Executive Summary

Objective: The project's primary goal is to analyze and forecast residential electric power usage, with a particular focus on the global active power variable, utilizing the "Household Electric Power Consumption" dataset from the UCI Machine Learning repository.

Data Analysis: Analysis was performed on detailed electric power consumption metrics for households, captured in a dataset featuring variables such as global active and reactive power, voltage, intensity, and sub-metering data. The processed dataset included weekly averages of active power consumption and temperatures, facilitating a deeper understanding of consumption patterns over time.

Methodology: To achieve accurate forecasting, the project employed both Seasonal Autoregressive Integrated Moving Average with eXogenous factors (SARIMAX) and Additive Seasonal Exponential Smoothing models, in addition to the SARIMA model. These models were selected for their robustness in handling complex seasonal patterns and incorporating external variables that could influence power consumption trends. The methodology encompassed:

Data Preparation: Transformation and cleaning of raw data to a format suitable for analysis.

Model Selection: Evaluation of various statistical models to identify those most capable of accurately forecasting power consumption. The SARIMAX model allowed for the inclusion of external variable temperature, providing a nuanced understanding of their impact on power usage. The Additive Seasonal Exponential Smoothing model was employed as there is no trend in data.

Model Estimation and Validation: Detailed estimation of model parameters followed by rigorous validation to ensure accuracy and reliability of forecasts. This included outlier detection, residual analysis, and evaluation of forecast accuracy through measures such as MAPE and RMSE.

Results: The application of SARIMAX and Additive Seasonal Exponential Smoothing models, alongside the SARIMA model, significantly enhanced the forecasting accuracy. The SARIMAX model incorporated temperature variations, a key external factor affecting power consumption. The forecasting outputs of SARIMA(1,0,0)(0,0,0) revealed the models' capacity to closely match actual data and provide reliable predictions for future energy usage.

Conclusion: The comprehensive approach, incorporating SARIMA, SARIMAX, and Additive Seasonal Exponential Smoothing models, offered a detailed and accurate forecast of residential electric power consumption.

Dataset

The dataset "Household Electric Power Consumption" was obtained from the UCI Machine Learning repository. The objective of this project is to analyze and forecast residential electric power usage, with particular emphasis on the global active power variable during the analysis.

Raw Data:

The Raw data typically consists of several columns representing different aspects of electric power consumption in households.

The column description and details in the dataset include:

- **Date**: The date on which the data was recorded.
- **Time**: The specific time of day at which the data was recorded.
- **Global_active_power**: The total active power consumed by the household (in kilowatts).
- **Global_reactive_power**: The total reactive power consumed by the household (in kilowatts).
- **Voltage**: The average voltage (in volts) measured during the minute.
- **Global_intensity**: The total current intensity (in amperes) of electricity consumption.
- **Sub_metering_1**: The active energy (in watt-hours) measured in sub-metering zone 1, typically for specific rooms or circuits.
- **Sub_metering_2**: The active energy (in watt-hours) measured in sub-metering zone 2.
- **Sub_metering_3**: The active energy (in watt-hours) measured in sub-metering zone 3, often representing the energy consumed by an electric water heater and an air-conditioner.

Raw Data
(household_power_co

Processed Data:

The processed data was derived from the raw data, which encompasses:

- **week_start_date**: This column represents the start date of each week for which data is recorded. The dates are in the format YYYY-MM-DD.
- **avgactivepower**: This column shows the average active power consumption for the week, measured in some unit (likely kilowatts or a similar unit of power).

- **Weekly Avgtemp**: This column indicates the average temperature for the week in degrees Fahrenheit.

- The average active power was calculated using the formula

  (global_active_power*1000/60)

  which represents the active energy consumed per minute (in watt-hours) within the household by electrical equipment covered by sub-meterings 1, 2, and 3.

- The average temperature data was derived by calculating the weekly average temperatures from historical weather records available at Wunderground, specifically by filtering through France's weekly historical data for each corresponding week recorded in the dataset.

Processed Data
(POWER CONSUMPTI

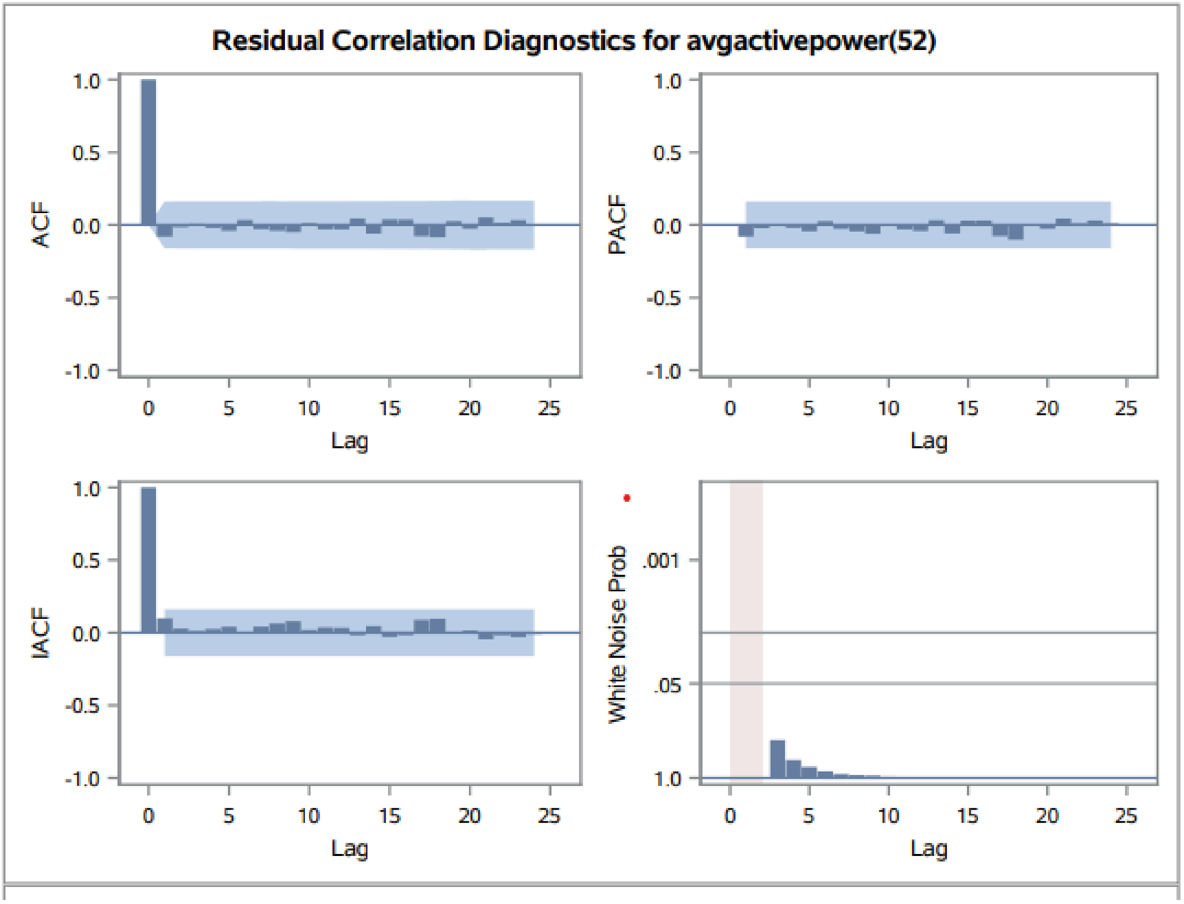
Best Model:

The SARIMA (1,0,0)(0,1,0) model was deemed the most suitable choice based on several compelling factors:

