

Short-term Gas Demand Forecasting

Keyang Liang 2989789L

Contents

1	Introduction	2
2	Data and Exploratory Analysis	2
3	Methodologies	7
4	Formal Analysis	8
5	Conclusions	15
	Reference	16

1 Introduction

Natural gas is an important energy resource, which is used across all regions and in various sectors, including industry and electricity generation (IEA, 2025). To satisfy its demands, a sufficient and economic supply is important, and requires accurate forecasts.

Multiple factors can influence gas demand and forecast accuracy, including temperature, wind speed, seasonal activities, and special events.

This study aims to forecast natural gas demands in Scotland from several days ahead, evaluate prediction intervals, and furthermore determine what factors affect forecast uncertainty.

2 Data and Exploratory Analysis

2.1 Data Source

Historic natural gas demand data and weather data for Scotland from 2021 to 2024 were retrieved from National Gas, which is the owner of the National Transmission System that transports natural gas across Great Britain (National Gas, n.d.). Bank holiday data were taken from UK government API (UK Government, n.d.).

Data from National Gas are partially shown in Table 1. Column **Applicable For** refers to the date to which the observation applies. Column **Quality Indicator** indicates quality of the observation, where value *A* means amended, and value *L* means late.

Table 1: Part of the Data

Applicable For	Data Item	Value	Quality Indicator
31/12/2024	Demand Actual, LDZ (SC), D+6	13.34	
30/12/2024	Demand Actual, LDZ (SC), D+6	13.18	
29/12/2024	Demand Actual, LDZ (SC), D+6	12.60	
28/12/2024	Demand Actual, LDZ (SC), D+6	12.11	
27/12/2024	Demand Actual, LDZ (SC), D+6	12.16	
31/12/2024	Composite Weather Variable, Actual, LDZ(SC), D+1	5.60	

Data item *Demand Actual, LDZ (SC), D+6* refers to natural gas demand for the Local Distribution Zone of Scotland for the given date, which is measured in mega standard cubed metre (MSm³).

Figure 1 plots natural gas demand against date, and highlights observations with quality indicator labels. It can be observed that the data exhibit annual seasonality, and several outliers locate out of the pattern. To be specific, late data items with value 5.4, and amended data items were considered as outliers, and would be removed in further analysis.

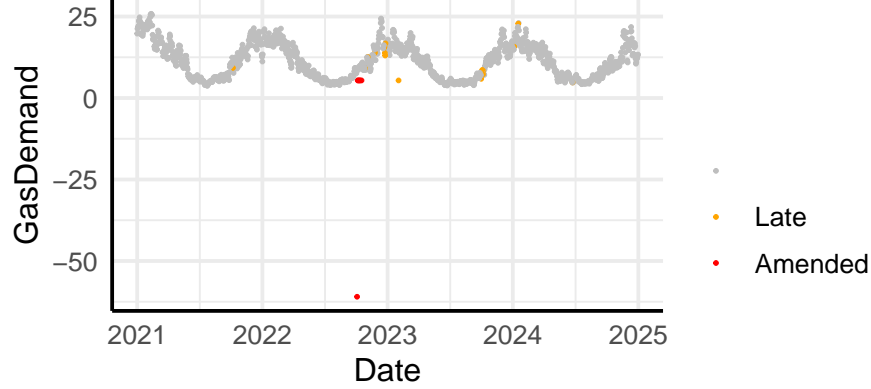


Figure 1: Scatter Plot of Gas Demand against Date

Data item *Composite Weather Variable, Actual, LDZ(SC), D+1* refers to composite weather variable (CWV) for the Local Distribution Zone of Scotland for the given date, which is calculated from temperature and wind speed.

Figure 2 plots composite weather variable against date. No obvious outliers are observed through the figure.

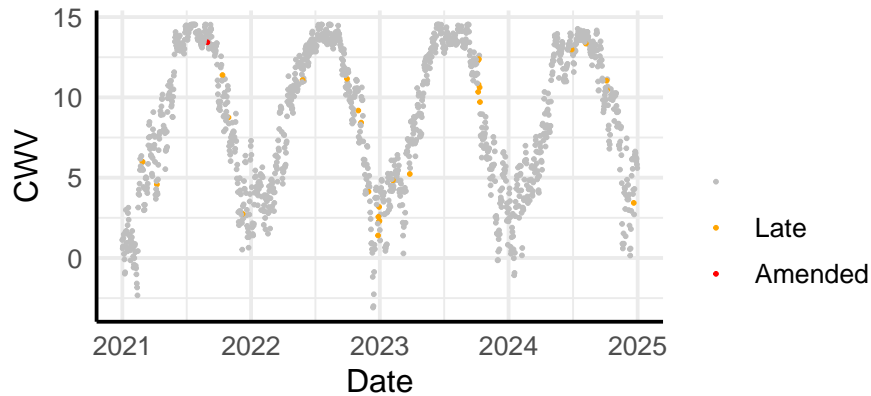


Figure 2: Scatter Plot of Composite Weather Variable against Date

2.2 Univariate Analysis

After removing the outliers, further exploratory analysis were applied to describe the data.

