Gaming Isn't Homework: Predicting Demand at the Neighborhood Level

A Technical Paper prepared for SCTE by

Christopher Stump
Senior Engineer, Predictive Analysis
Comcast Cable
1701 JFK Blvd Philadelphia, PA
christopher_stump@cable.comcast.com

May Tan
Senior Manager, Traffic Forecasting
Comcast Cable
1701 JFK Blvd Philadelphia, PA
may_tan@cable.comcast.com

Adam Gruchevsky
Senior Manager, Network Planning
Comcast Cable
1701 JFK Blvd Philadelphia, PA
Adam_gruchevsky@cable.comcast.com

Table of Contents

List of Figures

1. Introduction

Efficient network planning requires accurate forecasting of customer behaviors in the near and long term. The goal is to anticipate the location and magnitude of capacity augments months, and even years, into the future using forecasted monthly bits per second demand (Mbps). This prevents outages, increases reliability and prepares the network to handle faster speeds. However, no longer does a blanket growth rate apply to every neighborhood. Network architecture is transitioning to run fiber deeper into the network and make Service Groups (SGs) smaller, decreasing households passed while increasing speeds. This adds further challenges as smaller SGs are more sensitive to individual household patterns and faster speeds create more usage spikes. As an example, single releases of popular games downloading to a console are enough to create network congestion as are increasingly frequent live sports streaming events.

In the existing analog node world, an office park could hang off the same SG as a housing complex. The difference between day versus night peak, and weekday versus weekend, are washed away due to the size of the SG and augment decisions could be made by the macro trend to consume more traffic at faster speeds. Now, in the digital node world, SGs are smaller. An office park may have a dedicated SG and a residential community may have its own dedicated SG. If we sum their traffic at the headend level, we will certainly see the macro trend, but now the individual SGs have their own driving forces. The office park is driven by weekday, daytime peaks with seasonal dips for vacations and holidays, not to mention pandemics. The residential component is driven by nighttime and weekend peaks that spike heavily in work-from-home situations or when popular games have new releases (and homework lay forgotten). In fact, popular gaming releases can more than double the number of SGs that peak on a single day. With the popularity of gaming and the increase of sporting events dedicated to streaming, these spikes are anticipated to become more frequent. As such, the trends of these SGs cannot be accurately predicted using a headend or national level CAGR (Compound Annual Growth Rate

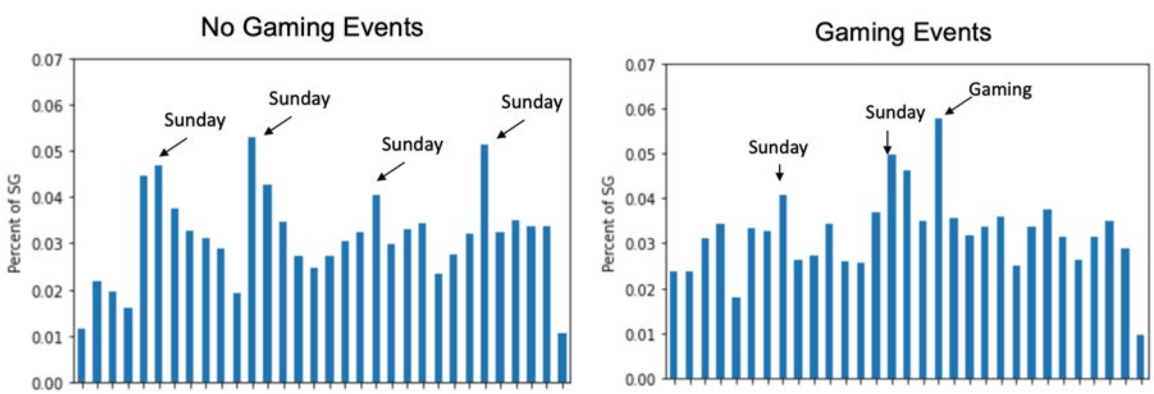

Figure 1 - Gaming is changing when SGs peak

The share of SGs experiencing their monthly peak on a given day is predictable when there are not large gaming download events.

