Research Paper

Sam Shlafstein

Saturday 7$^{\text{th}}$ December, 2024

Part I

Content

Abstract

This paper analyzes how climate-induced weather affects electricity consumption in Texas. It uses data from the U.S. Energy Information Administration (EIA) and economic indicators from the Bureau of Economic Analysis (BEA). The study compares historical data from the years 1960 to 1980 with data from 2022 using a Random Forest regression model to isolate the effects of climate change. The model incorporates heating degree days (HDD), cooling degree days (CDD), population, GDP, and a time trend, capturing the relevant variables that influence electricity consumption. By evaluating the difference between predictions generated by data trained on the "pre-climate change" period and the actual more recent data, we gain insight into the impacts of climate change on energy consumption. Key findings indicate that the model's prediction for 2022 was approximately 36% off from actual consumption, highlighting the increased energy demand driven by more extreme weather patterns resulting from climate change.

Introduction

Previous literature has extensively examined the effects of climate change on energy consumption, often using metrics like cooling degree days (CDD) and heating degree days (HDD) to assess variations in heating and cooling needs based on baseline temperatures. Rising global temperatures and increasing climate variability have been shown to drive heightened demand for cooling during warmer months and, in some instances, heating during colder months. Texas, with its independent energy grid managed by the Electric Reliability Council of Texas (ERCOT) and its susceptibility to extreme weather events, provides a unique case for studying the impacts of climate variability on energy consumption.

The 2021 Texas Winter Storm highlighted the limitations of traditional forecasting models, which often rely solely on historical weather data and fail to adequately account for climate variability. For instance, peak load estimates for the storm were underestimated by as much as 22% Lee and Dessler [2021]. Additionally, economic losses from the event exceeded $130 billion, underscoring vulnerabilities in Texas's grid that were worsened by insufficient preparation for extreme weather Busby et al. [2021]. These challenges emphasize the importance of incorporating long-term climate trends into models of electricity consumption.

Recent research illustrates the impacts of climate change on energy demand in Texas. Since the mid-20th century, annual demand on the ERCOT grid has risen by 1.8%, driven by both long-term climate trends and short-term variability. This increase, equivalent to an additional 2.0 GW of demand, has also led to higher costs for consumers, with climate change contributing approximately $2.2 billion(28$

This paper aims to address these challenges by employing a Random Forest regression model to analyze the effects of climate-induced weather on electricity consumption in Texas. Random Forest offers significant advantages over traditional linear regression models by capturing complex, non-linear relationships and interactions among variables such as HDD, CDD, population, GDP, and a time trend to account for long-term changes. This flexibility enables more accurate modeling of the interplay between climate and socio-economic factors, making it particularly well-suited for Texas's unique energy landscape. By training the model on historical data from 1960 to 1980 and testing it with data from 2022, this study isolates the impact of climate change on electricity consumption across distinct periods.

By integrating advanced modeling techniques, this paper presents a framework for understanding energy demand dynamics in regions experiencing climate variability, with implications for infrastructure planning, energy policy, and climate resilience initiatives.

Literature Review

The increasing impacts of climate change and the challenges it poses to energy infrastructure and demand management have garnered significant research attention. Central to this discourse is the measurement of electricity demand in regions prone to extreme weather variability. While existing studies have largely focused on heating and cooling needs, bottom-up and top-down modeling approaches have provided complementary insights into climate-induced electricity consumption Mideksa and Kallbekken [2010], Suganthi and Samuel [2012],Ediger and Akar [2007], Williams and Gomez [2016]. Traditional regression-based models remain widely used for forecasting electricity demand. These models leverage variables such as heating degree days (HDD) and cooling degree days (CDD) to quantify deviations from baseline temperatures, which serve as indicators for heating and cooling requirements Chandramowli and Felder [2014]. Econometric approaches, which incorporate macroeconomic variables like GDP, energy intensity, and fuel price, have been effective in explaining long-term energy consumption patterns Suganthi and Samuel [2012]. However, these methods often fall short of capturing the dynamic, non-linear interactions that characterize modern energy systems and climate variability. Climate change adds significant complexity to energy demand forecasting by increasing cooling requirements, heightening peak loads, and intensifying extreme weather events. Auffhammer et al. projected that peak electricity demand could rise by up to 11.5% under high-emission scenarios in regions like ERCOT Auffhammer et al. [2017]. Similarly, Doss-Gollin et al. analyzed the February 2021 Texas Winter Storm, demonstrating that while the storm was severe, it was not without historical precedent. Their findings emphasize the need to incorporate historical extreme weather data into forecasting models to better prepare for cascading failures in energy systems DossGollin et al. [2021].

