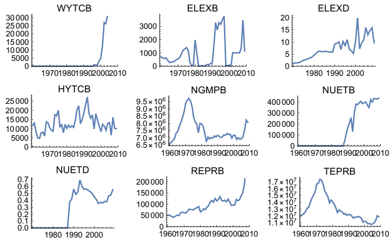

Figure 5: Texas produced energy

top row being $n = 0$, the n^{th} row shows the n^{th} derivative of the function, which can the substituted into the expression $f^{(n)}(m - n)$.

In fact, we can understand and view the coefficients of the Taylor polynomial as nodes in a directed graph, G. We assert the *source* vertices are the numeric values of our initial dataset, which we denote $\{x_1, x_2, \ldots, x_n\}$ and the *sink* vertex, v, is the only value remaining after n iterations of the subtraction process. By calculating a *shortest path* **cite** between any of the x_i and v, we now have the coefficients for a modified Taylor-like power series centered about $i - 1$. This gives us a polynomial of the form

$$\sum_{n=0}^{m} f^{(n)}(m-n)_j \times (t-m)^n \times \frac{1}{\Gamma(n+1)} \tag{2}$$

where m is the approximated highest order derivative, and $(m-n)_j$ may change due to the path finding. As we can see in figure 6 this allows us a rather intuitive approach towards calculating the coefficients of our polynomials. By utilizing either Dijkstra's Algorithm, or the Bellman-Ford, [10] can determine that the coefficients of our polynomials will be as low as possible; as the presence of negative vertex weights is allowed.

This allows us to generate a system of Taylor polynomials for each variable, each a reliable approximation of the original data. We can then adjust the original model accordingly.

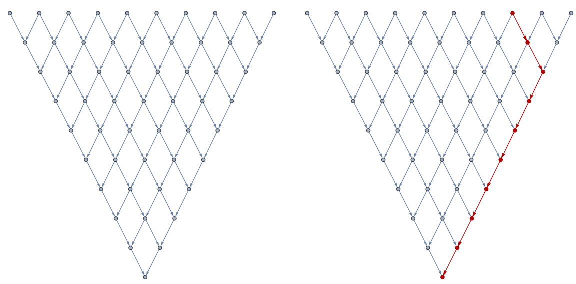

Figure 6: Repeated Subtraction Process

4.2. Adjusted Model

We propose the following method towards generating a system of piecewise continuous functions to better estimate and understand the systematic change of energy in California, Arizona, New Mexico, and Texas. By utilizing our method of approximating rates of change, we define,

$$p(t) = \begin{cases} \sum_{n=0}^{m} f^{(n)}(m-n)_k \times (t-1)^n \times \frac{1}{\Gamma(n+1)} & 0 < x \leq 1 \\ \dots & \\ \sum_{n=0}^{m} f^{(n)}(m-n)_h \times (t-n)^n \times \frac{1}{\Gamma(n+1)} & x \geq n-1 \end{cases}$$

which allows us to determine the general tendency of these data. Furthermore, with the added benefit of piecewise continuity, we can evaluate the instantaneous rate of change for these functions. Refining our intervals for an arbitrary value M, such that $p'(t_0) \geq M$ requires a reduction in interval length; likewise, for an arbitrary m if $p'(t) \leq m$ we decrease our interval as well.

We should note that by sampling our piecewise functions, we can retrieve more data points to approximate a function. Using the method of least squares, we may be able to ascertain a continuous function that allows for more accurate metrics to be run. Likewise, we can use these data points, or $p(t)$ itself, to find the average slope of the data across the interval. This gives us a rudimentary prediction for where these data might go, but we suggest a different approach.

4.3. Predictive Model

After converting the provided data to Time Series, we used the built in Mathematica [11] command TimeSeriesModelFit to generate a set of models to use in our forecast. A sample of these models for a subset of our selected variables can be seen below 7. Please note they only represent data for the state of Arizona.

Upon some additional research, we utilized the following processes from Mathematica: Autoregressive Models (AR), Moving Average (MA), Autoregressive Moving Average (ARMA), Autoregressive Integrated Moving Average (ARIMA), Seasonal Autoregressive Moving Average (SARMA), Seasonal Autoregressive Integrated Moving Average (SARIMA), Autoregressive Conditional Heteroskedasticity (ARCH), and Generalized Autoregressive Conditional Heteroskedasticity (GARCH) [12]. Please note that for all of these processes, we make the basic assumption that our data is linear and can be fit to a distribution (Gaussian or otherwise).