Figure 3 shows the trends and distribution of natural gas demand. Overall, there is an decreasing trend over time. Also, an annual seasonality exists in the data, such that gas demand increases from August to January, and decreases from February to July.

Furthermore, it can be observed that the variance in gas demand turns large in winter days, and turns small in summer days. As for the density plot, the distribution is bimodal and right-skewed, with the right mode smaller than the left mode. In summary, these discoveries indicate that a transformation to the data is plausible, in order to make the distribution better for modelling.

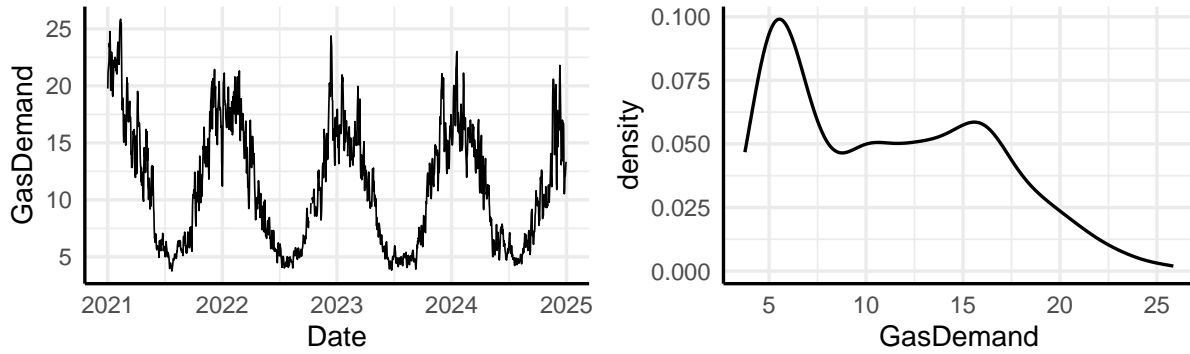


Figure 3: Plots of Gas Demand

Figure 4 describes the log-transformed natural gas demand data. The variance turns stable, and the distribution is less skewed.

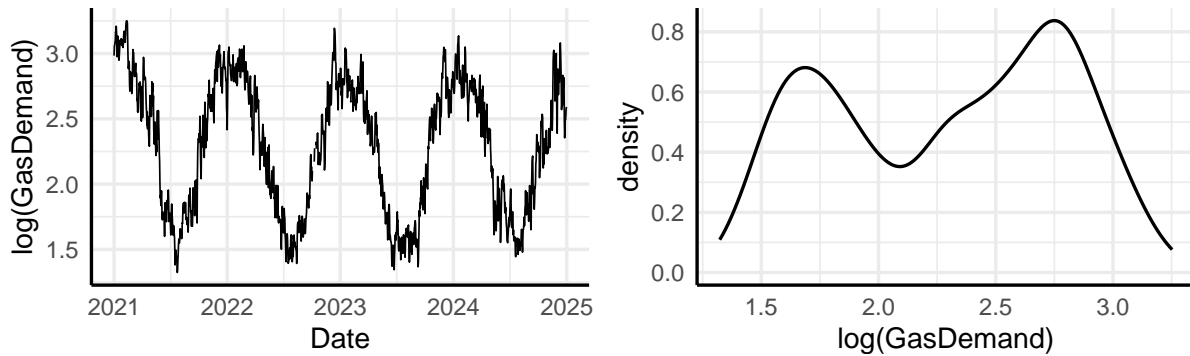


Figure 4: Plots of log-transformed Gas Demand

As for the composite weather variable, Figure 5 shows its seasonality and distribution. It increases from February to July and decreases from August to January annually, and has larger variance in winter days and smaller variance in summer days. The distribution is bimodal and left-skewed, with the left mode smaller than the right mode.

