A PNM System Using Artificial Intelligence, HFC Network Impairment, Atmospheric and Weather Data to Predict HFC Network Degradation and Avert Customer Impact

A Technical Paper Prepared For SCTE/ISBE By

Larry Wolcott
Comcast Fellow, Next Generation Operations Technology
Comcast
1401 Wynkoop Suite 300, Denver, CO 80202
(303) 726-1596
Larry_Wolcott@cable.comcast.com

Michael O'Dell
Director, Network Maintenance
Next Generation Access Networks
Comcast
215 East North St, New Castle, PA
(724) 856-3074
Michael_ODell@cable.comcast.com

Peter Kuykendall
Principal Engineer
Video Infrastructure and Security Engineering (VISE)
Comcast
4100 E. Dry Creek Road, Centennial, CO 80122
(720) 663-7581
Peter_Kuykendall@cable.comcast.com

Vishnu Gopal
Senior Engineer
Software Dev & Engineering, Engineering & Operations
Comcast
183 Inverness Drive West, Englewood, CO 80112
(303) 658-7049
Vishnu_Gopal@cable.comcast.com

Jason Woodrich
Senior Engineer
Software Dev & Engineering, System Engineering
Comcast
183 Inverness Drive West, Englewood, CO 80112
(303) 658-7123
Jason_Woodrich@cable.comcast.com

Nick Pinckernell
Senior Principal Researcher
Technical Research and Development, Product & AI
Comcast
183 Inverness Drive West, Englewood, CO 80112
(303) 658-7305
Nicholas_Pinckernell@cable.comcast.com

Table of Contents

Title **Page Number**

List of Figures

Title	Page Number

Introduction

Proactive network maintenance (PNM) has become a cornerstone technology within the Data Over Cable Service Interface Specification (DOCSIS®), providing tremendous benefit to cable operators and their customers. From adaptive equalization to full band capture, a rich and extensive data model exists to proactively maintain our valuable networks and reduce operational costs. However, due to the complexity, financial and cultural barriers, many operators have been unable to gain traction with implementing such systems. This paper will examine PNM capabilities, weather information, artificial intelligence, machine learning, operational practice and financial implications to provide a meaningful approach for implementation.

Physical networks that are exposed to the environment are subject to environmental influences which can affect their performance, especially, but not limited to, coaxial and HFC networks and also including fiber optic, satellite, and wireless networking solutions. This is due in part to physics and the simple fact that physical things respond to environmental conditions in predictable ways. Some of the most common factors that influence these networks are atmospheric conditions such as heat, cold, wind, rain, snow, humidity, and freezing. The core premise of this work is to correlate the predictable nature of weather and seemingly unpredictability of weather related outages. While it is generally accepted that these things affect the network, very little work has been done to quantify in an empirical way. Thus, having predictable correlation with faulty network elements allows for operators to proactively prioritize and repair the problems before they impact customers.

Further, by understanding the environmental influence throughout the life span of a network, better decisions can be made about design and construction. This allows operators to evaluate the true cost of ownership over the entire life of the network including support and maintenance, which are generally very expensive.

Acknowledgements

This work would not have been possible without a significant contribution of the individuals listed as authors and contributors. This unique collaboration was facilitated by the Comcast CLEAR engineering program, of which we are all grateful. Thank you to Andrew Frederick, Denice Loud, mentors, and the CLEAR committee members for creating this opportunity.

The Society of Cable Telecommunications Engineers (SCTE/ISBE) has also been instrumental in the ongoing support and promotion of PNM. Especially Chris Bastian, Dean Stoneback, the network operations subcommittee for proactive network maintenance and its members.

Finally, but not least important, special thanks to CableLabs, Alberto Campos, Tom Williams, Jason Rupe, Robert Cruickshank and Curtis Knittle for the steadfast industrial support of PNM. Without their support and innovation, PNM would not exist. Also, Justin Menard and Wil Colon (retired) from Comcast for their previous work on the impact of temperature to the customer experience.

History and Background of Environmental Influence on HFC

Weather has always informed CATV plant maintenance activities. From the earliest days of antenna systems, to more modern hybrid fiber coax (HFC) architectures, atmospheric conditions have always had an influence on the performance and maintenance of CATV networks. Heat, cold, wind, rain, snow and ice. Wherever there are cable network components exposed to the open air and elements, operator's behaviors, and network performance will be influenced by some aspect of weather. All telecommunication and utility teams have developed maintenance cultures and "tribal knowledge" that have been influenced by localized weather patterns. These cultures are ingrained in the practices, habits, and annual maintenance timelines as defined by the historical weather cycles.

