

Trends and Predictions for Renewable & Non-Renewable Energy Production

Jamie Ho & Yi Zhou
Carnegie Mellon University

Abstract- During the COP26 conference in November 2021, the Glasgow Climate Pact was negotiated with representatives from 197 countries, which included a pact to phase out the use of coal for energy production and consumption [1]. The use of coal to generate energy produces carbon dioxide emissions as a byproduct, which is detrimental to the environment, thus making coal the biggest contributor to climate change [2]. Worse yet, coal is a nonrenewable resource. Therefore, in recent years, there has been a drive to use renewable resources to generate and consume clean energy through policy changes and promised targets. This project designs and deploys a Streamlit app that allows users to explore and compare existing trends in energy production and renewable energy targets for each US state between 1960 to 2019. In addition to exploration purposes, the Streamli app uses historical data to predict the changes in renewable versus non-renewable energy in the future. Ultimately, this project is intended to allow users to understand what the current progress is in the transition towards renewable energy and when nonrenewable energy might be fully phased out.

INTRODUCTION

During the COP26 conference in November 2021, the Glasgow Climate Pact was negotiated with representatives from 197 countries, which included a pact to phase out the use of coal for energy production and consumption [1]. The use of coal to generate energy produces carbon dioxide emissions as a byproduct, which is detrimental to the environment, thus making coal the biggest contributor to climate change [2].

In recent years, there has been an increase in the use of renewable energy through solar, wind, hydro,

biomass, and geothermal. In particular, California has been recognized as a leader among the US states for promoting renewable energy and policies that support renewable and carbon-free technologies as a solution for reducing greenhouse gas emissions. To balance sustainability and development, California has a diverse renewable energy portfolio. Additionally, the 100 Percent Clean Energy Act of 2018, also known as the Senate Bill 100, sets the goal that by 2045, all retail electricity sold in California and state agency electricity needs to be powered with renewable and zero-carbon resources. Thus, it would be interesting to investigate how the rest of the US compares to California in its initiatives to fully generate and consume clean energy.

In this project, we plan to focus on the different US states and understand the changes in renewable energy and nonrenewable energy over time relative to the state's promised renewable energy goals. Specifically, three questions will be investigated through data visualisations:

1. Which state in the US had the largest increase in renewable energy and the largest reduction in the nonrenewable energy over time?
2. How close is each state in reaching their renewable energy goals?
3. How long would it take for renewable energy to replace fossil fuels?

RELATED WORK

Energy production and consumption in the US has been recorded and visualised by the government and other organisations. The visualisations of the energy production and consumption breakdown come in different forms. The Lawrence Livermore National Laboratory developed static Sankey diagrams (Fig. 1) for the US and for each US state [3]. This type of diagram visualises the amount of energy produced from

each source and how that energy is being consumed as end uses. Aside from static visualisations, Our World in Data developed dynamic and interactive charts and maps of different energy data (Fig. 2) [4].

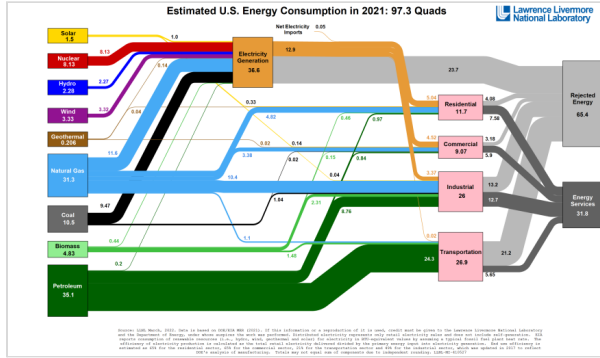


Fig. 1: A Sankey diagram developed by the Lawrence Livermore National Laboratory that visualises the amount of energy produced from each source and how the energy is being used [6].

