Library Progress International *Print version ISSN 0970 1052*
Vol.44 No.3, Jul-Dec 2024: P.6020-6028 *Online version ISSN 2320 317X*

Original Article *Available online at www.bpasjournals.com*

Predictive Modelling of Renewable Energy Growth with Machine Learning and Evidence Markov Chain

Vatsalam Krishna Jha[1], Tanishq Sachdeva[2], Dr. Janardan Prasad Kesari[3]

1Department of Electrical engineering, Delhi Technological University, New Delhi, India
vatsalamjha.dtu@gmail.com
2Department of Mechanical engineering, Delhi Technological University, New Delhi, India
tanishqsach1@gmail.com
3Adjunct faculty, Department of Mechanical engineering, Delhi Technological University, New Delhi, India
drjpkesari@gmail.com

How to cite this article: Vatsalam Krishna Jha, Tanishq Sachdeva, Janardan Prasad Kesari (2024) Predictive Modelling of Renewable Energy Growth with Machine Learning and Evidence Markov Chain. *Library Progress International*, 44(3), 6020-6028.

Abstract
This paper presents an advanced computational framework that integrates advanced machine learning curve fitting algorithms with an enhanced Markov Chain model to simulate and forecast the growth of the renewable energy sector in the United States. The proposed model incorporates Dempster-Shafer Evidence Theory to manage the inherent uncertainties in renewable energy production, producing refined probability distributions across diverse energy sources, such as biomass, wind, solar, and hydroelectric power. These distributions, used within a Monte Carlo simulation, enable robust and accurate predictions of future renewable energy trends. The application of machine learning plays a pivotal role in optimizing the model's performance, allowing for precise adjustments to the probabilistic structure and improving prediction accuracy. The model's efficacy was demonstrated through accurate forecasts for the years 2017 and 2018. This study offers a detailed examination of the machine learning algorithms employed, the probabilistic reasoning framework, and the simulation results, providing valuable insights into the future trajectory of renewable energy growth and highlighting the transformative role of machine learning in energy forecasting.
Keywords: Renewable energy; Markov chain; Simulation; Dempster Shafer evidence theory; Probability

1. 1. Introduction

Since the past few years, there has been a great shift from non-renewable energy sources to renewable energy sources. Due to their massive, and ever-growing consumption, non-renewable energy sources are depleting at a rapid pace. Other sources of energy, like solar or wind are becoming quite popular. U.S. is moving towards a quick adoption of greener sources of energy across its power, industry and transport sectors, aiming to minimize dependence on the fossil-fuel energy economy. The United States produces electricity using several sources and technologies. These have evolved over the years, leading to the dominance of some sources over the others. According to the data released by the U.S Energy Administration Information, the sources of energy for electricity generation can be divided into three major categories: fossil fuels (coal, natural gas, petroleum etc.), nuclear energy, and renewable energy. According to the study conducted in 2017, the three major fossil fuels – coal, natural gas, and petroleum – summed up to account for around 63% of the primary energy production in US.

Fossil fuels contribute the most in the generation of electricity.

- About 32% of the energy was generated through Natural Gas, which is used to produce electricity through steam turbines and gas turbines.

- About 30% of the energy was generated through Coal-fired power plants, almost all of which use steam turbines to produce electricity.
- Less than 1% of the energy was generated with the use of Petroleum.

Nuclear power plants contribute to about 20% of the total electricity produced in U.S. Nuclear fission heats up water to produce steam, and steam turbines are used to produce electricity.

Renewable energy plants contribute about 17% of total U.S. energy production.

- Hydropower contributes around 7% to total U.S. energy production, and around 44% of electricity generated from renewable sources.
- Wind energy contributes about 6% to total U.S. energy production, and around 37% of electricity generated from renewable sources.
- Biomass, solar energy and geothermal power plants together contribute around 4% to the total U.S. energy production.