Looking at the earliest CATV systems like the ones built by Ed Parson in Astoria Oregon, or Bob Tarlton in Lansford Pennsylvania, their antenna structures were necessarily placed in elevated positions to get the best possible line of site to broadcast antenna locations. Those elevated positions can be subject to winds, moisture, and temperature changes requiring maintenance to keep them properly aligned, and the RF connectors tight and free from corrosion. The maintenance of the transmission lines from those elevated positions, connecting the receive antennas to the community served, are also subject to the influences of weather. Coaxial or other cables ran on aerial pole structures can be influenced by wind, temperature changes, tree limbs, snow and ice loading, and other weather influenced circumstances. Fast forward to newer networks, which carry a broader array of services and signal types, and while they are constructed with more modern materials and components, they are still significant influenced by localized weather.

Satellite dishes, microwave links, long amplifier cascades, fiber and coaxial cables, and distribution passive components: Though the components may have changed, the behaviors remain the same. Northern climates or higher altitudes that are subject to heavy wet snow in the winter have adapted their maintenance cultures to deal with the accumulation on aerial cable spans and satellite dishes. Shaking spans and sweeping dishes in the winter have become common practice born through necessity over time. Nor'easters, in the New England states, lake effect snows from the Great Lakes in the upper Midwest, and wildfires in the dry seasons are all localized weather phenomena that have informed maintenance behaviors. High Plains states can see temperature changes exceeding 30° Fahrenheit from daytime highs, to nighttime lows. Mountainous areas have developed procedures to deal with flash flooding, when heavy rains run off the hills and pool in low lying areas. These areas can experience mudslides/hill slides when there is insufficient foliage cover to hold topsoils in place during the runoff. Other climates, such as Florida, have developed their own cultures based on their own localized weather influences. The peninsula of Florida experiences large volumes of lightning strikes, locally heavy thunderstorms, and tropical storms such as hurricanes on an annual basis. Plains states experience annual tornadoes and high wind events that can be devastating to utility and telecommunications infrastructures. Across the vast and varied landscapes of America, maintenance behaviors have always been informed by weather.

Evolution of HFC Network Architecture

The earliest systems, like those previously referenced in Astoria, Oregon and Lansford, Pennsylvania, were designed and constructed to retransmit or extend the reach of over-the-air television signals to communities that didn't have a direct line of site to the television broadcaster in the community or adjacent community. Whether because of geographic terrestrial land formations, buildings, or simply distance, the available over-the-air broadcast signals were not sufficiently strong to provide for good reception in the community of interest. From the 1940s through the 1960s, these types of (re)transmission systems were built out to serve the growing video entertainment appetite of the country.

The evolution of cable television systems infrastructure naturally followed the public right of ways, and previously established placement and attachment routes developed by the telephone and electric companies. Telephone companies had already been deploying services for nearly 50 years before CATV systems were deployed. The first transatlantic telephone lines were laid in the mid 1950s[1]. Aerial cable attachments to telephone or electric service poles are still the most common coaxial cable placements, however underground facilities have become increasingly more common.

The components used in those early CATV delivery systems were relatively few and simple. There were antennas, twin lead or coaxial cables, combiners, RF amplifiers, and distribution splitters or passives. The fledgling bandwidth limited systems often only carried the 12 VHF (Very High Frequency) channels, and were building the foundational engineering practices necessary to intermodulate video signals and transmit them over long distances while maintaining signal quality and integrity. As the popularity of, and demand for "Cable Television" grew, providers expanded the reach of the one-way video delivery system into what would become the Tree-and-Branch architecture.

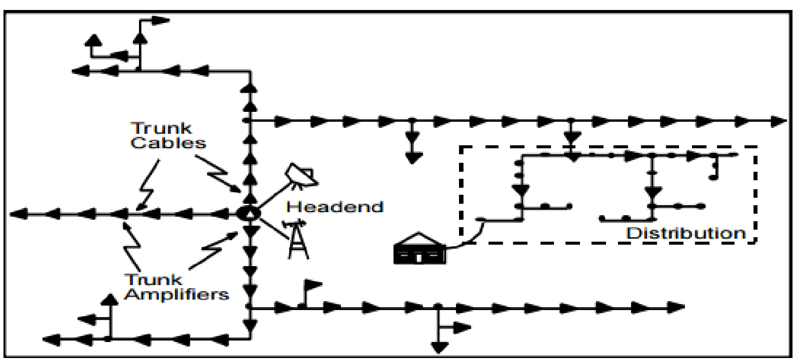

Figure 1 – HFC Amplifier Cascade

The tree-and-branch architecture was comprised of a signal receiving and processing center, called a headend, trunk cables and amplifiers, distribution cables, bridging and line extending amplifiers, distribution passives and "taps" as the drop cable system to deliver the signals to the subscriber television, or terminal. Trunk cable and amplifiers were built to carry the signals long distances from

