Project Title: London Smart Meter Forecasting Report

Group Member:

Rui Chen

Xinran Chen

Yinqi Liu

Date of Submission: 8/8/2025

Content

Abstract

This project analyses household electricity usage patterns using the London Smart Meters dataset, which records high-frequency electricity consumption data for over 5,000 London households between November 2011 and February 2014. To meet the course requirement for univariate time series analysis, we aggregated the original multivariate data into daily average electricity consumption series and constructed a stationary time series for modelling and analysis.

This project applied various time series analysis methods introduced in the course, including ADF and KPSS stationarity tests, ACF/PACF autocorrelation diagnostics, and spectral analysis to identify periodic structures. We constructed and compared various models, including AR, MA, ARMA, ARIMA, SARIMA, and linear regression models with trend terms and weekday components. We also introduced auto.arima() as a baseline model.

Model performance was evaluated using metrics such as AIC, BIC, RMSE, MAE, and a residual white noise test. The results showed that the SARIMA model with weekly seasonality performed best. This project deepened our understanding of time series analysis and highlighted the importance of incorporating known seasonal structures into modeling to improve forecast accuracy.

Introduction

As people pay more and more attention to energy efficiency and sustainable development, accurately predicting household electricity consumption is becoming increasingly crucial for energy suppliers and urban planners. This project uses the dataset of smart meters in London to analyse and model household electricity consumption patterns. This dataset records the energy consumption (in kilowatt-hours) of London households every half hour from November 2011 to February 2014. Our aim is to make reliable short-term predictions by exploring and comparing different time series models, while gaining a deeper understanding of the temporal variation patterns of residential electricity demand.

We selected a representative household electricity consumption time series and summarized the data into daily frequencies. Such preprocessing enables us to capture more obvious consumption trends and seasonal patterns, especially the common weekly cycle patterns in household electricity usage. The processed time series contains 48 daily observations and is highly suitable for analysis using time series modelling techniques such as ARIMA, SARIMA, and regression-based methods.

The main objective of this project is to evaluate and compare the performance of various prediction models, including manual and automatic ARIMA models, seasonal ARIMA models based on spectral analysis, and time series linear regression models. We will use conventional statistical indicators such as RMSE, AIC, and BIC, as well as the Ljung-Box test of residual autocorrelation, to evaluate the fitting effect and predictive ability of the model. In addition, we will also combine time-domain and frequency-domain technologies to make the selection and interpretation of the model more based on evidence.

The structure of this report is as follows: Section 4 introduces the dataset and elaborates on the preprocessing work done to convert the original.TSF format into usable daily sequences; Section 5 Outlines the methods used, including stationarity tests, ACF/PACF diagnosis, and spectral analysis; Section 6 presents the results of model fitting and prediction; Section 7 discusses the performance, interpretation and challenges encountered of the model; Finally, Section 8 summarizes the main findings and reflects on the modelling process.

Data Description

The dataset used for this project is the London Smart Meters dataset, sourced from the Monash Time Series Forecasting Repository. This dataset, which belongs to the energy sector, contains electricity usage data from 5,560 smart meters. The dataset provides electricity usage readings from each smart meter over 39,648 half-hourly time points, sampled 288 times daily, with a half-hourly granularity. This dataset is univariate and does not include competition background. You can choose between officially provided versions with and without missing values.

This data was collected between November 2011 and February 2014 as part of the Low Carbon London project, led by UK Power Networks. The project aims to assess the potential of smart meters to reduce carbon emissions and improve consumer electricity usage. The project, a key component of the UK's response to EU climate policy and the renewal of its energy infrastructure, covers real-world electricity usage data from over 5,000 London households.

The original data is structured as a wide table (each column represents a user, and each row represents a time point). To adapt this data for univariate time series analysis, we constructed a series based on daily average electricity consumption, covering approximately 800 days, by shaping, aggregating, and removing missing values. The resulting time series has daily observations and was converted into a ts object with weekly seasonal frequency (frequency = 7), and observations with missing values were removed.

After data processing, we also constructed time variables (t) and weekday variables (weekday) for linear trend modelling and seasonality modelling. Although the original data provides weather data information, these additional variables are not included in this project to meet the requirements of univariate analysis in the course.