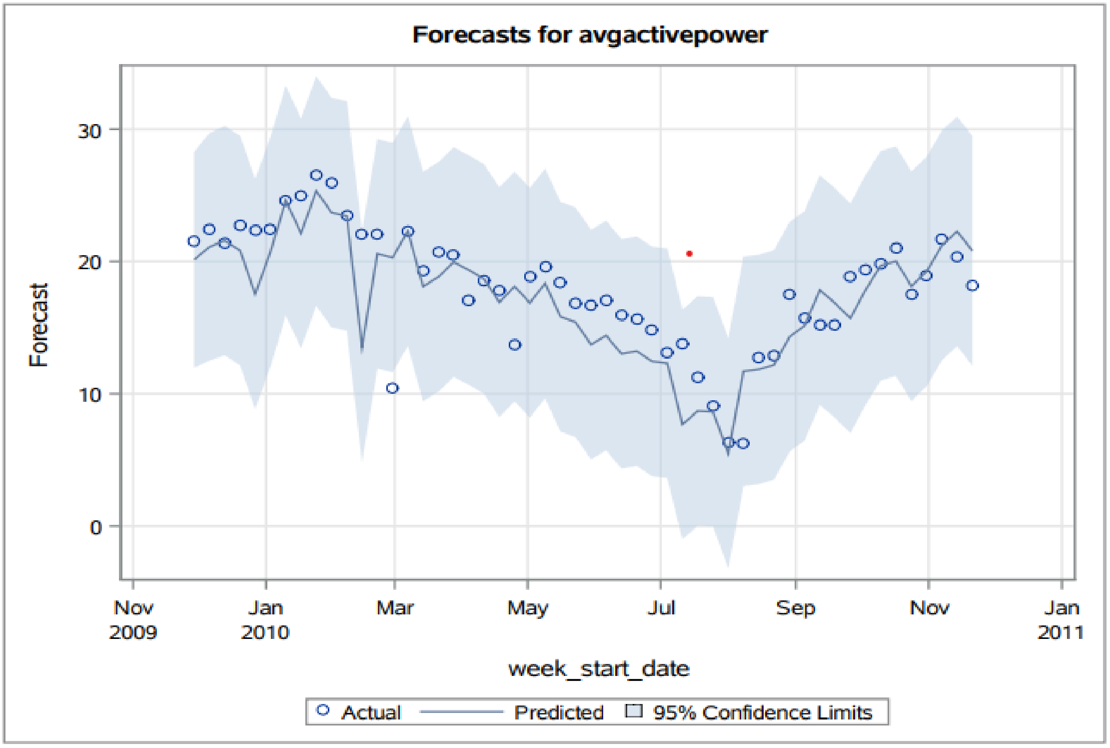
This section outlines the methodology and results of using an ARIMA model for forecasting average active power consumption. This model was chosen for its high accuracy, demonstrated through various statistical measures such as the Akaike Information Criterion (AIC) and the Schwarz Bayesian Criterion (SBC) and the inclusion of a Mean Absolute Percentage Error (MAPE) value of 13.64 and RMSE value of 1.34, which indicates a reasonable level of accuracy for predictions.

The model's parameters, including autoregressive factors and differencing periods, were carefully selected based on the dataset's characteristics, including seasonality and trends. Technical details such as the model's ability to account for white noise and its residual analysis indicate robust performance in forecasting. This Model provides a comprehensive explanation of the modeling process, including estimation techniques, outlier detection, and forecast accuracy, justifying its selection as the best model for this forecasting project.

Maximum Likelihood Estimation					
Parameter	Estimate	Standard Error	t Value	Approx Pr > |t|	Lag
MU	-0.45758	0.50150	-0.91	0.3615	0
AR1,1	0.33889	0.07607	4.46	<.0001	1

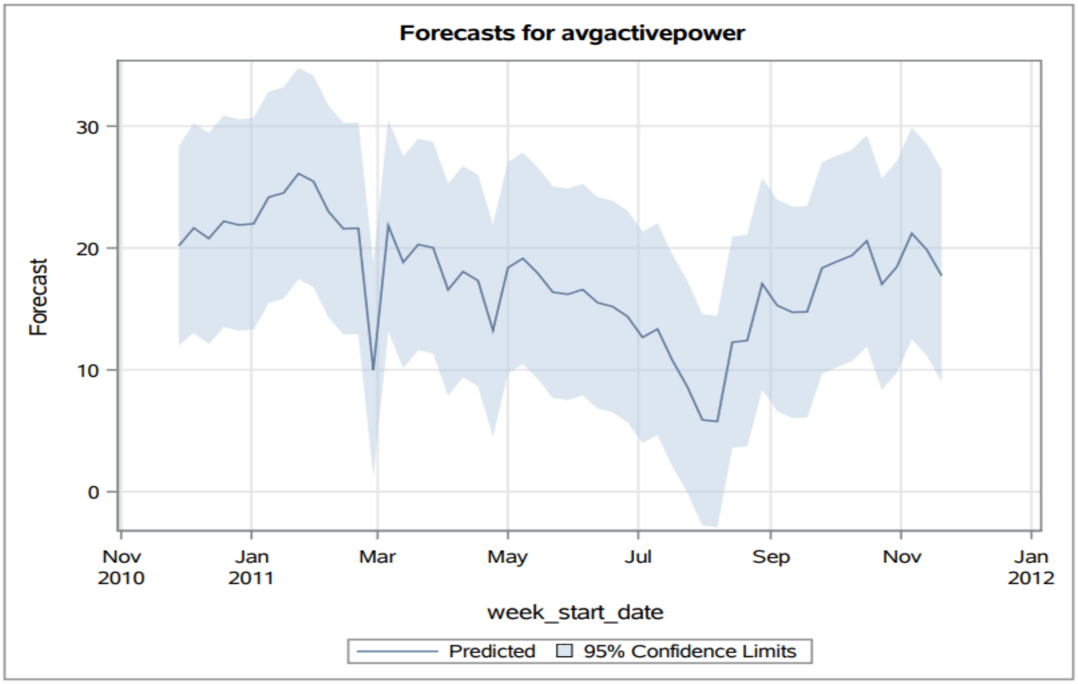


The forecasting graph typically shows the actual vs. predicted values over time, highlighting the model's ability to closely track the observed data and its performance in capturing the underlying data dynamics. The choice of this model for forecasting is justified by its statistical efficiency, the ability to minimize forecast errors, and its suitability for the project's specific data characteristics and modeling objectives.



Since the SARIMA(1,0,0) model demonstrated superior forecasting ability, indicated by a low RMSE value of 1.34, it was selected for projecting future values.

The provided graph illustrates the effectiveness of the ARIMA model in forecasting, highlighting the anticipated future trajectory of 'avgactivepower' based on the model's predictions.



Comparison and Justification

We have shortlisted Additive seasonal exponential smoothing, SARIMA(1,0,0)(0,0,0) and SARIMAX(1,0,0)(0,1,0) from all the models we performed.

We have selected SARIMA(1,0,0)(0,1,0) as the best model by looking at MAPE and RMSE values. When we compared MAPE, SRIMA AND SARIMAX had a tie. So, we looked at RMSE and concluded SARIMA(1,0,0)(0,0,0) as the best model.

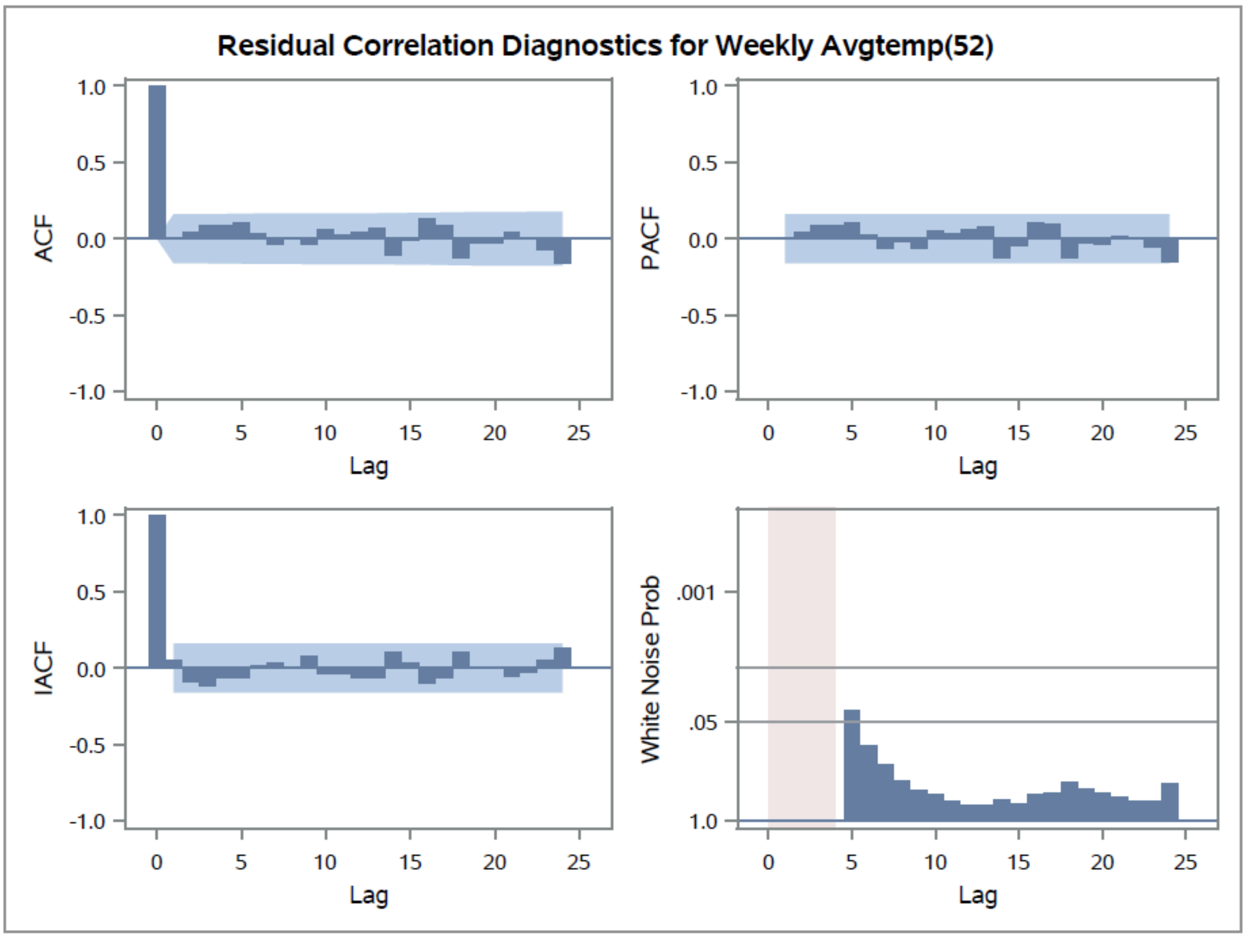
MODEL	RMSE	MAPE
Additive Seasonal	3.24	14
SARIMA(1,0,0)(0,1,0)	1.34	13.62
SARIMAX(1,0,0)(0,1,0)	1.8	13.62

