Electricity Consumption and Renewable Energy Generation in Colorado

Yaquan Yang, Weixiao Li, Haocheng Zhu

2023-04-17

Abstract:

To analyze the role of consumer behavior in driving the use of renewable energy and the need for energy efficiency practices in Colorado. This project uses a variety of machine learning models, such as multiple linear regression, decision trees, random forests, and support vector machines to analyze patterns and trends in electricity consumption as a consumer behavior. In addition, time series models are used to analyze the use of renewable energy in Colorado in recent years, and ARIMA and Holt-Winters models are used to forecast the total renewable energy production, wind power consumption, and coal power consumption in Colorado for the next seven years.

Introduction:

Increasing energy demand has led to a greater emphasis on renewable energy. This project attempts to examine the impact of consumer behavior on energy consumption and renewable energy conservation. It has been observed that electricity consumption in Colorado has been increasing in recent years. To address this consumption phenomenon, we apply various models to analyze the impact of consumer behavior on renewable energy conservation and energy consumption. A time series model is also used to predict the future use of renewable energy. The results show that renewable energy production and consumption show an upward trend, while non-renewable energy consumption shows a downward trend. The results of this analysis will provide insight into the potential impact of consumer behavior on energy consumption and renewable energy conservation.

Data source:

Electricity_Revenue_by_Utility_in_Colorado.csv:https://data.colorado.gov/Business/Electricity-Revenue-by-Utility-in-Colorado/gdh8-8pg4
Electricity_Revenue_in_Colorado.csv:https://data.colorado.gov/Business/Electricity-Revenue-in-Colorado/q6sk-tjm9
prod_btu_re_te:https://www.eia.gov/state/seds/seds-data-complete.php?sid=US#StatisticsIndicators

Prod_dataset:https://www.eia.gov/state/seds/seds-data-complete.php?sid=AL#CompleteDataFile
use_all_btu:https://www.eia.gov/state/seds/seds-data-complete.php?sid=AL#Consumption

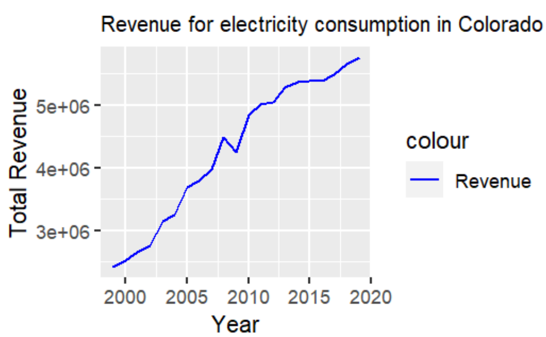
Electricity consumption

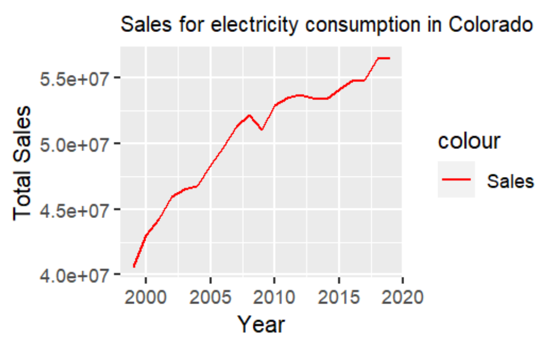
First, we will analyze the revenue and pricing of monthly electricity consumption in Colorado since 1990. The analysis of these data will allow us to understand the size and trends of the electricity market and explore the impact of pricing strategies on revenues.

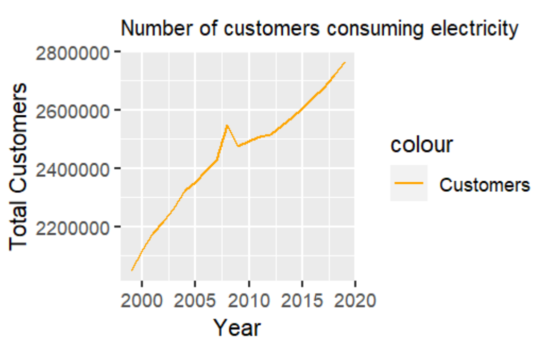
Electricity Revenue

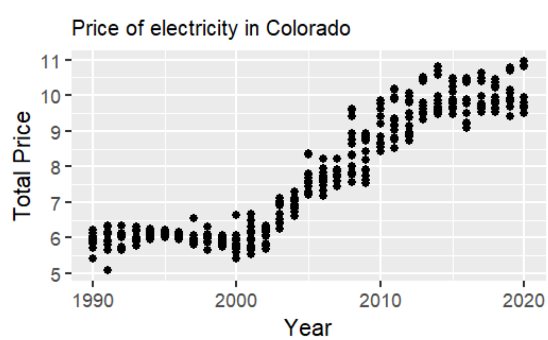
According to the data, the electricity market in Colorado has grown year by year since 1990, and the monthly electricity consumption has gradually increased. Especially in recent years, with the increase in population and economic development, the size of the electricity market has shown a rapid growth trend. At the same time, the level of electricity prices has also changed with the size of the market, with prices fluctuating, but generally showing an upward trend.

Rising demand may be driving renewable energy generation.