At Comcast we are enhancing our forecast efforts to predict traffic deeper into the network. A small team within the CONNECT organization has worked for over one and a half years to develop a machine learning (ML) solution that produces individual patterns for SGs and Bonding Groups (BGs). The resulting Neural Network uses the growth patterns of SGs and BGs to accurately predict monthly 98th percentile bps regardless of age or technology across the entire network representing hundreds of thousands of individual time series allowing for targeted network planning months and years in advance .

2. Current State

For more than 10 years, Comcast has been developing a comprehensive process to provide network traffic forecasting. Models are developed at the national, regional, and site level. Our access network component model is built to take the site-level aggregate and apply it to each active SG on the network. The model forecasts when SG augments will be required to maintain healthy SG utilization.

At the national level, there are macro level growth analysis models which look at different contributors, drivers, and industry trends. We also evaluate various assumptions and insights across Comcast product categories.

At the regional and site levels, time series focused Machine Learning (ML) has been developed to predict DOCSIS 3.0 (D3.0) and DOCSIS 3.1 (D3.1) traffic growth. We also apply reconciliation and a quality control process to reconcile the bottom-up (site level) model outputs with top-down (national level, regional level) models to ensure consistent forecasting output throughout. A Unified Demand Model (UDM) is the last step in the process to integrate bottom-up and top-down, reconciliation, quality control model outputs, products, device assumptions, and facility information. The output of the UDM model for access network is site level D3.0 and D3.1 traffic.

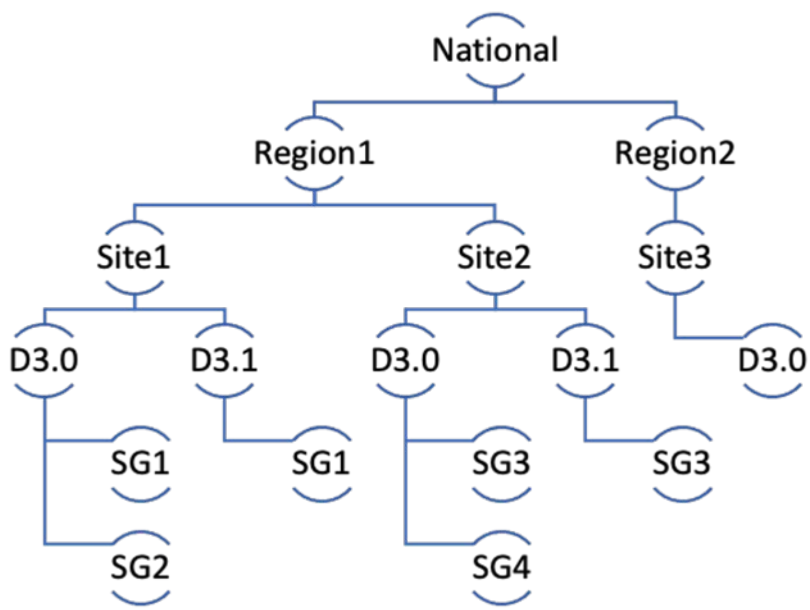

Figure 2 - Forecasting layers

National, Regional and Site level models are reconciled, then product, device and facility assumptions are added through the UDM to produce D3.1 and D3.0 traffic forecasts .

The next layer is SG. However, due to the number of the SGs, the scale and volatility of the data, we were not able to generate SG level forecast from the ML models which were used in site-level forecasting. At the SG level, our traffic growth process applies ratio-based site-level utilization to the individual SG (e.g., SG-01's last known actual bps utilization was 30% of the site-level aggregate, so in the next forecast period we will add 30% of the incremental bps site-level traffic to SG-01). We then look at the number of DOCSIS QAMs currently configured on each SG. If the number is less than the current year's target, we will add D3.0 or D3.1 channels to ensure the SGs will have the targeted capacity for that time period. QAM configurations will also be adjusted based on the amount of D3.0 versus D3.1 traffic. If an augment needs to occur in the model, a new SG is created; traffic is split among old and new SGs, and the traffic growth process continues. The ratio-based traffic growth process applies to upstream SGs as well.

	n98th Mbps Aggregate	
	Site Actual	Site Forecast Month 1
Site A	5000	6000

	Service Group Actual	Percent of Actual	Service Group Forecast Month 1
SG-01	1500	30%	1800
SG-02	2000	40%	2400
SG-03	1500	30%	1800
Total	5000		6000