Full Model Development Details:

SARIMA MODELS (PRE WHITENING TEMPERATURE VARIABLE FOR SARIMAX MODELING)

1) **SARIMA(2,0,2)(0,1,0)**

Maximum Likelihood Estimation					
Parameter	Estimate	Standard Error	t Value	Approx Pr > |t|	Lag
MU	-0.42751	1.05741	-0.40	0.6860	0
MA1,1	1.11046	5.19819	0.21	0.8308	1
MA1,2	-0.11056	0.62868	-0.18	0.8604	2
AR1,1	1.36120	0.35770	3.81	0.0001	1
AR1,2	-0.36398	0.29185	-1.25	0.2123	2



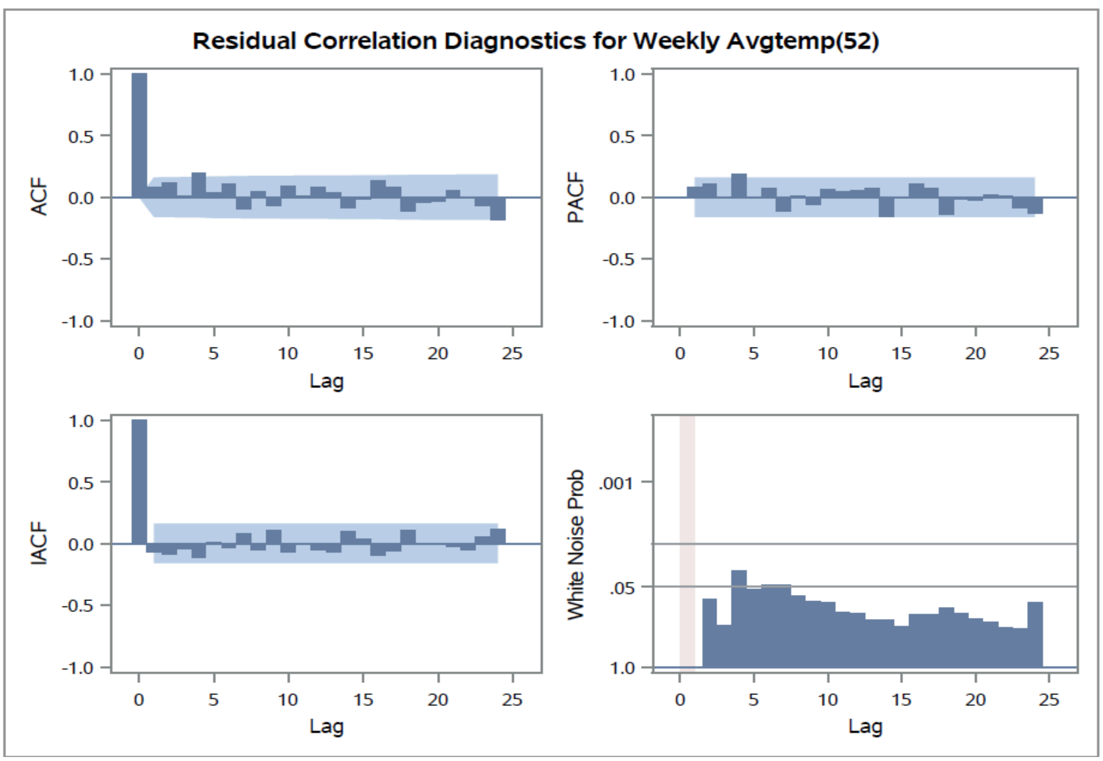

```
proc sort data=STSM.POWERCONSUMPTION
out=Work.preProcessedData;
```

```
        by week_start_date;

run;


proc arima data=Work.preProcessedData plots

    (only)=(series(corr crosscorr) residual(corr normal) ) out=work.out;

        identify var='Weekly Avgtemp'n(52);

        estimate q=(1) method=ML outstat=work.outstat;

        forecast lead=52 back=52 alpha=0.05 id=week_start_date
interval=week;

        run;
```

2) **SARIMA(0,0,1)(0,1,0)**

The table and graphs presented indicate the results of an **SARIMA(0,0,1)(0,1,0)** model's maximum likelihood estimation and diagnostic checks. The estimation results show the parameters MU and MA1,1, with their estimates, standard errors, and t-values, which suggest the significance of these parameters. The Residual Correlation Diagnostics include Autocorrelation Function (ACF) and Partial Autocorrelation Function (PACF) plots, which do not show significant correlations at various lags, indicating a good model fit. The last graph shows the ACF of squared residuals to check for non-linearity or heteroscedasticity, with no apparent patterns suggesting a good model fit.

Maximum Likelihood Estimation					
Parameter	Estimate	Standard Error	t Value	Approx Pr > |t|	Lag
MU	-1.02678	0.61467	-1.67	0.0948	0
MA1,1	-0.47946	0.07110	-6.74	<.0001	1



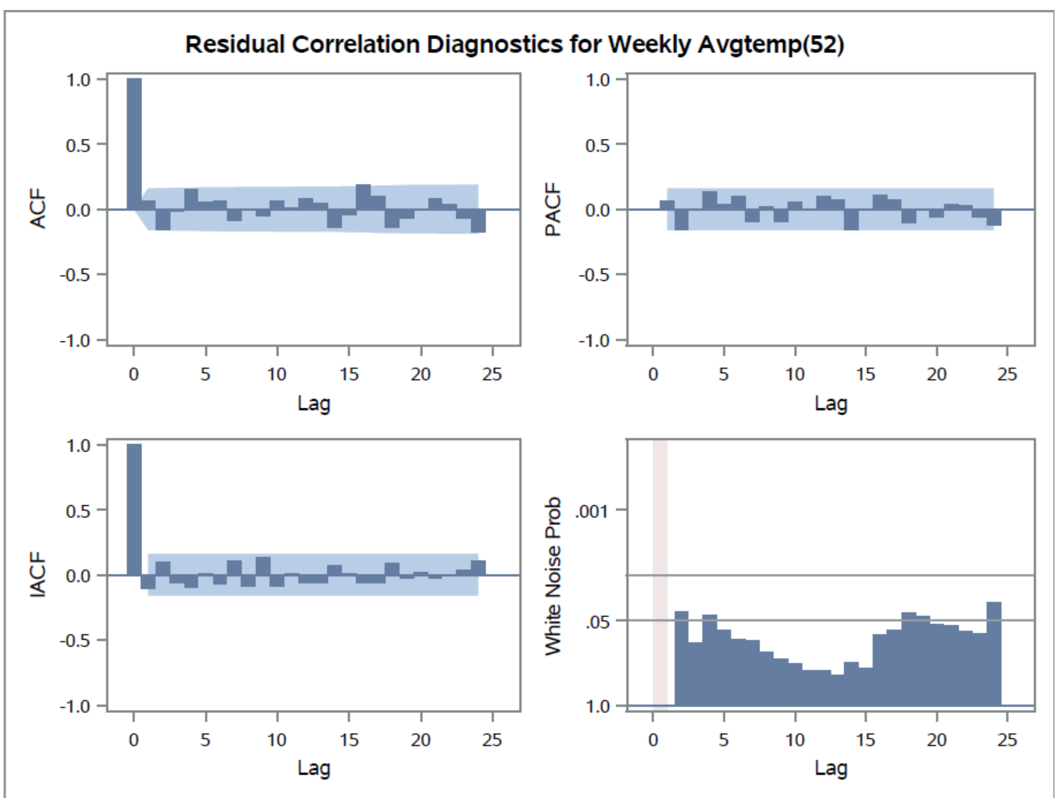
```
proc sort data=STSM.POWERCONSUMPTION
out=Work.preProcessedData;

        by week_start_date;

run;


proc arima data=Work.preProcessedData plots

    (only)=(series(corr crosscorr) residual(corr normal) ) out=work.out;

        identify var='Weekly Avgtemp'n(52);

        estimate q=(1) method=ML outstat=work.outstat;

        forecast lead=52 back=52 alpha=0.05 id=week_start_date
interval=week;

        run;

quit;


proc delete data=Work.preProcessedData;

run;
```

3) SARIMA(1,0,0)(0,1,0)

The estimation table shows the parameter MU (mean of the series) and AR1,1 (the coefficient for the first autoregressive term), including their estimates, standard errors, and t-values. The AR coefficient is significant, indicating a relationship between the current value and its immediate past value.