As the citizens of a highly developed and industrialized country, Americans consume a lot of energy in their households, businesses and in industry. In 2017, U.S. consumed over 36 quadrillion British Thermal Units of energy generated from petroleum, around 28 quadrillion BTUs of energy generated from natural gas, and around 14 quadrillion BTUs of energy generated from coal, and a little over 8 quadrillion BTUs of energy generated from nuclear energy sources. There are 5 basic sectors where energy is consumed:

- Industrial Sector: This sector includes facilities, plants, and equipment used for manufacturing, agriculture, construction and mining.
- Transportation Sector: This sector includes vehicles for transportation of goods as well as passengers. Eg. Cars, trucks, boats, and ships.
- Residential Sector: This sector includes houses, apartments, condos etc.
- Commercial Sector: This sector includes malls, offices, schools, hospital, restaurant, places of public assembly and worship.
- Electric Power Sector: This sector consumes energy for generating a major chunk of the electricity consumed by the other four sectors.

From the information given above, we can conclude that the present U.S. energy sector depends heavily upon fossil fuels. Minimizing the usage of the non-renewable fossil fuels and utilising renewable sources to save resources and the environment is an important concern. However, these alternative sources are investment-intensive, and owing to uncertainty of the climate and other factors may provide low benefit.

2. 2. Methodology

2.1 2.1 Discrete Time Markov Chain

The Discrete-Time Markov Chain model is based on Markov property. Let a sequence of random variables be denoted by $\{X_n : n > 0\}$. When n is a positive integer, and discrete states are $i_1, i_2, \cdots$, and $P\{X_n = i_n, X_{n-1} = i_{n-1}, \cdots, X_1 = i_1\} > 0$, the sequence $\{X_n : n > 0\}$ is a discrete-time Markov Chain, if and only if,

$$P\{X_{n+1} = i_{n+1}|X_n = i_n, X_{n-1} = i_{n-1}, \cdots, X_1 = i_1\} = P\{X_{n+1} = i_{n+1}|X_n = i_n\}$$

is true, in which $X_1, X_2, \cdots, X_n, X_{n+1}$, are random variables, while $P\{X_{n+1} = i_{n+1}|X_n = i_n\}$ is the conditional probability of X_{n+1} belonging to state i_{n+1} which is followed by X_n. All the possible values of X_i are used to form a set $S = \{i_1, i_2, \cdots, i_n, i_{n+1}, \cdots\}$, which is referred to as its state space. The above equation is known as Markov property. It states that, if we are have the present as well as the past states, the probability distribution of the consequent future state does not depend on its past states, and only depends upon the present state.

The traditional method of carrying out predictions for a time series using discrete time Markov chain is by the use of Transfer matrix. To construct a transfer matrix, the following procedure is used:

First: The observed transition frequencies between two states are counted from the past data. The transition frequency from state i to state j at times m and (m+1) respectively, is indicated by N_{ij}.

Second: The probability of single-step transition, which is represented by p_{ij} is calculated by the formula,

$$P_{ij} = \frac{N_{ij}}{\sum_j N_{ij}} \tag{1}$$

Third: As the transition probabilities are calculated, the transition probability matrix is obtained as $P = [p_{ij}]_{ij}$

Fourth: Based on the present state, and the obtained transition probability matrix P, we can obtain the probability distribution about the future state. Using it we can find the most probable and least probable state.

Hence when using the transfer matrix method, we get the transition probability matrix P, which is dependent on the frequency of transition between two consecutive states. It is observed that when using this method, the predicted outcomes maybe unstable, especially when the number of possible states is small.