Figure 3 - SG traffic allocation

	n98th Mbps Aggregte	
	Site Actual	Site Forecast Month 1
Site A	5,000	6,000

	Service Group Actual	%'age of Actual Mbps	Service Group Forecast Month 1
SG-01	1,500	30%	1,800

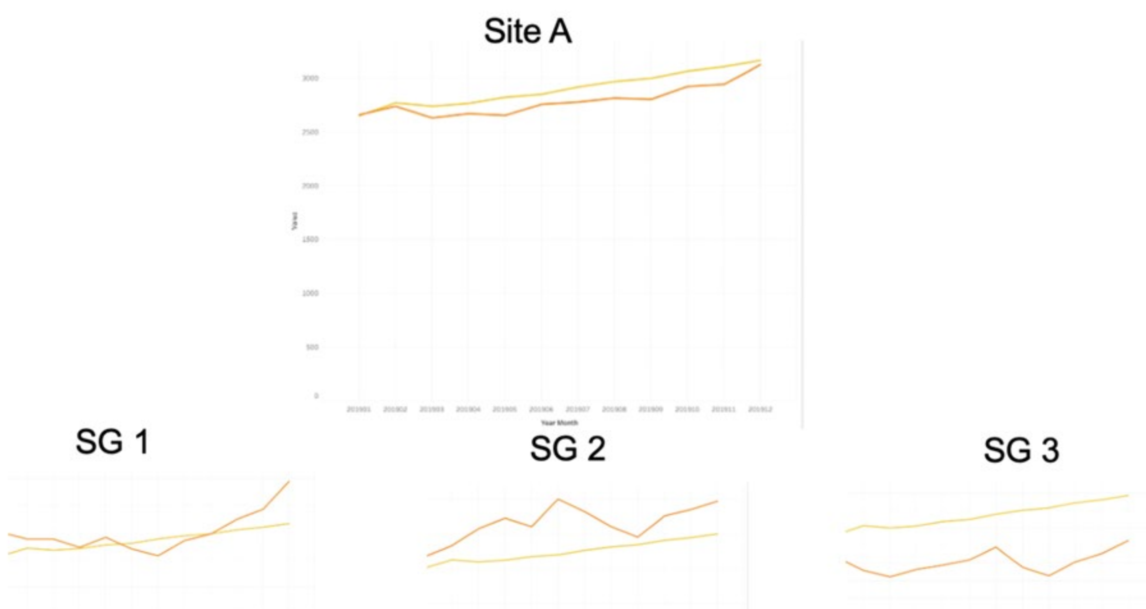

Figure 4 - Site to SG example

The site level forecast is very accurate, but when applied to the SGs underneath it the pattern no longer holds This is because each individual SG has its own growth drivers and the smaller sample size leads to higher variance in the pattern.

This network simulation method has effectively helped with our long-range planning and budget needs for the next calendar year. The challenge comes as we get closer to

and into those budget years – how do we help our partners in the field find the right SGs to augment six to twelve months out? Our site-level-aggregate forecasts are only run at certain times of the year which leads to stale data. Customers churn, and release events for gaming and streaming video occur, leading to site-level forecasts that do not translate to in-year planning for the individual SGs. As we focus on proactive network augments, how can we direct construction teams to the right areas before utilization exceeds capacity and impacts customers?

3. Model Evolution

There is currently a time-series solution at both the regional and site (headend) levels. These solutions use ARIMA (Autoregressive integrated moving average) and Exponential Smoothing models to produce results for each unique region and site. So, the first step in the SG modeling effort is to see if there is an application of what we are already familiar with at the SG level. Using R on Spark and later Databricks we built a solution that produces unique ARIMA and Exponential Smoothing Models for each SG. This model produces results for all SG regardless of age and the sMAPE (Symmetric mean absolute percentage error) for the DS traffic was improved.

