

Comprehensive forecasting of California's energy consumption: A multi-source and sectoral analysis using ARIMA and ARIMAX models

Zahra Moslemi ^{1,*}, Samantha Rehome ², Logan Clark ², Sarah Kernal ³, Scott Sprengel ², Ahoora Tamizifar ¹, Shawna Tuli ⁴, Vish Chokshi⁴, Mo Nomeli ⁴, Ella Liang ⁴, Moury Bidgoli ⁴, Jeff Lu ⁴, Manish Dasaur ⁴ and Marty Hodgett ⁴

¹ Departments of Statistics and Mathematics, University of California, Irvine, Irvine, CA, USA.

² Departments of Computer Science and Mathematics, California State University, Fullerton, Fullerton, CA, USA.

³ Department of Mechanical Engineering, Cypress College, Cypress, CA, USA.

⁴ Applied Intelligence, Accenture California, CA, USA

World Journal of Advanced Research and Reviews, 2024, 22(02), 484–497

Publication history: Received on 30 March 2024; revised on 05 May 2024; accepted on 07 May 2024

Article DOI: <https://doi.org/10.30574/wjarr.2024.22.2.1367>

Abstract

California's significant role as the second-largest consumer of energy in the United States underscores the importance of accurate energy consumption predictions. With a thriving industrial sector, a burgeoning population, and ambitious environmental goals, the state's energy landscape is dynamic and complex. This paper presents a comprehensive analysis of California's energy consumption trends and provides detailed forecasting models for different energy sources and sectors. The study leverages ARIMA and ARIMAX models, considering both historical consumption data and exogenous variables. We address the unique challenges posed by the COVID-19 pandemic and the limited data for 2022, highlighting the resilience of these models in the face of uncertainty. Our analysis reveals that while fossil fuels continue to dominate California's energy landscape, renewable energy sources, particularly solar and biomass, are experiencing substantial growth. Hydroelectric power, while sensitive to precipitation, remains a significant contributor to renewable energy consumption. Furthermore, we anticipate ongoing efforts to reduce fossil fuel consumption. The forecasts for energy consumption by sector suggest some decline in the commercial and residential sectors, reflecting California's recently declining population and the shift away from brick-and-mortar shops and offices to online websites and remote work. In contrast, the industrial and transportation sectors are expected to experience some growth until they return to more constant pre-COVID levels.

Keywords: Energy Forecasting; California Energy; ARIMA; Sectoral Analysis; Predictive Models

1. Introduction

California is the second largest consumer of energy in the United States [1]. This is due to rapid urbanization and a thriving industrial sector. California is also the most populous state in the nation and has the largest economy in the United States. Thus, accurate predictions of the state's energy demands are crucial for policymakers to make decisions to help shape a sustainable future [1].

California has long been a trailblazer in the shift toward cleaner and more sustainable energy sources [2]. The state ranks as the second-highest producer of electricity from renewable energy in the United States, underscoring its commitment to a greener future. Recent policy initiatives further emphasize this transition, promoting the use of renewable energy while imposing limitations on fossil fuel consumption [3] California's ambitious targets include

* Corresponding author: Zahra Moslemi

becoming 60% reliant on renewable energy by 2030 and striving for 100% clean energy by 2045, a testament to its dedication to environmental sustainability.

In spite of the growing popularity of renewable energy, fossil fuels continue to dominate California's energy landscape. The state stands as the second-largest consumer of natural gas and petroleum in the nation [4], contributing to approximately 8% of the total consumption of refined petroleum products in the United States [1]. This persistent reliance on fossil fuels can be attributed, in part, to California's extensive transportation infrastructure. With more registered motor vehicles and a higher volume of vehicle miles traveled by residents than any other state, the demand for traditional fuels remains substantial.

1.1. Dataset Description

Our data come from various public institutions, including the *United States Energy Information Administration (EIA)*, *State of California Department of Finance (DOF)*, and *National Centers for Environmental Information: National Oceanic and Atmospheric Administration (NOAA)*. This study is based on annual data, with values running from 1960 to 2021 for some variables and 1970 to 2021 for others.

Total yearly energy consumption for California is given in trillion British thermal units (Btu), where one Btu is "the quantity of heat required to raise the temperature of one pound of liquid water by 1° Fahrenheit (F) at the temperature that water has its greatest density (approximately 39° F)" [1]. In this dataset, yearly energy consumption is also broken down by energy sources (coal, natural gas, petroleum, biomass, geothermal, solar [5, 6], wind [7], and hydroelectric) and sectors (commercial, industrial, residential, and transportation).

Additional numerical variables include California's population; energy production for California, broken down by source (trillion Btu); average annual price of energy in California, broken down by source and sector (dollars per million Btu); annual precipitation in California (inches), and average maximum temperature in California (degrees Fahrenheit).

The one categorical variable utilized in this study is providing the Mnemonic Series Name (MSN) for a recorded observation of energy consumption. Each MSN is created in terms of several classifiers, including energy source, the sector or activity to which the corresponding observation of consumption is dedicated, and the units this value is to be expressed in. The MSN variable allows us to break down energy consumption in California by these classifiers.

1.2. Goals

- Our project seeks to predict energy consumption in California in the following way:
- Predict fossil fuels and renewable energy consumption for the next 10 years.
- Predict energy consumption by sector for the next 10 years.

2. Materials and Methods

2.1. Exploratory Data Analysis

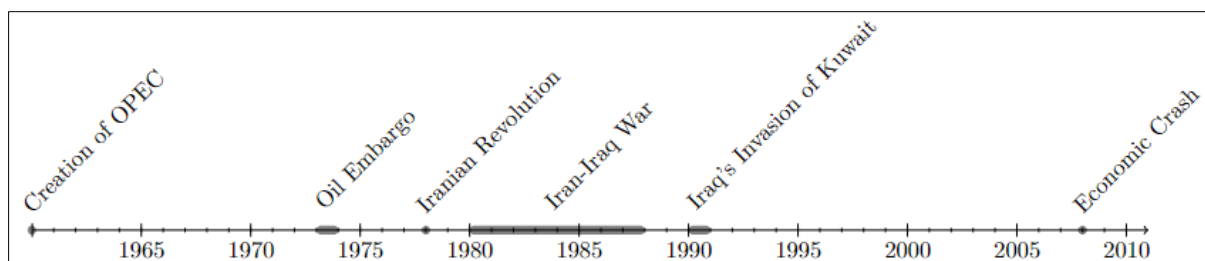


Figure 1 Timeline of Key Historical Events Impacting Fossil Fuel Consumption (1960-2010) in California and Beyond. Events include the creation of OPEC, the 1973 Oil Embargo, the Iranian Revolution in 1978, the Iran-Iraq War (1980-1988), Iraq's Invasion of Kuwait in 1990, and the Economic Crash of 2008. Gray lines represent significant events, while gray-shaded regions indicate the duration of certain impactful events.

We began our exploratory data analysis by looking into California's fossil fuel consumption over time, taking note of the largest sources of consumables as well as significant peaks in our exploratory plots. Upon first glance, it was clear natural gas and petroleum were California's primary sources of fossil fuel consumption. Before going further, we

decided to take a step back and look at historical events that have influenced oil in California that could explain the volatility of our plots [8].