Methodology

This section details the process of constructing a univariate time series model from raw London smart meter data, applying modeling techniques, and evaluating forecasting performance. All analyses were performed using the R.

1. Data Preprocessing

The original dataset was provided in wide format, with each column representing electricity consumption from an individual household and rows corresponding to half-hour intervals. To align with the univariate analysis requirement, the data was reshaped using pivot_longer() to a long format containing three key columns: date, time, and consumption.

After converting the time column into a proper HMS object, we created a complete timestamp （datetime = date + time） and removed all missing values to ensure continuity. The data was then aggregated by date to calculate the average daily consumption across all households. This resulted in a single time series with one observation per day.

The final daily time series was converted into a ts object in R, using frequency = 7 to reflect weekly periodicity. This frequency assumption is further supported by the spectral analysis discussed below.

2. Stationarity Tests

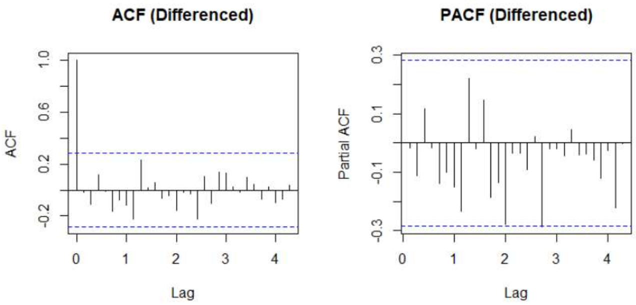
To determine the need for differencing, we conducted two standard tests:

- **Augmented Dickey-Fuller (ADF) Test:** A p-value of 0.325 indicated non-stationarity.
- **Kwiatkowski-Phillips-Schmidt-Shin (KPSS) Test:** A p-value of 0.10 suggested marginal stationarity.

Given these results, we computed the optimal number of differences using ndiffs() and nsdiffs(), which provided the values of d (non-seasonal differences) and D (seasonal differences) required for ARIMA/SARIMA modelling.

3. ACF and PACF Analysis

According to the above differencing requirements, the time series is differentiated and the ACF an Based on the stationary test results, we first perform a non-seasonal difference on the original daily mean series to remove the linear trend (d=1), and then perform a 7-day seasonal difference to eliminate the weekly cycle effect (D=1). Subsequently, we plot the autocorrelation function (ACF) and partial autocorrelation function (PACF) of the double difference series. From the ACF plot, we can see that there is a significant peak at lag 1, which then decays rapidly to within the confidence interval, indicating that there may be a non-seasonal moving average component (q ≈ 1 or 2). At the same time, the PACF plot also shows obvious peaks at lags 1 and 2, suggesting that there may be a non-seasonal autoregressive component (p ≈ 2). Therefore, we manually specified and tested the SARIMA model with a parameter configuration of (2, d, 2) (2, D, 2) [7] to capture both short-term dependence and weekly periodic structure. The applicability of the model was further verified by the AIC/BIC criterion and residual diagnostics.



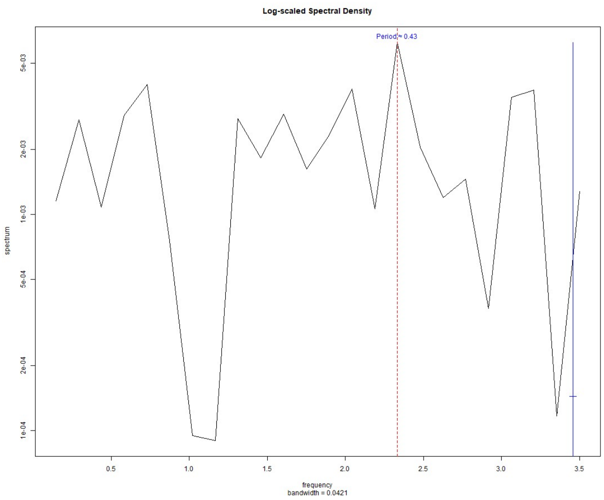
4. Spectral Density Analysis

We perform a detailed spectral density analysis[1] on the original daily mean series using R's spectrum(log="yes") function, stabilizing the energy at different frequencies on a logarithmic scale[2]. The periodogram exhibits a dominant peak near 0.14 cycles/day, indicating a strong weekly seasonality in the data. We experimented with different smoothing bandwidths (e.g., span=5) to verify the robustness of the primary and secondary peaks. The weekly pattern is consistent with literature showing distinct differences between weekday and weekend electricity usage. We retained a 7-day seasonal period in the SARIMA model (seasonal.period=7). Incorporating a 7-day seasonal period not only improves the model's interpretability but also ensures that key weekly cyclical features are effectively captured.