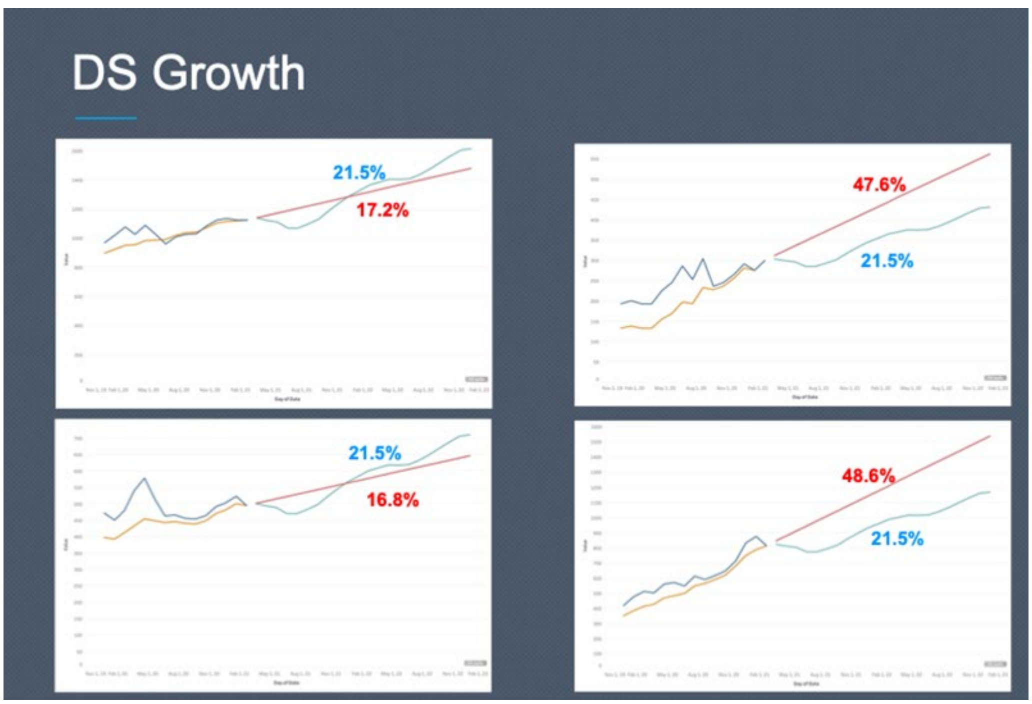

Figure 5 - Time Series model results

This demonstrated improvement in the forecasting process, but the model still did not produce a scalable solution because: (1) the sMAPE nationally was not improved, (2) the model behaved very poorly for SGs with inconsistent trends, and (3) it took 28 hours to run on Spark. So, our first pass was not our final answer, but we did learn how to handle the massive data sets and gained an appreciation for variance in the observed data.

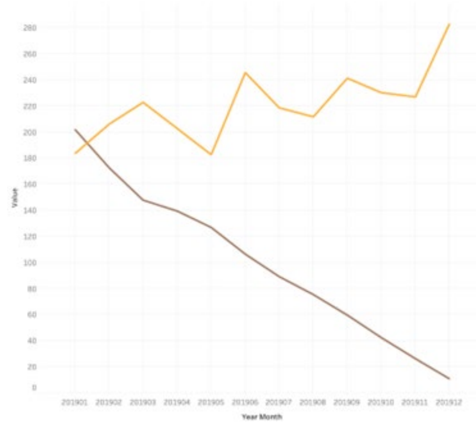

Figure 6 - Time Series model issues with irregular patterns

The next step was to explore more advanced machine learning models. Based on the deficiencies mentioned above, we needed a solution that could scale, handle high variance, and run in less than a day. We experimented with multiple options of ensemble methods and Neural Networks and landed on a Sequential Neural Network. We trained the neural network on 2 years of raw bps data per SG. The results included improvement in the sMAPE for downstream (DS), D3.1 and upstream (US) demand.