[1] History of the Telephone - Jason Morris

headend with minimal signal degradation due to amplification. The bridger amplifiers and distribution systems then deliver the signals the "last mile" to subscribing homes via taps and drop cables. Trunk amplifiers can be added one after another in a "cascade" of amplifiers to extend the reach of a cable system, provided that proper engineering and plant management practices are followed to maintain quality signals all the way to the last subscriber terminal. Amplifying intermodulated signals has some potential consequences, however. There are limitations to how many amplifiers can be "cascaded" before signal quality deteriorates to an unsatisfactory level. This tree-and-branch style of architecture was the prevailing build preference for one-way video delivery systems for many years.

According to Roger Hughes, Director of Plant Architectures and Technologies at Armstrong Group of Companies, "*upgrading the amplifiers to push-pull hybrids allowed us to expand the channel plan from 12 channels, up to 22 channels, and extend the reach of the cable plant to longer cascades. Improvements in amplifier technology allowed us to push the bandwidth out to 450 MHz, and cascades longer than 50 amps. That bandwidth expansion resulted in the ability to offer additional video services to our customers, drive penetration, and increase revenue streams. Then in the mid-1980's fiber optic cables allowed us to cut that cascade in half, and reduce the impact of temperature and weather to that extended trunk cascade*." In his 40 years with Armstrong, Huges has seen the architecture grow, and then begin to retract, with the direction of his organization heading toward reduced cascades, smaller data-driven service groups, and ultimately toward passive optical networks (PON).

Cable television programming expanded dramatically in the late 1970s with the introduction of specialized content providers using satellites to make their channels readily available to a great many CATV operators. That technological advancement started a race for expanded downstream bandwidth, and the ability to offer such premium channels as Home Box Office (HBO), the Entertainment and Sports Programming Network (ESPN), and Turner Broadcasting System (TBS). These expanded programming options required more bandwidth to deliver, and upgrades to the cable network amplifier technology. They did not, however, require significant changes to the tree-and-branch architecture, nor the depth of cascade.

The next significant technological advancement to influence cable architecture and design was the introduction of internet services over cable. Cable (internet) modems and cable modem termination systems by companies like Lan City, Com21 and Zenith required that cable systems have operational return paths, and be segmented geographically into smaller footprints or "service groups." [2] Fiber optic technology, while still fairly expensive, was already deployed, and becoming more commonplace, and the tree-and-branch architectures were being redesigned using cascade reduction techniques. Fiber optic cables were run toward the further reaches of the cable footprints, along the same routes as the trunk cables, and connected to fiber nodes. These nodes allowed for the long distance transmission of video signals with low losses from attenuation, and reduced distortions attributable to repeated amplification. It also allowed for return path signals to be sent back to the headend, and smaller two-way footprints, or nodes to be established. In this manner, amplifier cascades of 50 or more amplifiers could be reduced to fewer than 10 amplifiers in a single cascade. In 1997, the first iteration of the Data-over-Cable Interface Specification (DOCSIS®) was released, which provided a set of specifications from which vendors could design and build cable modem termination systems (CMTs) and cable modems that would be interoperable for Multiple System Operators (MSOs). DOCSIS® Architecture model: Figure 2.

[2] How DOCSIS Revolutionized the Cable Industry, 2016 - The Volpe Firm, Brady Volpe