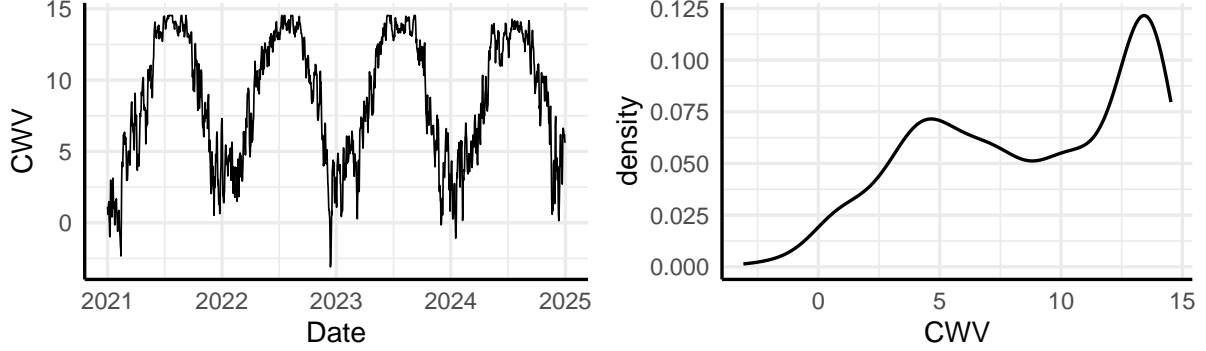


Figure 5: Plots of Composite Weather Variable

2.3 Multivariate Analysis

After univariate analysis, relationship between variables were evaluated.

Figure 6 shows a strong linear relationship between natural gas demand and composite weather variable. Also, it can be observed that when CWV becomes smaller, the variance in natural gas demand becomes larger.

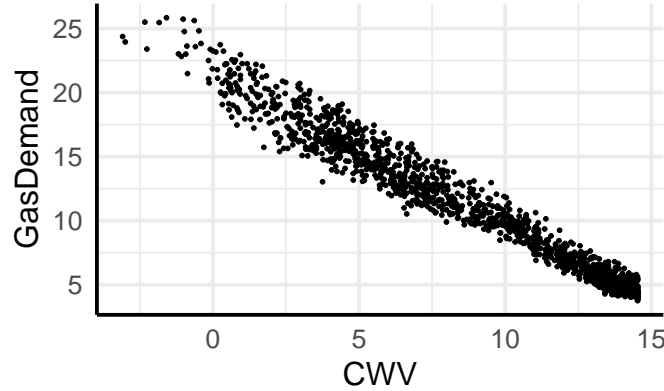


Figure 6: Scatter Plot of Gas Demand against Composite Weather Variable

Since this study aims to give prediction of natural gas demand several days ahead, the weather data on the date in question would not be available when making prediction.

Therefore, evaluating the relationship between composite weather variable several days before and natural gas demand would be an approach.

Figure 7 illustrates the relationship between CWV several days ahead and natural gas demand. There exists a negative linear relationship through all the plots. However, as the date gap becomes larger, the variance in natural gas demand becomes larger.