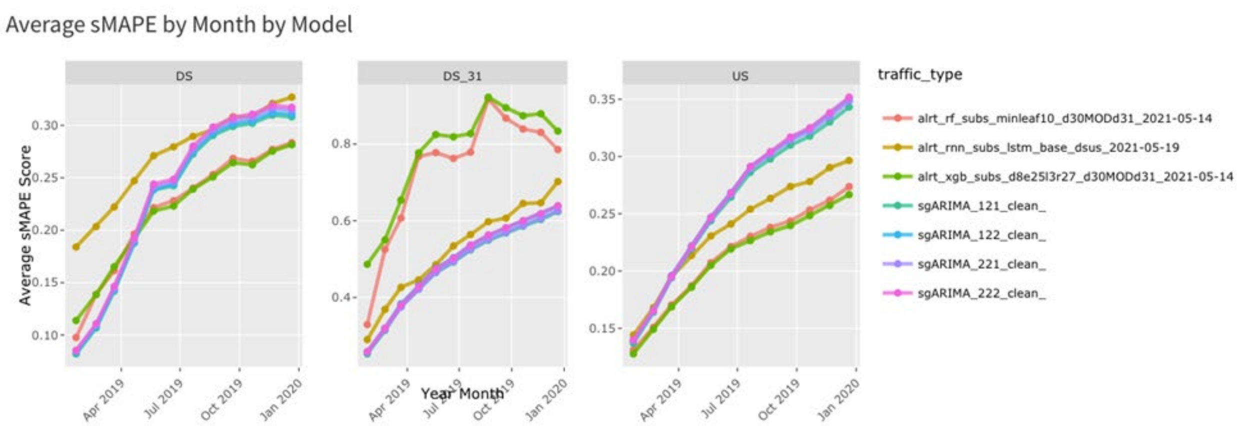

Figure 7 - sMAPE of NN on bps

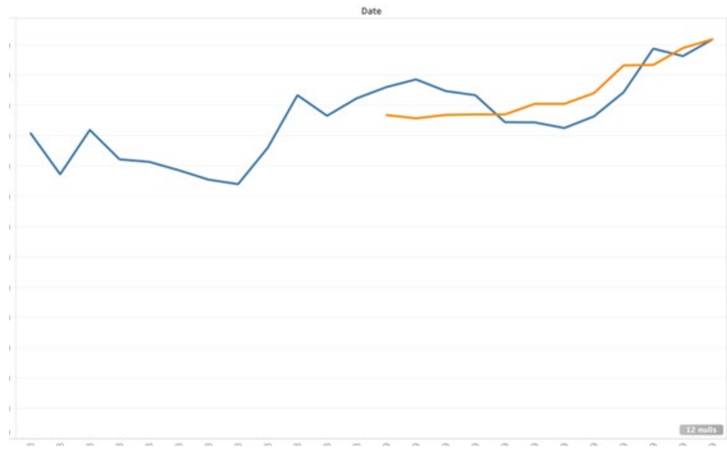

Figure 8 - NN on bps example output

However, once again certain inconsistent trends, like SGs being collapsed and seasonal patterns did not perform well. Furthermore, the model could only be run on SGs with perfect data. This meant that we could not produce forecasts for SGs with data issues or seasonality, nor provide guidance on newer SGs including traffic over RPDs (Remote PHY Devices) on the new vCMTS locations that only have a few months of observed data.

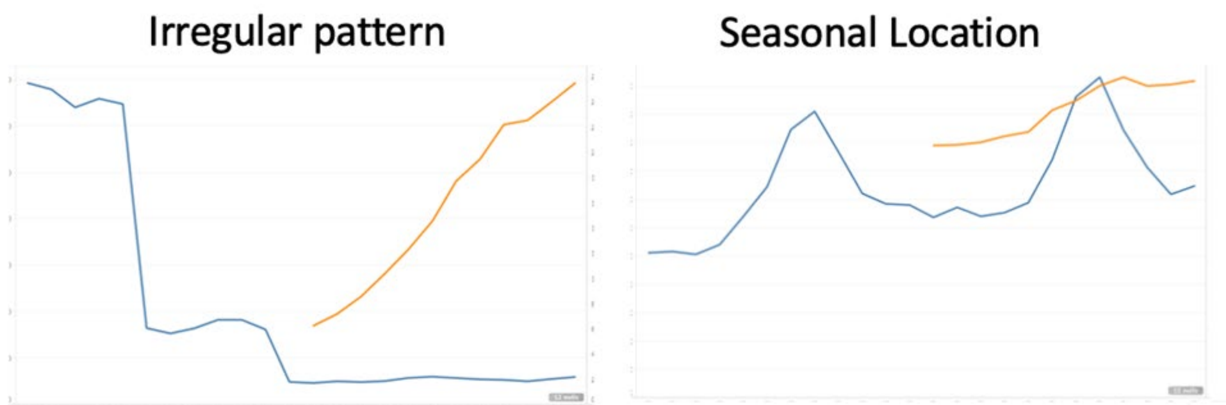

Figure 9 - Where NN on bps does not forecast well.