It's for this reason that we suggest sampling the system $p(t)$, and using convolution [13] with a least squares fit, f, which is defined as

$$[p \times f](t) = \int_{-\infty}^{\infty} p(\tau) f(t-\tau) d\tau \tag{3}$$

and fitting the time series process to the linear interpolation of this re-sampling. This would provide the most accurate prediction of data, and could be generalized to the whole of the dataset given our adjusted model.

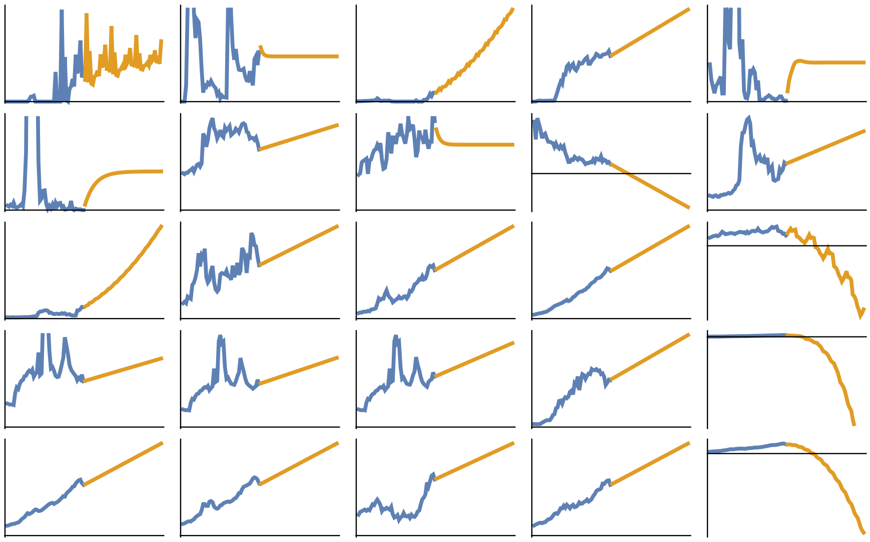

Figure 7: Time Series with Processes given in orange

Based on the comparison between the four states, the renewable energy potential of each state, and the predictions, renewable energy targets for 2025 and 2050 are seen in the time series in figure 7. These targets should be goals for the new four-state energy compact. Some actions that the four states might take to meet their energy compact goals include more usage of renewable energy sources such as solar panels and windmills and less usage of nonrenewable energy such as oil and coal. These states would benefit from an increase in the number of solar panels since the climate is mostly dry and hot in these areas of the US. The states could increase funding for renewable energy sources. The states could also implement laws that would penalize companies for using nonrenewable sources of energy and reward companies for using renewable sources of energy.

5. Model Assessment

5.1. Strengths

- For recent data (approx. 10 years), the primary model, which is significantly less taxing computationally, is more than sufficient.
- Our path finding algorithm approach produces remarkably accurate power series approximations similar to a Taylor polynomial.
- Left or right shift is easily modified by vertex index, or arbitrary scaling factor
- Increased accuracy in re-sampling data for use in predictive analysis
- Generalizable to all variables in the dataset, and automatically performs constructs a near exact representation of sample data

5.2. Weaknesses

- Our primary model, while providing accurate approximations about data near 2009, doesn't suggest that index choice and centering result in satisfactory approximations; a stark contradiction to the results of the path finding approach. In all honesty, we're not sure why our path finding approach works.
- Dijkstra's Algorithm tends to produce polynomials with a negative inclination, which may make approximation along different intervals less accurate.
- Time Series processes assume linearity of data, which may be problematic for variables with many leading or trailing zeros, as well as processes (like population growth) that may be exponential.