---

[1] Hyndman, R. J., & Athanasopoulos, G. (2021). Forecasting: Principles and Practice. 3rd ed. https://otexts.com/fpp3/

[2] Shumway, Robert H and David S Stoffer, Time Series Analysis and Its Applications : With R Examples (Springer Nature Switzerland, 5th ed. 2025., 2025)



5. Time Series Modelling

We explored both manually specified and automatically selected models:
Manual Model:

- AR (2), MA (2), ARMA (2,2): to assess basic model structures
- ARIMA (1, d,1): based on differencing tests
- SARIMA (2, d,2) (2, D,2) [7]: incorporating seasonality

Automatic Model:

- auto. Arima(): used to select the best model via AICc minimization with exhaustive search (stepwise = FALSE, approximation = FALSE)

Spectrum-guided SARIMA:

- Construct a SARIMA model using the period obtained from spectral analysis as seasonal. Period; if spectral fitting fails, we fall back to a default SARIMA(2,d,1)(2,D,1)[7] configuration.

6. Time Series Regression

To enhance model diversity, we also built a linear regression model:

daily_consumption ~ t + weekday

Here, t represents the time trend term, and weekday represents the day of the week （factor variable） . This model is highly interpretable and serves as a baseline for other models.

We use a constructed data frame of dates for the next 30 days, use predict () to make predictions and plot the prediction curves for comparison.

7. Model Diagnostics

All models were evaluated using:

- **AIC and BIC:** Used to compare the balance between model fitting effect and complexity. The smaller the two information criteria, the better the model. They are particularly suitable for selecting models with fewer parameters but better fitting.
- **RMSE and MAE:** They measure the point prediction error of the model respectively. RMSE is more sensitive to larger errors, while MAE provides a more stable average error assessment, which helps to compare the prediction accuracy of each model.
- **Ljung-Box Test**: Used to determine whether the model residual is white noise. If the test passes, it means that the model is well fitted and there is no missing autocorrelation structure.

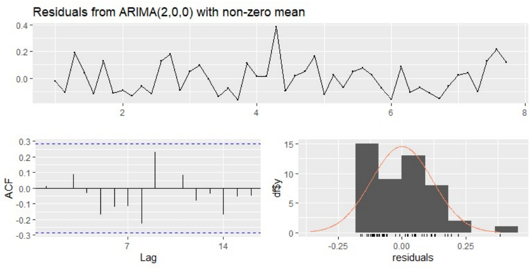
8. Forecasting

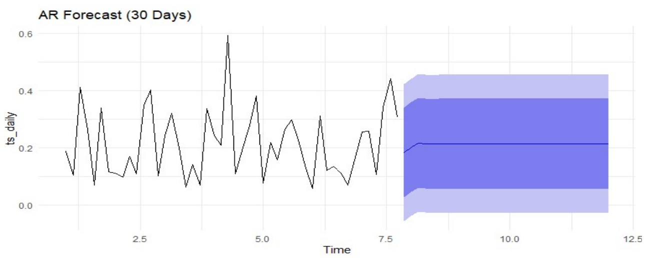
We generated 30-day forecasts for each model and plotted them using autoplot() , including confidence intervals. The visualization shows that the SARIMA model outperformed the other models in terms of RMSE and MAE, achieving the best performance. While auto.arima() had a clear advantage in AIC/BIC, it lagged slightly behind in residual tests.