Our exploration of California's historical oil landscape [9, 10] begins with the establishment of the Organization of the Petroleum Exporting Nations (OPEC) in 1960 [11]. OPEC's primary objective was to stabilize and defend global oil prices. Over a decade later, we have the oil embargo beginning in 1973, following President Nixon's request for Congress to make 2.2 billion dollars in emergency aid to Israel for the Yom Kippur War. Because of this, the Organization of Arab Petroleum Exporting Countries initiated an oil embargo on the United States, limiting our imports of oil products. This lasted around two years.

The Iranian Revolution of 1978 and the subsequent Iran-Iraq war in the 1980s caused substantial disruptions to global oil production [12, 13]. The Iranian Revolution of 1978 caused the oil output in Iran to drop over 5 million barrels a day, leading to a 5 percent global loss in production. Much like the Iranian Revolution, the Iran-Iraq war caused a sharp drop in oil production for most of the eighties. Iraq invaded Kuwait in 1990 causing the Bush administration to release 34 million barrels of oil from the U.S Strategic Petroleum Reserve anticipating an oil crisis. To conclude the timeline, oil prices steadily began to increase from 2006 to a record high in 2008, leading into the economic crash of 2008. The crash caused oil prices to plummet, resulting in a sharp decrease in profits for oil and gas companies [14].

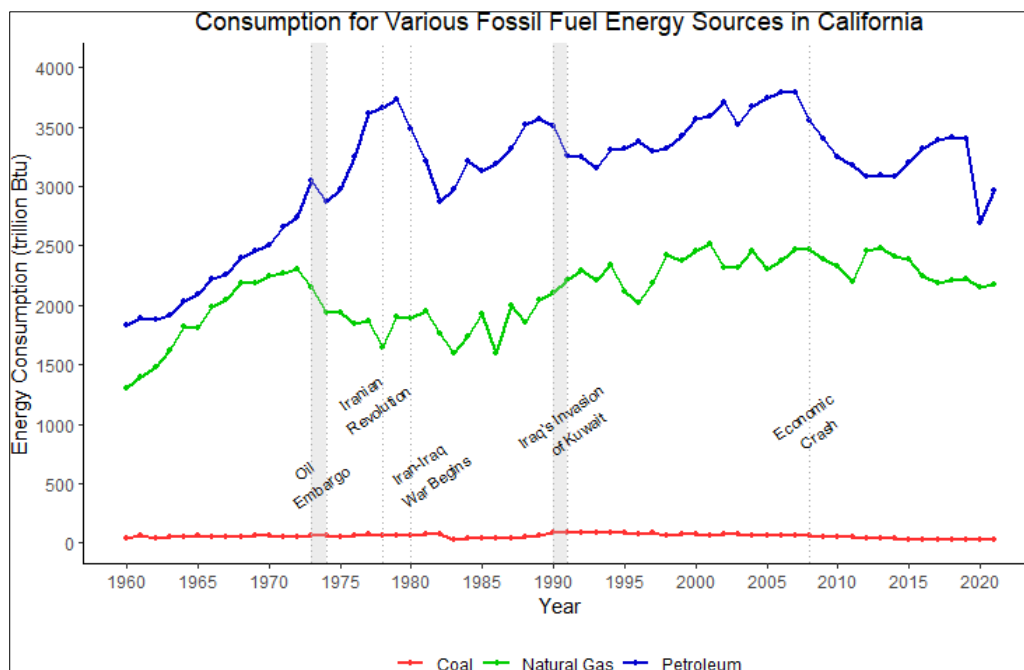


Figure 2 Fossil fuel energy consumption in trillion Btu from 1960 to 2021. Gray lines indicate events that had an impact on one or more fossil fuels in California or the rest of the US while gray-shaded regions represent the duration of certain events [15].

We can see from the plot above that coal is consumed on a considerably smaller scale in California relative to natural gas and petroleum. When looking at a plot of California's coal production [16], it was nearly nonexistent, which is a major reason why the state doesn't consume much of it. In addition to coal, looking at the plot's timeline, a direct effect can be seen on the petroleum line following the major historical events of the Middle East. It is clear those events caused a momentary decline in the consumption of petroleum during those years.

After reviewing the history of fossil fuels, we expanded our research to include the major policies that impacted the growth of renewable energy sources in California.

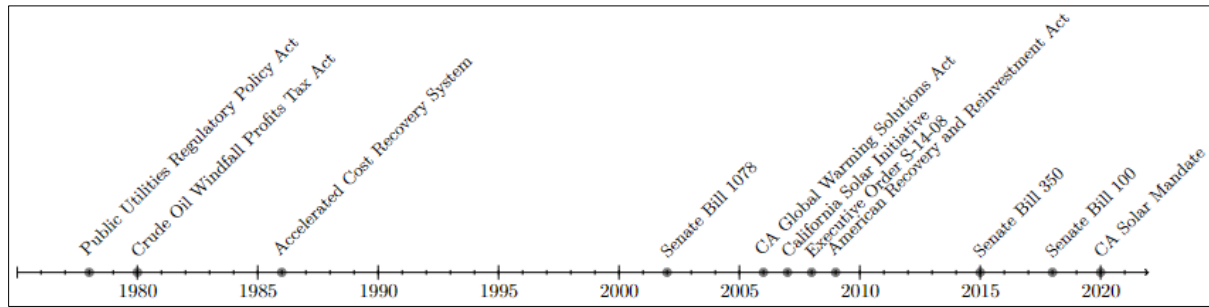


Figure 3 Chronology of Key Policies Shaping California's Transition to Renewable Energy (1975-2021)

The Public Utilities Regulatory Policy Act (PURPA) of 1978 played a significant role in opening doors for non-utility electricity generators, particularly renewable energy producers [17]. It allowed them to sell electricity into a utility's power grid, breaking the monopoly of power companies on electricity generation.

Following PURPA, the Crude Oil Windfall Profits Tax Act of 1980 was instrumental in promoting renewable energy. It authorized an increase in the nonresidential energy tax credit for solar and wind systems from 10% to 15% and raised the residential credit from around 22% to 40%. The Economic Recovery Tax Act of 1981 [18] established an Accelerated Cost Recovery System (ACRS) [19], enabling wind, solar, and geothermal property placed in service after 1986 to be depreciated for tax purposes over a five-year period. These early acts laid the foundation for solar and wind energy to begin contributing to the power grid in the late 1980s.

In 2002, Senate Bill (SB) 1078 established California's renewable portfolio standard (RPS) program, aiming to increase the share of electric power sales from renewable energy sources. The program initially required 20% of electricity retail sales to be served by renewables by 2017. California's commitment to environmental goals continued with the implementation of the Global Warming Solutions Act (GWSA) in 2006 [20]. This act mandated a reduction in greenhouse gas emissions to 1990 levels by 2020 and the adoption of regulations to achieve that goal [21].

Further progress came with the California Solar Initiative (CSI) in 2007, which provided incentives for low-income customers to install solar photovoltaic systems [22]. Executive Order S-14-08 increased RPS requirements, aiming for 33% renewable energy by 2020.