The diagnostic plots, including the Autocorrelation Function (ACF) and Partial Autocorrelation Function (PACF), show that correlations are within the confidence bounds for most lags, suggesting that the model residuals are random (good fit). The Ljung-Box test (White Noise Prob) indicates p-values above the significance level, suggesting the residuals are white noise.

Maximum Likelihood Estimation					
Parameter	Estimate	Standard Error	t Value	Approx Pr > |t|	Lag
MU	-1.07304	0.80052	-1.34	0.1801	0
AR1,1	0.48323	0.07101	6.80	<.0001	1



```
proc sort data=STSM.POWERCONSUMPTION out=Work.preProcessedData;
        by week_start_date;
run;

proc arima data=Work.preProcessedData plots
   (only)=(series(corr crosscorr) residual(corr normal) ) out=work.out;
        identify var='Weekly Avgtemp'n(52);
        estimate p=(1) method=ML outstat=work.outstat;
        forecast lead=52 back=52 alpha=0.05 id=week_start_date
interval=week;
        run;
quit;

proc delete data=Work.preProcessedData;
run;
```

4) **SARIMA(1,0,2)(0,1,0)**

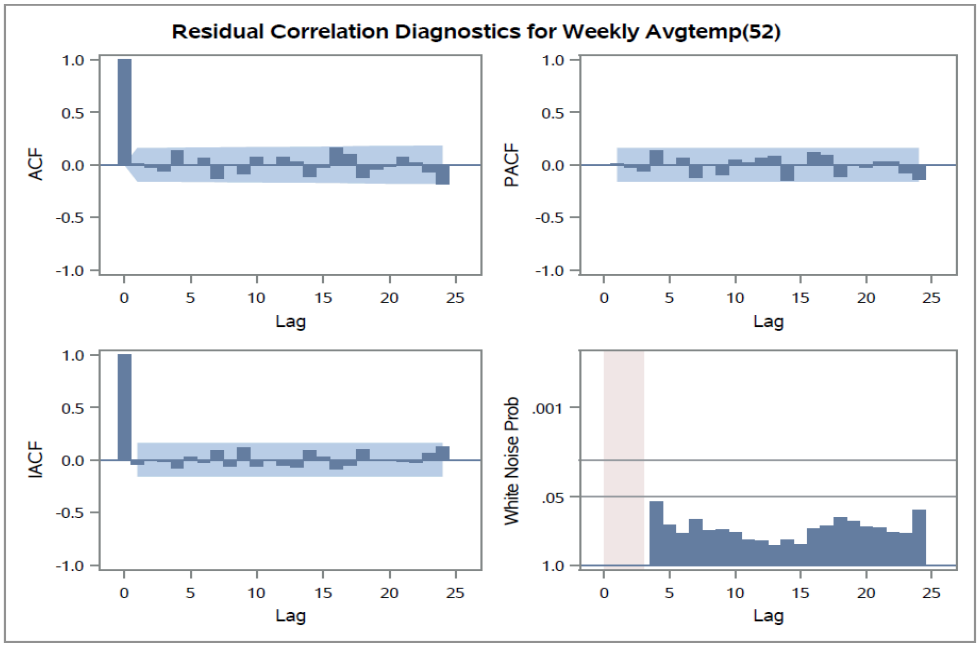
The section shows the Maximum Likelihood Estimation results for an **SARIMA(1,0,2)(0,1,0)** model, with parameters for MU (mean), MA1,1 and MA1,2 (moving average terms), and AR1,1 (autoregressive term). The AR term has a significant t-value, indicating its influence in the model.

The diagnostic plots include ACF and PACF, which assess the residuals' correlation at different lags. The lack of significant spikes beyond the confidence bands suggests that the model captures the data's autocorrelation structure well.

The bottom right plot likely represents a check for white noise residuals. The p-values from this test appear to be above the significance level (usually 0.05), suggesting the residuals do not exhibit autocorrelation and that the model is adequate.

Maximum Likelihood Estimation

Parameter	Estimate	Standard Error	t Value	Approx Pr > |t|	Lag
MU	-1.08158	0.91251	-1.19	0.2359	0
MA1,1	0.27786	0.23506	1.18	0.2372	1
MA1,2	0.28738	0.15570	1.85	0.0649	2
AR1,1	0.80592	0.20744	3.89	0.0001	1




```
proc sort data=STSM.POWERCONSUMPTION out=Work.preProcessedData;

        by week_start_date;

run;


proc arima data=Work.preProcessedData plots

    (only)=(series(corr crosscorr) residual(corr normal) ) out=work.out;

        identify var='Weekly Avgtemp'n(52);

        estimate p=(1) q=(1 2) method=ML outstat=work.outstat;

        forecast lead=52 back=52 alpha=0.05 id=week_start_date
interval=week;

        run;

quit;
```

```
proc delete data=Work.preProcessedData;

run;
```

5) SARIMA(2,0,1)(0,1,0)

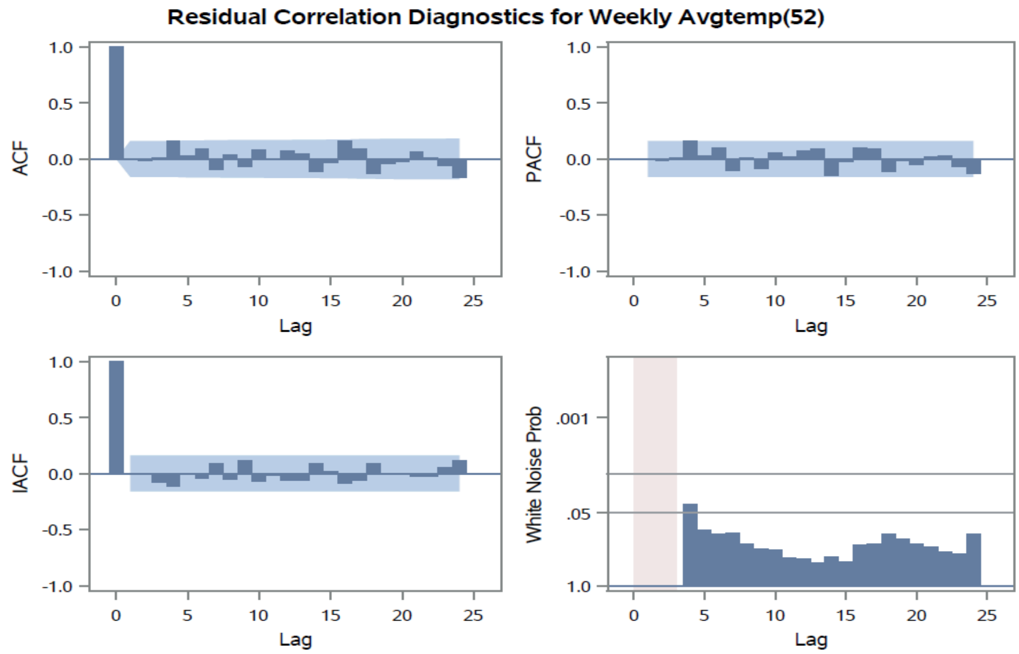
This section shows the Maximum Likelihood Estimation results and Residual Correlation Diagnostics for an **SARIMA(2,0,1)(0,1,0)** model. The parameter estimates for MU, MA1,1, AR1,1, and AR1,2 are provided with their standard errors and t-values, which suggest the parameters are not statistically significant since their p-values are above 0.05.

The diagnostic graphs show ACF and PACF for the residuals, which are used to identify any remaining autocorrelation after fitting the model. The ACF and PACF values are mostly within the confidence bounds, indicating no significant autocorrelation.

The bottom graphs are likely a test for white noise in the residuals, such as the Ljung-Box test, showing p-values which suggest the residuals are random, pointing to a well-fitting model.