Results

AR Model

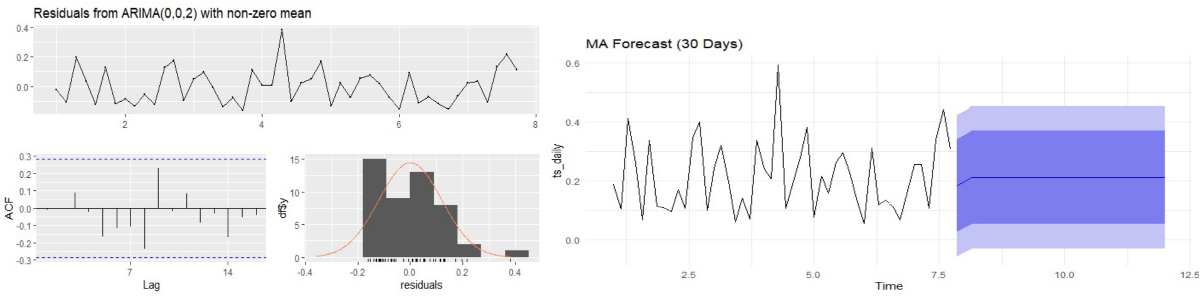
The AR(2) model performs quite well in data fitting. It can capture potential temporal dynamics without overfitting. The residuals fluctuate randomly around zero, indicating that the model has explained all the obvious structures in the data. The autocorrelation function graph shows no significant autocorrelation. The P-value of the Ljung-Box test is 0.2635, which also confirms this point, indicating that the residuals are close to white noise. The histogram of the residuals roughly conforms to the normal distribution, indicating that the assumed conditions of the model have been satisfied. The 30-day prediction results show a stable trend, and the prediction range is relatively narrow, indicating that it has reliable short-term prediction capabilities. Overall, for this data, AR(2) is a reasonable and easily interpretable model.





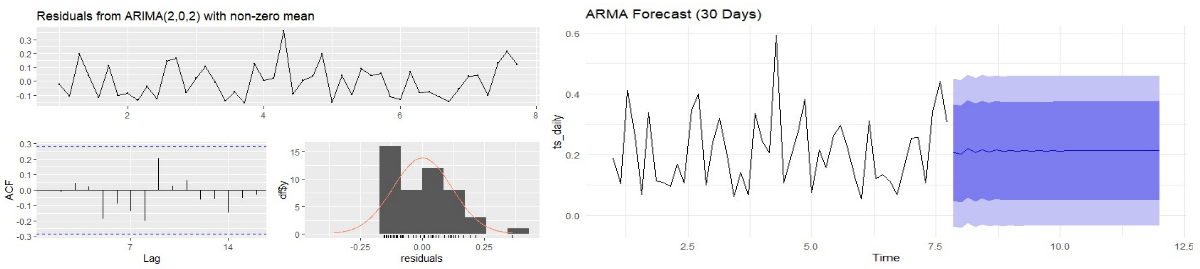
MA Model

After the MA(2) model was fitted, the residual distribution was approximately normal. The ACF plot did not show significant autocorrelation. The P-value of the Ljung-Box test was 0.269 (> 0.05), indicating that the residual was white noise and the model was well-fitted. In the 30-day forecast chart, the predicted values tend to be stable, and the confidence intervals are relatively reasonable, indicating that the model has a certain predictive ability for future trends.