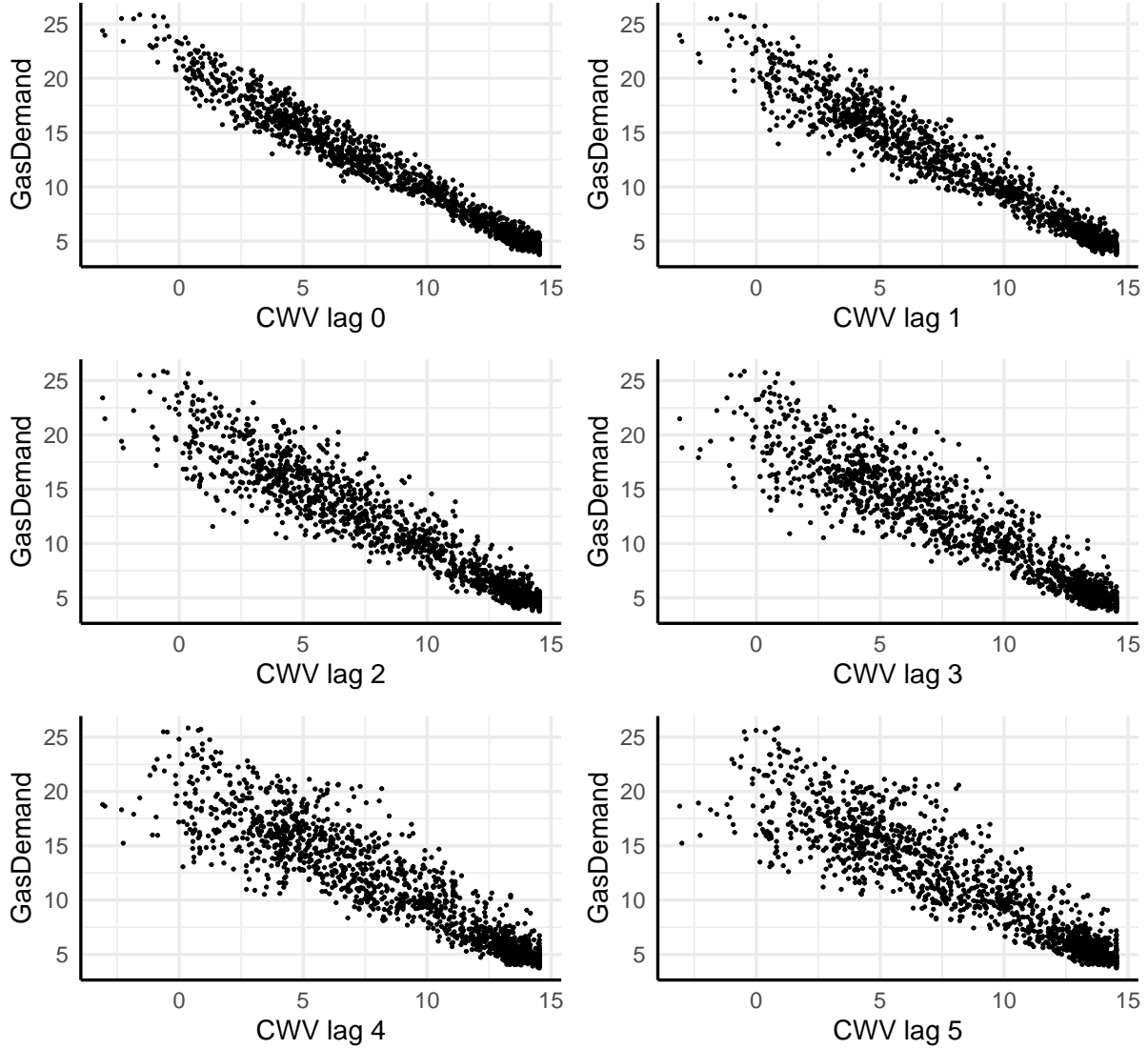


Figure 7: Plots of Gas Demand against lagged Composite Weather Variable

Holiday is also an important factor affecting natural gas demand. Figure 8 shows the differences in natural gas demand levels between each day of the week. The gas demand

level in weekends are lower than that in weekdays. Therefore, it is plausible to combine weekends and weekdays respectively, and transform day of the week into a binary variable. In further analysis, weekends and bank holidays would be categorized as holidays, with the remaining categorized as non-holidays.

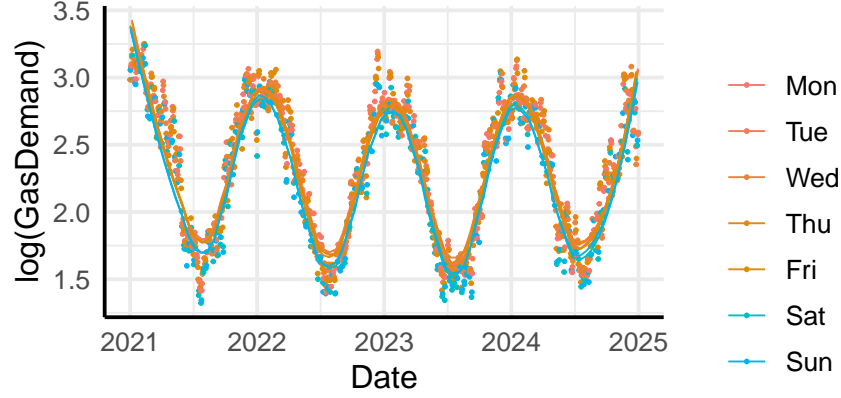


Figure 8: Plot of Gas Demand against Date by Week Day

The interaction of composite weather variable and holiday was also considered. Figure 9 shows that the linear effect of CWV in holidays is smaller than that in non-holidays.

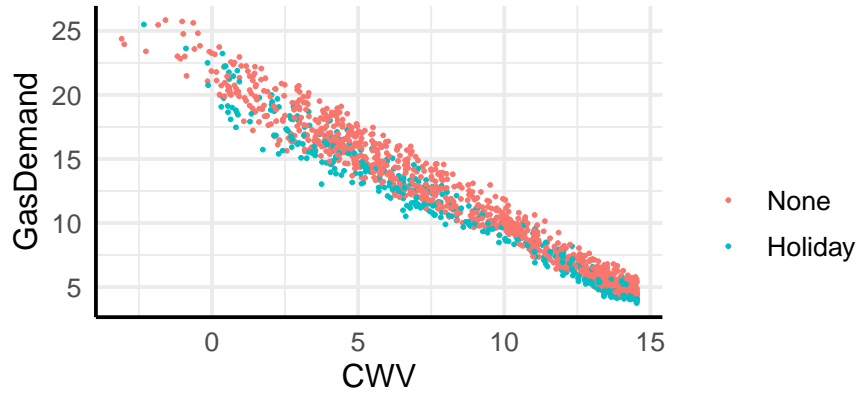


Figure 9: Scatter Plot of Gas Demand against Composite Weather Variable by Holiday

3 Methodologies

Generalised Additive Models (GAM) and Autoregressive Integrated Moving Average (ARIMA) models were used in this study.