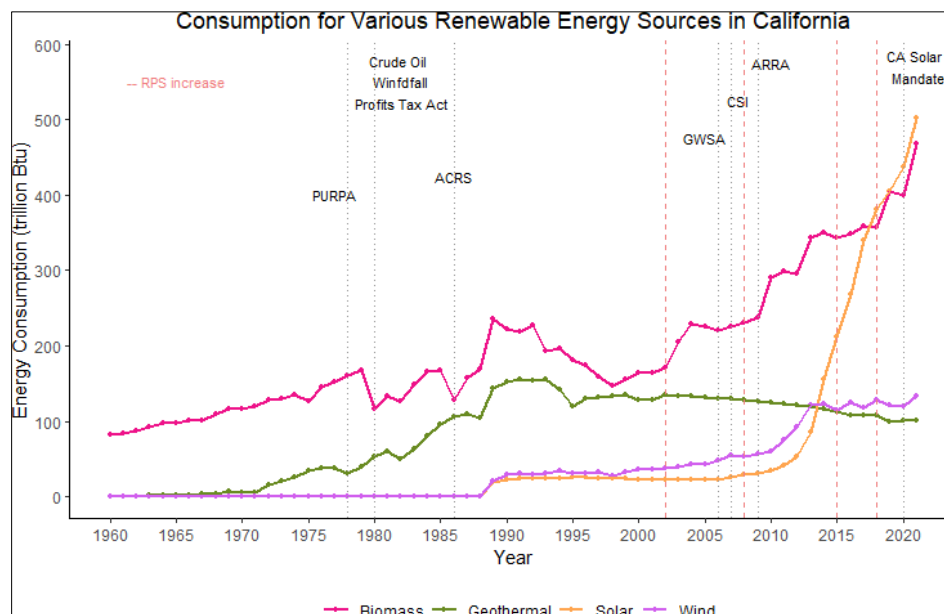


Figure 4 Renewable energy sources consumption in trillion Btu from 1960 to 2021. Gray lines indicate policies that had an impact on one or more renewable energy power sources in California or the rest of the US while light red dashed lines represent an increase in the RPS.

These early 2000s policies fostered investment in clean energy technology, making renewable energy generation more efficient and widespread. This laid the groundwork for a significant increase in renewable energy consumption, particularly evident in 2010. In 2015, Senate Bill 350 mandated a 50% RPS by 2030, and in 2018, Senate Bill 100 increased the RPS to 60% by 2030 and 100% by 2045.

Effective from January 1, 2020, the California Solar Mandate is a building code requiring all new residential construction projects, including single-family homes, condominiums, and three-story apartment buildings, to have solar photovoltaic (PV) systems [23]. While the impact of the Solar Mandate may not yet be apparent in our data, it is likely to be a significant factor in the future growth of solar energy consumption [24].

Additionally, although not represented in the timeline, research by GTM (Greentech Media) [25] indicated a remarkable reduction of 97.2% in the average price of photovoltaic systems from 1975 to 2012, and the National Renewable Energy Laboratory's 2020 report [26] revealed a drop of 64-82% in solar system prices from 2010 to 2020. These price reductions, coupled with the array of clean energy policies, likely played a pivotal role in the exponential growth of solar energy consumption in California.

We can see from the plot above that until 2018 when solar energy surpassed it, biomass was the renewable energy source that California consumed the most. It is also apparent that until 1989 the amount of solar and wind energy consumed was negligible enough to be recorded as zero. Solar and wind energy both had a sudden increase around 2010, but while solar consumption continued to increase exponentially, wind power consumption leveled off after 2013. The plot also reveals that geothermal energy consumption has been slightly declining since about 1996.

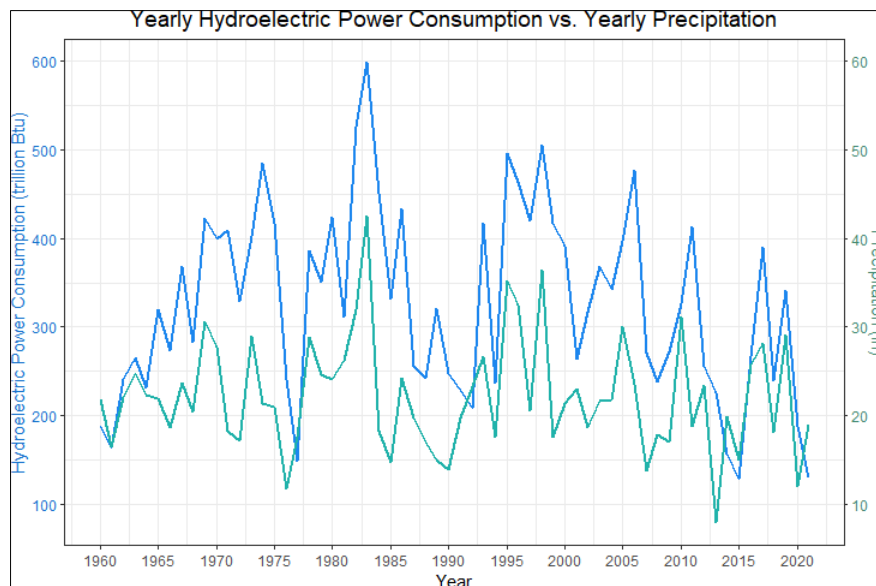


Figure 5 Hydroelectric power consumption in trillion Btu and average yearly precipitation in inches from 1960 to 2021. The blue line and left axis represent hydroelectric power consumption while the teal line and right axis represent average yearly precipitation.

Not included in the various renewable energy source consumption plots above was the line displaying hydroelectric power consumption. We did not include hydroelectric power because the volatility in hydroelectric power consumption from one year to the next created a very jagged line that distracted from the other renewable sources. The volatility in the consumption values did intrigue us though, and in our search for understanding, we discovered that the pattern of peaks and lows in hydroelectric power consumption seemed to correspond to the peaks and lows of the average yearly precipitation in California.

After researching the history of energy sources in California and plotting them to see their change over the years, we created a density plot with all the energy sources to see how much of California's total energy consumption comes from each source.

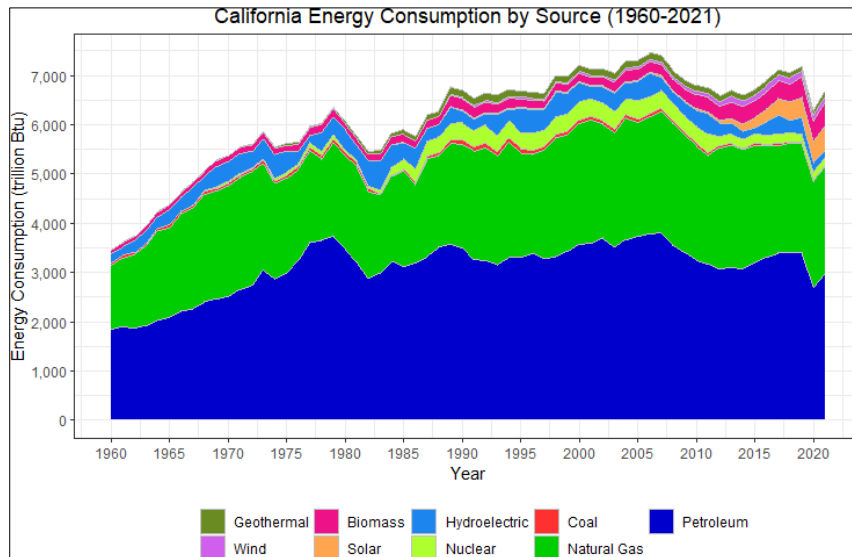


Figure 6 Portion of energy consumption in California in trillion Btu by source (geothermal, wind, biomass, solar, hydroelectric, nuclear, coal, natural gas, and petroleum) from 1960 to 2021.