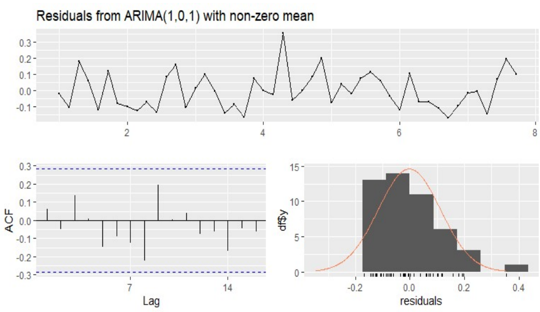
ARMA Model

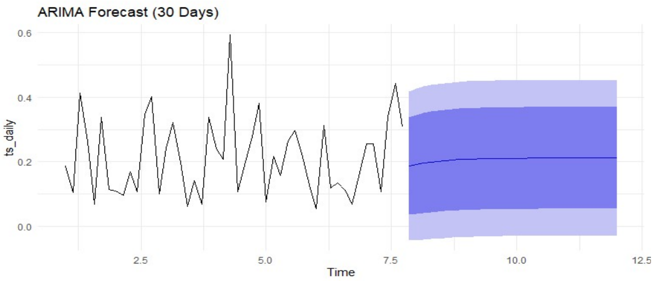
The residuals of the ARMA(2,0,2) model do not closely revolve around the zero value distribution and fluctuate significantly. The autocorrelation function graph has a small peak at the 6-period lag. Although most of the lag terms are within the confidence interval, the autocorrelation has not been completely eliminated. The histogram of the residuals also deviates from the ideal normal distribution, showing a right-skewed feature. In addition, in the 30-day prediction results, the confidence interval has been continuously widening, indicating that uncertainty is increasing over time. This means that although ARMA(2,0,2) can capture some short-term dynamics, it may not be the most robust model for this time series.



ARIMA Model

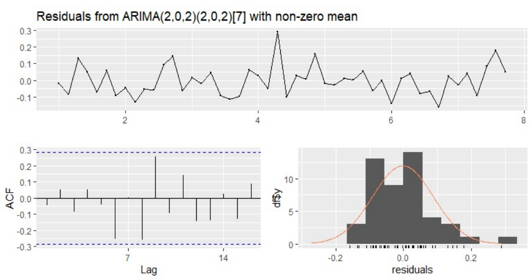
The ARIMA(1,0,1) model was used to model the stabilized data. The residual plot of the model shows an overall random distribution without significant autocorrelation. The P-value of the Ljung-Box test was 0.3211, which was higher than 0.05, indicating that there was no significant autocorrelation in the residuals and the model fit well. The 30-day forecast chart shows that the future power load is tending to stabilize, with a moderate confidence interval and relatively low prediction uncertainty. Comprehensive assessment indicates that the ARIMA model can better capture the fluctuation characteristics of non-seasonal time series.

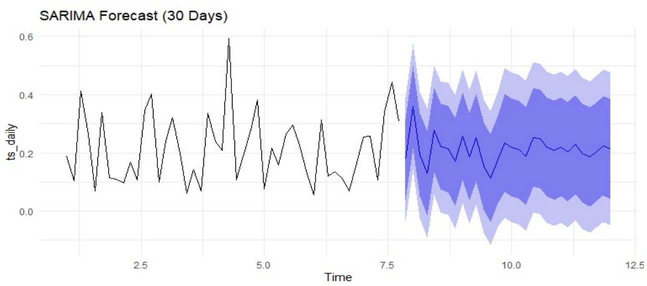




SARIMA Model

The SARIMA model is set as ARIMA(2,0,2)(2,0,2)[7]. The residual plot shows that it is roughly stable and the distribution is close to normal, but in the ACF plot, some lags still have significant autocorrelation. The P-value of the Ljung-Box test was 0.0023, indicating significance that there was still structural information in the residuals and the model did not fully capture the sequence features. Despite this, the 30-day forecast chart shows a relatively sensitive portrayal of the trend, with a moderate confidence interval and certain practicality.

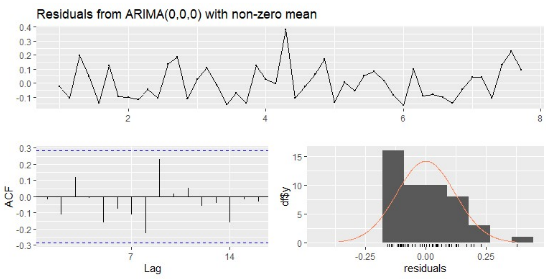


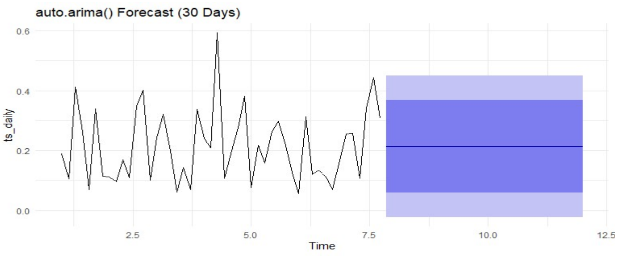


auto.ARIMA Model

auto.arima() automatically selects the simplest arima(0,0,0) model, meaning that the sequence behaves approximately as white noise. The residual plot and ACF plot did not show significant patterns. The P-value of the Ljung-Box test was 0.4189, indicating that there was no significant autocorrelation in the residuals. Although the model structure is extremely simple, due to automated seasonal and trend detection, as implemented in the `forecast` package in R[3], the prediction graph presents as a horizontal straight line with a relatively wide confidence interval, indicating that its predictive ability is not good.

