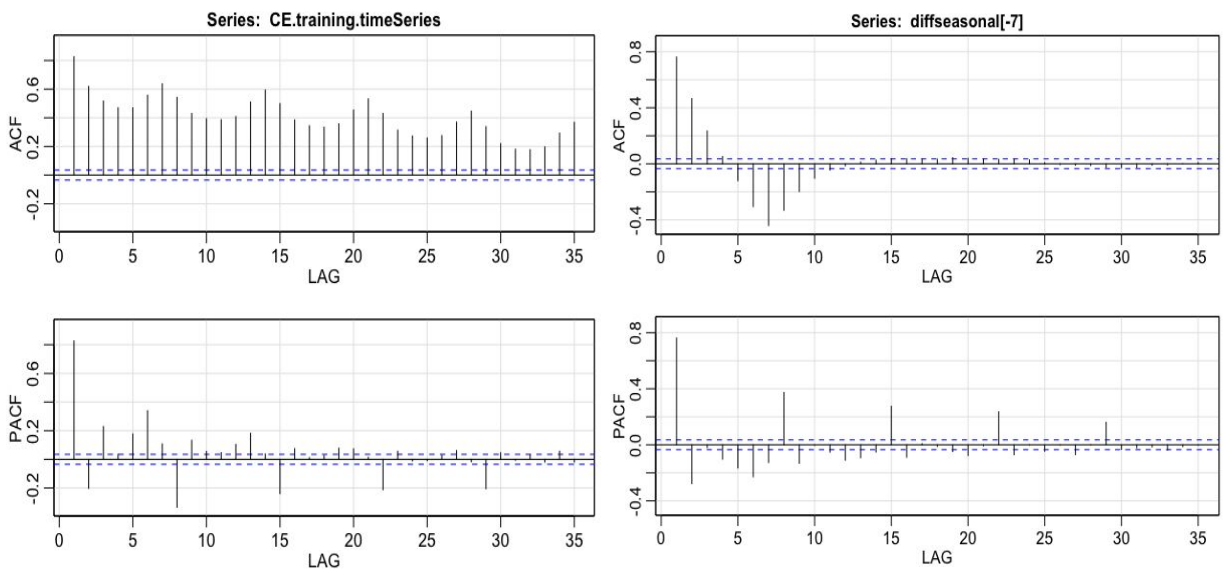
Group 5
Blosher Brar
Junyuan Lin
Miao Hu
Felipe Gonzalez

Evaluation of ARIMA/SARIMA model

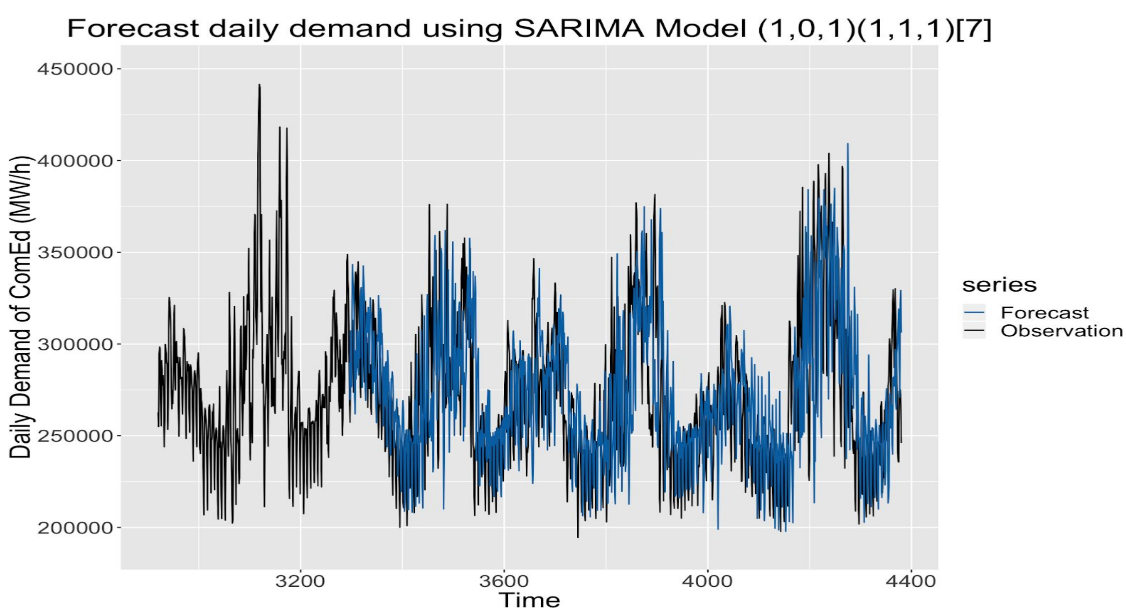
To begin with, we examined the ACF and PACF plot of the original series and found that in our dataset there is a clear 7-day seasonality. In that sense, we decided not to use ARIMA model but to use SARIMA instead given that our dataset is not stationary and has a seasonal component.