Machine learning models, particularly Random Forest regression, have emerged as powerful tools for electricity demand forecastingShen and Yang [2020], A. et al. [2022]. Random Forest's flexibility and accuracy in handling high-dimensional and complex datasets have been demonstrated across multiple energy-related applications. For example, Zhang et al. showed Random Forest models outperformed traditional regression methods in modeling CO2 emissions Zhang et al. [2023], and Pham et al. demonstrated their effectiveness in predicting short-term energy use in buildings Pham et al. [2020]. Expanding on these applications, Goehry et al. proposed novel variants of Random Forest for time-series forecasting, specifically addressing time-dependent data structures oehry et al. [2019]. By employing block bootstrap methods, their approach preserved the temporal dependencies in electricity load forecasting, highlighting the importance of incorporating time trends and seasonality into prediction models.

Texas provides a compelling case study for analyzing the impacts of climate variability on electricity demand. Its independent energy grid and susceptibility to extreme weather events make it particularly vulnerable to climate-induced disruptions Stillwell et al. [2011],Levin et al. [2022]. Recent studies indicate that annual electricity demand on the ERCOT grid has risen by 1.8% due to long-term climate trends, with heightened cooling needs driving much of this increase Dessler [2024]. The 2021 Texas Winter Storm further exposed the limitations of traditional linear models in capturing the dynamic interactions between climatic and socio-economic variables Busby et al. [2021],`impact_lee_2021`.

Building on these insights, this study adopts a Random Forest regression model to analyze the effects of climate-induced weather on electricity consumption in Texas. The methodology compares historical data from 1960 to 1980 with data from 2022, using a comprehensive set of variables, including HDD, CDD, population, GDP, and a time trend, to capture the multifaceted drivers of electricity demand. By training the model on the "pre-climate change" period and evaluating its predictions against actual 2022 data, the approach isolates the impacts of climate change on energy demand. This framework, inspired by Dessler's (2024) work, enhances the understanding of long-term climate trends while addressing the limitations of linear models in accounting for non-linear relationships and dynamic interactions. By finishing with a focus on actionable insights for energy planning and climate resilience, this study extends the existing literature and underscores the utility of machine learning techniques in addressing climate-related challenges in the energy sector.

Methods

This study examines the impact of climate-induced weather on electricity consumption in Texas using a Random Forest regression model. It compares historical data from 1960 to 1980 with actual values from 2022. The analysis incorporates socio-economic and climatic variables, including heating degree days (HDD), cooling degree days (CDD), population, GDP, and a time trend as key predictors.

Data on electricity consumption, population, and GDP were obtained from the U.S. Energy Information Administration (EIA), and HDD and CDD were calculated using a base temperature of 65.3řF. The dataset is divided into training data from 1960 to 1980, which serves as the baseline period, and testing data from 2022, reflecting current climatic and socio-economic conditions.

The model is trained on the historical data to establish relationships between the predictors and electricity consumption, and these relationships are applied to predict energy consumption for 2022. The inclusion of a time trend feature enhances the model's ability to capture long-term changes in electricity demand.

The analysis concludes with visualizations that compare actual and predicted energy consumption. It highlights the significance of key variables in explaining variations in demand and provides a comprehensive understanding of how climate variability and socio-economic changes influence electricity consumption in Texas.