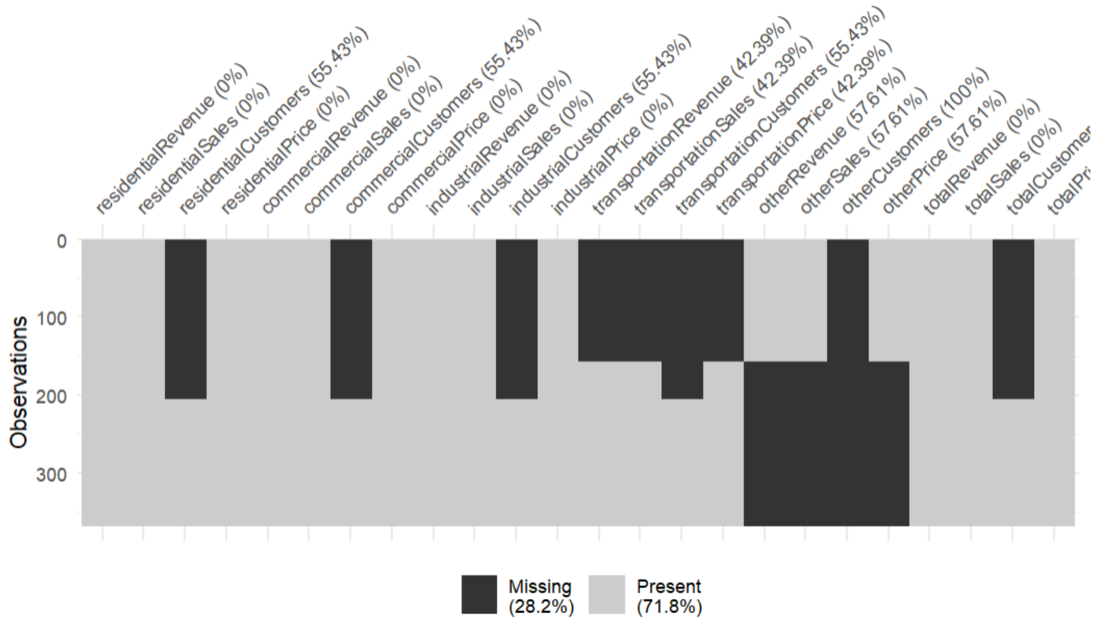






Now we build models with different machine learning algorithms to regress and predict the total electricity sales in Colorado to determine how electricity demand has changed over the last three decades.

Checking missing value



Model fitting

We fitted different machine learning models with total electricity sales as a response variable and made predictions.

Multiple linear regression

```
Call:
lm(formula = .outcome ~ ., data = dat)

Residuals:
     Min       1Q   Median       3Q      Max 
-1.28484 -0.12692  0.01091  0.14208  1.27401 

Coefficients: (1 not defined because of singularities)
                          Estimate Std. Error    t value Pr(>|t|)    
(Intercept)              2.117e-01  2.284e+00      0.093   0.9262    
residentialRevenue       3.097e-02  5.239e-02      0.591   0.5550    
residentialSales         1.000e+00  1.566e-06 638684.892   <2e-16 ***
residentialCustomers    -4.477e-06  2.406e-06     -1.861   0.0639 .  
residentialPrice         8.395e-01  9.835e-01      0.854   0.3941    
commercialRevenue        3.094e-02  5.239e-02      0.591   0.5553    
commercialSales          1.000e+00  1.913e-06 522874.965   <2e-16 ***
commercialCustomers      2.775e-05  1.579e-05      1.757   0.0800 .  
commercialPrice          1.367e+00  1.033e+00      1.323   0.1871    
industrialRevenue        3.090e-02  5.239e-02      0.590   0.5558    
industrialSales          1.000e+00  2.207e-06 453060.095   <2e-16 ***
industrialCustomers      1.969e-05  3.838e-05      0.513   0.6083    
industrialPrice          1.378e+00  9.324e-01      1.478   0.1406    
transportationRevenue    3.068e-02  5.244e-02      0.585   0.5590    
transportationSales      1.000e+00  1.998e-04   5006.032   <2e-16 ***
transportationCustomers -2.851e-01  1.191e+00     -0.239   0.8111    
transportationPrice     -6.150e-02  7.600e-02     -0.809   0.4191    
otherRevenue             3.113e-02  5.240e-02      0.594   0.5529    
otherSales               1.000e+00  1.100e-05  90938.307   <2e-16 ***
otherPrice              -1.114e-01  8.308e-02     -1.340   0.1813    
totalRevenue            -3.095e-02  5.239e-02     -0.591   0.5552    
totalCustomers                  NA         NA         NA       NA    
totalPrice              -3.244e+00  2.793e+00     -1.161   0.2465    
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 0.524 on 272 degrees of freedom
Multiple R-squared:      1,	Adjusted R-squared:      1 
F-statistic: 2.881e+13 on 21 and 272 DF,  p-value: < 2.2e-16
```

Decision Tree

```
CART

294 samples
 22 predictor

No pre-processing
Resampling: Cross-Validated (5 fold)
Summary of sample sizes: 235, 235, 236, 235, 235
Resampling results across tuning parameters:

  cp          RMSE      Rsquared   MAE
  0.07262643  289138.8  0.8546107  237341.7
  0.12616616  399998.1  0.7227249  330940.4
  0.69168734  556964.3  0.6629876  461579.7

RMSE was used to select the optimal model using the smallest value.
The final value used for the model was cp = 0.07262643.
```

Random Forest

```
Random Forest

294 samples
 22 predictor

No pre-processing
Resampling: Cross-Validated (5 fold)
Summary of sample sizes: 234, 235, 236, 236, 235
Resampling results across tuning parameters:

  mtry  RMSE       Rsquared   MAE
   2    131221.87  0.9722230  97911.90
  12     66706.54  0.9923153  50299.17
  22     67329.15  0.9919648  50910.04

RMSE was used to select the optimal model using the smallest value.
The final value used for the model was mtry = 12.
```

Support vector machine

```
Support Vector Machines with Radial Basis Function Kernel

294 samples
 22 predictor

No pre-processing
Resampling: Cross-Validated (5 fold)
Summary of sample sizes: 235, 234, 235, 236, 236
Resampling results across tuning parameters:

  C     RMSE       Rsquared   MAE
  0.25  138013.67  0.9710832  77450.79
  0.50  101257.92  0.9832036  62378.70
  1.00   78406.52  0.9899682  53656.12

Tuning parameter 'sigma' was held constant at a value of 0.2068405
RMSE was used to select the optimal model using the smallest value.
The final values used for the model were sigma = 0.2068405 and C = 1.
```