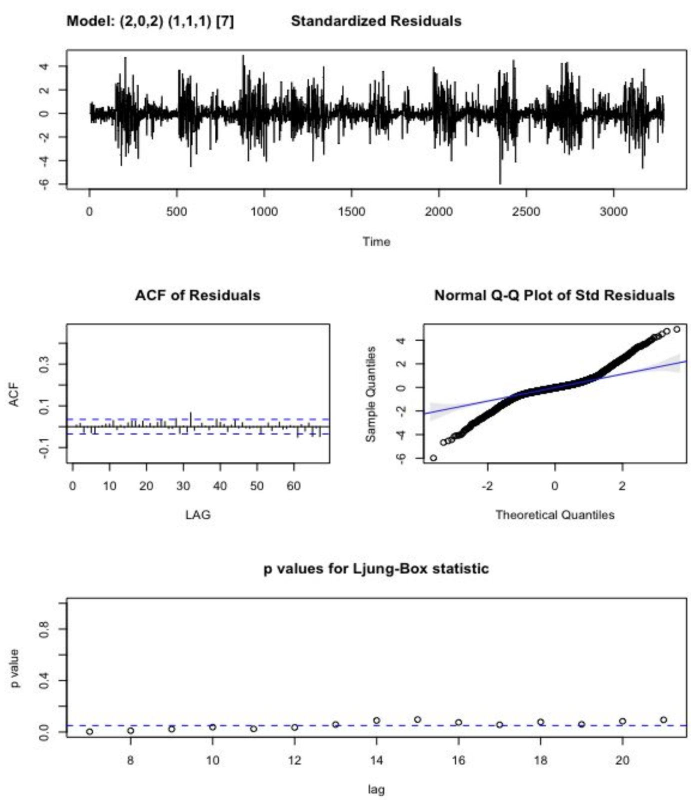
As seen in Figure 1, there is clear weekly seasonality, indicating us to take lag 7 differences. Figure 2 shows the ACF and PACF after taking lag 7 difference, showing a decaying correlation at lag-7's with no 'pollution' around these points. This prompted us to set Q equal to 1 and so our starting point was (0,0,0)(0,1,1)[7].

We then iteratively fitted larger parameters for p, q, P, and Q for the SARIMA model because our starting point diagnostic plot indicated that residuals do not have constant variance, the normal Q-Q plot indicates the errors are not normal distributed, and the Ljung-Box test shows autocorrelation in the errors. However, even after trying various combinations, we still did not manage to pass the Ljung-Box test, have normality in the normal Q-Q plot, and the residuals still did not have constant variance. However, further investigation of the residuals showed significantly larger residuals during the summer time, which meant our model failed to capture the behavior of our data during summer. This prompted us to consider using a hybrid model as we will discuss further below. Our top three SARIMA models in terms of MAPE, AIC, and how far off each model was from our assumptions are shown below in Table 1.

As table 1 indicates, Model 2 gives the lowest MAPE while Model 3 gives the lowest AIC and AICc. However, since Model 1 is the least complex, and the difference in MAPE is only 2.32%, while the difference in AIC, AICc, and BIC is minimal, we choose Model 1 as our SARIMA model. Looking at the residuals, we noted a tendency of the SARIMA models to have increased residuals in the summer months compared to non-summer months. Given this tendency, we considered having 'hybrid' model using another model that performed well during the summer months.



		MAPE	Pbias	Bias	RMSE	AIC	AICc	BIC
Model 1 (1,0,1)	(1,1,1)[7]	4.30%	0.33	78.57	17281	20.553	20.554	19.562
Model 2 (2,0,2)	(1,1,1)[7]	4.20%	0.29	58.93	16909	20.522	20.523	19.535
Model 3 (4,0,2)	(1,1,1)[7]	4.23%	0.22	27.92	16994	20.521	20.522	19.538