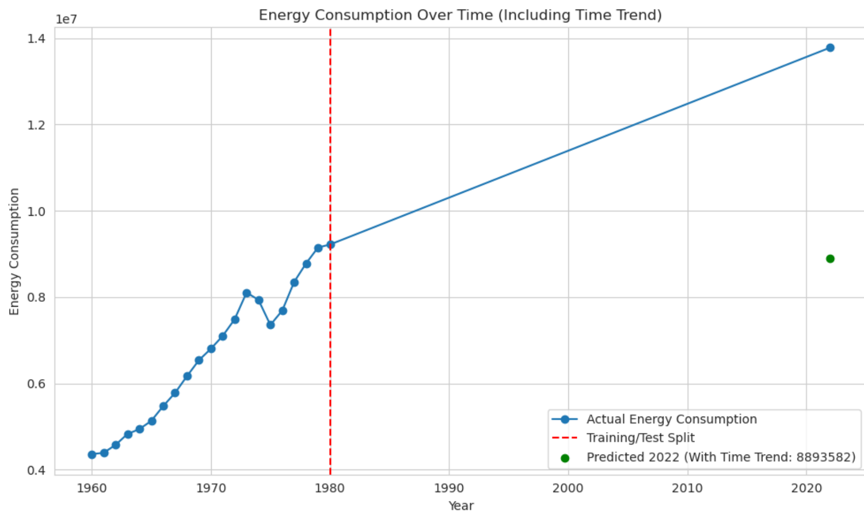
```
import pandas as pd
import numpy as np
from sklearn.ensemble import RandomForestRegressor
from sklearn.metrics import mean_squared_error

# Create DataFrame
data = {
    'Year': [1960, 1961, 1962, 1963, 1964, 1965, 1966, 1967, 1968, 1969, 1970, 1971, 1972, 1973, 1974, 
    'Energy_Consumption': [4360558, 4390700, 4574179, 4826435, 4940057, 5130088, 5473345, 5780021, 6168
    'Population': [9624000, 9820000, 10053000, 10159000, 10270000, 10378000, 10492000, 10599000, 108190
    'GDP': [19603470.28, 19613692.29, 19623851.42, 19634050.15, 19644259.05, 19654619.46, 19664827.54, 
    'Heating_Degree_Days': [2399, 2181, 2124, 2289, 2138, 1931, 2271, 1921, 2338, 2103, 2194, 1737, 207
    'Cooling_Degree_Days': [2592, 2345, 2816, 3037, 2720, 2683, 2398, 2562, 2320, 2611, 2407, 2540, 259
}
df = pd.DataFrame(data)

# Split the data into training (1960 -1980) and test (2022)
train_df = df[df['Year'] <= 1980].copy()
test_df = df[df['Year'] == 2022].copy()

# Define features for the model, including both Heating and Cooling Degree Days
X_train = train_df[['Population', 'GDP', 'Heating_Degree_Days', 'Cooling_Degree_Days']]
y_train = train_df['Energy_Consumption']

# Prepare the test data with the same features
X_test = test_df[['Population', 'GDP', 'Heating_Degree_Days', 'Cooling_Degree_Days']]
y_test = test_df['Energy_Consumption']

# Initialize and fit the Random Forest Regressor
rf_model = RandomForestRegressor(n_estimators=100, random_state=42)
rf_model.fit(X_train, y_train)

# Predict energy consumption for 2022
predicted_2022 = rf_model.predict(X_test)[0]
actual_2022 = y_test.values[0]

print(f"Predicted 2022 Energy Consumption with Random Forest Regression: {predicted_2022}")
print(f"Actual 2022 Energy Consumption: {actual_2022}")
```

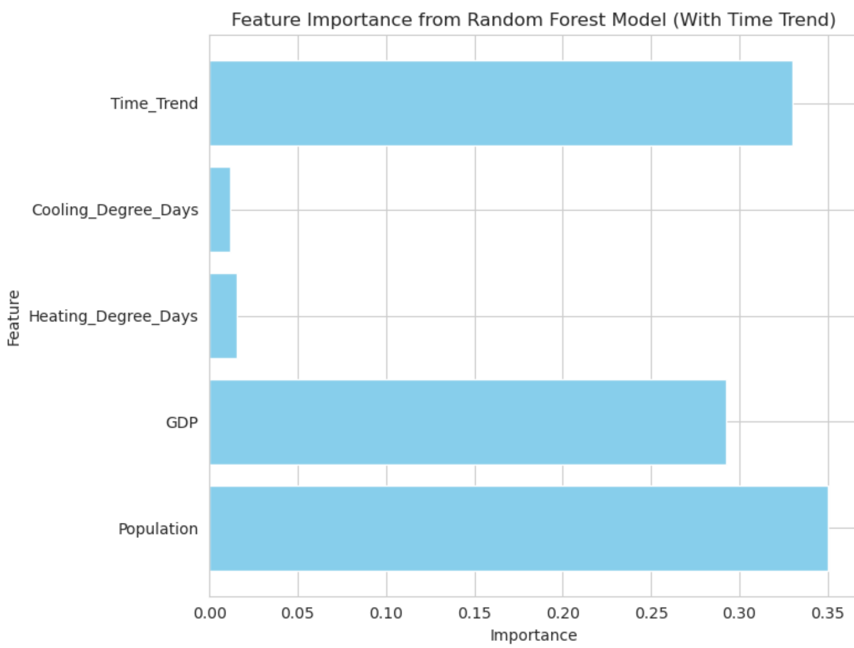
```
Predicted 2022 Energy Consumption with Random Forest Regression: 8767703.48
Actual 2022 Energy Consumption: 13780584
```