Collapsing sites and seasonal locations did not produce valid models.

To produce a model that meets our needs we need improved accuracy, fast runtime, and solutions for all SGs regardless of age. The discussion then veered to isolating problematic SGs since they were the gap in the Neural Network solution. The largest class of problematic SGs is seasonal locations. That is, SGs serving a geography that is primarily a vacation area such as shore and mountain locations. Identifying an appropriate model for these SGs first requires identifying them at scale. We currently identify site level seasonality through a mix of time series analysis and exception lists developed over years of site level modeling. However, seasonal sites can have non-seasonal SGs, and we need to be able to capture that in an automated way.

We know a seasonal site has a ‘look’ when you are scanning visually - you can see the peaks and valleys. This pattern can be described as a monthly growth rate with high positive growth in the season ramp-up followed by large negative growth as the season ends. Using this approach, we can put all the SGs on the same scale - size will not be visible. We can identify similar growth patterns and find seasonal locations. We can also more universally describe all SG growth regardless of size and pattern. This highlights customer behavior as opposed to size driven by engineering decisions. We can increase forecast accuracy by feeding all SG data into the model as a monthly growth rate. Since the scale of the input is the same, the neural network can focus on learning a pattern, not both size and pattern.

4. The Model

Zooming back out to the national view we can create monthly growth rates from the monthly 98th traffic for each SG/BG. To accurately predict seasonal trends, it is ideal to have 24 monthly growth rates. Additionally, we will need a target growth rate for our first prediction month. This means that we need 26 months of data to produce 25 growth rates. Many of our SGs have missing data so we can fill in those gaps with 1s. We also have the issue of hyper-growth or data driving high monthly increases. Although it is impractical to perfectly clean a huge data set, we can cap monthly growth and decline to anywhere between 15%-50%. These options seek to balance variance vs natural seasonality. The most recent month is used as the target and the previous 24 months are the features. We then train that on a sequential neural network with 4 layers. Twenty-four months of data predicts 1 month. To reach a full year, we must modify the model input to include recent forecasts as features in the trained model.

P_month1: X includes Growth Rate 1 - 24

P_month2: X includes Growth Rate 2-24, P_month1

P_month3: X includes Growth Rate 3-24, P_month1, P_month2

These growth rates are then restated as bps by referencing the most recent month's bps and multiplying by the following month's predicted growth rate.

Once the result is stated as bps, sMAPE is used to guide model tuning and showed that breaking out the model by geography (North, South, East, West) and direction (downstream, upstream, OFDM) produced the most accurate result. Hyperparameter tuning used sMAPE and Loss to select layer count, optimization functions and epoch count to optimize the model.

The resulting model predicts 12 months of traffic for each SG and BG regardless of age.

Figure 10 - NN on growth rate results

To determine if this is a superior model to our previous SG model iterations or the legacy solution, we once again look at the sMAPE which shows that the NN produces the lowest error of any model for both the US and DS. So, this NN solution produces a more accurate model for all SG/BG regardless of age and can be run in a few hours. We can now move to the step of implementing these results in network planning models to inform future budgeting.