6. Conclusion

From these models, and the general evaluation of these data, we can determine that the overall renewable energy potential of these states can be shown by the following equations.

$$\textit{renewable energy potential} = \frac{\textit{renewable energy produced in 2009}}{\textit{total energy consumed in 2009}} \tag{4}$$

$$\textit{renewable energy potential of California} = \frac{635062.3653}{8005515.051} \times 100 = 7.932810834\% \tag{5}$$

$$\textit{renewable energy potential of Arizona} = \frac{88571.38442}{1454313.457} \times 100 = 6.090254064\% \tag{6}$$

$$\textit{renewable energy potential of New Mexico} = \frac{33785.17435}{670094.5064} \times 100 = 5.041852161\% \tag{7}$$

$$\textit{renewable energy potential of Texas} = \frac{303697.0626}{11297410.59} \times 100 = 2.68820063\% \tag{8}$$

As we can see by equations 4 and 5, California had the best energy profile of the states. In fact, California appeared to have the best renewable energy outlook according to our model, which can be seen in figure 8.

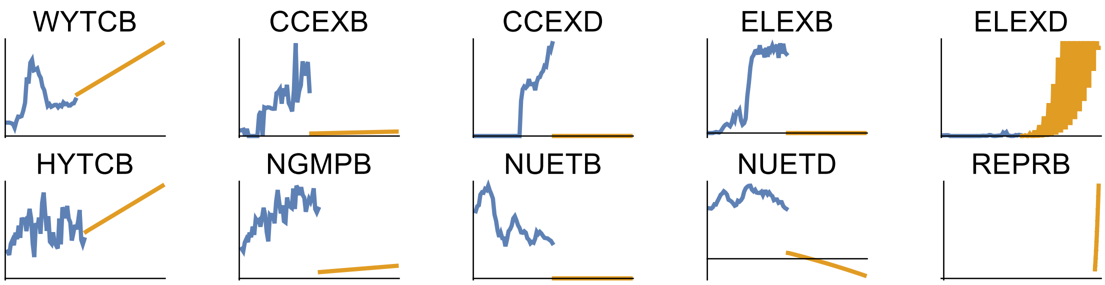

Figure 8: Forecast of Energy Production in California

7. A Letter to the group of Governors

Dear Governors Edmund G. Brown Jr., Doug Ducey, Susana Martinez, and Greg Abbott,

Our team has developed a model to analyze and predict how renewable energy has changed and will change in your states. We came up with a method to rank the states based on total renewable energy produced and total energy consumed in each state. According to our calculations, the states ranked in terms of renewable energy potential from highest to lowest are: California, Arizona, New Mexico, and Texas. California has a renewable energy potential of 7.93%, Arizona has a renewable energy potential of 6.09%, New Mexico has a renewable energy potential of 5.04%, and Texas has a renewable energy potential of 2.69%.

For the future, we recommend that each of these states increase its renewable energy potential to above 10%. Some realistic actions these states could to meet this new goal are to use more renewable energy sources such as solar panels and windmills and less nonrenewable energy such as oil, coal, and wood. These states would benefit from an increase in the number of solar panels since the climate is mostly arid and hot in these areas of the United States.

In regard to Texas' low renewable energy potential, we conclude that reduced petroleum consumption, as is already the tendency shown in 9, may be beneficial and increase the renewable energy potential.

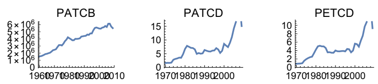

Figure 9: Texas Petroleum Consumption

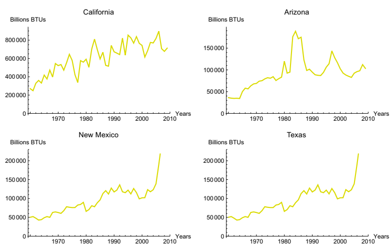

Figure 10: Renewable Energy Total Consumption

Yours Sincerely,
Team #92609

References

[1] *California Climate*. [Online]. Available: http://baydeltaoffice.water.ca.gov/climatechange/200610_ClimateChangeHistorical_CALFEDScience_manderso.pdf.

[2] *Texas Climate*. [Online]. Available: https://www.texasobserver.org/climate-change-means-texas-11-charts/.

[3] *New Mexio Climate*. [Online]. Available: https://weather.nmsu.edu/climate/about/.

[4] *Arizona Climate*. [Online]. Available: https://azclimate.asu.edu/climate/.

[5] *California Population (Census)*. [Online]. Available: https://www.google.com/publicdata/explore?ds=kf7tgg1uo9ude_&met_y=population&hl=en&dl=en.

[6] *New Mexico Population (Census)*. [Online]. Available: https://www.google.com/publicdata/explore?ds=kf7tgg1uo9ude_&met_y=population&hl=en&dl=en.