2.2 2.2 Evidence Markov Chain

DS Evidence theory: Let us assume that U, known as the frame of discernment, is a finite set and has N mutually exclusive and exhaustive elements, and is given by $U=\{\theta, \theta_2 \ldots, \theta_N\}$. The power set for U is given by:

$$2^U= \{\emptyset, \{\theta_1\} \{\theta_2\}, \ldots, \{\theta_N\}, \{\theta_1 \cup \theta_2\}, \{\theta_1 \cup \theta_3\}, \ldots, U\}$$

Now a mass function is constructed which maps from U to [0,1], which is defined by $m : 2^U \rightarrow [0, 1]$, and satisfies:

$$\sum_{A\in 2^U} m(A) = 1 \tag{2}$$

$$m(\emptyset) = 0 \tag{3}$$

It is known as the basic probability assignment or BPA. Here, if $m(A) > 0$ is satisfied, then A will be the element of 2U termed as the focal element. m(A) returns the supporting degree for the focal element A.

The mass function corresponds to a belief (Bel) function and a plausibility (Pl) function.

For $m : 2^U \rightarrow [0, 1]$, Bel(A) denotes the degree of belief for the supporting focal element A, and Pl(A) represents the degree of belief for not denying A, and is defined as:

$$Bel(A) = \sum_{B\subseteq A} m(B), \forall A \subseteq 2^U \tag{4}$$

$$Pl(A) = \sum_{B\cap A=\emptyset} m(B), \forall A, B \subseteq 2^u \tag{5}$$

If every focal element for m is a singleton, the BPA degenerates into classical probability. Consequently, the belief and Plausibility functions also degenerate into the same measure of probability.

Pignistic probability transform: Dempster-Shafer theory can be used to assign belief to all the subsets of the frame of discernment. Belief assignment of a multi element subset which can have uncertainty can be displayed through BPA, though making decisions directly via using BPAs can be difficult. Hence we convert BPA to a probability distribution. Pignistic probability transformation is used to obtain singleton sets by taking the average of the BPAs of a multi-element set, and it can be defined as:

$$Bet\ P(B) = \sum_{B\subseteq A, B\in U} \frac{m(A)}{|A|} \tag{6}$$

Here, m(A) is mass function for focal element A with cardinality |A|.

Interval number: For given lower and upper limits a^- and a^+ it is given by $\tilde{a} = [a^-, a^+] = \{x | a^- \le x \le a^+\}$. For $a^+ = a^-$, ã degenerates into a crisp value.

Let $A = [a_1, a_2]$ and $B = [b_1, b_2]$ be two interval numbers. The square of the distance between A and B is represented by $D^2(A, B)$ which is defined as:

$$\begin{aligned} D^2(A,B) &= \int_{-\frac{1}{2}}^{\frac{1}{2}} \left\{\left[\frac{(a_1+a_2)}{2} + x\,(a_2-a_1)\right] - \left[\frac{(b_1+b_2)}{2} + x\,(b_2-b_1)\right]\right\}^2 \mathrm{d}x \\ &= \left[\frac{(a_1+a_2)}{2} - \frac{(b_1+b_2)}{2}\right]^2 + \frac{[(a_2-a_1)+(b_2-b_1)]^2}{12}. \end{aligned} \tag{7}$$

New Evidence Markov Chain Model: To construct an evidence Markov chain model, the following steps are carried out:

First: Extending the state space to an evidential framework.

U is the frame of discernment, which is found out using collected data and the user's requirement. The states are determined in the evidential framework which is taken as a subset of the power set of U (ie. 2^U). The scope of these states can be adjusted or tuned by the user based on practical conditions.

Second: Generating the BPAs of the data collected.

In this model, we use interval numbers for the collected data, which helps in handling uncertain information. Optimisation based on distance between the intervals is used to generate BPAs of the existing data. For a number I (interval number) and n possible states in the frame U, all states have a analogous scope, while I is given a BPA m_i designated to the the state i. Then the distance between the data collected and the state I (termed as D_i) can be obtained. Discounted distance (DD_i) is then calculated using $DDi = m_i \cdot D_i$ where i = 1, 2, 3, · · · , n. Since BPA denotes the degree of belongingness to a state, a shorter D_i leads to a larger m_i value. We can minimise the RMS of discounted distances for this optimisation.

$$\arg\min_{m} \sqrt{\frac{\sum_{i=1}^{n}(m_i \,.D_i)^2}{n}} \tag{8}$$

This becomes a problem solvable through quadratic programming, with a positive definite Hessian matrix. This step computes the m_i (BPA):

Third: Obtaining the transition belief matrix.