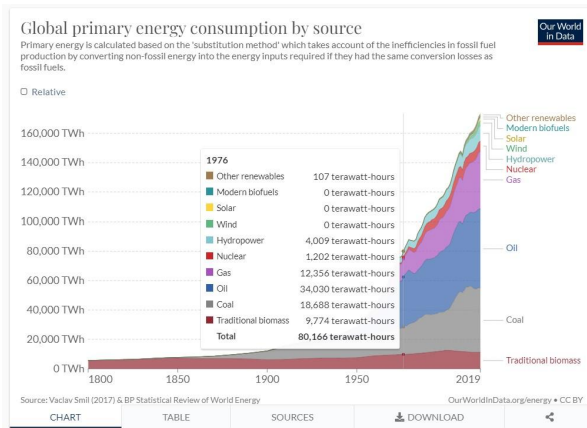


Fig. 2: An interactive chart developed by Our World in Data that shows the global primary energy consumption by source [7].

METHODS

I. Data Sources

In this project, we used data from two sources. Firstly, we downloaded the 1960-2019 state energy production estimate data published by the U.S. Energy Information Administration (EIA) [3]. This dataset has the breakdown of energy production for each state for every year in that timeframe. Secondly, we extracted the text data about each state's current Renewable Portfolio Standards (RPS) data from the National Conference of State Legislatures (NCSL) website [4]. This NCSL dataset includes the name of the renewable

energy standard, establishment year of the standard, the requirements, the applicable sectors, as well as the enabling statute/order/code for each state. The states are categorised into four groups based on their stance: no requirements, expired requirements, voluntary requirements, or with RPS requirements. This dataset will help us understand how policies are linked to actions, especially how effective policies are in driving renewable energy.

II. Data Cleaning & Preparation

For the EIA dataset, we had to clean the data by removing certain energy categories that were not the focus for this project. For each state, there are 29 rows of data for each type of energy production or consumption. Since we are only interested in the energy production sources, we have identified the following energy type codes to use in our project: Biomass inputs to the production of biodiesel (BDFDB), Biomass inputs to the production of biofuels (BFFDB), coal production (CLPRB), noncombustible renewable energy production (NCPRB), natural gas marketed production (NGMPB), crude oil production (PAPRB), and wood energy production (WDPRB). In particular, BDFDB, BFFDB, and WDPRB will be combined into one category of "biomass". After cleaning the rows, there will be 5 categories of energy production sources.

Furthermore, since the dataset from EIA had the years organised in columns, we had to transpose the data such that each row represented the different energy production data for each year. Once the data was transposed, we added additional columns of data that represented the energy production type as a percentage of the total energy produced. By doing so, we can use the percentages to make comparisons and understand how close each state is to reaching their RPS goals, which is typically stated as a percentage to be met by a certain year. This will help us see which states are more proactive to convert fully into clean energy and see whether the states are working towards their set goals at a reasonable rate.

For the NCSL dataset, the text data from the website was downloaded as an HTML, cleaned using Python, and exported as a CSV file to be used in the Streamlit app.

III. Data Visualisation

In this Streamlit app, we utilised three main types of interactive visualisations: a choropleth map, tables, and line charts. For the exploratory data analysis section, using a choropleth map that colour codes the states, we can differentiate each state's stance on the RPS targets (Fig. 3). The map has a tooltip feature that shows information about the name of the state, the type of target, and the name of the RPS target if applicable. A dropdown selection with the four different RPS stances filters the choropleth map and provides a table that includes further information about the state's requirements based on the NCSL data (Fig. 4).

Aside from the choropleth map and table, we also used line charts for further explorations of the historical data. In this section, there are two options: single state or multiple state explorations. For the single state exploration, users can select the state of interest from the dropdown menu and see the breakdown of the different energy productions (in %) in each state (Fig. 5). The lines are colour coded in shades of green or grey/black. Energy types that are renewable are green, whereas non-renewable energy types are grey or black. Furthermore, there is also the option to show the US baseline for each energy production type.

On the other hand, the multiple state exploration allows users to compare the renewable energy and non-renewable energy productions (in %) with a maximum of three states. If a user selects more than three states, the Streamlit app will present the user with a message that says: "ERROR. You have elected more than 3 states." The number of three states was chosen to maintain visual clarity while comparing between states. If a user selects less than two states, there will be a different message that tells the user to add another state to the selection or switch to the single state option. For both the single and multiple state options, the tooltip shows information about the state name, year, and percentages for each type of energy production.