[7] *Arizona Population (Census)*. [Online]. Available: https://www.google.com/publicdata/explore?ds=kf7tgg1uo9ude_&met_y=population&hl=en&dl=en.

[8] *Texas Population (Census)*, https://www.google.com/publicdata/explore?ds=kf7tgg1uo9ude_&met_ [Online]. Available: https://www.tsl.texas.gov/ref/abouttx/census.html.

[9] C. Remani, “Numerical methods for solving systems of nonlinear equations,” *Lakehead University, Thunder Bay, Ontario, Canada*, vol. 13, 2012.

[10] E. W. Weisstein, *"dijkstra's algorithm." from mathworld–a wolfram web resource.* [Online]. Available: http://mathworld.wolfram.com/DijkstrasAlgorithm.html.

[11] Wolfram Research, Inc., *Mathematica 8.0*, version 0.8, 2010. [Online]. Available: https://www.wolfram.com.

[12] R. Adhikari and R. K. Agrawal, “An introductory study on time series modeling and forecasting,” *ArXiv preprint arXiv:1302.6613*, 2013.

[13] E. W. Weisstein. [Online]. Available: http://mathworld.wolfram.com/Convolution.html.

Code

```
Clear[MethodOfDescent];
MethodOfDescent[list_] := Module[{l = list},
  Clear[Overlap, slope, edges, edgeset, Tree];
  Overlap = Partition[l, 2, 1];
  slope[{a_, b_}] := b - a;
  edges[{a_, b_}] := {a \[DirectedEdge] (b - a),
    b \[DirectedEdge] (b - a)};
  edgeset = {};
  Tree = {list};
  While[
   Length[Overlap] != 0,
   {
     edgeset = Union[edgeset, Flatten[Map[edges, Overlap]]],
     Overlap = Map[slope, Overlap],
     If[Length[Overlap] == 1, RootVertex = Overlap, Continue],
     Overlap = Partition[Overlap, 2, 1]
     };
   ];
  Return[{edgeset, l, RootVertex}];
  ]

  (*Define Functions for calculation of slope using method of descent*)

Clear[RateOfChange, TableOfChange, OverlappedLists];
RateOfChange[{a_, b_}] := b - a;
TableOfChange[data_] := Table[RateOfChange[i], {i, data}];
OverlappedLists[data_] := Map[Partition[#, 2, 1] &, data];

Clear[ModelInitialization];
(*Initialize the model, this command must be run first using \
{AZPart,TXPart,NMPart,CAPart} or {AZPart0,TXPart0,NMPart0,CAPart0}*)

ModelInitialization[data_, OptionsForData_] :=
  Module[{d = data, o = OptionsForData},
   {PSAZ, PSTX, PSNM, PSCA} = Map[OverlappedLists, d];
   {PSAZ, PSTX, PSNM, PSCA} =
    Map[Table[TableOfChange[i], {i, #}] &, {PSAZ, PSTX, PSNM, PSCA}];
   Slopes =
    Append[Slopes, Table[Map[o, i], {i, {PSAZ, PSTX, PSNM, PSCA}}]];
   ];

Clear[Model];
(*Modeling the data, m is the center of the Taylor Series, \
OptionForData can be Mean, Median, First, Last, etc...*)

Model[dataset_, m_, OptionForData_] := Module[
  {data0 = dataset, o = OptionForData},

  Clear[AZPart, TXPart, NMPart, CAPart];
  {AZPart, TXPart, NMPart, CAPart} =
   Table[Map[DataFunc, i, {2}], {i, data0}];
```

```

ModelInitialization[{PSAZ, PSTX, PSNM, PSCA}, o];

n = 0;
Print[ProgressIndicator[Dynamic[n], {0, 49}]];

Slopes = {};
Slopes =
 Append[Slopes,
   Table[Map[o, i], {i, {AZPart, TXPart, NMPart, CAPart}}]];
ModelInitialization[{AZPart, TXPart, NMPart, CAPart}, o];
While[
 n < 49,
 {ModelInitialization[{PSAZ, PSTX, PSNM, PSCA}, o]}
 ; n++];

CenterOfSeries = m;

Clear[TaylorSeries];
TaylorSeries[x_, {n0_}] :=
 x*(t - CenterOfSeries)^(n0 - 1)*(1/Factorial[n0]);