Maximum Likelihood Estimation

Parameter	Estimate	Standard Error	t Value	Approx Pr > |t|	Lag
MU	-1.04636	0.72605	-1.44	0.1495	0
MA1,1	-0.30478	0.52390	-0.58	0.5607	1
AR1,1	0.24985	0.53100	0.47	0.6380	1
AR1,2	0.0079807	0.28278	0.03	0.9775	2




```
proc sort data=STSM.POWERCONSUMPTION out=Work.preProcessedData;
        by week_start_date;
run;


proc arima data=Work.preProcessedData plots
    (only)=(series(corr crosscorr) residual(corr normal) ) out=work.out;
        identify var='Weekly Avgtemp'n(52);
        estimate p=(1 2) q=(1) method=ML outstat=work.outstat;
        forecast lead=52 back=52 alpha=0.05 id=week_start_date
interval=week;
        run;
quit;
```

```
proc delete data=Work.preProcessedData;

run;
```

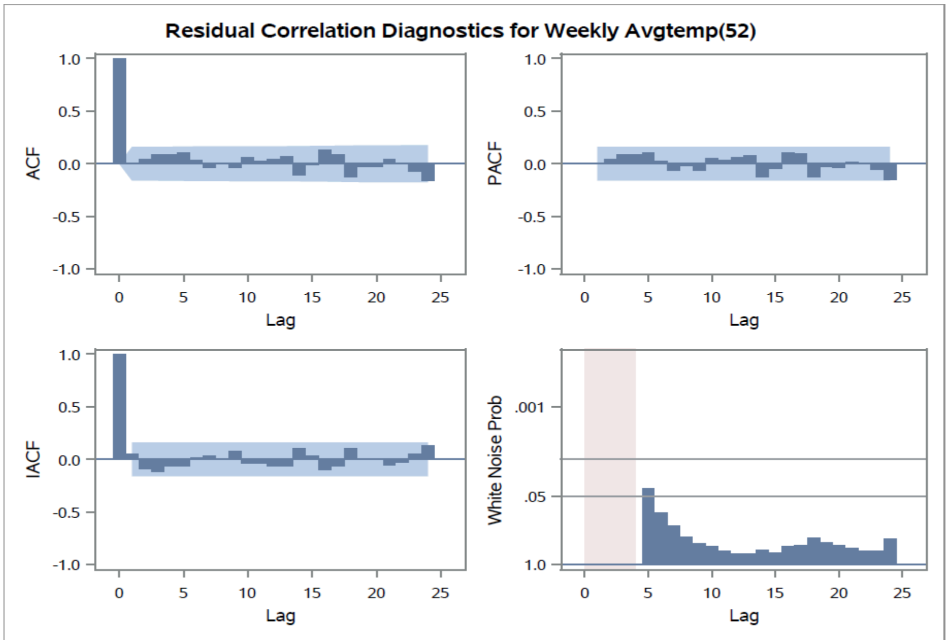
6) SARIMA(2,0,2)(0,1,0)

The section shows Maximum Likelihood Estimation table lists the estimated coefficients for the model's parameters (MU, MA1,1, MA1,2, AR1,1, AR1,2), including their standard errors, t-values, and p-values. The significant p-values for MA1,1, MA1,2, and AR1,1 suggest these parameters are statistically significant.

The Residual Correlation Diagnostics include the ACF and PACF plots, which help identify any remaining autocorrelation in the residuals after fitting the model. The ACF and PACF values within the confidence bounds suggest a good fit, as there are no significant spikes indicating autocorrelations.

The bottom right graph seems to be a Ljung-Box test, assessing the randomness of the residuals over time. The p-values well above the significance level suggest the residuals do not exhibit patterns, indicating an adequate model fit.

Maximum Likelihood Estimation					
Parameter	Estimate	Standard Error	t Value	Approx Pr > |t|	Lag
MU	-1.03047	0.67768	-1.52	0.1284	0
MA1,1	-1.39518	0.13588	-10.27	<.0001	1
MA1,2	-0.52518	0.11976	-4.39	<.0001	2
AR1,1	-0.80661	0.15617	-5.16	<.0001	1
AR1,2	0.04302	0.14952	0.29	0.7736	2




```
proc sort data=STSM.POWERCONSUMPTION out=Work.preProcessedData;
        by week_start_date;
run;


proc arima data=Work.preProcessedData plots
    (only)=(series(corr crosscorr) residual(corr normal) ) out=work.out;
        identify var='Weekly Avgtemp'n(52);
        estimate p=(1 2) q=(1 2) method=ML outstat=work.outstat;
```

```
    forecast lead=52 back=52 alpha=0.05 id=week_start_date
interval=week;

    run;

quit;



proc delete data=Work.preProcessedData;

run;
```

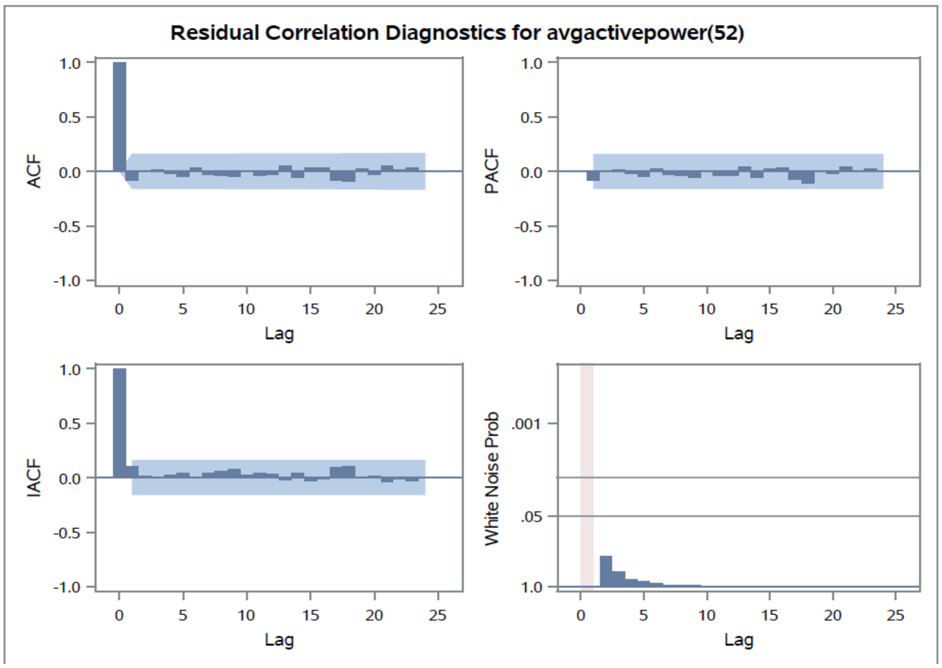
SARIMAX MODELS(WITH PREWHITENED TEMERATURE VARIABLE INCORPORATED)

1) SARIMAX(1,0,0)(010)

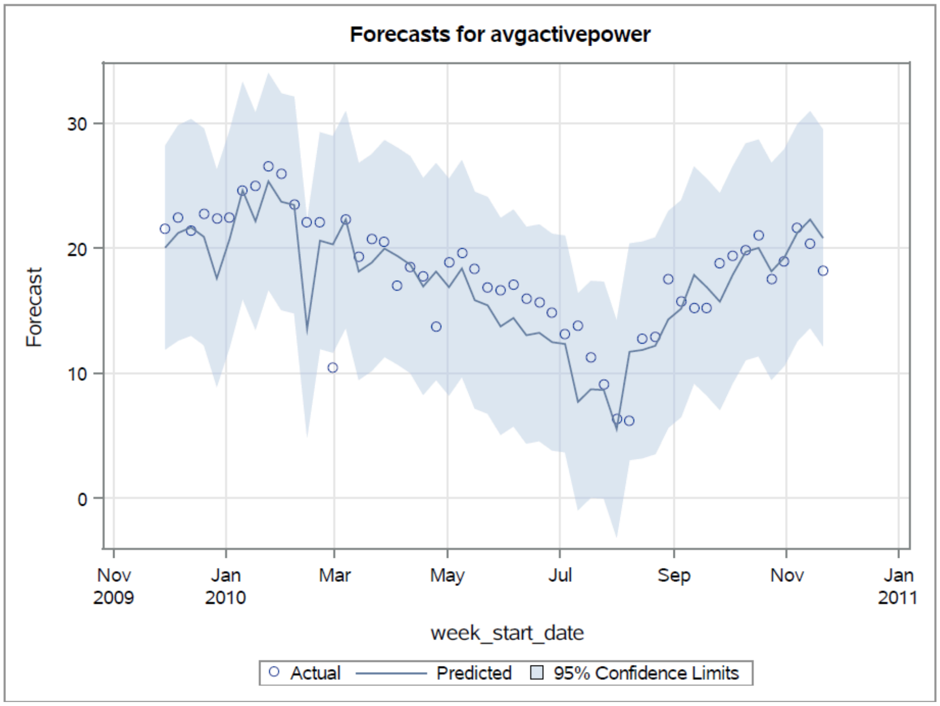
The image displays Maximum Likelihood Estimation results for an **SARIMAX(1,0,0)(010)** model along with Residual Correlation Diagnostics. The table shows the estimates, standard errors, t-values, and p-values for parameters MU (intercept), AR1,1 (autoregressive term), and NUM1 (possibly an exogenous variable or intervention). AR1,1 is statistically significant with a p-value less than 0.05.

The diagnostic plots (ACF, PACF) show no significant autocorrelations in the residuals, suggesting a good model fit. The bottom right plot indicates the p-values from a test for white noise in the residuals, such as the Ljung-Box test, which also suggest that the residuals are random, confirming the model's adequacy.