GAM extends the generalised linear models, by replacing the linear term by an additive smooth function (Hastie and Tibshirani, 1986). It provides more flexibility for capturing non-linear relationship between explanatory variables and response variable. In the case of this study, it helps identifying the non-linear effects of date on the log-transformed natural gas demand.

Model assumption checks followed model fitting, where residual independence, zero-mean and constant variance assumptions were checked.

ARIMA model is a time series model that can model the trend, seasonal variation, and correlation simultaneously (Anon, 2022). It is a composition of autoregressive process, moving average process, and differencing, hence it can better fit the data with more flexibility. By checking autocorrelation plots of the residuals, the parameters used in ARIMA model were determined.

Bootstrap methods were applied in calculating the prediction intervals (Hyndman and Athanasopoulos, 2021).

4 Formal Analysis

4.1 Model Fitting and Model Checking

As is shown in Section 2, there exists a seasonal trend in log-transformed natural gas demand. The trigonometric function $\sin(\frac{2\pi}{365}(x + \phi))$ would be used in modelling, where x is the date index, and ϕ is the phase shift.

Since

$$\sin(\frac{2\pi}{365}(x + \phi)) = \sin \frac{2\pi}{365} \phi \cdot \cos \frac{2\pi}{365} x + \cos \frac{2\pi}{365} \phi \cdot \sin \frac{2\pi}{365} x,$$

the phase shift can be approximately determined by fitting the linear model

$$y_i = \beta_0 + \beta_1 \cos \frac{2\pi}{365} x_i + \beta_2 \sin \frac{2\pi}{365} x_i + \varepsilon_i, \quad (1)$$

where

y_i is log-transformed natural gas demand,

x_i is date index,

$\varepsilon \sim N(0, \sigma^2)$ independently.

Figure 10 shows the results of Equation 1. The estimated coefficients are $\hat{\beta}_1 = 0.566$ and $\hat{\beta}_2 = 0.394$. Therefore, the phase shift ϕ is 56 days to the left of 2021 Jan 1st.

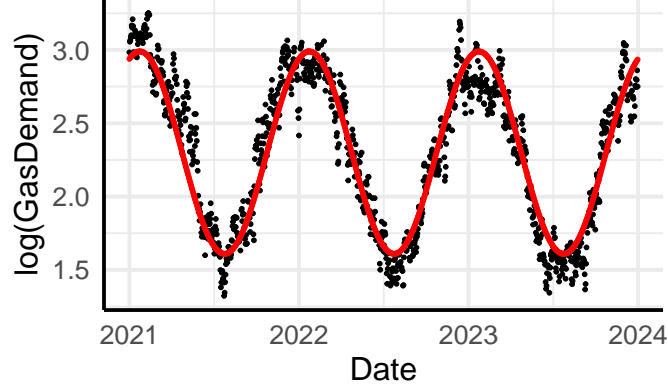


Figure 10: Scatter Plot of Fitted Values in Equation 1

The generalised additive model

$$g(E[y_i]) = \beta_0 + f_1(x_{1i}) + f_2(x_{2i}) + f_3(x_{3i}) + \beta_4 x_{4i} \quad (2)$$

was applied, where

g is the identity link function,

y_i is log-transformed natural gas demand,

β_i 's are intercepts,

f_i 's are smooth functions,

x_{1i} is date index,

x_{2i} is trigonometric function $\sin(\frac{2\pi}{365}(x_{1i} + \phi))$ such that $\phi = \pi/2$,

x_{3i} is lag 7 composite weather variable,

x_{4i} is indicator function of holiday.

After fitting Equation 2, model assumption checks were carried out.

Figure 11 shows that residuals are evenly scattered around zero, and have constant variance across all levels of the fitted values.

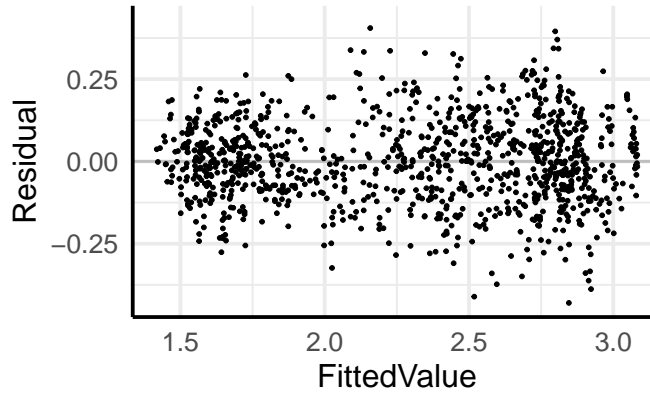


Figure 11: Scatter Plot of Residual against Fitted Value in Equation 2

In Figure 12, the points fall closely along the diagonal line. Therefore, it is safe to assume that the residuals are normally distributed.