For the prediction section, line charts are used to illustrate the historical and predicted future data. In particular, solid lines represent the historical data and dashed lines represent the future trends based on the results of a prediction model.

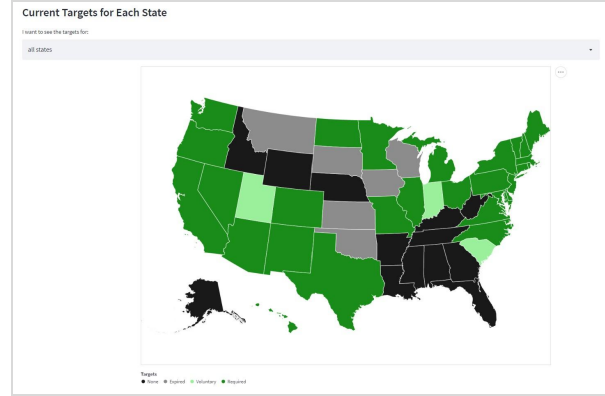


Fig. 3: Choropleth map colour-coded with each state's current state of RPS target (Dark green for with RPS requirements; light green for with voluntary requirements; grey for expired targets; and black for no targets).

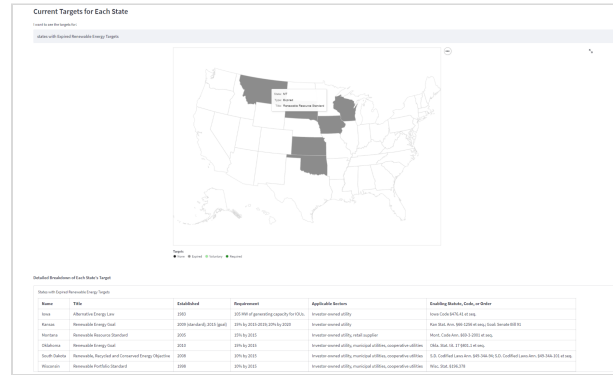


Fig. 4: Choropleth map with states that do not have any RPS targets highlighted and more information shown in a table format below.

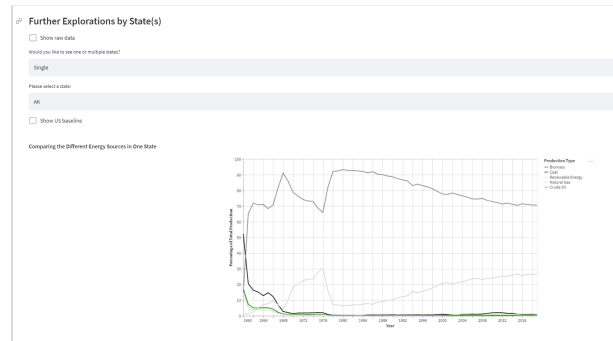


Fig. 5: Line chart with the breakdowns of the percentage of energy production in Alaska (single-state exploration).

IV. Prediction Model

For the prediction model, we choose multi-step time series forecasting model and refer to skforecast for information. We use existing data to extrapolate trends about the future to see whether the current rate is reasonable to reach the state-defined RPS goals.

For the data preparation part, we change the data format to timestamp by `df.to_datetime()` to enable the timeline chart to display. By splitting the data by 50 for training and 10 for testing, the machine learning model is created and trained from a `RandomForestRegressor` regressor with a time window of 15 lags which is intended to create an autoregressive model capable of predicting future annual energy consumption.

For the hyperparameter tuning, we modify the steps, lags, `max_depth` and `n_estimators`. It's interesting to see that the results seem to repeat a pattern when the step is set to 10.

However, the lines seem to be mismatched with the ground truth when the step is set to 5, and the lines seem to be a bit too smooth when the step is set to 20. Therefore for the final model, we set step at 10 for the prediction.