Maximum Likelihood Estimation							
Parameter	Estimate	Standard Error	t Value	Approx Pr > |t|	Lag	Variable	Shift
MU	-0.44091	0.50834	-0.87	0.3858	0	avgactivepower	0
AR1,1	0.33973	0.07633	4.45	<.0001	1	avgactivepower	0
NUM1	0.01624	0.06395	0.25	0.7995	0	Weekly Avgtemp	0



Given that this model proved to be superior among all evaluated models, we proceeded with the analysis of its forecasting graph.




```
ods noproctitle;

ods graphics / imagemap=on;


proc sort data=STSM.POWERCONSUMPTION_TEST
out=Work.preProcessedData;

        by week_start_date;

run;


proc arima data=Work.preProcessedData plots

    (only)=(series(corr crosscorr) residual(corr normal) forecast(forecast
forecastonly) );

        identify var='Weekly Avgtemp'n(52);

        estimate p=(1) q=(1) method=ML;

        identify var=avgactivepower(52) crosscorr=('Weekly Avgtemp'n(52)
);
```

```
estimate p=(1) input=('Weekly Avgtemp'n) method=ML;
forecast lead=52 back=52 alpha=0.05 id=week_start_date
interval=week out=work.ARIMAX;
run;
quit;

proc delete data=Work.preProcessedData;
run;
```

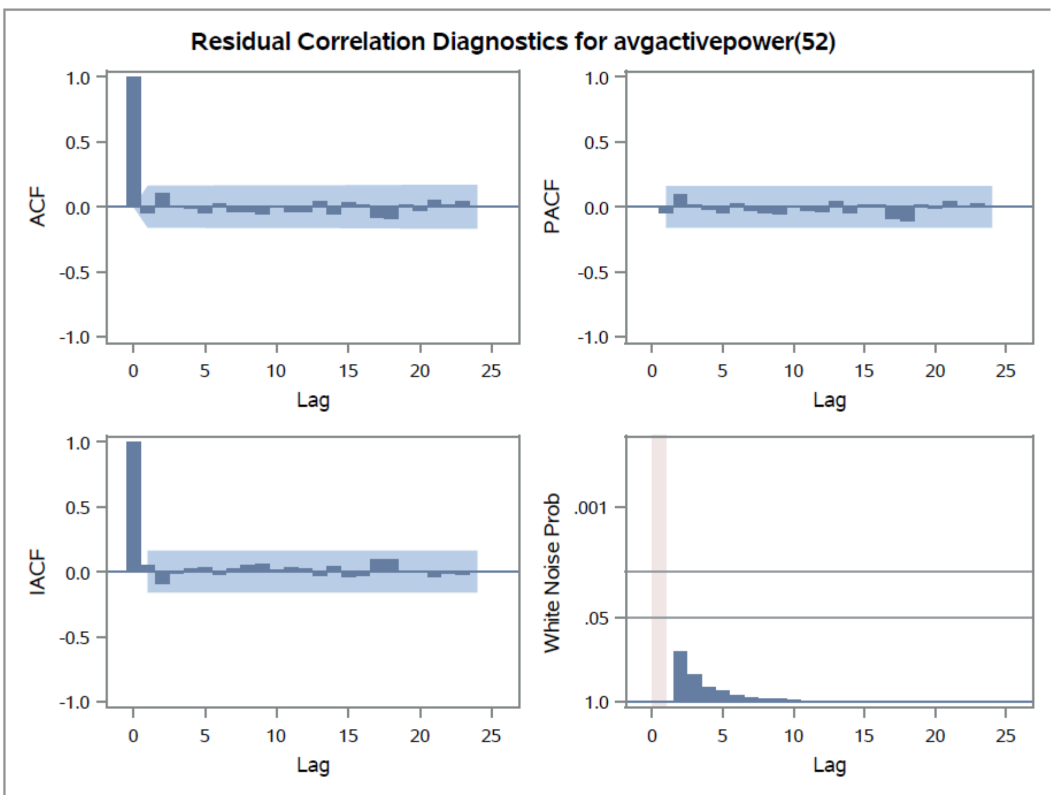
2) SARIMAX(0,0,1)(0,1,0)

The section includes Maximum Likelihood Estimation results and Residual Correlation Diagnostics for an **SARIMAX(0,0,1)(0,1,0)** model. The table shows the model's parameters with their estimates, standard errors, and t-values, indicating the significance of each parameter. The MA1,1 parameter is statistically significant.

The diagnostic graphs for ACF and PACF suggest the residuals have little to no autocorrelation, as most lag values are within the confidence intervals. The bottom graph, likely a Ljung-Box test, assesses the randomness of residuals and shows p-values that are not significant, indicating no autocorrelation, which is desirable for a well-fitting model.

Maximum Likelihood Estimation

Parameter	Estimate	Standard Error	t Value	Approx Pr > |t|	Lag	Variable	Shift
MU	-0.39643	0.44043	-0.90	0.3681	0	avgactivepower	0
MA1,1	-0.29909	0.07887	-3.79	0.0001	1	avgactivepower	0
NUM1	0.01744	0.06360	0.27	0.7839	0	Weekly Avgtemp	0



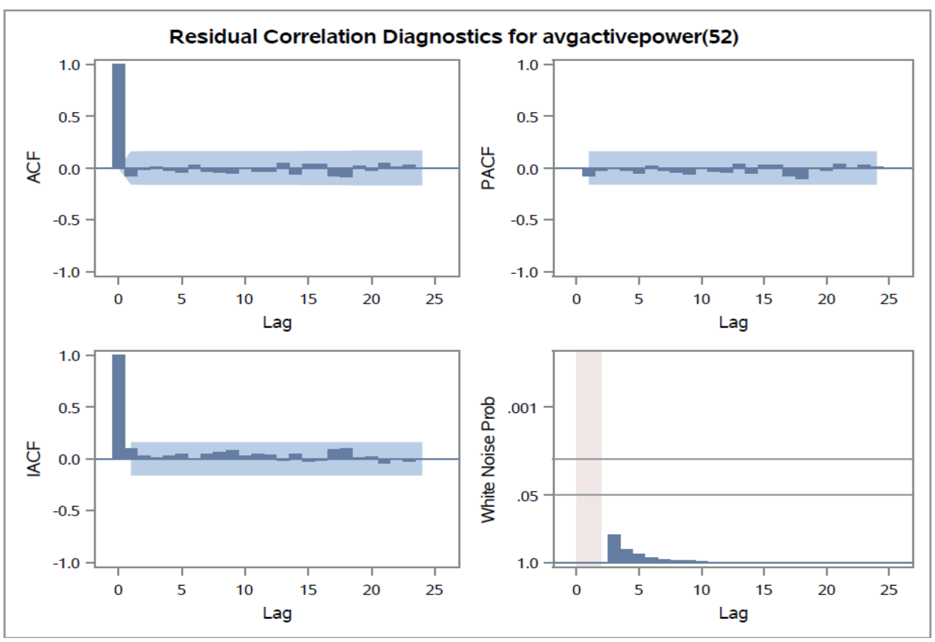
```
ods noproctitle;

ods graphics / imagemap=on;


proc sort data=STSM.POWERCONSUMPTION_TEST
out=Work.preProcessedData;

        by week_start_date;

run;


proc arima data=Work.preProcessedData plots

    (only)=(series(corr crosscorr) residual(corr normal) forecast(forecast
forecastonly) );

        identify var='Weekly Avgtemp'n(52);

        estimate p=(1) q=(1) method=ML;

        identify var=avgactivepower(52) crosscorr=('Weekly Avgtemp'n(52)
);

        estimate q=(1) input=('Weekly Avgtemp'n) method=ML;

        forecast lead=52 back=52 alpha=0.05 id=week_start_date
interval=week out=work.ARIMAX;

        run;

quit;


proc delete data=Work.preProcessedData;

run;
```

3) SARIMAX(1,0,1)(0,1,0)

The section shows a correlation matrix of parameter estimates for an ARIMA model and corresponding Residual Correlation Diagnostics graphs. The matrix indicates low correlation between the model's parameters, which is generally a good sign of model stability. The diagnostic graphs, including the ACF and PACF, suggest that the residuals do not exhibit significant autocorrelation, implying an adequate model fit. The Ljung-Box test graph indicates the residuals are likely random, reinforcing the model's adequacy.

Correlations of Parameter Estimates

Variable Parameter		avgactivepower MU	avgactivepower MA1,1	avgactivepower AR1,1	Weekly Avgtemp NUM1
avgactivepower	MU	1.000	-0.013	-0.015	0.130
avgactivepower	MA1,1	-0.013	1.000	0.941	-0.042
avgactivepower	AR1,1	-0.015	0.941	1.000	-0.053
Weekly Avgtemp	NUM1	0.130	-0.042	-0.053	1.000



```
ods noproctitle;

ods graphics / imagemap=on;


proc sort data=STSM.POWERCONSUMPTION_TEST
out=Work.preProcessedData;

          by week_start_date;

run;


proc arima data=Work.preProcessedData plots

    (only)=(series(corr crosscorr) residual(corr normal) forecast(forecast
forecastonly) );

          identify var='Weekly Avgtemp'n(52);

          estimate p=(1) q=(1) method=ML;

          identify var=avgactivepower(52) crosscorr=('Weekly Avgtemp'n(52)
);

          estimate p=(1) q=(1) input=('Weekly Avgtemp'n) method=ML;

          forecast lead=52 back=52 alpha=0.05 id=week_start_date
interval=week out=work.ARIMAX;

          run;

quit;


proc delete data=Work.preProcessedData;

run;
```