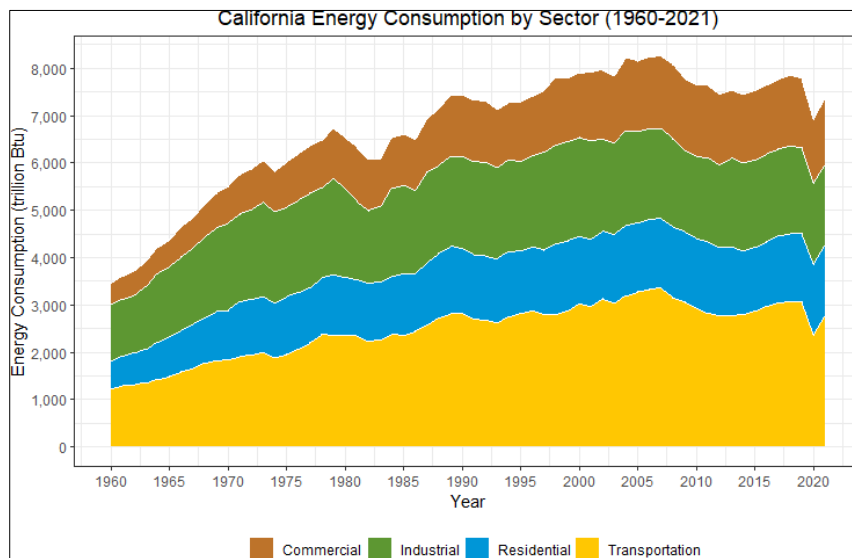


Figure 7 Portion of energy consumption in California in trillion Btu by sector (commercial, industrial, residential, and transportation) from 1960 to 2021.

The density plot clearly shows that energy from petroleum and natural gas accounts for about two-thirds of energy consumption in California while coal is barely a sliver of our consumption [27]. Nuclear energy accounted for a decent portion of California's consumption for a little over three decades, but it currently accounts for a very small portion. Renewable energy sources appear to only account for approximately 30% of California's energy consumption. Historically, hydroelectric accounted for most of California's renewable energy consumption, but now it is only a very small portion. Solar is currently California's largest source of renewable energy consumption with biomass not far behind, but even combined, wind and geothermal energy account for very little of the total energy consumption.

We then created a density plot with all the energy sectors to see how much of California's total energy consumption comes from each sector.

The density plot clearly shows that California's transportation sector has historically been the largest energy consumer and currently accounts for a little over a third of California's consumption. The industrial sector is and has always been the next largest consumer and currently accounts for nearly a quarter of California's consumption. The commercial and

residential sectors are the smallest energy consumers, each currently accounting for about a fifth of California's total energy consumption.

In this research, an R Markdown document authored by Logan Clark was employed to conduct a comprehensive analysis of energy consumption trends in California's industrial and transportation sectors. The document utilized various R packages, including tidyverse, janitor, dplyr, readr, tseries, and forecast, for data manipulation, analysis, and time series forecasting.

2.2. Modeling & Implementation

To properly model our data, we decided to use an ARIMA (Auto-regressive Integrated Moving Average) time series model. The ARIMA model takes in three separate parameters (p, d, q) [28]:

p: The order of the auto-regressive (AR) component, which represents the number of lag observations included in the model. The AR component is based on the idea that the past values in the time series can be used to predict future values.

d: The order of differencing, which represents the number of times the raw observations are differenced to make the time series stationary. Stationarity is often required for time series analysis, and differencing is a common technique to achieve this.

q: The order of the Moving Average (MA) component, which represents the number of lagged forecast errors in the prediction equation [29]. The MA component is based on the idea that past forecast errors can help predict future values.

The equation given below is a general equation for all ARIMA time series models:

$$Y_t = \alpha + \beta_1 Y_{t-1} + \beta_2 Y_{t-2} + \dots + \beta_p Y_{t-p} + \epsilon_t + \phi_1 \epsilon_{t-1} + \phi_2 \epsilon_{t-2} + \dots + \phi_q \epsilon_{t-q} \quad (1)$$

While this equation may seem complex and confusing, below is a quick breakdown of each of the variables and what they mean in the context of the equation

Y_t is the response variable at the current time period

α is the constant or intercept term of the equation

β_p is the correlation of the value Y_{t-p} has to the response variable. The values of the coefficient are between -1 and 1

ϕ_p is the correlation of the value ϵ_{t-p} has to the response variable. The values of the coefficient are between -1 and 1

Y_{t-p} is the response variable p periods ago

ϵ_{t-q} is the error q periods ago

Using the three parameters of the ARIMA model, we can make a forecasting model to predict future values of energy consumption based on energy consumption values from previous years. We can also include other time series of the same time frame and frequency to account for other variables, such as California's population, energy production, energy prices, etc. Using the `auto.arima()` function in R, we're able to find the best parameters to fit our model by identifying the parameters with the lowest AICc values. We then test those models by excluding the last 5-10 years of consumption values to use as a testing set and training our model on the remaining values.

We tested the top few (p, d, q) parameters and included various exogenous predictors to see which would obtain the lowest mean-square error (MSE) [30] compared to the true testing data. We then chose the model parameters and variables with the lowest MSE to build an ARIMA model using the full dataset, which we then used to predict the next 10 years of values.

On the other hand, ARIMAX, or Auto-Regressive Integrated Moving Average with exogenous variables, is an extension of the ARIMA model that includes additional explanatory variables, also known as exogenous variables or covariates. The inclusion of these exogenous variables allows ARIMAX to capture the influence of external factors on the time series being analyzed.

The ARIMAX model is denoted as ARIMAX(p, d, q) with exogenous variables, where:

p: The order of the Auto-Regressive (AR) component, representing the number of lag observations included in the model. AR relies on past values in the time series to predict future values.

d: The order of differencing, indicating the number of times raw observations are differenced to achieve stationarity. Stationarity is crucial for time series analysis, and differencing is a common technique.

q: The order of the Moving Average (MA) component, indicating the number of lagged forecast errors. MA incorporates past forecast errors to predict future values.

X: Exogenous variables that are external to the time series but can influence it. These variables are not predicted by the model but are included to improve prediction accuracy.

The general equation for ARIMAX time series models is as follows:

$$Y_t = \alpha + \beta_1 Y_{t-1} + \beta_2 Y_{t-2} + \dots + \beta_p Y_{t-p} + \epsilon_t + \phi_1 \epsilon_{t-1} + \phi_2 \epsilon_{t-2} + \dots + \phi_q \epsilon_{t-q} + \gamma_1 X_{t-1} + \gamma_2 X_{t-2} + \dots + \gamma_r X_{t-r} \quad (2)$$

Here's a breakdown of the variables in the ARIMAX equation:

Y_t : The response variable at the current time period.

α : The constant or intercept term of the equation.

β_p : The correlation of the value Y_{t-p} with the response variable, ranging from -1 to 1.

ϕ_p : The correlation of the value ϵ_{t-p} with the response variable, ranging from -1 to 1 [31].

Y_{t-p} : The response variable p periods ago.

ϵ_{t-q} : The error q periods ago.

X_{t-r} : Exogenous variable r periods ago.

γ_r : The coefficient representing the impact of the exogenous variable X_{t-r} on the response variable.