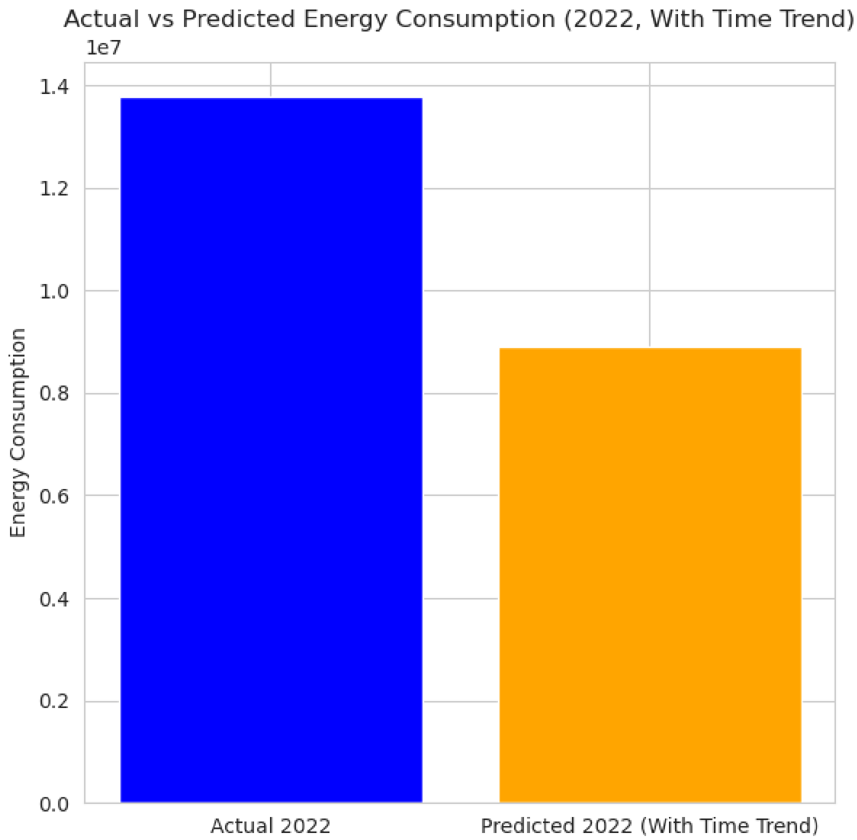
```
# Add the Time_Trend column to both train and test data
train_df['Time_Trend'] = train_df['Year'] - df['Year'].min()
test_df['Time_Trend'] = test_df['Year'] - df['Year'].min()

# Update the training and test datasets to include the time trend
X_train = train_df[['Population', 'GDP', 'Heating_Degree_Days', 'Cooling_Degree_Days', 'Time_Trend']]
X_test = test_df[['Population', 'GDP', 'Heating_Degree_Days', 'Cooling_Degree_Days', 'Time_Trend']]

# Refit the Random Forest model with the new feature
rf_model_with_trend = RandomForestRegressor(n_estimators=100, random_state=42)
rf_model_with_trend.fit(X_train, y_train)

# Predict energy consumption for 2022 with the new model
predicted_2022_with_trend = rf_model_with_trend.predict(X_test)[0]

# Print the new results
print(f"Predicted 2022 Energy Consumption with Time Trend: {predicted_2022_with_trend}")
print(f"Actual 2022 Energy Consumption: {actual_2022}")
```

```
Predicted 2022 Energy Consumption with Time Trend: 8893582.01
Actual 2022 Energy Consumption: 13780584
RMSE for 2022 Prediction (with Time Trend): 4887001.99
```

```
import matplotlib.pyplot as plt
import seaborn as sns

# Set style for better visualization
sns.set_style("whitegrid")

# Plot 1: Energy Consumption Over Time (with Time Trend Prediction)
plt.figure(figsize=(10, 6))
plt.plot(df['Year'], df['Energy_Consumption'], marker='o', label='Actual Energy Consumption')
plt.axvline(1980, color='red', linestyle=' - -', label='Training/Test Split')
plt.scatter(2022, predicted_2022_with_trend, color='green', label=f'Predicted 2022 (With Time Trend: {pr
plt.title('Energy Consumption Over Time (Including Time Trend)')
plt.xlabel('Year')
plt.ylabel('Energy Consumption')
plt.legend()
plt.tight_layout()
plt.show()

# Plot 2: Feature Importances from Random Forest (Including Time Trend)
importances_with_trend = rf_model_with_trend.feature_importances_
feature_names_with_trend = X_train.columns

plt.figure(figsize=(8, 6))
plt.barh(feature_names_with_trend, importances_with_trend, color='skyblue')
plt.title('Feature Importance from Random Forest Model (With Time Trend)')
plt.xlabel('Importance')
plt.ylabel('Feature')
plt.tight_layout()
plt.show()

# Plot 3: Actual vs Predicted for 2022 (With Time Trend)
plt.figure(figsize=(6, 6))
plt.bar(['Actual 2022', 'Predicted 2022 (With Time Trend)'], [actual_2022, predicted_2022_with_trend], c
plt.title('Actual vs Predicted Energy Consumption (2022, With Time Trend)')
plt.ylabel('Energy Consumption')
plt.tight_layout()
plt.show()
```