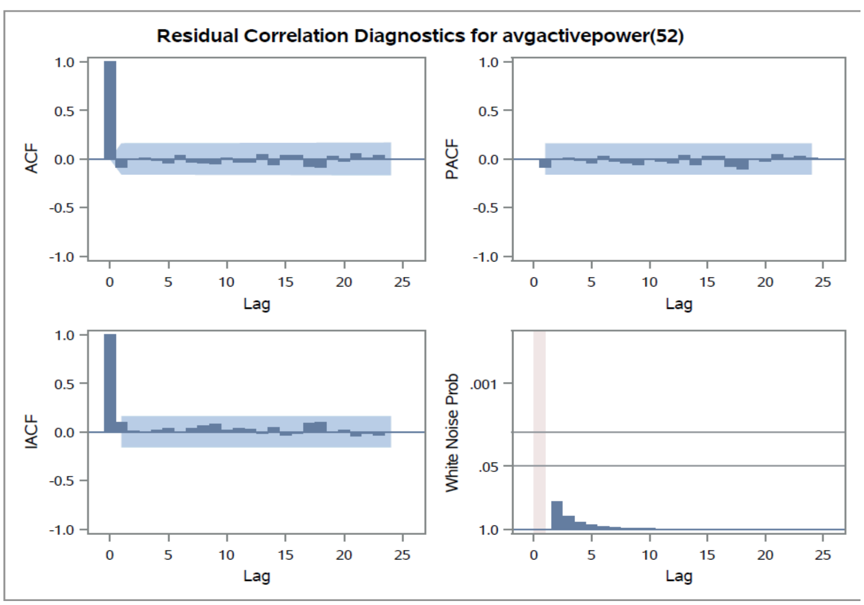
SARIMA MODELS

1) SARIMA(1,0,0)(0,1,0)

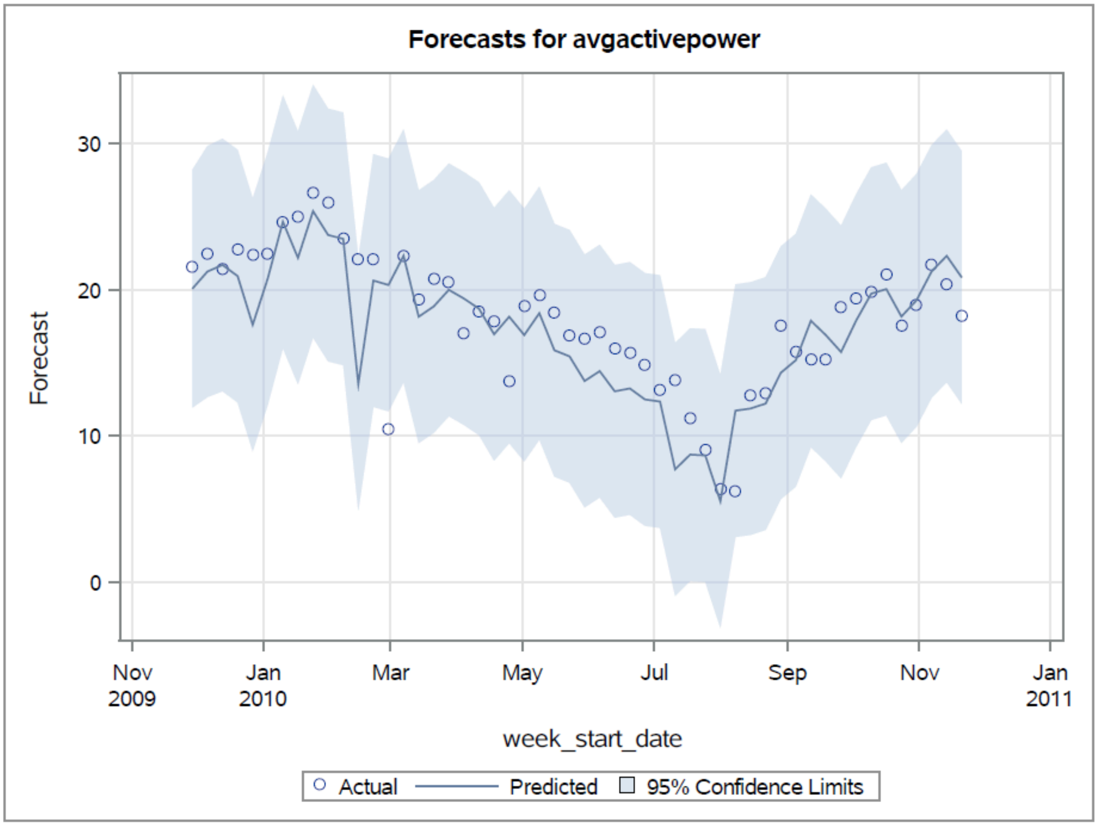
The section contains results from an **SARIMA(1,0,0)(0,1,0)** model analysis. The Maximum Likelihood Estimation table indicates the estimate, standard error, and t-value for the intercept (MU) and the autoregressive term (AR1,1). AR1,1 is statistically significant, as suggested by its p-value.

The diagnostic graphs include the Autocorrelation Function (ACF) and Partial Autocorrelation Function (PACF), which show that the residuals do not exhibit significant autocorrelation at most lags. This suggests that the model is capturing the data's structure well. The Ljung-Box test graph at the bottom indicates the residuals are likely white noise, as the p-values are not significant, which is an indicator of a good model fit.

Maximum Likelihood Estimation					
Parameter	Estimate	Standard Error	t Value	Approx Pr > |t|	Lag
MU	-0.45758	0.50150	-0.91	0.3615	0
AR1,1	0.33889	0.07607	4.46	<.0001	1



Given that this model proved to be superior among all evaluated models, we proceeded with the analysis of its forecasting graph.



```
ods noproctitle;
ods graphics / imagemap=on;

proc sort data=STSM.POWERCONSUMPTION_TEST
out=Work.preProcessedData;
	by week_start_date;
run;

proc arima data=Work.preProcessedData plots
   (only)=(series(corr crosscorr) residual(corr normal)
		forecast(forecastonly));
	identify var=avgactivepower(52);
	estimate p=(1) method=ML;
	forecast lead=52 back=52 alpha=0.05 id=week_start_date
interval=week;
	outlier;
	run;
quit;

proc delete data=Work.preProcessedData;
```

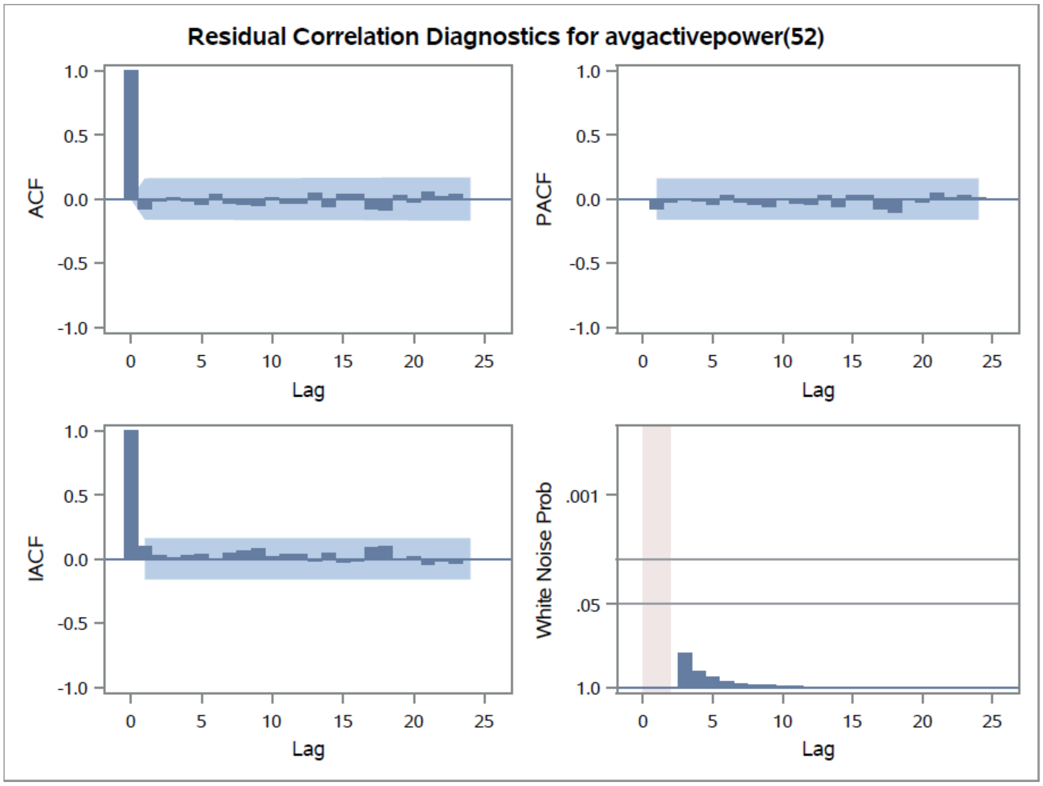
2) SARIMA(1,0,1)(0,1,0)

The image displays the Maximum Likelihood Estimation results and Residual Correlation Diagnostics for an ARIMA model. The parameter estimates include MU, MA1,1, and AR1,1 with their corresponding standard errors, t-values, and p-values. The AR1,1 parameter is nearing statistical significance, suggesting a potential autoregressive effect in the model.