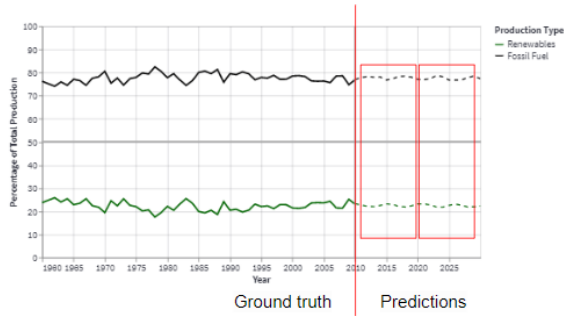


Fig. 6: Prediction result
(Step: 10 Lags: 20 Max_depth: 300 N_estimators: 400)

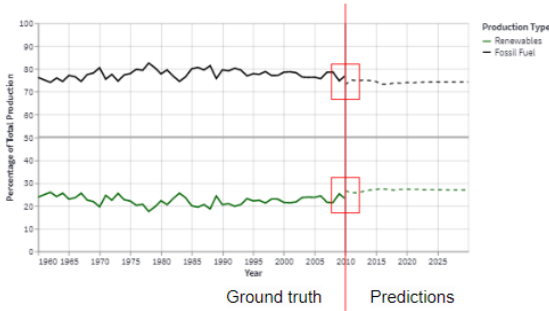


Fig. 7: Prediction result
(Step: 5 Lags: 20 Max_depth: 300 N_estimators: 400)

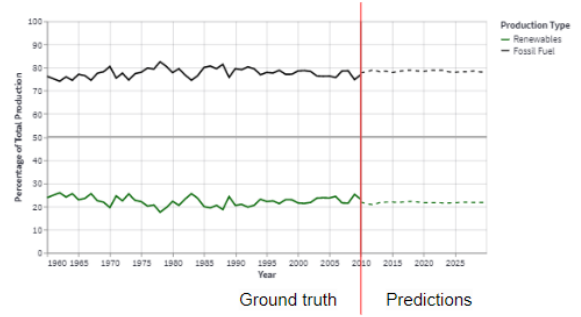


Fig. 8: Prediction result
(Step: 20 Lags: 20 Max_depth: 300 N_estimators: 400)

The test data is predicted 20 years into the future, the energy consumption is calculated as percentage; the vertical bar stands for the splitting time for ground truth existing data and prediction; the horizontal grey rule stands for the datum where renewables and fossil fuel productions are equal, and will possibly move towards the trend where renewables production exceeds fossil fuel production.

V. Uncertainty assessment and visualization

In order to assess the performance of the model, we backtest the data with prediction intervals using `backtesting_forecaster` and look back 20 years prior to 2010 which are separated for backtesting. Predictions for renewable energy and the fossil fuel are separately visualized in which the orange prediction lines with bands shows the amount of uncertainty in our predictions.

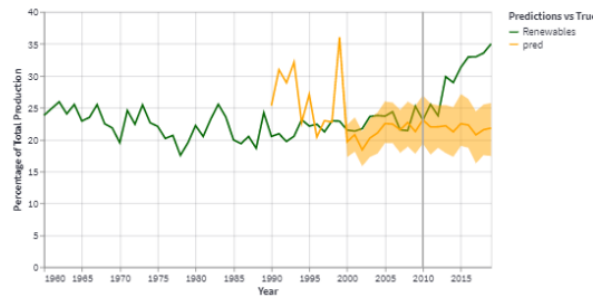


Fig. 9: Model performance visualization showing uncertainty

RESULTS

Looking at the graphs, prediction of future trends, and state targets, we can observe some really interesting trends. In this section, we present a summary of our findings for six different states with varying stances on renewable energy.

Arkansas (AR), No RPS:

The percentage of natural gas increased from the 1960s, with an especially large increase between 2007-2012. The curve for biomass and renewable energy had fluctuations, but showed a steady increasing trend until 2007. It is important to keep in mind that Arkansas does not have a target for RPS. This may explain the changes in the state's energy profile. However, it is noteworthy that the renewable energy production percentages are increasing again in recent years, whereas the non-renewable natural gas energy source is decreasing. Based on the future projections from the prediction model, the energy profile is trending in an encouraging manner, but it is not expected for the percentage of renewable energy to overtake that of non-renewable energy in Arkansas anytime soon.