Model check

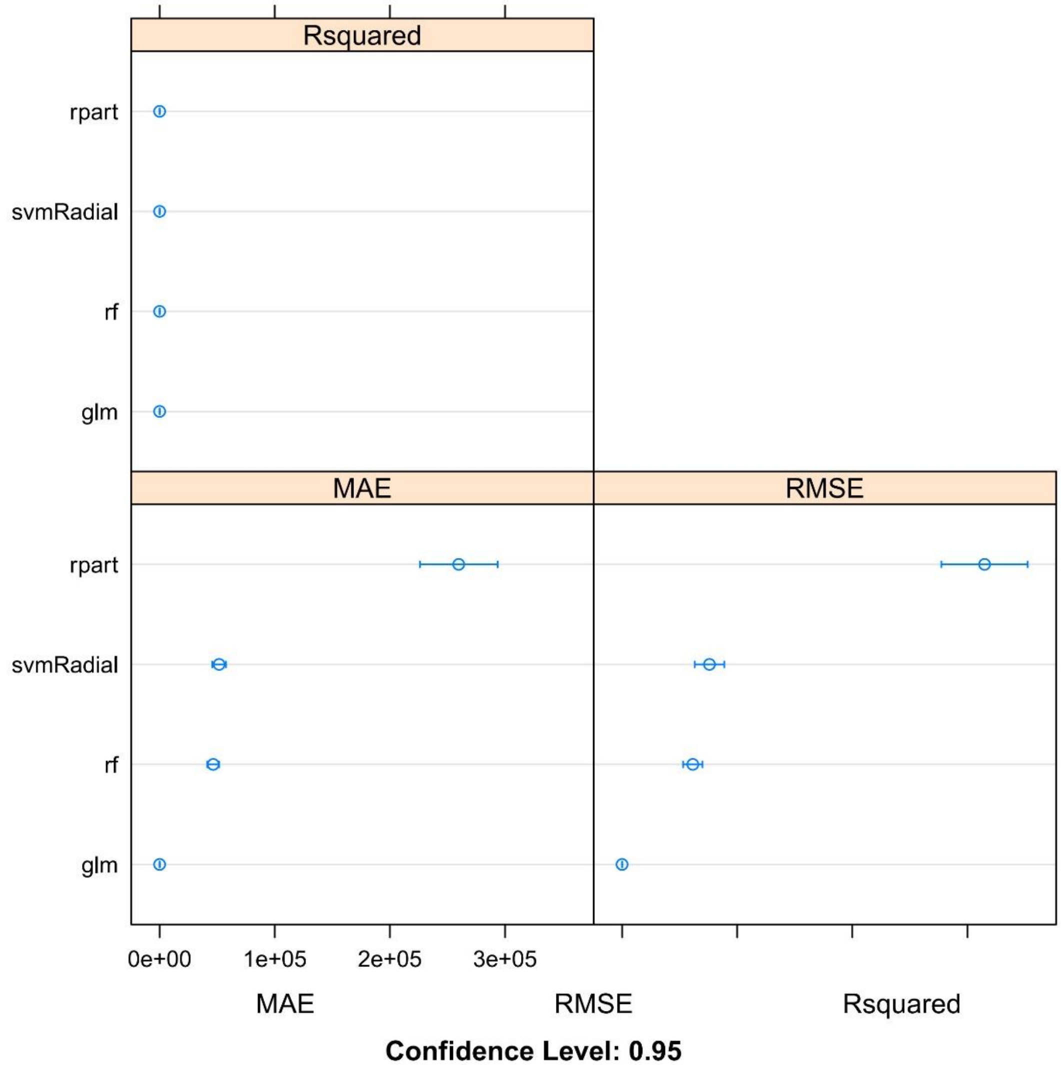
From the root mean square error (RMSE) of these four models, we can see that the multiple linear regression model performs best in predicting electricity sales, which may be related to the linear relationship between the response variable and the predictor variable.

```
[1] "RMSE of Multiple linear regression:  0.509245703449566"
[1] "RMSE of Decision Trees:  348140.414348679"
[1] "RMSE of Random Forest:  69334.3925098738"
[1] "RMSE of Support vector machine:  97473.7998019751"
```

Ensemble

Now we train multiple models simultaneously, cross-validate them and calculate evaluation metrics that help us choose the best model.

```
Call:
summary.resamples(object = results)

Models: glm, rf, rpart, svmRadial 
Number of resamples: 10 

MAE 
                  Min.      1st Qu.       Median         Mean      3rd Qu.         Max. NA's
glm       2.483054e-01 3.251156e-01 3.525566e-01 3.656808e-01 4.101310e-01 5.035633e-01    0
rf        3.856633e+04 4.121489e+04 4.527360e+04 4.655648e+04 4.977156e+04 6.046319e+04    0
rpart     1.831003e+05 2.199420e+05 2.771358e+05 2.597715e+05 2.865094e+05 3.249328e+05    0
svmRadial 3.808238e+04 4.929526e+04 5.213941e+04 5.169144e+04 5.420214e+04 6.703784e+04    0

RMSE 
                  Min.      1st Qu.       Median         Mean      3rd Qu.         Max. NA's
glm       3.916270e-01 5.098089e-01 5.432182e-01 5.564860e-01 6.173846e-01 6.927776e-01    0
rf        4.757392e+04 5.409260e+04 5.896705e+04 6.142299e+04 6.762439e+04 8.299571e+04    0
rpart     2.366738e+05 2.702184e+05 3.216412e+05 3.148162e+05 3.490270e+05 3.963666e+05    0
svmRadial 4.643964e+04 6.920321e+04 7.091314e+04 7.586467e+04 7.871333e+04 1.098335e+05    0

Rsquared 
               Min.   1st Qu.    Median      Mean   3rd Qu.      Max. NA's
glm       1.0000000 1.0000000 1.0000000 1.0000000 1.0000000 1.0000000    0
rf        0.9828249 0.9923234 0.9939790 0.9932573 0.9959403 0.9969025    0
rpart     0.7574158 0.7968386 0.8217899 0.8321542 0.8772453 0.9089569    0
svmRadial 0.9791394 0.9892371 0.9922436 0.9899597 0.9932365 0.9962272    0

null device 
          1 
```