The exogenous variables can represent external factors that are believed to have an impact on the time series. For example, in the context of forecasting sales, the ARIMAX model might include economic indicators, marketing expenses, or other relevant variables as exogenous inputs.

The parameters of the ARIMAX model, including the autoregressive parameters (ϕ), moving average parameters (θ), and coefficients associated with exogenous variables (β), are estimated using statistical methods, and the model performance is evaluated based on its ability to accurately predict future values of the time series.

We also used Holt's damped linear trend method, a form of exponential smoothing, to forecast each renewable energy source separately. Exponential smoothing methods use weighted observations of past values to make forecasts. Holt's linear trend method is comprised of two smoothing equations, one for the level, and one for the trend. The dampening parameter is added to this method to help prevent over-casting future values by flattening out the curve as time goes on. This method was chosen for its ease of use and ability to handle nonlinear and non-stationary data.

Holt's Damped Trend Equation:

$$\hat{y}_{t+h|t} = l_t + (\phi + \phi^2 + \dots + \phi^h) b_t \quad (3)$$

$$l_t = \alpha y_t + (1 - \alpha)(l_{t-1} + \phi b_{t-1}) \quad (4)$$

$$b_t = \beta^\oplus (l_t - l_{t-1}) + (1 - \beta^\oplus) \phi b_{t-1} \quad (5)$$

Equation (3) is the entire forecast equation of Holt's damped trend.

\hat{y} is the forecast value

l_t is the level estimate at time t

ϕ is the dampening term

b_t is the trend estimate at time t

Equation (4) is the level equation.

α is the level smoothing parameter

Equation (5) is the trend equation.

β^{\oplus} is the trend smoothing parameter

3. Results

In the pursuit of a thorough examination of energy consumption in California, a comprehensive strategy was employed involving a total of nine ARIMAX models. The predictive models comprised a (2,0,0) ARIMAX model specifically designed for total energy consumption forecasting. Additionally, two (2,1,0) ARIMAX models were crafted to analyze petroleum and natural gas consumption, accounting for scenarios with and without the impact of COVID-19 ([32]; [33]). A (1,1,1) ARIMAX model was deployed to predict hydroelectric power consumption, while a (2,1,2) ARIMA model was applied to forecast consumption trends in Solar, Wind, Biomass, and Geothermal sources.

Furthermore, a sector-specific focus was incorporated into the analysis, with four distinct ARIMAX models tailored to different segments of the energy landscape. These included (4,1,0) ARIMAX models for both the commercial and residential sectors, as well as (2,1,2) ARIMAX models for the industrial and transportation sectors. This meticulous multi-model framework, encompassing nine ARIMAX models in total, was designed to provide a comprehensive and nuanced understanding of energy consumption dynamics across diverse sectors in California.

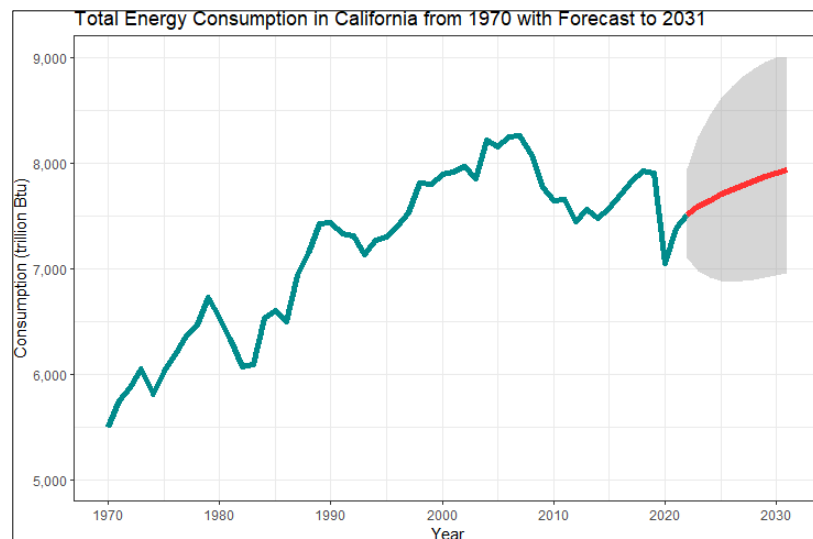


Figure 8 Total energy consumption in California in trillion Btu from 1970 to 2021. Dark cyan line represents the observed values, the red line represents the forecast values, gray shaded area represents a 95% confidence interval for the forecast values.

To begin our various resulting predictive models, we have included a model of total energy consumption in California with a 10-year consumption forecast based on a (2,0,0) ARIMAX model which used California's population, total primary energy production, gas prices (adjusted for inflation), and the average price of energy (adjusted for inflation) as exogenous variables.

Based on the predictions of this model, California is expected to reach a consumption level of 7936.225 trillion Btu by 2031 with a 95% confidence interval that spans from about 6952 to 9060 trillion Btu.

Next, we look closer at each source, beginning with fossil fuels. Forecasting consumption of petroleum and natural gas utilized a (2,1,0) ARIMAX model that used population and production as exogenous predictors. Taking into consideration COVID's anomalous effect on energy consumption habits, we built another model on pre-COVID data and predicted the next 12 years.

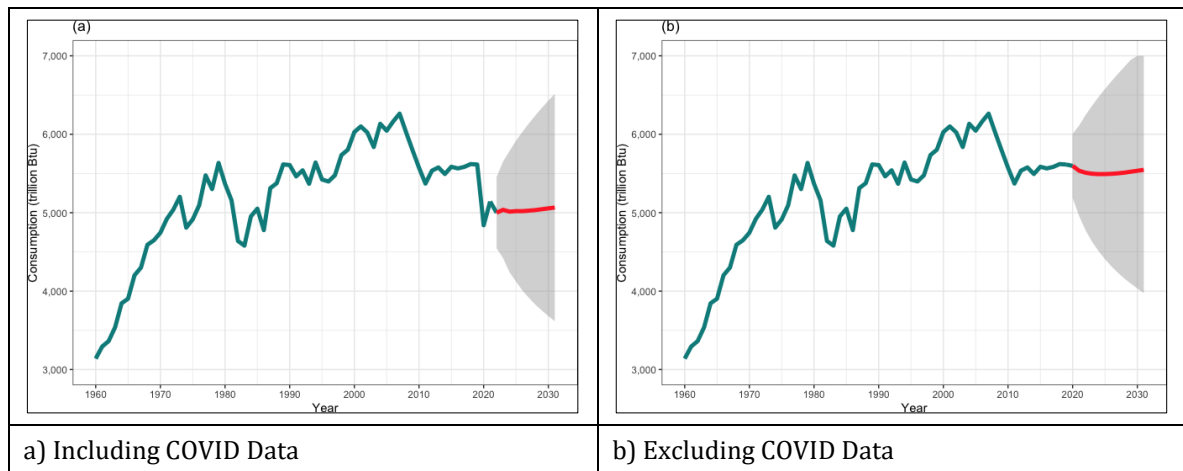


Figure 9 10 Year Fossil Fuel Consumption Forecasts. The blue line represents observed values, the red line represents forecast values, gray shaded area represents a 95% confidence interval for forecast values.