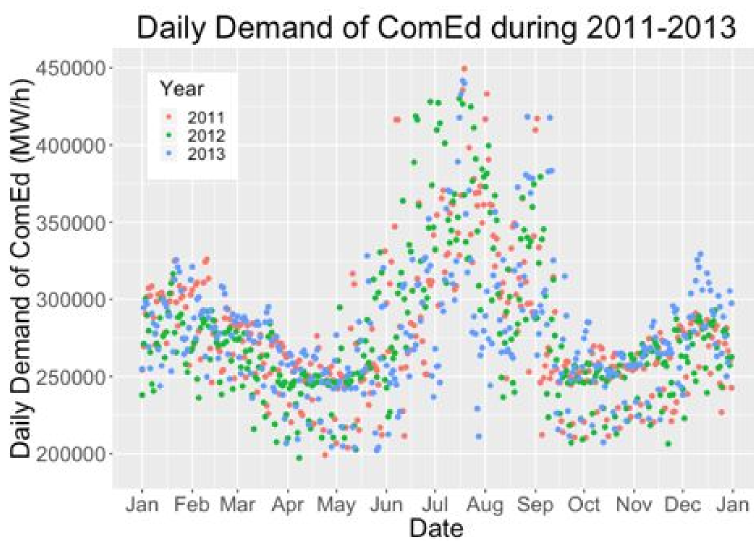


Overall comparisons

Using the Naïve Method (MAPE = 6.41%) as our benchmark, we compared the best models from Extrapolation, Multiple Regression, and ARIMA, by forecasting on the validation set. As shown in table 1, our Regression model had the lowest summer MAPE while the SARIMA model had the lowest non-summer MAPE. As such, we also included a ‘hybrid’ model using a combination of Multiple Regression and SARIMA models.

Summer	Regression	SARIMA	TBATS
MAPE	3.29	5.70	6.16
Non-Summer	Regression	SARIMA	TBATS
MAPE	3.67	3.42	3.77

	Naïve	Linear Regression	TBATS	SARIMA	DSHW	ARX(1)	Hybrid
MAPE	6.41%	3.54%	4.58%	4.30%	5.08%	2.43%	3.20%
Pbias	0.39	0.54	0.23	0.33	0.36	1.17	0.51
Bias	24.5	505.8	-219.4	78.57	292.1	2998.4	654.93
RMSE	24242	11832	17757	17281	20570	8663	12618



We split the data into two summer and non-summer on the basis of figure#. We figured that demand because as we can see, the demand starts climbing beginning in May, peaks at roughly July, and then descends until roughly end of August. As such, we defined summer as beginning from May 1st to August 31st.

As can be seen above, the Hybrid model performed the best on the validation set and therefore chosen as our recommended model to be tested on the test set. One must keep in mind however, that this model failed to pass the normal Q-Q plot, Durbin-Watson Test, and

Ljung-Box test. As such, one cannot create reliable forecasting intervals with this model unless advanced methods such as bootstrapping are used. Moreover, similar to our regression model, this MAPE is best-case scenario since our regression assumed perfect forecasts of our temperature dependent explanatory variables.

We split the data into two summer and non-summer on the basis of figure#. We figured that demand because as we can see, the demand starts climbing beginning in May, peaks at roughly July, and then descends until roughly end of August. As such, we defined summer as beginning from May 1st to August 31st.

Nonetheless, the series seems fairly stationary and the point forecasts can still be useful even without forecasting intervals. Looking at the validation period forecast, we noted a tendency to underestimate during fall and springs months while capturing the summer months with greater accuracy. Given this tendency, we decided to include it in our hybrid model to make predictions in the summer.

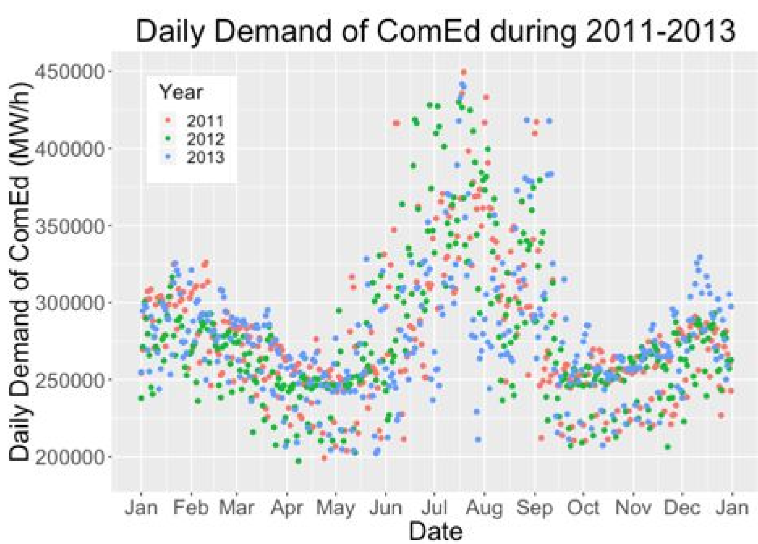
1) ***TBATS:*** This model can handle more than one seasonality, and also accounts for ARMA errors. This method theoretically performs better than other smoothing methods, and in our case, TBATS is the winner in the exponential smoothing categories.