The transition belief assignment for the single step, M_{ij} , $i,j \in 2^U$ represents the assignment of belief for transitioning from state i to state j.

$$M_{ij} = \frac{\sum_{t=1}^{n-1}\left(m(i)_t \cdot m(j)_{t+1}\right)}{\sum_{k\in 2^U}\sum_{t=1}^{n-1}\left(m(i)_t \cdot m(k)_{t+1}\right)} \tag{9}$$

With the help of the above fomula transition belief matrix $[M_{ij}]$ is derived. Here unlike DTMC, the transition is among the states represented in an evidential framework and not the basic states.

Fourth: Prediction of BPA for the subsequent period.

The prediction of the state assignments for the subsequent period takes place using the transition belief matrix. If the BPA for last period is m', then for the next period, it is given by:

$$m = m'.[M_{ij}] \tag{10}$$

Fifth: Getting the ultimate prediction.

As we have already discussed, the decision-making is difficult based on BPA, hence we use the predicted BPA into a state probability assignment using classical pignistic probability transformation.

2.3 2.3 Proposed Model

We propose an approach as an extension to Monte Carlo Simulation in which rather than determining the random variable using a probability distribution about a fixed central tendency (here, mean), we use the probability distribution about a variable mean which can be estimated using techniques such as linear regression, LOESS, curve fitting, etc.

First we will use any curve fitting model to obtain with time stamps as in input. Using this model we will calculate the deviation of the data points from the fitted curve. Now, we will use evidence Markov chain model on the deviation data to find out the probability distribution of the points about the curve. Now we will perform Monte Carlo simulations on the probability distribution about the curve to simulate the uncertainty in the deviation.

3. 3. Dataset

The production and consumption data for the energy was obtained from U.S. Energy Information Administration's (EIA) released outlook document. It consisted of monthly energy data for our required location (United States). This dataset includes statistics on total energy production and consumption (in Trillion Btu) for all different sectors from the year 1973 to 2018, i.e. for 552 data points in a time series. The data consisted of primary energy production from different sources (fossil fuels, nuclear power and renewable sources) and primary, and total energy consumption for various sectors such as residential, industrial, commercial, and transport, along with electric power.

Using this data, we trained 5 models described in the methodology for all the renewable resources and total consumption.

The models were trained using this data for the year 1973-2004 and tested for year 2005-2016.

4. 4. Energy Consumption Model

As discussed earlier, we have used the proposed simulation model on monthly energy consumption and renewable energy production data.

The energy consumption data consisted of monthly energy consumption (in Trillion Btu) in the United States from 1973 to 2018, i.e. for 552 data points in a time series. We used data from 1973 to 2004 to train the model, and tested the data for years 2005 to 2018. Hence, for training 70% of the data was used and for testing 30% of the data was used in the model building.

First a curve was fitted between the consumption and timestamps. Now, the deviation from the fitted curve was found out as shown in the figure.