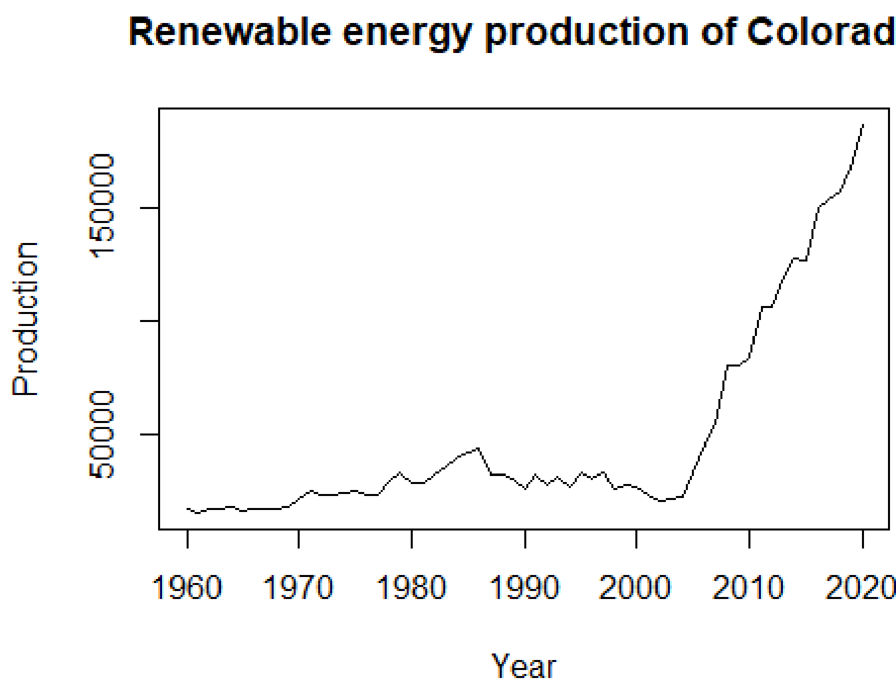
From the Mean Absolute Error(MAE) and root mean square error (RMSE) in the model evaluation above we can conclude that the best performing model is multiple linear regression, followed by random forest and support vector machine, and the decision tree algorithm does not perform as well as it should in this problem.

Renewable Energy Analysis

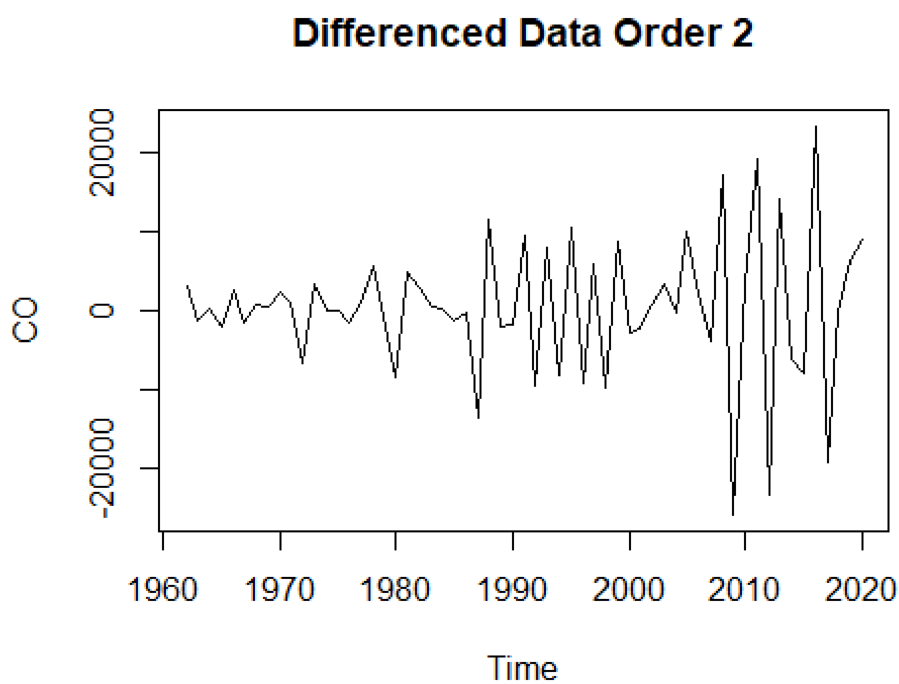
Next, we need to analyze the annual renewable energy generation in Colorado since 1960. By analyzing this data, we can understand the use and trends of renewable energy in Colorado.

U.S. Energy Information Administration collects yearly Total Renewable Energy Production data of Colorado, and units are billion Bu.

From the time series plot, there is an obvious upward trend generally, so it seems not to be stationary. In order to do the following analysis and forecast works. we need to do transformations.