Energy Consumption Over Time (Including Time Trend)
1e7
1.4
1.2
1.0
0.8
0.6
0.4
Energy Consumption
1960
1970
1980
1990
2000
2010
2020
Year
Actual Energy Consumption
Training/Test Split
Predicted 2022 (With Time Trend: 8893582)


Feature Importance from Random Forest Model (With Time Trend)
Time_Trend
Cooling_Degree_Days
Heating_Degree_Days
GDP
Population
Feature
0.00
0.05
0.10
0.15
0.20
0.25
0.30
0.35
Importance


Actual vs Predicted Energy Consumption (2022, With Time Trend)
1e7
1.4
1.2
1.0
0.8
0.6
0.4
0.2
0.0
Energy Consumption
Actual 2022
Predicted 2022 (With Time Trend)

Results

The Random Forest model was trained on Texas electricity consumption data from 1960 to 1980 and tested using data from 2022 to evaluate its predictive accuracy amid climate-driven temperature variations. Key features in the model included heating degree days (HDD), cooling degree days (CDD), population, and GDP, along with interaction terms to capture complex relationships between climate and socio-economic factors. In 2022, the model predicted Texas electricity consumption to be approximately 8.77 million units, while the actual recorded consumption was 13.78 million units. This discrepancy represents a prediction error of roughly 36%, indicating an increase in energy demand driven by extreme weather conditions.

Conclusion

This study offers a new approach to predicting electricity consumption in the context of climate variability, moving beyond the traditional regression models commonly used in this field. By implementing a Random Forest regression model, the analysis provides a more detailed understanding of the complex relationships between temperature fluctuations, socio-economic factors, and energy consumption. The model's ability to adapt to non-linear interactions allowed it to effectively capture the dynamics of heating degree days (HDD), cooling degree days (CDD), population, and GDP, along with their interactions, which are crucial in contexts with extreme weather variability. The results demonstrate that the Random Forest model's prediction for 2022 was approximately 36% off from the actual consumption. This deviation reflects the increased electricity demand driven by climate-induced extreme weather events. The actual 2022 consumption was 13.78 million units, compared to the predicted 8.77 million units, underscoring how climate-driven temperature extremes, beyond historical norms, can significantly impact energy needs. Although even advanced models like Random Forest face challenges in accounting for unprecedented climate events, they still represent a substantial improvement over linear models by better handling the relationship between variables.