Maximum positive deviation=m^+=848.284

Minimum negative deviation=m^-=-1726.327

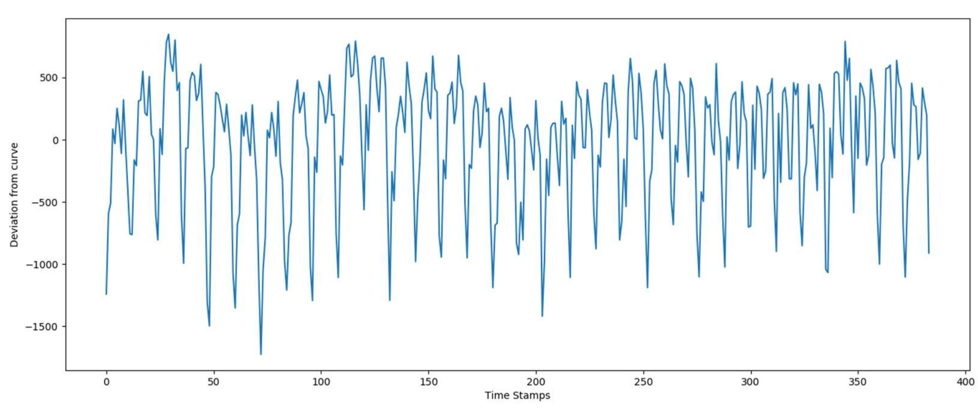

Figure 1: Deviation from mean line

Now we calculate the intervals for given m^+ and m^-, and 'n' number of data points using the formula:

$$Interval\ Size = \left|\frac{m^+ - m^-}{\sqrt{n}}\right| \quad (11)$$

Hence, we obtain 19 intervals with a size of 135 units. We used these intervals as states.

Now we used Evidence Markov Chain on deviation to find out its probability distribution about the fitted curve with the help of the transition matrix. This probability distribution was used in Monte Carlo simulation to simulate the uncertainty in the deviation from the fitted curve.