We predict fossil fuel usage will remain constant but may decrease under new state policies. California's fossil fuel reduction efforts face challenges due to extensive transportation infrastructure and the ongoing demand for energy from various industries. Between the two models, we would expect consumption to fall closer in line with the one that excludes COVID data, as the extreme drop from COVID would skew predictions for future years, especially when using auto-regressive terms.

Now we move to address renewable energy sources. Below are the resulting 10-year consumption forecasts of the (1,1,1) ARIMAX model that used precipitation as an exogenous predictor for hydroelectric power, and the (2,1,2) ARIMA model that was used to model the rest of the renewable energy sources.

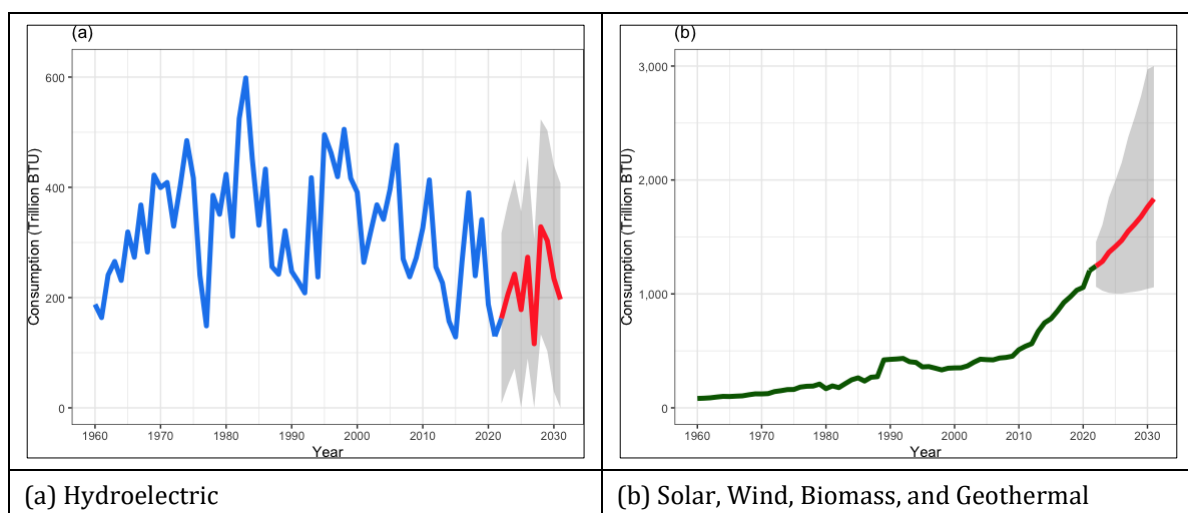


Figure 10 10 Year Renewable Energy Consumption Forecasts. The blue/Green line represents observed values, the red line represents forecast values, gray-shaded area represents a 95% confidence interval for forecast values.

We anticipate that hydroelectric power will continue to be a major contributor to renewable energy consumption. However, it remains particularly sensitive to precipitation levels, making long-term predictions challenging. For hydroelectric power, more accurate forecasts are possible on a shorter timescale, ideally looking ahead just one year. By using short-term precipitation forecasts, we can provide a closer estimate for the upcoming year. Hydroelectric power has the potential to rise well above 350 trillion Btu during rainy seasons, but it could also drop below 200 trillion Btu during periods of drought.

Conversely, for other forms of renewable energy, we project a consistent increase in consumption as California strives to achieve its ambitious goal of 100% renewable consumption by 2045. Among these renewable sources, we expect solar and biomass to be the primary contributors to this ongoing growth. This trend can be visualized in the subsequent

plot, which employs Holt's dampened linear trend to offer a general forecast for the consumption of each distinct source in the near future. These forecasts were generated using a damped trend parameter of 0.95, and the entire process was automated using the forecast package in R.

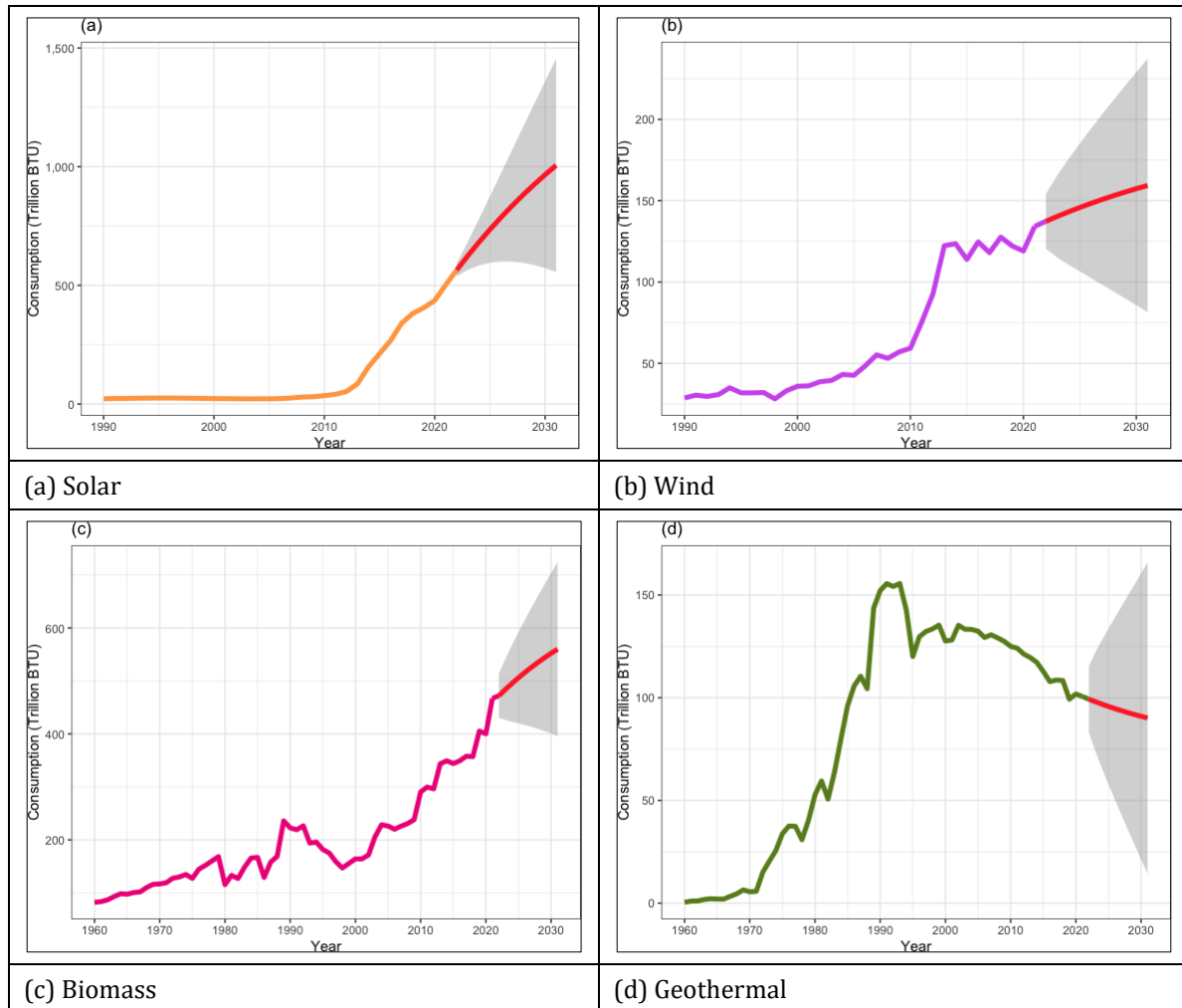


Figure 11 Red lines represent predicted values, colored lines represent actual values, and grey shaded areas represent 95% confidence intervals for the forecasts.