Clear[AZSlopes, TXSlopes, NMSlopes, CASlopes];
{AZSlopes,
   TXSlopes,
   NMSlopes,
   CASlopes} =
  Table[
   Transpose[
    PadRight[
     Map[
       Part[#, 1] &,
       Table[
        Map[
          DeleteCases[#, _Last] &,
          i],
        {i,
         Drop[
           Slopes,
           -1]
         }
        ]
       ]
      ]
    ],
   {j,
    {1,
     2,
     3,
     4}
    }
   ];

{AZMonomials,
```

```
      TXMonomials,
      NMMonomials,
      CAMonomials} =
     Table[
      Table[
       MapIndexed[
        TaylorSeries,
        i],
       {i,
        j}
       ],
      {j,
       {AZSlopes,
        TXSlopes,
        NMSlopes,
        CASlopes}
       }
      ];

  {AZSystemConsumption,
      TXSystemConsumption,
      NMSystemConsumption,
      CASystemConsumption} =
     Table[
      Map[Total[#] &,
       i],
      {i,
       {AZMonomials,
        TXMonomials,
        NMMonomials,
        CAMonomials}
       }
      ];

  Return[
   {AZSystemConsumption,
    TXSystemConsumption,
    NMSystemConsumption,
    CASystemConsumption}
   ];
  ]

  Clear[pathModel];
pathModel[dataset_, {indx1_, indx2_, indx3_}] :=
  MethodOfDescent[Transpose[dataset[[indx1]][[indx2]]][[indx3]]];

Clear[TaylorSeries0];
TaylorSeries0[x_, {n0_}, m_] :=
  x*(t - m + 1)^(n0 - 1)*(1/Factorial[n0]);

Clear[PathingTaylorSeries]
PathingTaylorSeries[dataset_, {indx1_, indx2_, indx3_}, m_,
  VertexNum_] := Show[
  ListLinePlot[
```

```
   TimeSeries[
    Transpose[
      dataset[[indx1]][[indx2]]][[indx3 + 1]],
    {Transpose[dataset[[indx1]][[indx2]]][[indx3]]} - 1960]
   ],
  Plot[
   Total[
    MapIndexed[
     TaylorSeries0[#1, {#2}, m] &,
     FindShortestPath[
      graph1,
      g1[[2]][[VertexNum]],
      g1[[3]][[1]]
      ]
     ]
    ],
   {t, 0, 55},
   PlotStyle -> Orange]
  ]

  Clear[PathingTaylorSeriesLinkedbyVertexIndx]
PathingTaylorSeriesLinkedbyVertexIndx[
  dataset_, {indx1_, indx2_, indx3_}] := Module[{},
  g1 = pathModel[dataset, {indx1, indx2, indx3 + 1}];
  graph1 =
   Graph[g1[[1]], GraphLayout -> "LayeredDigraphEmbedding",
    ImageSize -> Medium];
  MinPaths = Table[
    FindShortestPath[
     graph1,
     g1[[2]][[i]],
     g1[[3]][[1]]
     ], {i, 1, 50}
    ];
  ShortestPaths = Table[Total[
     MapIndexed[
      TaylorSeries0[#1, {#2}, i - 1] &,
      MinPaths[[i]]
      ]
     ],
    {i, 1, 50}];
  Print["Making Plots, this might take awhile..."];
  count = 0;
  Print[ProgressIndicator[Dynamic[count], {0, 49}]];
  Plots = Map[{Plot[
       ShortestPaths[[#]],
       {t, 0, 50},
       PlotStyle -> Orange], count++} &, Range[50]];
  Deploy[
   Manipulate[GraphicsRow[{
      Show[
       ListLinePlot[
        TimeSeries[
         Transpose[
```

```
            dataset[[indx1]][[indx2]]][[indx3 + 1]],
          {Transpose[dataset[[indx1]][[indx2]]][[indx3]]} - 1960]
         , ImageSize -> Medium],
        Plots[[VertexIndex]][[1]]
        ],
       HighlightGraph[graph1,
        PathGraph[MinPaths[[VertexIndex]], DirectedEdges -> True]]}]
   , {VertexIndex}, {{VertexIndex, 1, "Vertex Index"}, 1, 50, 1}
   , FrameMargins -> 50]
  ]
 ]
```