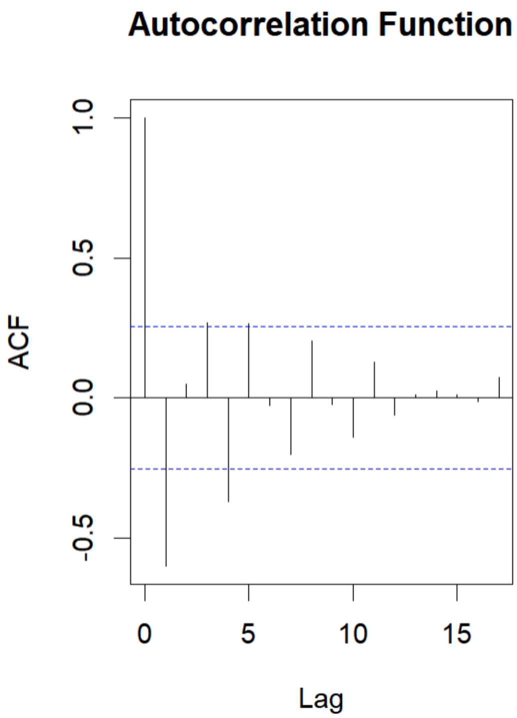


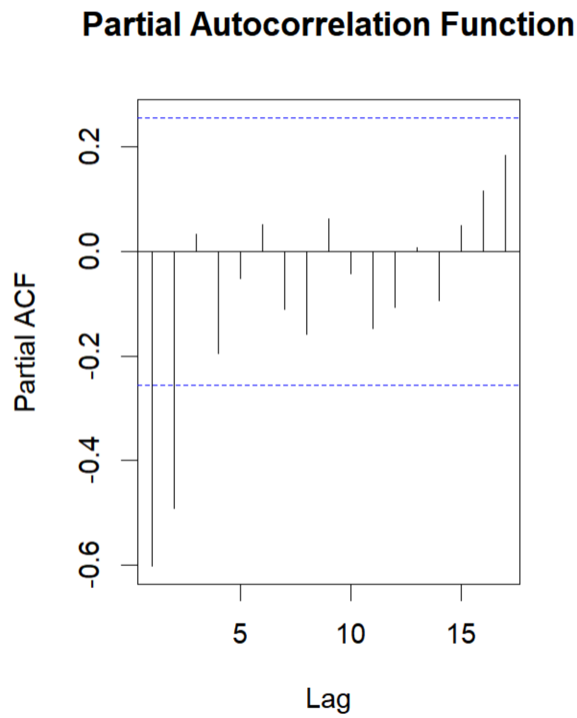
To eliminate the trend, we performed two differentials on the data. We take Dicky-Fuller (DF) Test, which is a type of unit root test that tests the null hypothesis that a time series has a unit root against the alternative hypothesis that it does not have a unit root. The results show that the p- value < 0.05, rejecting the original hypothesis and the data is stationary.




```
##
##  Augmented Dickey-Fuller Test
##
## data:  diff_ts
## Dickey-Fuller = -5.3715, Lag order = 3, p-value = 0.01
## alternative hypothesis: stationary
```

To forecast the future renewable energy production of Colorado. We apply the ARIMA model. From the ACF and PACF plots, we can see that PACF cuts off after lag 2 and then decayed to zero, while ACF exponential decaying to zero. So for the initial model, I choose ARIMA(2,2,0).





We apply auto.arima() function as a reference Since there are many fitting models, I chose the three most likely models based on the ACF and PACF plots. Through the results, we can see that the second model has the smallest AIC and BIC values, so I choose ARIMA(2,2,2) as my final model to do the following forecasts

```
Series: diff_ts
ARIMA(5,2,0)

Coefficients:
          ar1      ar2      ar3      ar4      ar5
      -2.0837  -2.5361  -2.1173  -1.3151  -0.4785
s.e.   0.1252   0.2693   0.3392   0.2685   0.1270

sigma^2 = 79389077:  log likelihood = -599.39
AIC=1210.78   AICc=1212.46   BIC=1223.03

Training set error measures:
                   ME     RMSE      MAE      MPE     MAPE      MASE       ACF1
Training set 379.3758 8364.809 5881.161 37.64382 385.8945 0.5256636 -0.1802985
```

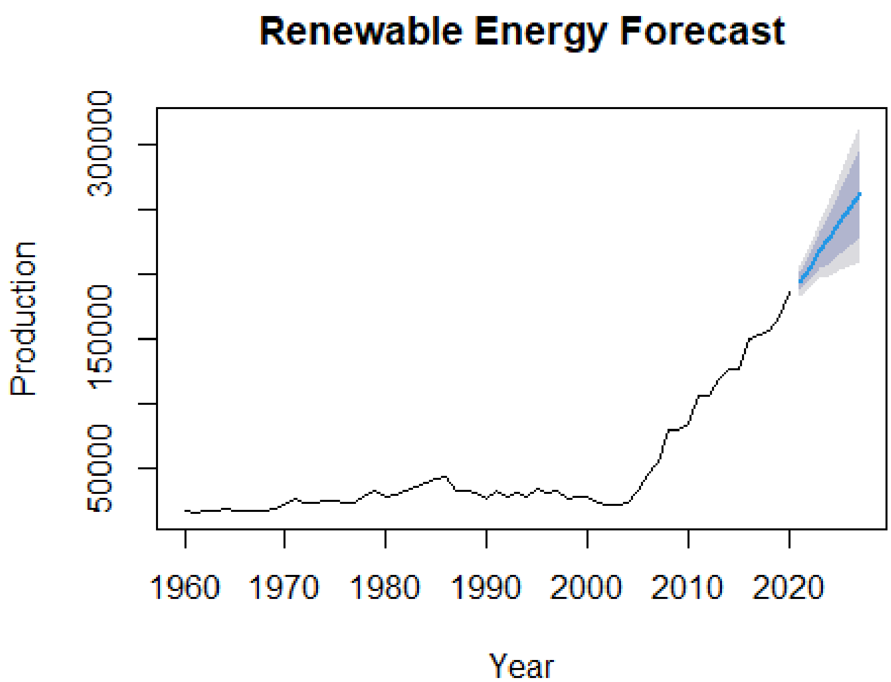
	Model <chr>	AIC <dbl>	BIC <dbl>
fit1	fit1	1210.776	1223.034
fit2	fit2	1181.732	1191.948
fit3	fit3	1237.102	1243.231

According to the result, Colorado's renewable energy generation has been increasing year by year since 1960. Especially after 2005, with the encouragement and support of the government, the development of renewable energy has shown a rapid growth trend.