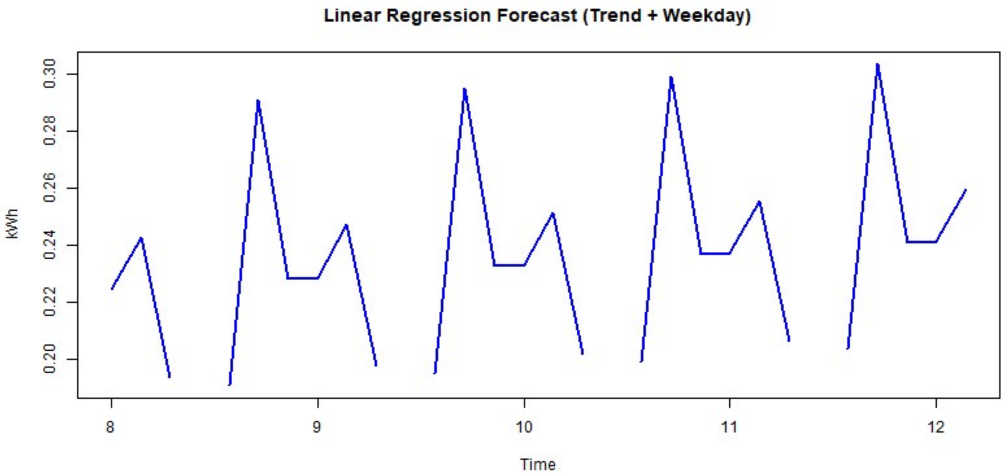
[3] R Core Team. (2025). *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing. https://www.r-project.org





Time Series Regression Model

We fitted a time series regression model, which is interpretable and flexible, but tends to underperform in short-term accuracy compared to SARIMA model[4], incorporating linear trend components and working day indicators, to simulate daily electricity consumption. The residuals of the model are randomly dispersed and have no obvious autocorrelation. The prediction results clearly reveal the potential weekly seasonality - peaks and troughs will recur within a fixed 7-day cycle. In addition, a clear upward trend can be observed, indicating that electricity consumption will increase over time.



Compare

We compared seven different time series models using four evaluation metrics: Root Mean Squared Error (RMSE), Mean Absolute Error (MAE), Akaike Information Criterion (AIC), and Bayesian Information Criterion (BIC). The results are as follows:

Model	RMSE	MAE	AIC	BIC
AR	0.12	0.10	-60.42	-52.93
MA	0.12	0.10	-60.43	-52.94
ARMA	0.12	0.10	-57.29	-46.06
ARIMA	0.11	0.09	**-63.28**	**-55.79**
SARIMA	**0.09**	**0.07**	-59.54	-40.82

4Cowpertwait, P. S., & Metcalfe, A. V. (2009). Introductory Time Series with R.

Model	RMSE	MAE	AIC	BIC
auto.arima()	0.12	0.10	-63.75	-60.00
TS_Regression	0.12	0.10	-55.26	-40.29

From this table, we can see that:

The ARIMA model performs best on AIC and BIC, indicating that the model has a good balance between fitting ability and complexity.

The SARIMA model is superior in terms of prediction accuracy (RMSE and MAE).

Although the auto.arima() model automatically selects parameters, it is slightly inferior to the manually set ARIMA in terms of fitting metrics.

The performance of the TS_Regression model is relatively weak, indicating that simple linear trends and day-week factors cannot fully capture the structural features in the data.

Discussion

Our project mainly focuses on the time series modelling of the average daily electricity consumption. It adopts multiple models including AR, MA, ARMA, ARIMA, SARIMA, auto.arima and time series regression for modelling and prediction. The model performance was systematically evaluated through indicators such as RMSE, MAE, AIC, and BIC.

