Justin Vernieri

Final Project – Forecasting Solar Energy Consumption

**Forecast and Conclusion**

This project forecasts solar energy consumption in the United States using monthly data from 1984 to 2024. The original time series shows a strong upward trend in solar energy use, with seasonal peaks in May, June, and July. Based on the two best models—ETS and ARIMA(2,1,2)(0,1,1)[12]—solar energy use is expected to continue rising over the next two years while following its seasonal pattern. ETS forecasts a peak of ~0.1385 quadrillion BTUs of solar use in June 2025, while ARIMA forecasts ~0.1291 for the same month. These projections reflect solar energy’s growing role in meeting climate goals and enhancing national energy security, underscoring the need for continued investment, policy support, and energy grid modernization.

**Methodology**

Several models were evaluated, including Mean, Naïve, Random Walk, Seasonal Naïve, Holt-Winters, ETS, and ARIMA. Model performance was assessed using MAPE (Mean Absolute Percent Error), which provides easy-to-interpret, scale-independent percentage accuracy. Given that the original data ranges from 0 to ~0.12, MAPE avoids the misinterpretation that can happen with RMSE (e.g., RMSE of 0.02 appears small but is actually large relative to the data scale). In addition to MAPE, residual diagnostics—particularly ACF plots—were also used to assess model fit and randomness in forecast errors.

**Exploration and Insights**

The data, sourced from the U.S. Energy Information Administration (EIA), records and updates monthly solar energy use in Quadrillion British Thermal Units (BTUs) from January 1984 to December 2024. I created a focused window from 2010 to 2024 to isolate and capture recent trends and seasonal dynamics. The autocorrelation plot (ACF) of the original data confirmed strong temporal relationships, validating the need for time series modeling. It also showed upticks in lags 12 and 24, supporting the presence of seasonality. Decomposition further showed a dominant trend component (0-0.10), modest but meaningful seasonality (±0.01), and residuals (±0.02) that increased after 2020, likely due to structural changes, external shocks, and increasing solar installation after COVID. In exploring the different models, ETS and ARIMA emerged as the best. ETS had the best MAPE of ~3.53%, while ARIMA had slightly higher MAPE of ~3.70% but cleaner residuals and less autocorrelation.

**Limitations and Recommendations**

While both ETS and ARIMA performed well, they have limitations. ETS assumes stable patterns over time, which may not hold true under rapid changes in energy tech or policy. ARIMA, while powerful, assumes linearity and is univariate, missing outside factors like solar panel adoption rates or energy incentives. Both models could be improved by incorporating external variables or re-tuning parameters. Future analysis could involve more rigorous residual checks to detect lingering autocorrelation or model misspecification, as well as an in-depth exploration of the confidence intervals for each forecast. Testing hybrid models such as ARIMAX or dynamic regression would also improve robustness and adaptability.