ARIMA Forecast

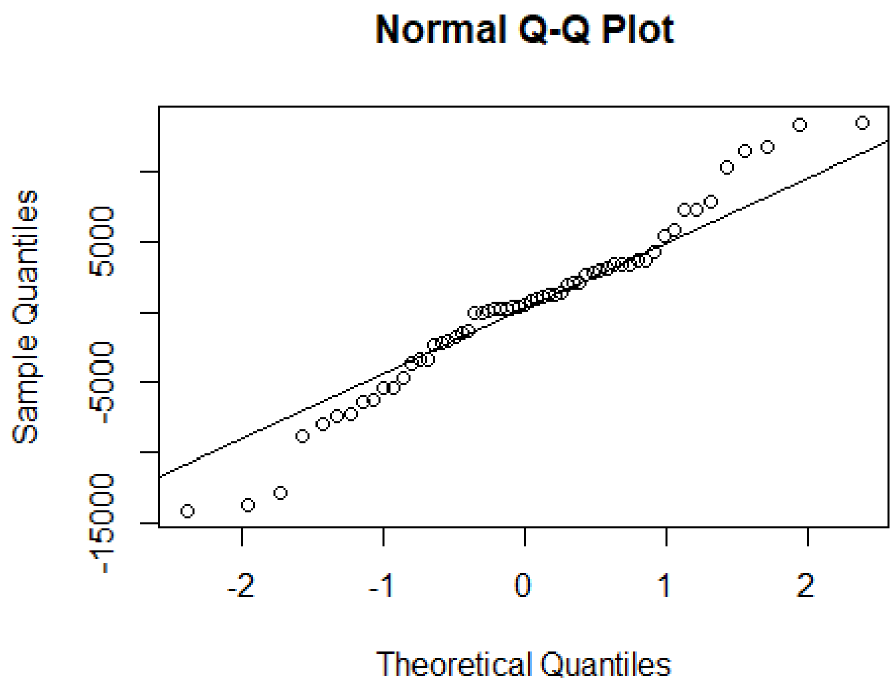
Through the picture we can see that the future energy production are in a clear upward trend in the following seven years.

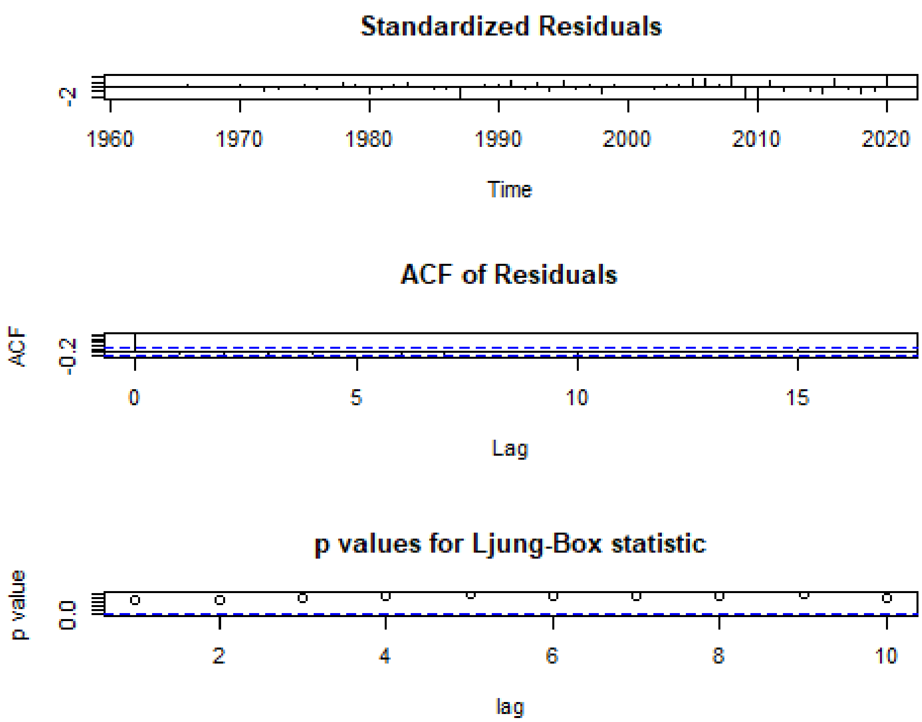
```
##      Point Forecast    Lo 80    Hi 80    Lo 95    Hi 95
## 2021       194376.4 186827.3 201925.4 182831.1 205921.7
## 2022       205369.0 194665.1 216072.9 188998.8 221739.3
## 2023       218500.6 204119.2 232881.9 196506.2 240494.9
## 2024       228119.3 208705.2 247533.4 198427.9 257810.6
## 2025       240417.6 216438.2 264397.0 203744.3 277090.9
## 2026       251545.6 222173.7 280917.6 206625.1 296466.1
## 2027       262652.8 227656.7 297649.0 209130.8 316174.8
```