Based on the accuracy index, the SARIMA model performs the best. The RMSE (0.09) and MAE (0.07) indicate that its prediction error is the smallest. ARIMA also has relatively low values on AIC and BIC, demonstrating a good degree of model fitting and relatively good balance. Although the AR, MA and ARMA models have relatively simple structures, they are slightly inferior in prediction accuracy. The TS Regression model has an advantage in interpretability and can clearly reflect trends and periodicity, but it is slightly weaker than the SARIMA and ARIMA models in overall accuracy.

From the residual plots and the results of the Ljung-Box test, it can be seen that the residual sequences of most models do not show significant autocorrelation, which is in line with the white noise hypothesis. Take AR (2) as an example. Its mean residual is close to 0, and the P-value of the Ljung-Box test is 0.26, indicating that the autocorrelation structure of the original sequence has been basically eliminated after fitting. Although the SARIMA model has a complex structure, it can simultaneously capture trends and seasonality. Its residual distribution is closest to normal, and the confidence interval is relatively narrow, demonstrating high predictive stability. In contrast, although auto.arima automatically selects the model order, its prediction volatility is slightly greater and may not be superior to manually set models.

During the data analysis process, the first thing to do is to convert the data format. The original data does not have a unified time specification and needs to be unified into the same format. When the time series was initially created, the initial code could not run because of improper frequency settings. Because the data in the dataset is recorded every 30 minutes, the initial frequency setting was 48. However, after a detailed analysis of the data, it was decided to use the daily means for forecasting and analysis to solve the problem of too many values and missing values, so the frequency was finally set to 7. When setting the model, because the parameters needed to be set manually, many different values were tried. When the parameters were set too small, the prediction was not very accurate, but if the settings were too large, it would lead to overfitting. In the end, after many attempts, we determined to use these parameters to build the model, which gives us a good fit and predicts results.

Over all, different models have their own advantages and disadvantages. If the primary goal is prediction accuracy, SARIMA is currently the best choice.

Conclusion

This project systematically completed a complete analytical pipeline, from raw data processing to time series modelling and forecasting, using real-world electricity consumption data from the London Smart Meters dataset. Through appropriate data cleaning, transformation, and modelling steps, we successfully transformed a high-frequency, high-dimensional, multivariate dataset into a clean, interpretable daily univariate time series, enabling us to conduct forecasting and inference analysis.

This project applied several core time series analysis methods taught in MATH5845, including ADF and KPSS stationarity tests, correlation diagnostics using ACF and PACF plots, spectral density analysis for identifying periodicity, and the construction and comparison of various models, such as AR, MA, ARMA, ARIMA, and SARIMA. Furthermore, we used a time series regression model based on a weekly factor and a trend term, using the auto.arima() automatic model as a baseline.

We compared the performance of various models using multiple evaluation metrics, including AIC, BIC, RMSE, MAE, and the Ljung-Box residual white noise test. The results showed that SARIMA models with weekly seasonality (e.g., (2,d,2)(2,D,2)[7]) performed best in terms of both forecast accuracy and residual properties, indicating that incorporating known cyclical features into modelling is an important way to improve model performance.

This project not only deepened our understanding of time series modelling theory but also enhanced our practical skills in handling missing values in real-world data,

selecting aggregation granularity, and evaluating models. The entire workflow—from exploratory data analysis to model validation—demonstrated how classical statistical approaches remain powerful tools for interpreting structured temporal data.
Future work could include introducing external variables such as weather data and demographic classifications (e.g., ACORN), exploring the use of Box-Cox transformation to stabilize variance, and further investigating more advanced forecasting methods such as exponential smoothing (ETS) and machine learning-based models to improve forecast performance.

Reference

1. Hyndman, R. J., & Athanasopoulos, G. (2021). Forecasting: Principles and Practice. 3rd ed. https://otexts.com/fpp3/
2. Shumway, Robert H and David S Stoffer, Time Series Analysis and Its Applications : With R Examples (Springer Nature Switzerland, 5th ed. 2025., 2025)
3. R Core Team. (2025). R: A Language and Environment for Statistical Computing. R Foundation for Statistical Computing. https://www.r-project.org.
4. Cowpertwait, P. S., & Metcalfe, A. V. (2009). Introductory Time Series with R.