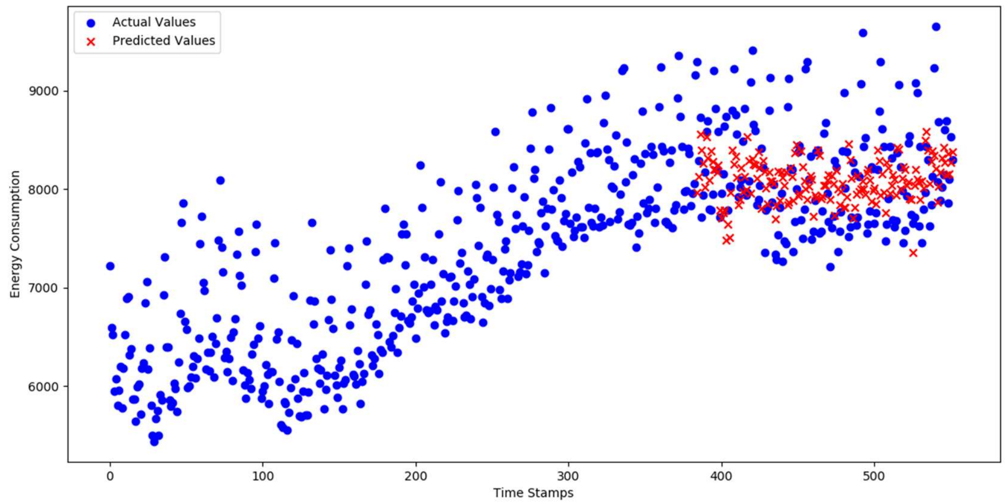

Figure 2: Monthly Energy Consumption Values

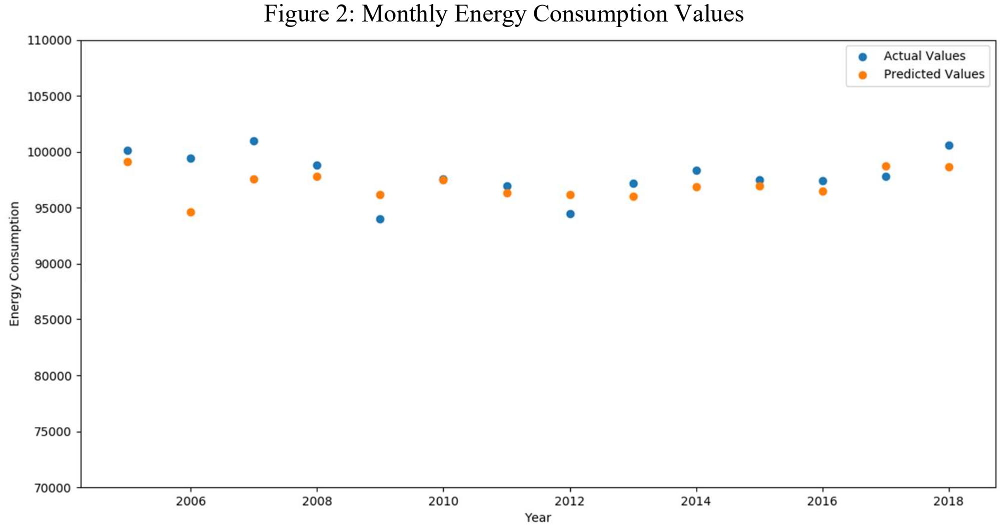

Figure 3: Annual Energy Consumption Values

Table 1: Comparison of actual and predicted values

Year	Actual Value of energy consumption	Predicted Value of energy consumption	%age logarithmic error
2005	100167.7	99160.2	0.088%
2006	99464.23	94652.87	0.431%
2007	100970.7	97554.81	0.299%
2008	98824.95	97835.14	0.088%
2009	94022.75	96183.95	-0.198%
2010	97607.7	97530.42	0.007%
2011	96948.8	96352.91	0.054%

2012	94477.46	96170.94	-0.155%
2013	97218.44	95989.29	0.111%
2014	98381.28	96911.2	0.131%
2015	97483.57	96925.69	0.050%
2016	97443.88	96454.65	0.089%
2017	97806.69	98716.21	-0.081%
2018	100565.4	98626.87	0.169%
Total	1371384	1359065	0.064%

Hence we are able to create a model which can simulate the energy consumption with a decent amount of accuracy. The same model was used to simulate the production through major renewable sources of energy such as geothermal energy, hydro energy, biomass, wind energy, and solar energy.

5. 5. Results

As discussed earlier, the model was used to simulate the power generated through renewable resources and total energy consumption for years 2005-2018**.** From the tables below, we can observe that predicted values are quite close that the actual values, considering that fact that this is a univariate model. The model was correctly able to simulate the growth in each source of renewable energy as well as in the energy consumption both in terms of absolute values and relative values. In the actual values the renewable energy sector grew from about 6% to about

Year	Wind	BioMass	GeoThermal	HydroElectric	Solar	Total_Production	Consumption	Percentage
2005	178.09	3101.19	180.7	2702.94	57.77	6220.69	100167.66	6%
2006	263.74	3211.51	181.2	2869.03	60.58	6586.06	99464.23	7%
2007	340.5	3472.08	185.77	2446.39	65.38	6510.12	100970.69	6%
2008	545.55	3868.25	192.43	2511.11	73.81	7191.15	98824.95	7%
2009	721.13	3956.62	200.19	2668.82	77.65	7624.41	94022.75	8%
2010	923.43	4552.53	207.98	2538.54	90.48	8312.96	97607.7	9%
2011	1167.64	4704.32	212.31	3102.85	111.13	9298.25	96948.8	10%
2012	1340.06	4546.7	211.59	2628.7	156.85	8883.9	94477.46	9%
2013	1601.36	4815.63	214.01	2562.38	224.52	9417.9	97218.44	10%
2014	1727.54	5020.37	214.49	2466.58	336.94	9765.92	98381.28	10%
2015	1777.31	4991.58	211.84	2321.18	425.73	9727.64	97483.57	10%
2016	2095.59	5080.5	209.6	2472.44	568.67	10426.8	97443.88	11%
2017	2342.89	5203.67	210.23	2766.97	774.47	11298.23	97806.69	12%
2018	2533.13	5329.11	217.31	2687.65	948.65	11715.85	100565.44	12%

12%, hence a positive trend was observed, which was also reflected by model.