The diagnostic plots, which include the Autocorrelation Function (ACF) and Partial Autocorrelation Function (PACF), show that the residuals are within the confidence bounds for most lags, indicating a good fit. The Ljung-Box test graph at the bottom right suggests the residuals are random, supporting the model's adequacy.

Maximum Likelihood Estimation

Parameter	Estimate	Standard Error	t Value	Approx Pr > |t|	Lag
MU	-0.46552	0.51512	-0.90	0.3661	0
MA1,1	0.06131	0.23835	0.26	0.7970	1
AR1,1	0.39402	0.21929	1.80	0.0724	1




```
ods noproctitle;
ods graphics / imagemap=on;


proc sort data=STSM.POWERCONSUMPTION_TEST
out=Work.preProcessedData;
        by week_start_date;
run;


proc arima data=Work.preProcessedData plots
```

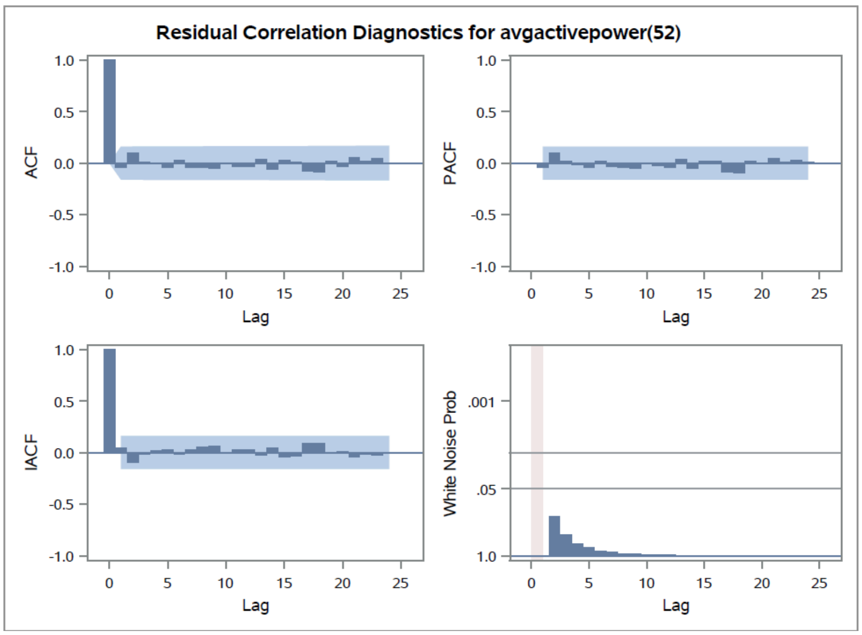
```
   (only)=(series(corr crosscorr) residual(corr normal)
              forecast(forecastonly));
        identify var=avgactivepower(52);
        estimate p=(1) q=(1) method=ML;
        forecast lead=52 back=52 alpha=0.05 id=week_start_date
interval=week;
        outlier;
        run;
quit;

proc delete data=Work.preProcessedData;
run;
```

3) SARIMA(0,0,1)(0,1,0)

The section displays the Maximum Likelihood Estimation results and Residual Correlation Diagnostics for an ARIMA model. The parameter estimates include MU, MA1,1, and AR1,1 with their corresponding standard errors, t-values, and p-values. The AR1,1 parameter is nearing statistical significance, suggesting a potential autoregressive effect in the model.

The diagnostic plots, which include the Autocorrelation Function (ACF) and Partial Autocorrelation Function (PACF), show that the residuals are within the confidence bounds for most lags, indicating a good fit. The Ljung-Box test graph at the bottom right suggests the residuals are random, supporting the model's adequacy.

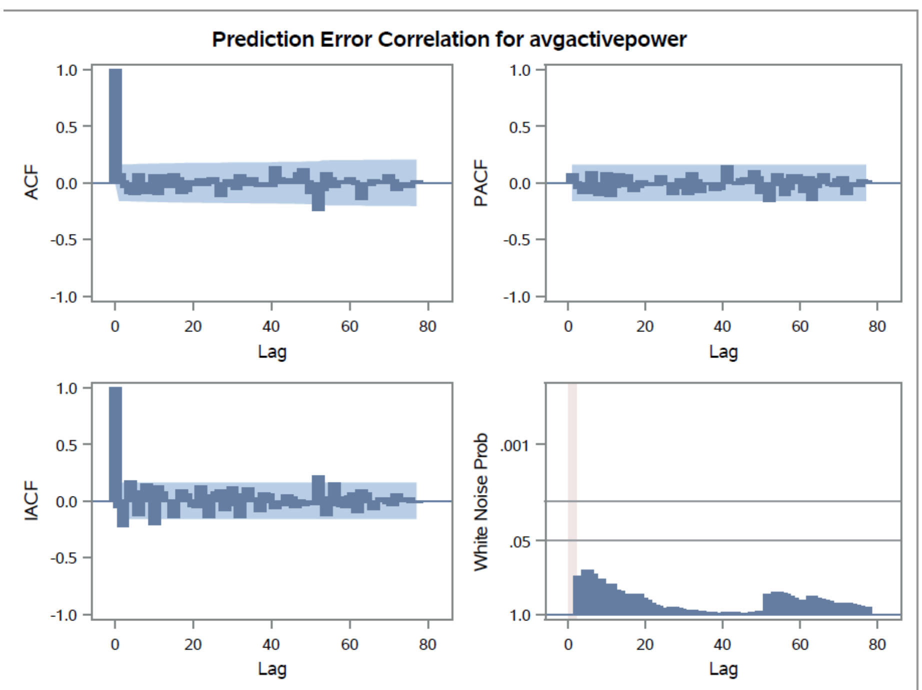
Maximum Likelihood Estimation					
Parameter	Estimate	Standard Error	t Value	Approx Pr > |t|	Lag
MU	-0.41407	0.43363	-0.95	0.3396	0
MA1,1	-0.29733	0.07865	-3.78	0.0002	1



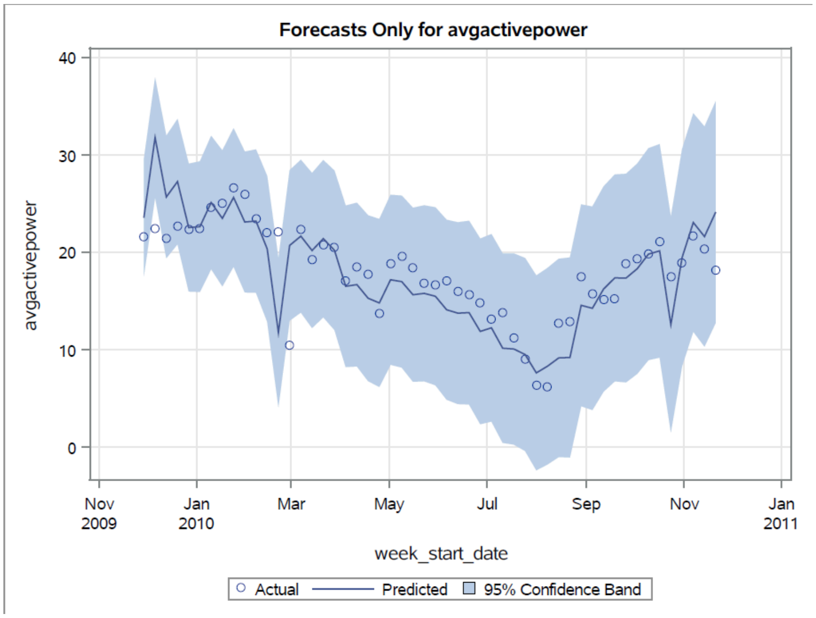

```
ods noproctitle;
ods graphics / imagemap=on;
```

```
proc sort data=STSM.POWERCONSUMPTION_TEST
out=Work.preProcessedData;

          by week_start_date;

run;


proc arima data=Work.preProcessedData plots

   (only)=(series(corr crosscorr) residual(corr normal)

                    forecast(forecastonly));

          identify var=avgactivepower(52);

          estimate q=(1) method=ML;

          forecast lead=52 back=52 alpha=0.05
id=week_start_date interval=week;

          outlier;

          run;

quit;


proc delete data=Work.preProcessedData;

run;
```

ADDITIVE SEASONAL EXPONENTIAL SMOOTHING



Forecasting Graph Analysis for the Additive Seasonal Exponential Smoothing




```
ods noproctitle;
ods graphics / imagemap=on;

proc sort data=STSM.POWERCONSUMPTION_TEST out=Work.preProcessedData;
	by week_start_date;
run;

proc esm data=Work.preProcessedData back=52 lead=52 seasonality=52 plot=(corr
		errors modelforecasts) outstat=work.outstat0001;
	id week_start_date interval=week;
	forecast avgactivepower / alpha=0.05 model=addseasonal transform=none;
run;

proc delete data=Work.preProcessedData;
run;
```