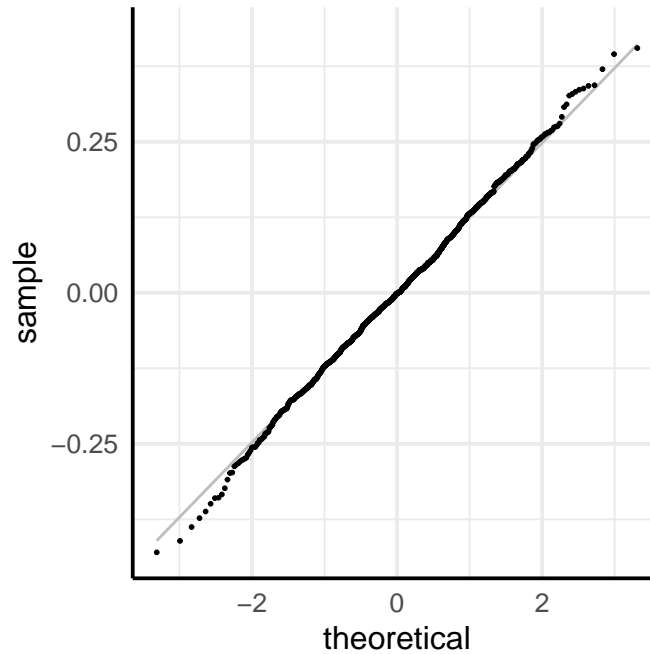


Figure 12: Quantile-Quantile Plot of Residual in Equation 2

Figure 13 displays the autocorrelation within the residuals. The ACF declines gradually, and the PACF is significantly large until lag 2. These results indicate the residuals are correlated, and potentially follow an autoregressive process of order 2.

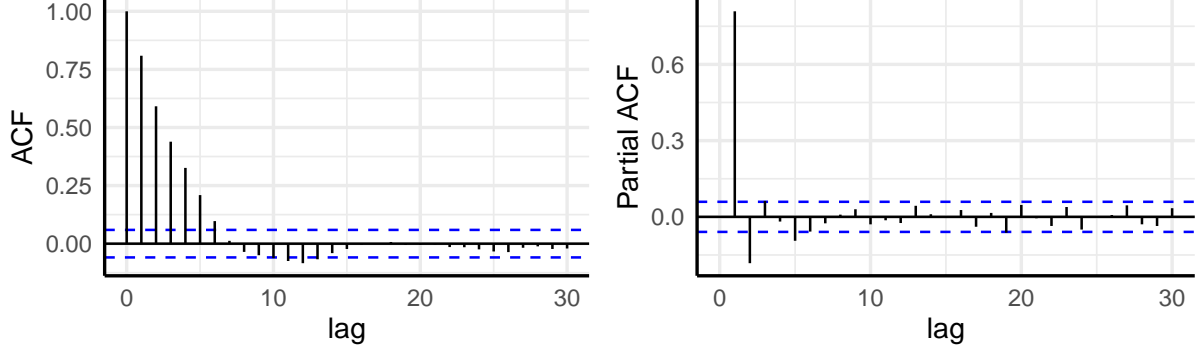


Figure 13: Autocorrelation Plots of Residual in Equation 2

According to above discussion, the AR(2) model

$$\varepsilon_t - \alpha_1 \varepsilon_{t-1} - \alpha_2 \varepsilon_{t-2} = Z_t, \quad (3)$$

is applied to the residuals of Equation 2, where

ε_t is the residual at date index t ,

α_i 's are coefficients,

$Z_t \sim N(0, \sigma^2)$ independently.

The model results of Equation 3 are $\hat{\alpha}_1 = 0.934$ and $\hat{\alpha}_2 = -0.162$.

The fitness of the model is evaluated by Figure 14. Overall, the ACF and PACF are small, although some of them slightly exceed the confidence interval. This indicates that the residuals are uncorrelated.

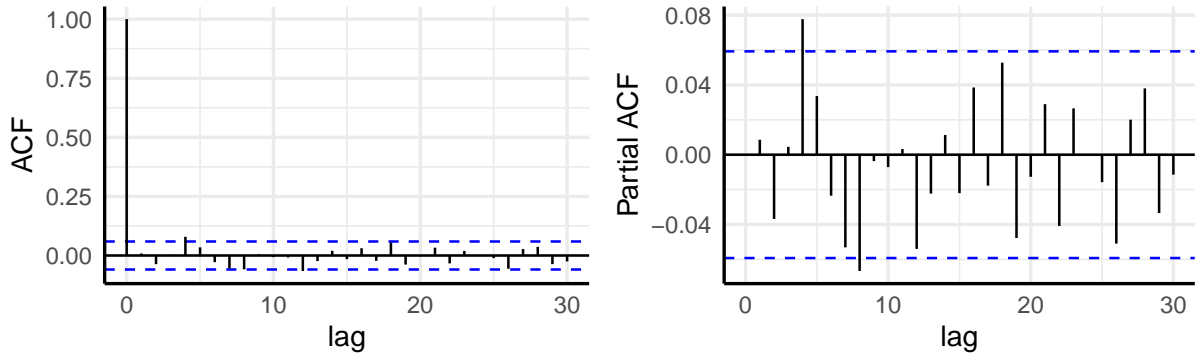


Figure 14: Autocorrelation Plots of Residual in Equation 3

4.2 Point Estimates and Prediction Intervals

The models which are used to give prediction of natural gas demand, are specified in Section 4.1.