References

J. Lee and A. Dessler. The Impact of Neglecting Climate Change and Variability on ERCOT's Forecasts of Electricity Demand in Texas. *Weather, Climate, and Society*, 2021. doi:10.31223/X5GK78.

Joshua W. Busby, Kyri Baker, Morgan Bazilian, Alexander Q. Gilbert, Emily Grubert, Varun Rai, Joshua D. Rhodes, Sarang Shidore, Caitlin A. Smith, and Michael E. Webber. Cascading risks: Understanding the 2021 winter blackout in Texas. *Energy research and social science*, 2021. doi:10.1016/J.ERSS.2021.102106.

Torben K. Mideksa and Steffen Kallbekken. The impact of climate change on the electricity market: A review. *Energy Policy*, 2010. doi:10.1016/J.ENPOL.2010.02.035.

L. Suganthi and Anand A. Samuel. Energy models for demand forecasting—A review. *Renewable Sustainable Energy Reviews*, 2012. doi:10.1016/J.RSER.2011.08.014.

Volkan Ş. Ediger and Sertac Akar. Arima forecasting of primary energy demand by fuel in Turkey. *Energy Policy*, 2007. doi:10.1016/J.ENPOL.2006.05.009.

Kristopher T. Williams and Juan D. Gomez. Predicting future monthly residential energy consumption using building characteristics and climate data: A statistical learning approach. *Energy and Buildings*, 2016. doi:10.1016/J.ENBUILD.2016.06.076.

Shankar Chandramowli and Frank A. Felder. Impact of climate change on electricity systems and markets – A review of models and forecasts. *Sustainable Energy Technologies and Assessments*, 2014. doi:10.1016/J.SETA.2013.11.003.

Maximilian Auffhammer, Patrick Baylis, and Catherine Hausman. Climate change is projected to have severe impacts on the frequency and intensity of peak electricity demand across the United States. *Proceedings of the National Academy of Sciences of the United States of America*, 2017. doi:10.1073/PNAS.1613193114.

James DossGollin, D. Farnham, U. Lall, and V. Modi. How unprecedented was the February 2021 Texas cold snap? *Environmental Research Letters*, 2021. doi:10.1088/1748-9326/AC0278.

Pengyuan Shen and Biao Yang. Projecting Texas energy use for residential sector under future climate and urbanization scenarios: A bottom-up method based on twenty-year regional energy use data. *Energy*, 2020. doi:10.1016/J.ENERGY.2019.116694.

Sujan Reddy A., Akashdeep S., Harshvardhan R., and S Sowmya Kamath. Stacking Deep learning and Machine learning models for short-term energy consumption forecasting. *Advanced Engineering Informatics*, 2022. doi:10.1016/J.AEI.2022.101542.

He Zhang, Jingyi Peng, Rui Wang, Mengxiao Zhang, Chang Gao, and Yang Yu. Use of random forest based on the effects of urban governance elements to forecast CO2 emissions in Chinese cities. *Heliyon*, 2023. doi:10.1016/J.HELIYON.2023.E16693.

Anh-Duc Pham, Ngoc-Tri Ngo, Thi Thu Ha Truong, Nhat-To Huynh, and Ngoc-Son Truong. Predicting energy consumption in multiple buildings using machine learning for improving energy efficiency and sustainability. *Journal of Cleaner Production*, 2020. doi:10.1016/J.JCLEPRO.2020.121082.

Benjamin G oehry, Hui Y an, Yannig G oude, §. PascalMassart, and Jean-Michel P oggi. Random Forests for Time Series. 2019.

A. Stillwell, C. King, M. Webber, I. Duncan, and Amy Hardberger. The energy-water nexus in Texas. 2011. doi:10.5751/ES-03781-160102.

Todd Levin, Audun Botterud, W. Neal Mann, Jonghwan Kwon, and Zhi Zhou. Extreme weather and electricity markets: Key lessons from the February 2021 Texas crisis. 2022. doi:10.1016/J.JOULE.2021.12.015.

Andrew E. Dessler. Climate change increases energy demand and cost in Texas. *Texas AM University*, 2024.