From the plot, we can see wind energy slowly increasing, which is to be expected as California continues to build more wind farms. The only decreasing forecast is geothermal energy, which has been steadily declining since 2000. Looking at the axes, we can again reemphasize the importance of solar and biomass in terms of how much we consume compared to the other renewable sources. Biomass and solar are almost quadruple the consumption of wind and geothermal energy.

The next set of plots display the energy consumption in California by sector, each with their 10-year forecast. Forecasting consumption in the commercial sector utilized a (4,1,0) ARIMAX model that used California's population and the average price of energy in the commercial sector as exogenous predictors. Forecasting consumption in the residential sector utilized a (4,1,0) ARIMAX model that used California's population and the average price of energy in the residential sector as exogenous predictors. Forecasting consumption in the industrial sector utilized a (2,1,2) ARIMAX model that used natural gas and petroleum prices as exogenous predictors. Forecasting consumption in the transportation sector utilized a (2,1,2) ARIMAX model that used population and average energy price in that sector as exogenous predictors.

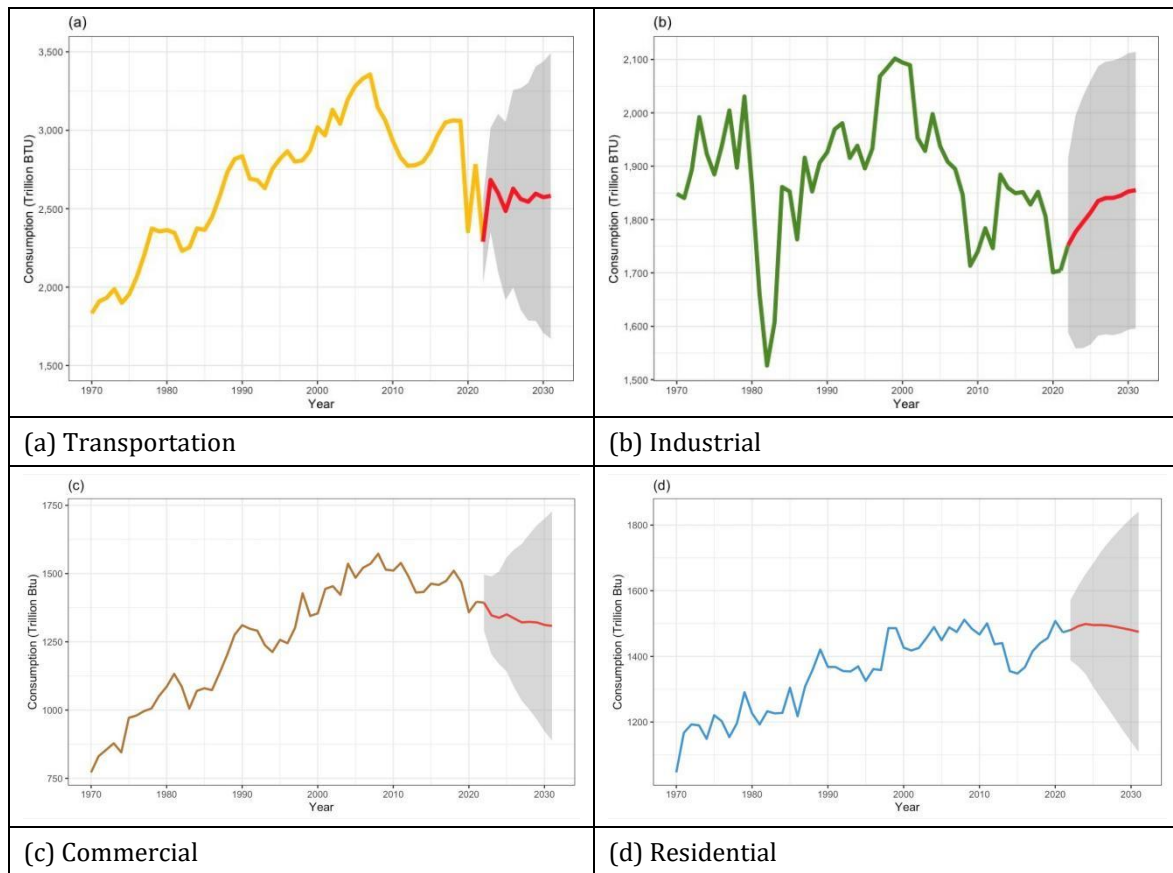


Figure 12 10 Year Energy Consumption Forecasts by Sector: (a) Transportation (b) Industrial (c) Commercial (d) Residential. Burgundy, blue, green, and yellow lines respectively represent the observed values for the previously listed sectors, red lines represent the forecast values, and grey-shaded areas represent 95% confidence intervals for the forecast values.

From the above set of plots, we first predict the transportation sector to continue to recover from the drop induced by the COVID-19 pandemic and then begin to flatten out as time goes on. For the industrial sector, we predict consumption will increase, however, the rather large confidence band indicates anything can happen. We expect consumption in the commercial sector to slowly decline over the next decade. This can be supported by the decline in office usage and increase in the popularity of working from home following the start of the COVID-19 pandemic. Lastly, we expect residential energy consumption to remain somewhat stagnant as the population in California has begun to slightly decrease in recent years.

4. Discussion

In all our modeling approaches, we make the fundamental assumption of sustained economic stability throughout the entire forecast period and that no major natural catastrophes or wars will befall the state. However, it is important to acknowledge that the disruptive and unprecedented effects of the COVID-19 pandemic pose a significant challenge for ARIMA and ARIMAX models in effectively accounting for their impact on energy consumption. This challenge is further compounded by the limited availability of data for the year 2022, which was characterized by a return to near normalcy.

Furthermore, certain exogenous variables in our analysis had fewer recorded years of data compared to the consumption variables. This limitation reduced the number of data points available for training our ARIMAX model. Despite these substantial challenges, our models exhibited commendable performance on the testing set, underscoring the accuracy and reliability of ARIMA and ARIMAX models in our forecasting framework.

5. Conclusion

Based on the predictions of our models, California is expected to consume just under 8000 trillion Btu by 2031. Fossil fuel consumption predictions show constant consumption for the next 10 years due to the demands of the industrial

and transportation sectors, and assuming 2022 returns to approximately pre-Covid levels, consumption for 2031 will reach over 5500 trillion Btu. The forecast for renewable energy sources, excluding hydroelectric, is expected to reach about 1800 trillion Btu due to the growth of solar and biomass, and if we assume a low level of precipitation affecting hydroelectric power, the total consumption for renewable energy, reaches about 2000 trillion Btu. Thus, according to these predictions, renewable energy will only account for about 25% of energy consumption by 2031, which means California's goal to reach 60% reliance on renewable energy by 2030 is likely unattainable unless drastic policies and major renewable energy installations are implemented in the next few years. California's transportation sector remains this state's largest consumer of energy while the commercial and residential sectors appear to be in decline as California's populace decreases and more businesses shift online after the pandemic.