To forecast natural gas demand h day ahead at date t , where $h = 1, \dots, 7$, the point estimate can be calculated by equation

$$g(\hat{y}_{t+h}) = \hat{\beta}_0 + \sum_{j=1}^3 \hat{f}_j(x_{j,t+h}) + \hat{\beta}_4 x_{4,t+h} + \hat{\varepsilon}_{t+h},$$

where

g is the identity link function,

$\hat{\beta}_i$'s and \hat{f}_j 's are the results of Equation 2, and

$$\hat{\varepsilon}_{t+k} = \hat{\alpha}_1 \hat{\varepsilon}_{t+k-1} + \hat{\alpha}_2 \hat{\varepsilon}_{t+k-2}, \quad k = 1, \dots, h,$$

and

$$\hat{\varepsilon}_{t+k} = g(y_{t+k}) - (\hat{\beta}_0 + \sum_{j=1}^3 \hat{f}_j(x_{j,t+k}) + \hat{\beta}_4 x_{4,t+k}), \quad k \leq 0,$$

where

$\hat{\alpha}_i$'s are the results of Equation 3.

Figure 15 shows the predicted log-transformed natural gas demand 1 day ahead on the test set.

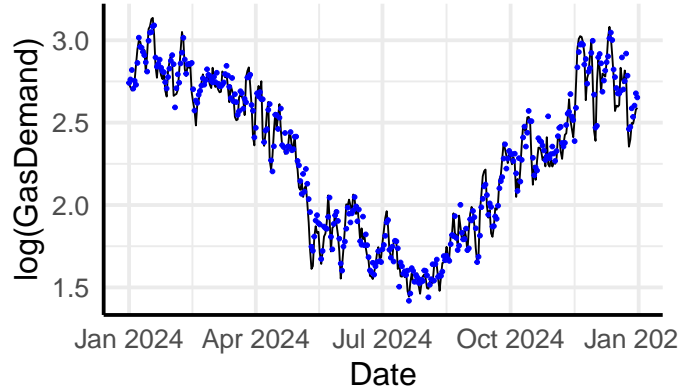


Figure 15: Plot of Predicted Value at $h = 1$ over True Value

Table 2 summarises model performances on the test set according to the number of days ahead. As the number increases, the mean absolute error and mean squared error

becomes larger. However, the increasing speed slows down, and the MAE and MSE converges, since the residuals exhibits stationary.

Table 2: Model Performance

Days Ahead	MAE	MSE
1	0.0671	0.0076
2	0.1000	0.0161
3	0.1167	0.0220
4	0.1260	0.0254
5	0.1311	0.0275
6	0.1347	0.0289
7	0.1355	0.0296

To further evaluate the prediction accuracy, the prediction intervals were calculated through bootstrap methods.

Figure 16 displays the 95% prediction intervals according to the number of days ahead. The width of the interval increases as the number becomes larger, but then converges.