Indiana (IN), Voluntary RPS:

Indiana's requirement of 10% by 2025 is voluntary for its applicable sectors. From the historical data graphed on a line chart, we can observe that there is a trend for reducing the use of nonrenewable energy, especially coal, and increasing the use of renewable energy. However, the result may be more effective if the requirement were to be required for all applicable sectors.

Kansas (KS), Expired RPS:

Kansas has expired RPS, of which the goal is to reach 20% renewable energy by 2020. Looking at the historical data, we can see that Kansas was able to meet the goal with biomass consisting of 12.32% and renewable energy consisting of 28.12% in 2019. We can also observe that the curve for natural gas and crude oil had significant changes over time. In particular, there were two periods of time between 1960-2019 that resulted in two pinch points in the graph. To further improve the energy profile and increase the use of renewable energy, Kansas should renew their RPS with new stringent targets.

Missouri (MO), Required RPS:

Unlike Arkansas, the percentage of renewable energy has overtaken that of non-renewable energy in Missouri. In particular, there was a significant decrease in the use of coal. Between 1992-1993, the use of coal for energy decreased by 42.07% and the use of renewable energy increased 36.17%. Looking at the RPS for Missouri, there is only one statute that has a requirement for renewable energy, which states a requirement of 15% renewable energy by 2021 for investor-owned utilities (IOUs). And based on our prediction model, it is estimated that renewable energy will maintain at 90% and non-renewable energy will maintain at 10% of the total energy production through 2030 and possibly beyond.

Nebraska (NE), No RPS:

It is interesting to note that despite having a significant increase in biomass percentage and significant decrease in crude oil percentage, Nebraska actually does not have any requirements. The use of crude oil was overtaken by biomass in 1995 and other forms of renewable energy in 2009. Our model predicts that this energy profile will remain consistent, with renewables comprising 90% of the total energy production. Thus, further investigation into the state of Nebraska will be necessary to get a better understanding of its stance on renewable energy. There may be additional policies or events that play into the current energy profile in Nebraska.

Ohio (OH), Required RPS:

Ohio made a significant effort in the reduction of coal to produce energy between 2012-2019. Specifically, coal comprised 72.04% of the total energy production in 2012 and decreased to 5.62% in 2019. The phasing out of coal is very critical to the reduction of carbon emissions, which is very beneficial to the environment. However, simultaneously, there was a significant increase in the use of natural gas, which shows that natural gas was the replacement for coal. Natural gas comprised 9.8% of the total energy in 2012, but increased to 84.84% in 2019. Although using natural gas to produce energy emits less carbon dioxide than coal [9], there are other environmental concerns with natural gas. Firstly, natural gas is a nonrenewable resource, which is not sustainable for the long-term [10]. Secondly, drilling for natural gas will destroy the environment and/or disturb the ecosystem (i.e., wildlife

habitats, people, water resources) [9]. And lastly, a leak in natural gas may risk damages to the environment, or worse, explosions. Thus, it is best to convert to using as much renewable energy as possible. The current renewable energy standard requires 8.5% by 2026. And our model predicts that the percentage of renewable versus nonrenewable energy will maintain at 7-8% for the former and 92-93% for the latter. It does not look like renewables will ever overtake non-renewables in the near future. Perhaps more stringent targets are needed to further reduce the use and impact of nonrenewable energy.

DISCUSSION

Since the target data is missing information for some states, and the data range only covers 1960 to 2019, our prediction is able to forecast till 2030. Overall we find there's a direct relationship between RPS goals and renewable energy production: more stringent and clear defined policy/target means higher possibility for the renewable energy trend to go upward and exceed the fossil fuel production. What we can observe that although trends vary a lot between states, fossil fuel companies seem politically powerful in the United States, their lobbying prowess is not the key reason that their fuels dominate the global energy system.