Future Work

In the future, we would like to work with monthly data instead of annual data for California so we can also analyze the seasonality of energy usage across individual calendar years. Additionally, we would also implement a recurrent neural network to aid in our forecasting because research in this field shows that a neural network might account for the anomalies in the energy data better than an ARIMA or ARIMAX model.

Compliance with ethical standards

Acknowledgments

This project was funded by National Science Foundation HDR Data Science Corps Awards \#2123366, \#2123380, \#2123384. We thank our wonderful faculty, Dr. Sam Behseta, Dr. Jessica Jaynes, Dr. Babak Shahbaba, Dr. Roberto Pelayo, Dr. Mine Dogucu, Professor Alma Castro, and teaching assistants, Brian Schetzle and Zahra Moslemi, for their support during our research.

Disclosure of conflict of interest

No conflict of interest to be disclosed.

References

- [1] Holtberg PD, Smith KA, Mayne L, Doman L, Boedecker EE, Skelly DH, et al. Annual Energy Outlook 2010 with projections to 2035. Annual Energy Outlook 2010. 2010 Apr 1;1–15.
- [2] Kirkpatrick AJ, Benneer LS. Promoting clean energy investment: An empirical analysis of property assessed Clean Energy. Journal of Environmental Economics and Management. 2014 Sep;68(2):357–75.
- [3] Ghanadan R, Koomey JG. Using energy scenarios to explore alternative energy pathways in California. Energy Policy. 2005 Jun;33(9):1117–42.
- [4] Hoots HW, Herold SC. Natural Gas Resources of California. Geology of Natural Gas. 1935;
- [5] Denholm P, O'Connell M, Brinkman G, Jorgenson J. Overgeneration from solar energy in California. A field guide to the duck chart. National Renewable Energy Lab (NREL). 2015 Nov 1;
- [6] Hamzeh M, Kachabi A, Andayesh M, Fadaei MH. Conceptual design of a cryocooler and investigation of the effect of design parameters on cryocooler performance. 2024 Mar 22;
- [7] Dvorak MJ, Archer CL, Jacobson MZ. California offshore wind energy potential. Renewable Energy. 2010 Jun;35(6):1244–54.
- [8] Tennyson ME. Growth history of oil reserves in major California oil fields during the twentieth century. 2005;
- [9] Leifer I, Luyendyk B, Broderick K. Tracking an oil slick from multiple natural sources, Coal Oil Point, California. Marine and Petroleum Geology. 2006 Jun;23(5):621–30.
- [10] Maugeri L. The age of oil: The mythology, history, and future of the world's Most controversial resource. Guilford, CT: Lyons Press; 2008.
- [11] Geboye Desta M. The Organization of Petroleum Exporting Countries, the World Trade Organization, and Regional Trade Agreements. Journal of World Trade. 2003 Jun 1; 37(Issue 3): 523–51.

- [12] Alnasrawi A. The economy of Iraq: Oil, wars, destruction of development and prospects, 1950-2010. Westport, Connecticut etc.: Greenwood Press; 1994.
- [13] Long DE. Oil and the iran-iraq war. In: The Iran-Iraq War (RLE Iran A). 1st ed. Routledge; 1984. p. 38–50.
- [14] Hamilton JD. Historical oil shocks. In: Routledge handbook of major events in economic history. 1st ed. Routledge; 2013. p. 239–65.
- [15] Ajanovic A, Hiesl A, Haas R. On the role of storage for electricity in Smart Energy Systems. *Energy*. 2020 Jun; 200: 117473.
- [16] Newcomb R. The American Coal Industry. *Current History*. 1978 Jun 1;74(437):206–9.
- [17] Saarman González G. Evolving Jurisdiction under the federal power act: Promoting clean energy policy. *UCLA Law Review*. 2016 Jun 4; 63.
- [18] Briner MG. Economic recovery tax act of 1981. *Akron Law Review*. 1982; 15(2): 325–68.
- [19] Accelerated cost recovery system. Deloitte, Haskins & Sells; 1981.
- [20] Visick M. If Not Now, When? The California Global Warming Solutions Act of 2006: California's Final Steps Toward Comprehensive Mandatory Greenhouse Gas Regulation. *Hastings Environmental Law Journal*. 2007;13(2).
- [21] Hanemann WM. How California Came to Pass AB 32, the Global Warming Solutions Act of 2006. *eScholarship*. 2007 Mar 1;
- [22] Hughes JE, Podolefsky M. Getting green with solar subsidies: Evidence from the california solar initiative. *Journal of the Association of Environmental and Resource Economists*. 2015 Jun;2(2):235–75.
- [23] Mani M, Pillai R. Impact of dust on solar photovoltaic (PV) performance: Research status, challenges and recommendations. *Renewable and Sustainable Energy Reviews*. 2010 Dec;14(9):3124–31.
- [24] Alikhani H, Kanduri A, Liljeberg P, Rahmani AM, Dutt N. Dynafuse: Dynamic Fusion for Resource Efficient Multimodal Machine Learning Inference. *IEEE Embedded Systems Letters*. 2023 Dec;15(4):222–5.
- [25] Hart D, Birson K. Deployment of Solar Photovoltaic Generation Capacity in the United States. Office of Energy Policy and Systems Analysis, U.S. Department of Energy; 2016.
- [26] El-Sheekh M, Abomohra AE-F, editors. *Handbook of Algal Biofuels: Aspects of cultivation, conversion, and Biorefinery*. Amsterdam: Elsevier; 2022. 599–613.
- [27] Quigley DC, Scott Hornafius J, Luyendyk BP, Francis RD, Clark J, Washburn L. Decrease in natural marine hydrocarbon seepage near coal oil point, California, associated with offshore oil production. *Geology*. 1999;27(11):1047–50.
- [28] Reisen VA. Estimation of the fractional difference parameter in the arima(p, D, q) model using the smoothed periodogram. *Journal of Time Series Analysis*. 1994 May; 15(3):335–50.
- [29] Hamzeh M, Kachabi A, Sipey MH, Ganji DD. New approach method for solving nonlinear differential equations of blood flow with nanoparticle in presence of Magnetic Field. 2024 Mar 5;
- [30] Prasad NG, Rao JN. The estimation of the mean squared error of small-area estimators. *Journal of the American Statistical Association*. 1990 Mar;85(409):163–71.
- [31] Rostami M, Kia S. Federated learning using variance reduced stochastic gradient for probabilistically activated agents. In: 2023 American Control Conference (ACC). IEEE; 2023. p. 861–6.
- [32] Kahn L, Najafi H. The impact of covid-19 on energy consumption in the United States: An overview. *ASME 2021 15th International Conference on Energy Sustainability*. 2021 Jun 16;
- [33] Wang Q, Li S, Zhang M, Li R. Impact of covid-19 pandemic on oil consumption in the United States: A new estimation approach. *Energy*. 2022 Jan;239:122280.
- [34] Soleymani S, Talebi A. Forecasting solar irradiance with geographical considerations: Integrating feature selection and learning algorithms. *Asian Journal of Social Sciences and Management Studies*. 2024 Jan;6(1):85–93.
- [35] Mirmiranpour H, Amjadi A, Khandani S, Shafae Y, Omid Sobhani S. Wavelength effect in laser therapy of diabetic rats on oxidants: Ages, AOPP, ox-LDL levels. *International Journal of Clinical and Experimental Medical Sciences*. 2020;6(2):17.