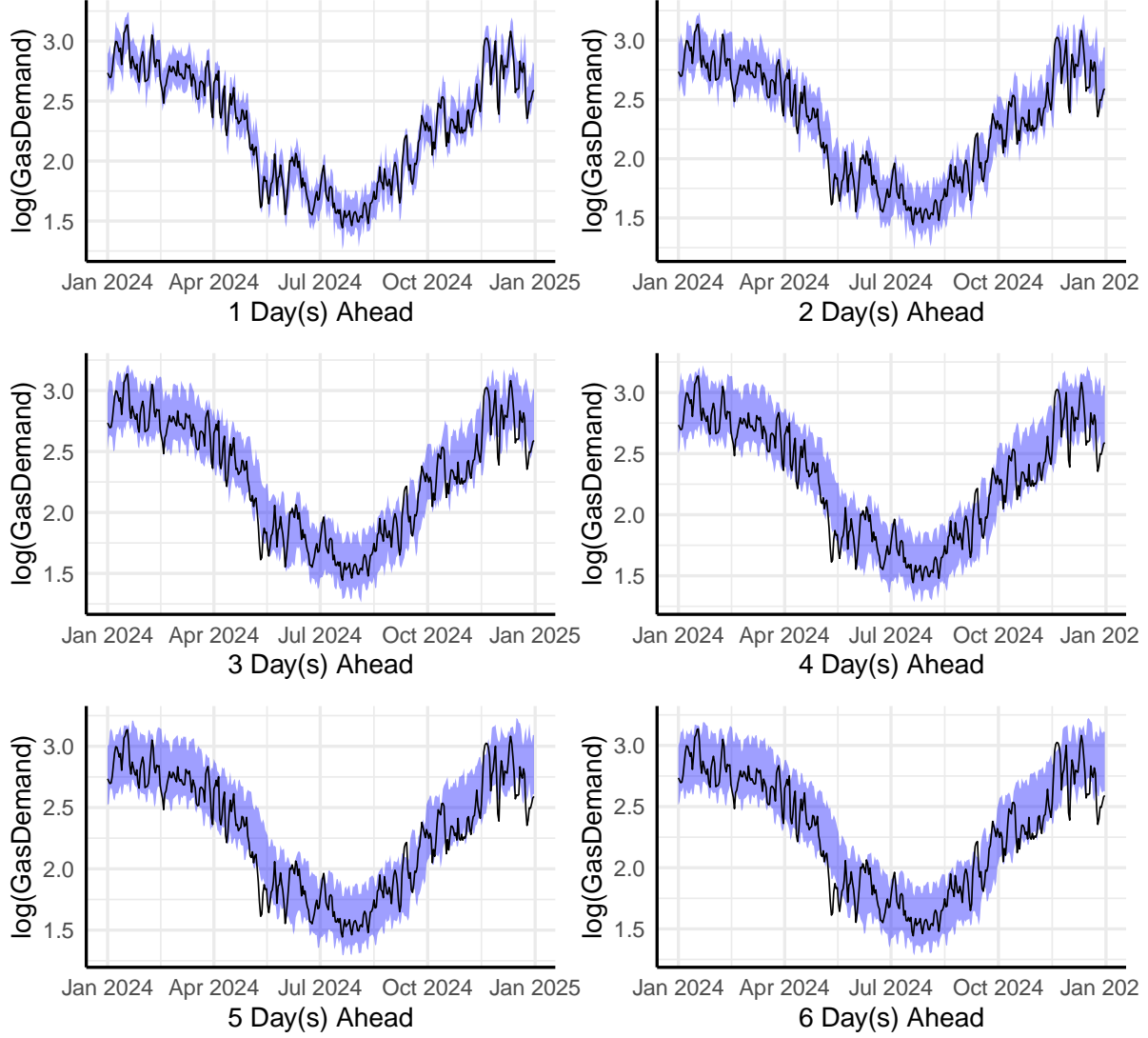


Figure 16: Plots of Prediction Intervals

Table 3 shows that the 95% prediction intervals fail to cover 95% of the true values, with the coverages between 83.6% and 94.0%.

Table 3: 95% Prediction Interval Coverage

Days Ahead	Coverage
1	0.9399
2	0.8989
3	0.9016

Table 3: 95% Prediction Interval Coverage

Days Ahead	Coverage
4	0.8716
5	0.8552
6	0.8361
7	0.8497

5 Conclusions

Natural gas demand can be affected by multiple factors, including temperature, wind speeds, holidays, and seasons. This study provides a modelling approach using GAM and ARIMA models to predict the trend of natural gas demand from several days ahead. Also, the prediction accuracy has been evaluated by estimating the prediction intervals. The study hence reveals that the increments in number of days ahead and the quantity of natural gas demand can increase forecast uncertainty.

The GAM and ARIMA model used in this study captured the non-linear relationship between explanatory variables and response variable well, and produced reasonable results that fit the model assumptions. It provides plausible short-term forecasts that can be potentially used in natural gas transmission and distribution. However, there lacks comparison and selection between different models. Also, the prediction intervals do not perform well for longer forecasts, which might be due to uncaptured variance. The future research could introduce more variables, more interaction terms, or different models, and evaluate model performance to compare these models, so that long-term forecasts could be more accurate.

Reference

Anon 2022. Advanced predictive models.

Hastie, T. and Tibshirani, R. 1986. Generalized additive models. *Statistical Science*. **1**(3), pp.297–310.

Hyndman, R.J. and Athanasopoulos, G. 2021. *Forecasting: Principles and practice* [Online] 3rd ed. Melbourne, Australia: OTexts. Available from: [OTexts.com/fpp3](https://otexts.com/fpp3).

IEA 2025. *Global energy review 2025* [Online]. Paris: IEA. Available from: <https://www.iea.org/reports/global-energy-review-2025>.

National Gas n.d. National gas transmission data portal. Available from: <https://data.nationalgas.com>.

UK Government n.d. Bank holidays. Available from: <https://www.api.gov.uk/gds/bank-holidays>.