2) ***DSHW***: This method from the exponential smoothing family can handle two seasonalities, a smaller one that repeats itself several times within a larger one, in our case these are weekly and annual seasonality. This method provided forecasts more accurate than the naïve method, however, it was discarded for final implementation since it cannot account for the structure of ARMA errors.

3) ***SARIMA***: SARIMA model can not only handle the seasonality but also accounts for ARMA errors. But in our case, the diagnostic plots showed us that we can't trust the prediction interval, since we can't meet the assumptions for SARIMA model.

When comparing all our models, we noticed that many models do not perform well during the summer, except for linear regression with ARMA errors (The winner among all our methods), although it during non-summer seasons. Therefore, we decided to calculate the MAPE for both summer and non-summer time for all methods in hand.

We split the data into two parts(Summer and non-summer) based on figure#, as we can see around May, the demand starts climbing and the peak ends at around July and then reach the bottom around end of August. So in our case, summer starts from May 1st to August 31st.



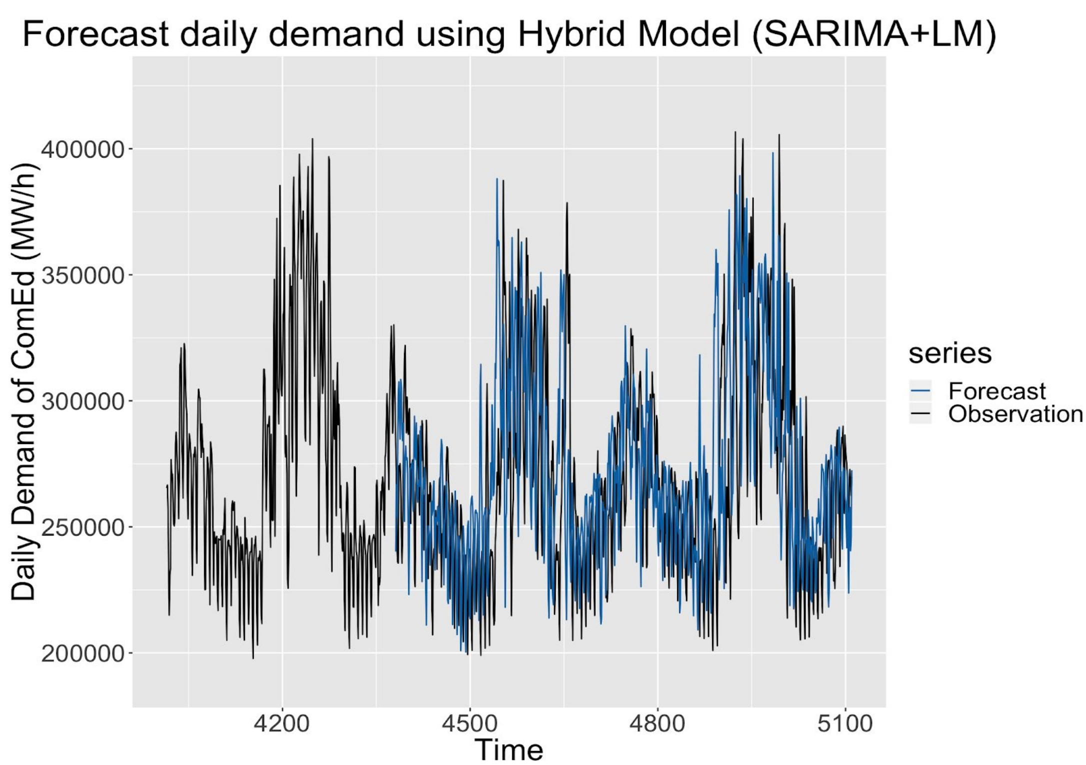
The result is given in table#. According to the table, we can see that Linear regression with ARMA errors performs the best during the summer, but SARIMA model performs better during non-summer season.

As a result, we decided to combine the two methods together, as we use SARIMA from January to April, September to December and use linear regression with ARMA errors from May to August. The overall MAPE for this method is lower than any other method we have in hand.

Recommendation for forecasters

Based on the result we obtained in the previous part, we are recommending the hybrid method to ComEd, as it has the best performance on validation dataset. The performance of this method on the test dataset(Day 4381 - 5110) is given below(Figure# and Table#)

We need to point out that we cannot get the weather data on t+1 when we are on t, so the performance of linear regression method is expected to be and also we cannot use the prediction interval of SARIMA method part as we failed to meet the assumption.



Hybrid Model Test Set	MAPE	Pbias	Bias	RMSE
	3.61%	1.65	3936.66	13618

Conclusion

Across the entire project, we used different methods for forecasting energy demand. The ComEd region in northern Illinois has very marked seasons that define the behavior of energy consumption for every period of the year. As mentioned before, the data also shows a clear weekly seasonal pattern. Those two conditions defined the set of models that we decided to evaluate.