What we can observe that although trends vary a lot between states, renewable energy production grows fast in states with stringent renewable target. Therefore, the biggest challenges are political: there is plenty of blame to go around, from fossil fuel companies that for years denied the problem to policymakers reluctant to enact the policies needed to force real change. It has been easier for everyone to stick with the status quo.

Our own thoughts about how we need to deal with climate change have certainly evolved over time, as we understand the climate system better and as time passes with emissions still increasing. So it's better for the politicians to take appropriate actions and set realistic goals sooner rather than later.

FUTURE WORK

Though the timeforecasting model can look into the future till 2030, there's more we can do with limited data by varying the training data we use. For future

improvements, we can resample the data so the model is trained on slightly different sets each time (bootstrapping). We can use XGBoost for implementation of the stochastic gradient boosting algorithm that has become a benchmark in the field of machine learning. XGBoost library includes the XGBRegressor class which is compatible with skforecast where we can set the number of bootstrapping iterations(n_boot) to estimate prediction intervals and improve the forecaster with a better grid search algorithm.

In addition, by adding realistic Gaussian noise to the training features, we can represent error in the measurements without producing huge deviations for features you know are fairly accurate. Below is the results for Energy demand prediction with fluctuating data following the real life daily trend.

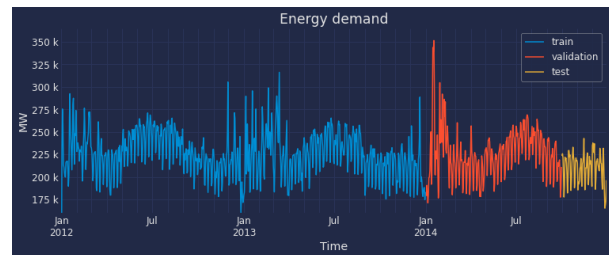


Fig. 10: Energy demand prediction with XGBoost (Joaquín Amat Rodrigo, 2022)

There is more we can do with the storytelling and to improve the logic of our project. Aside from the coding and prediction model aspects to improve upon, we can also consider expanding the dataset to look at the sources of consumed energy, similar to the Sankey diagram by LLNL. There may be some states that are not able to fully produce their own energy and need to buy energy from other states. It would be important to also consider how that energy is produced. Furthermore, it is crucial to understand that not all states can fully rely on their own produced renewable energy. For example, some states might not have consistent and adequate wind to generate energy through wind turbines. Geographical and climate limitations play a big role in each state's ability to fully convert to renewable energy. As such, it would be interesting to investigate how these limitations affect the state's timeline, policies, and collaboration with other states.

SOURCES

1. https://en.wikipedia.org/wiki/Glasgow_Climate_Pact
2. <https://endcoal.org/climate-change/#:~:text=Coal%20is%20the%20single%20biggest,emissions%20from%20the%20electricity%20sector>.
3. https://www.eia.gov/state/seds/sep_prod/xls/Prod_dataset.xlsx
4. <https://www.ncsl.org/research/energy/renewable-portfolio-standards.aspx>
5. Hannah Ritchie, Max Roser and Pablo Rosado (2020) - "Energy". Published online at OurWorldInData.org. Retrieved from: 'https://ourworldindata.org/energy' [Online Resource]
6. Flowcharts. (n.d.). Retrieved April 27, 2022, from <https://flowcharts.llnl.gov/>
7. <https://www.cienciadedatos.net/documentos/py27-time-series-forecasting-python-scikitlearn.html>
8. <https://www.cienciadedatos.net/documentos/py42-forecasting-prediction-intervals-machine-learning.html>
9. Natural gas and the environment—U.S. Energy Information Administration (EIA). (n.d.). Retrieved April 28, 2022, from <https://www.eia.gov/energyexplained/natural-gas/natural-gas-and-the-environment.php>
10. Society, N. G. (2019, October 24). Nonrenewable Resources. National Geographic Society. <http://www.nationalgeographic.org/encyclopedia/nonrenewable-resources/>