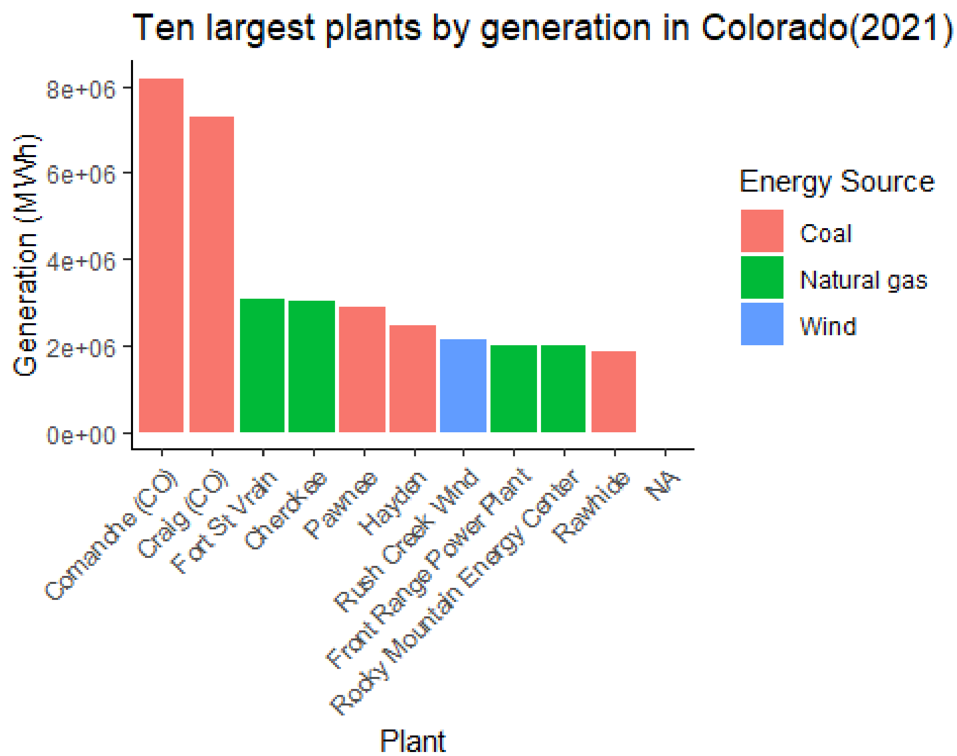
QQ-plot

The evaluation of the model is crucial in the forecasting process, thus I take normal QQ plot and diagnostic plots to do the work. The QQ residual plot shows that the residuals are almost normally distributed. And the Standardized Residuals plot shows there is no clear trend or seasonality, which indicates the model has captured these patterns in the data. In addition, the ACF plot is used to check the residuals for autocorrelation. While the ACF plot shows a significant spike at lag 1 and this may indicate that there is still some correlation between the residuals and the lagged values of the time series. But the p-values for the Ljung-Box statistic are all greater than the significance level(0.05), so the residuals are likely uncorrelated. In general, the ARIMA model is a good fit for the data





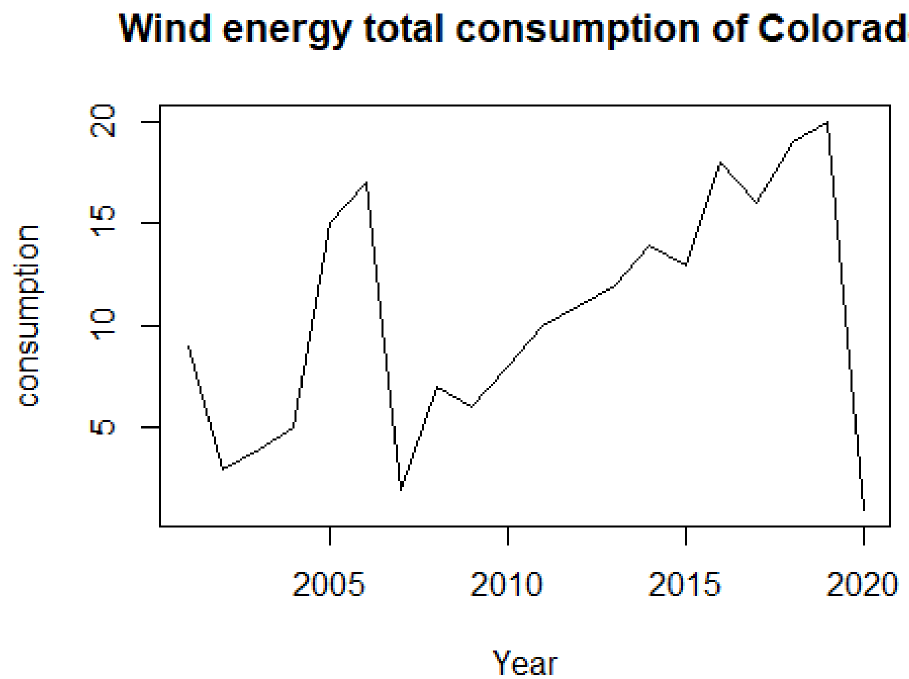
We found that one of the ten largest plants in Colorado by electricity generation is wind power and is ranked 7th, indicating that wind power is starting to become important in Colorado. Nevertheless, coal is still the main source of energy for power generation, so we focus our analysis below on wind and coal consumption.



Wind consumption

We also make forecasts of wind energy total consumption. The units are billion Bu.

This is the time series model of wind energy total consumption for the available

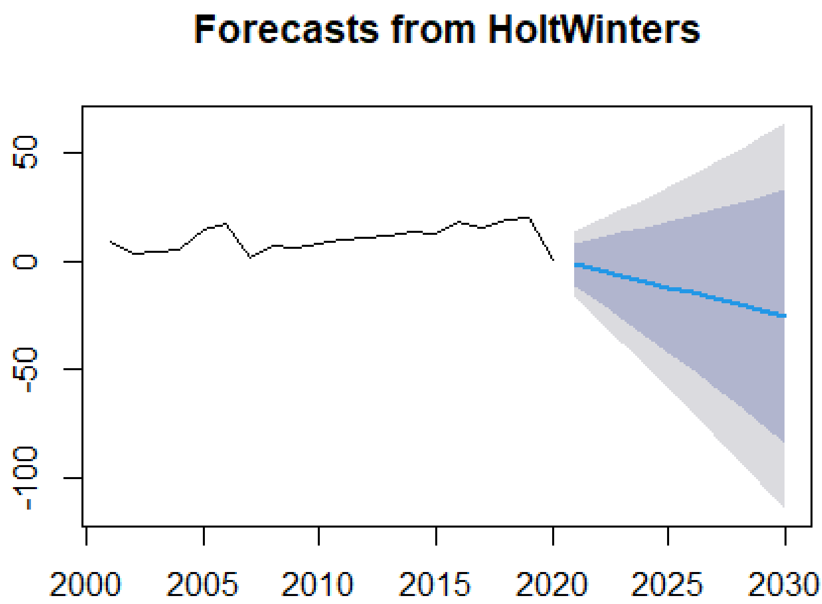


years.