Table 2: Actual values in Trillion Btu

Year	Wind	BioMass	GeoThermal	HydroElectric	Solar	Total_Production	Consumption	Percentage
2005	214.16	3002.6	181.98	3051.82	43.67	6494.23	97918.2	7%
2006	300.9	3190.6	183.38	2850.42	53.81	6579.11	97892.87	7%
2007	389.02	3457.57	186.08	2823.82	54.35	6910.84	99444.81	7%
2008	528.15	3751.48	192.29	2675.62	62.02	7209.56	98375.14	7%
2009	701.34	3863.51	196.7	2793.42	74.7	7629.67	97695.95	8%
2010	877.16	4373.66	205.85	3013.82	99.49	8569.98	97719.42	9%
2011	1072.15	4622.89	213.73	2766.82	131.72	8807.31	95785.91	9%
2012	1273.43	4590.77	216.91	2637.62	177.92	8896.65	96278.94	9%
2013	1492.88	4660.96	217.8	2694.62	241.82	9308.08	95854.29	10%
2014	1710.98	4793.52	215.76	3025.22	327.09	10072.57	97316.2	10%
2015	1920.25	4918.4	213.46	2892.22	430.7	10375.03	97330.69	11%
2016	2115.39	5002.47	211.61	2607.22	583.51	10520.2	97116.15	11%
2017	2294.22	5068.72	214.46	2466.62	751.98	10796	96745.21	11%
2018	2442.47	5263.94	218.31	3006.22	954.4	11885.34	99328.87	12%