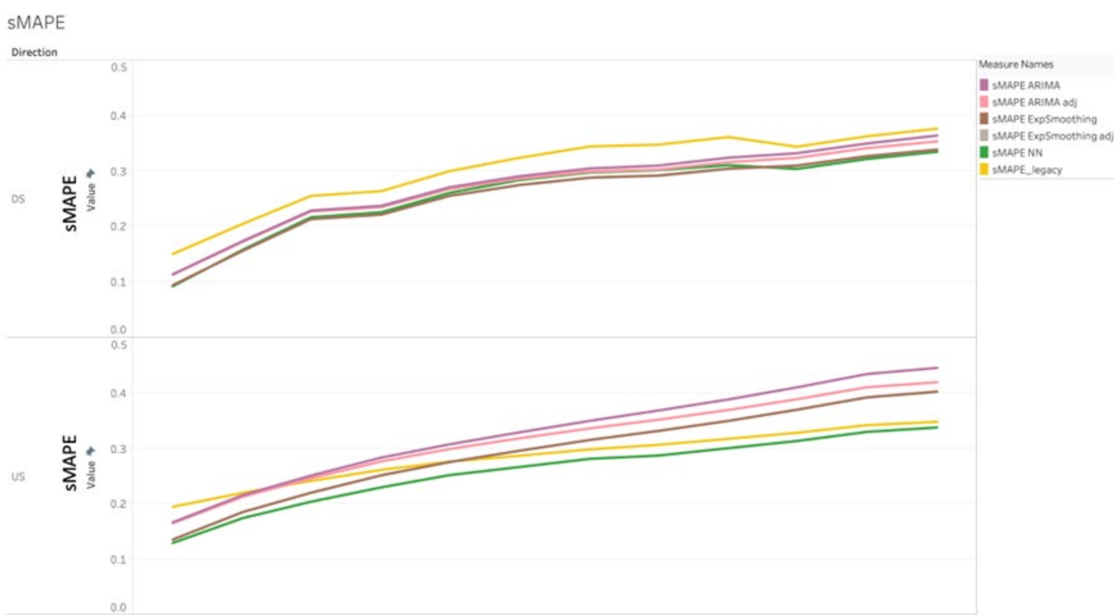

Figure 11. sMAPE graph comparing all models.

Lower is better for sMAPE. Across the time forecasted the NN using growth rate (green) outperforms all other models.

5. Future State

We now have a working time series model. Our immediate next step is to insert this model into the network planning process and facilitate decision-making from its output. We will have to slightly re-factor the network planning models as well as set up a quality control process. The quality control process will have multiple visualizations in 2 groups. The first group will be based on model accuracy to show how close we are in aggregate and compare model options to target the lowest error rate. The second group will highlight SGs with high variance and questionable forecasts.

Once we have a working model, we will continue to improve the neural network. We are striving to get more accurate with seasonal sites by incorporating Holt Winters seasonality in those specific locations. We are also looking at options to add other data into the model including geography, node type, customer mix, device type, and SG age. We will measure these additions by the level that sMAPE and MAPE improve when compared with the historical time series.

Additional improvements to the algorithm have already been identified. Currently, we are producing a point estimate for each SG for each month. In the future we can produce a range of values based on the probability of scenarios. Additionally, we can use this information to inform probabilistic risk of needing a split for each SG within a period. Being able to define algorithms and probabilities at this level also allows us to build scenarios using optimization modeling techniques. Improved forecast accuracy, probability scenarios and split risk can be extended to build cost functions for needing a split. We can then enter the next frontier of establishing the cost of NOT splitting by adding customer experience metrics like speed tests, latency, and packet loss. At which point we will have a set of ML driven functions defining risk of split/non-split which can be solved using a Linear Program.

6. Conclusion

Growth in internet demand is currently driven by residential gaming, streaming video, and business applications. To ensure the highest customer experience we need to forecast network traffic to ensure we maintain sufficient capacity to meet user demand. In the past national, regional and site level forecasts were sufficient and development cycles were dedicated to them. Now, in the digital node world we need more accurate and granular forecasting to the neighborhood level, using ML to predict trends at the SG level.

Our team has developed an ML solution that has demonstrated improvement in time series accuracy. Additionally, it can handle new SGs regardless of size because it relies on monthly growth rates instead of bps to predict demand. Currently we are working closely with our internal partners to add this level of detail to the network planning process. This includes visualizing results to both aid model selection and highlight SGs that need further analysis.

A working model that aids financial budgeting and long-range planning decisions will mark the completion of our first phase. Next, we need to enhance the accuracy of our forecasting models using seasonal models and include additional data about the SG's behavior. This will allow us to explore other modeling techniques to cluster based on device, geography, or customer behavior. A further goal is to incorporate optimization modeling techniques that build on our ML models to maximize both the customer experience and financial performance.

We are in the beginning phases of an evolution in forecasting and network planning. Our combination of ML, time series analysis and collaboration will ensure our success at accurately predicting network demand so that all games can continue to be streamed and homework can (eventually) be done.

Abbreviations

SG	Service Group
BG	Bonding Group
ML	Machine Learning
DS	Downstream
US	Upstream
bps	bits per second
DOCSIS	Data over cable service interface specification
UDM	Unified demand model
ARIMA	Autoregressive integrated moving average
sMAPE	Symmetric mean absolute percentage error
vCMTS	Virtual cable modem termination system
D3.0	DOCSIS 3.0
D3.1	DOCSIS 3.1