Holt-Winters Forecast

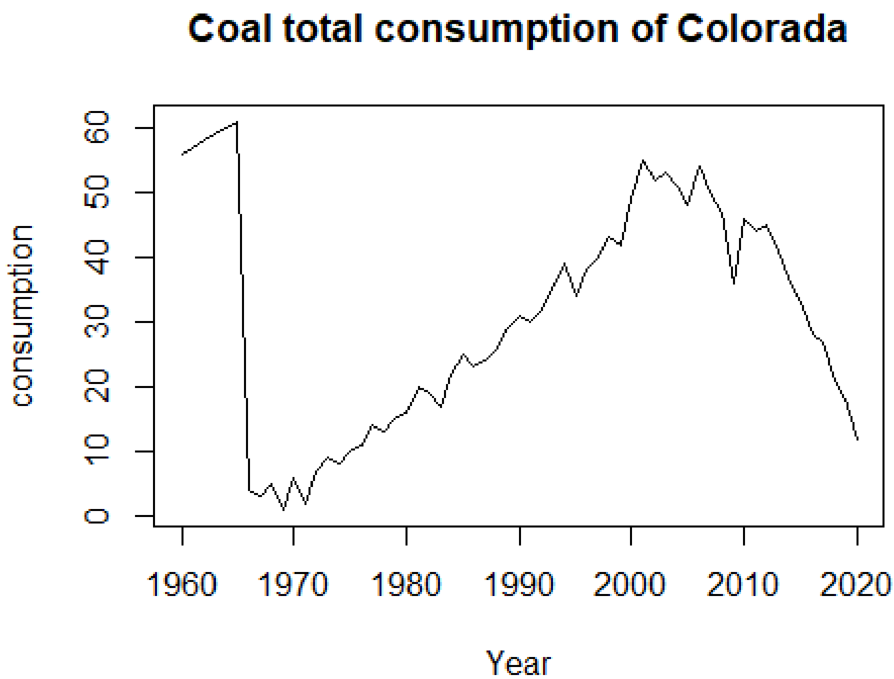
Another widely used forecast method is Holt-Winters Seasonal Forecast. It utilizes a form of double exponential smoother to compute the forecast. The table of predicted results is as follow. we can see that the future consumption of wind energy is almost flat with a slight downward trend.

```
##      Point Forecast     Lo 80     Hi 80      Lo 95    Hi 95
## 2021      -1.299012 -11.14852  8.550501  -16.36254 13.76452
## 2022      -3.986482 -19.10487 11.131904  -27.10806 19.13509
## 2023      -6.673951 -26.79747 13.449563  -37.45021 24.10231
## 2024      -9.361421 -34.51589 15.793045  -47.83186 29.10901
## 2025     -12.048891 -42.35571 18.257923  -58.39916 34.30138
## 2026     -14.736361 -50.35640 20.883673  -69.21250 39.73978
## 2027     -17.423831 -58.53541 23.687751  -80.29857 45.45091
## 2028     -20.111301 -66.90011 26.677506  -91.66861 51.44600
## 2029     -22.798771 -75.45275 29.855205 -103.32608 57.72854
## 2030     -25.486241 -84.19290 33.220419 -115.27033 64.29785
```



Coal(Non-renewable energy) consumption

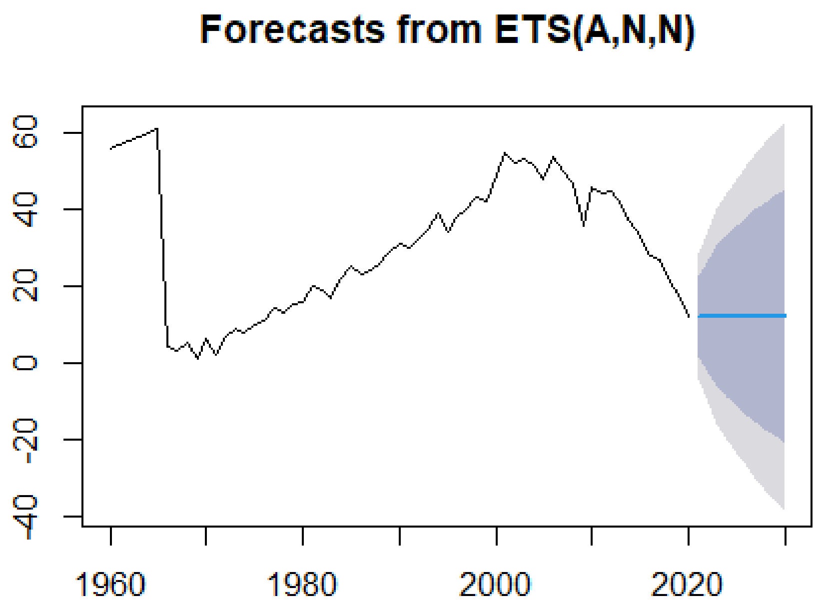
We have also forecast future non-renewable energy consumption for comparison, using coal as an example. After a significant increase in coal consumption from 1980 to 2020, coal consumption starts to decline from 2020.



Holt-Winters Forecast

Coal consumption is basically flat, with no upward trend in the future

```
##      Point Forecast      Lo 80    Hi 80       Lo 95    Hi 95
## 2021       12.11208   1.459409 22.76475   -4.179773 28.40393
## 2022       12.11208  -2.814353 27.03851  -10.715928 34.94008
## 2023       12.11208  -6.112008 30.33616  -15.759257 39.98341
## 2024       12.11208  -8.898307 33.12246  -20.020534 44.24469
## 2025       12.11208 -11.356098 35.58025  -23.779400 48.00356
## 2026       12.11208 -13.579832 37.80399  -27.180308 51.40446
## 2027       12.11208 -15.625861 39.85002  -30.309438 54.53359
## 2028       12.11208 -17.531002 41.75516  -33.223101 57.44726
## 2029       12.11208 -19.320885 43.54504  -35.960491 60.18465
## 2030       12.11208 -21.014198 45.23835  -38.550190 62.77435
```



Conclusion

By analyzing the data, we can find that Colorado's electricity market shows a trend of transitioning from traditional to renewable energy sources as the use of renewable energy increases. The development of renewable energy sources not only has a positive impact on the environment and climate change, but also brings good opportunities for economic and employment development in the state.

In summary, by analyzing the revenue and pricing of monthly electricity consumption in Colorado since 1990 versus annual renewable energy consumption since 1960, we can better understand the development and future trends of the electricity market and provide data to support better electricity planning and decision-making by government and energy companies.