Table 3: Predicted values in Trillion Btu

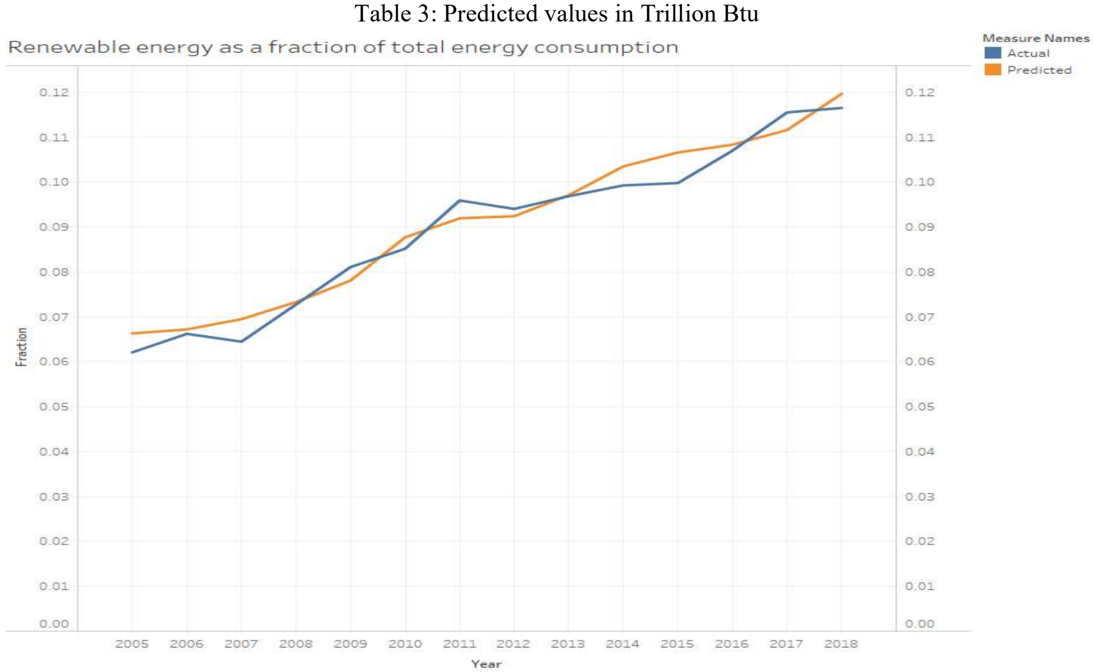

Figure 4: Renewable energy as a fraction of total energy consumption

6. 6. Conclusion

The primary objective of this research was to develop a robust framework to analyze and forecast the growth of the renewable energy sector, a critical factor influencing global energy markets due to increasing demand for resources and services. The trajectory of renewable energy development directly impacts a nation's economic and industrial sectors, with energy consumption patterns being closely linked to economic growth. As the United States emerges as a significant energy exporter, understanding these dynamics becomes ever more vital.

Previous studies have explored renewable energy forecasting through methods such as Kalman filtering, Monte Carlo simulations, artificial neural networks (ANN), and fuzzy logic systems. However, many of these models rely on the assumption of constant external factors, limiting their ability to respond to sudden disruptions or crisis events. These methods also struggle with random or highly volatile data, requiring time series data with observable trends to maintain accuracy. Moreover, these models typically provide accurate forecasts only for short-term predictions, with large errors accumulating over longer timeframes.

In this study, our proposed model successfully predicted the yearly growth in the percentage of energy generated from renewable sources for 2017 and 2018, demonstrating its practical efficacy. For 2017, the model forecasted an 11% annual increase in renewable energy production, closely aligning with the actual 12% growth reported by the U.S. Energy Information Administration (EIA). Similarly, for 2018, our model's prediction of a 12% growth was validated by the corresponding real-world data. Looking ahead, further improvements can be made by incorporating more dynamic external variables, such as climate conditions, including wind speed, wind direction, and dew point, to enhance prediction accuracy. Additionally, the integration of more sophisticated machine learning techniques, such as deep neural networks and fuzzy-based expert systems, could further refine the model, particularly when supplemented with historical data and real-time climate information.

1. **References**

2. He, Zichang, and Wen Jiang. “A new belief Markov chain model and its application in inventory prediction.” *International Journal of Production Research* 56, no. 8 (2017): 2800-2817.
3. Deng, Xinyang, Qi Liu Liu, and Yong Deng Deng. “Newborns prediction based on a belief Markov chain model.” *Applied Intelligence* 43, no. 3 (2015): 473–486.
4. Paul, V.K., Basu. C. *A Handbook for Construction Project Planning and Scheduling.* Copal Publishing Group, 2017.
5. Tüzüntürk, S., Eren Şenaras, A., Sezen, K. “FORECASTING WATER DEMAND BY USING MONTE CARLO SIMULATION .” 2015.
6. Shaver, Ben. *A Zero-Math Introduction to Markov Chain Monte Carlo Methods.* Towards Data Science. 22 December 2017. https://towardsdatascience.com/a-zero-math-introduction-to-markov-chain-monte-carlo-methods-dcba889e0c50 (accessed December 29, 2018).
7. *Monthly Energy Review.* U.S. Energy Information Administration. 1 January 2019. https://www.eia.gov/totalenergy/data/monthly/index.php (accessed January 15, 2019).
8. *OPEN DATA.* U.S. Energy Information Administration. 1 January 2019. https://www.eia.gov/opendata/ (accessed January 15, 2019).
9. Cassolaa, Federico, and Massimiliano Burlando. “Wind speed and wind energy forecast through Kalman filtering of Numerical Weather Prediction model output.” *Applied Energy* 99 (2012): 154-166.
10. *U.S. Energy Consumption - Statistics & Facts.* Statist. https://www.statista.com/topics/833/energy-consumption/ (accessed January 20, 2019).
11. *Electricity in the United States is produced with diverse energy sources and technologies.* U.S. Department of Energy. December 2018. https://www.eia.gov/energyexplained/index.php?page=electricity_in_the_united_states (accessed January 30, 2019).