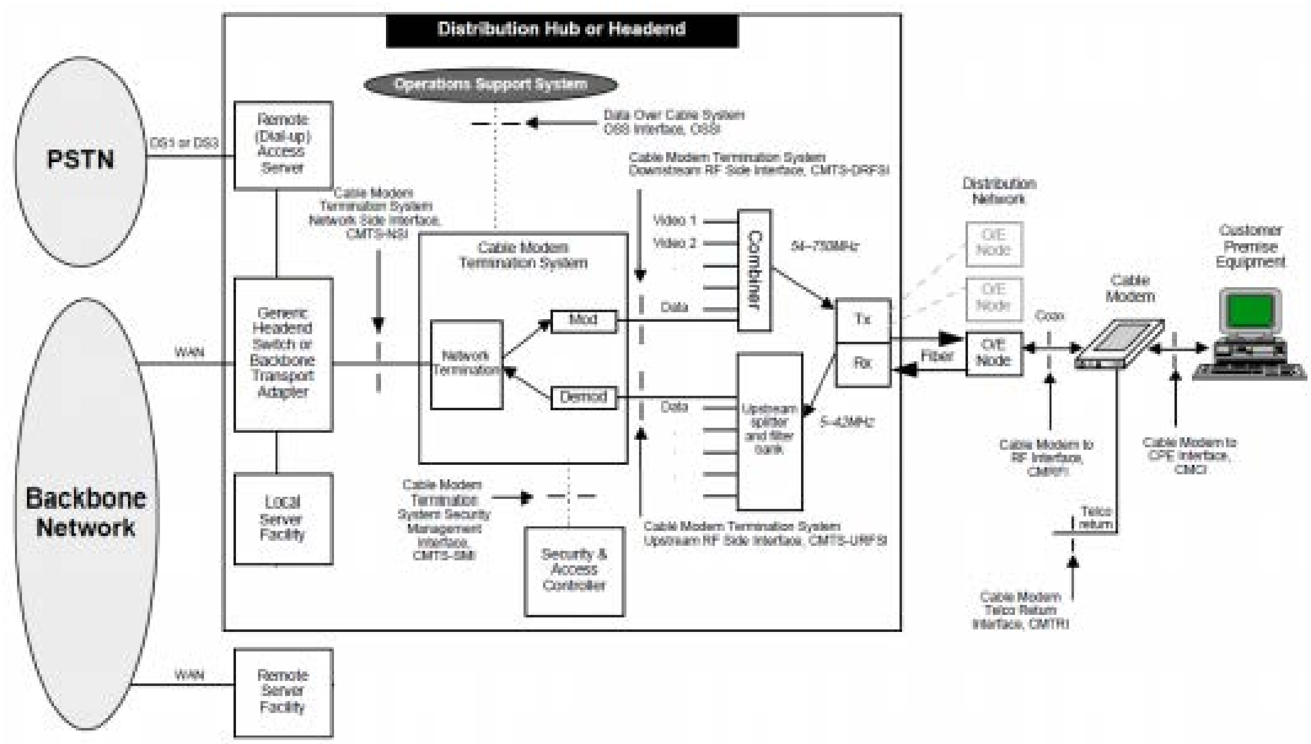

Figure 2 – HFC Logical Digaram

The DOCSIS® specified services, such as high speed (broadband) internet and Voice over Internet Protocol (VoIP) became significant income streams for the MSOs, provided that their network or node configurations were built to support the customers' increasing need for broadband services. This data-driven model currently informs both node size and amplifier cascades in the hybrid fiber coax (HFC) systems. From headend-fed, tree-and-branch architectures with 50+ amplifier cascades, to fiber nodes with nary a single amplifier, CATV systems continue to evolve to meet the needs of the subscribers they serve.

Traditional Network Maintenance Practices

As described by Dr. Walter Ciciora in the CableLabs paper *Cable Television in the United States - An Overview*, "Cable television is made possible by the technology of coaxial cable. Rigid coaxial cable has a solid aluminum outer tube and a center conductor of copper-clad aluminum. Flexible coaxial cable's outer conductor is a combination of metal foil and braided wire, with a copper-clad, steel center conductor. The characteristic impedance of the coaxial cable used in cable television is 75 ohms." [3] While there have been improvements in the materials and manufacturing processes, the fundamental physical and electrical properties have remained largely unchanged. As such, the traditional network maintenance practices used to find and fix impairments on coaxial cable networks have also remained largely unchanged. Many of the symptoms of service degradations in the coaxial network can be boiled down to a very few actual causes. Impedance mismatches, shielding integrity issues, and noise are among the chief contributors to service impairments in coaxial networks.

[3] Cable Television in the United States - An Overview, 2005 - Walter S Ciciora, Ph. D.

Active sweep and balance, and signal leakage detection remain extremely valuable practices in any architecture type that still relies on coaxial cable as the primary transmission medium. More recently, DOCSIS® protocols include Proactive Network Maintenance (PNM) information derived from Customer Premise Equipment (CPE), which can be included in the symptoms and can provide very detailed information on where network maintenance is required. PNM data can inform where there are reflective cavities from impedance mismatches, frequency suckouts, and standing waves, among other things. These symptoms are indicative of the need to visit the coaxial portion of the plant, and restore the physical and electrical properties of the cable plant.

Beyond the find and fix (Demand Maintenance) functions of a network maintenance team, there are, and have been, preventive maintenance functions that are crucial to the quality service delivery of a cable system. Chief among these, sweep and balance has been integral to keeping the forward and return portions of the plant operating properly. As further explained in Dr. Ciciora's *Cable Television in the United States - An Overview*, "The principal negative of coaxial cable is its relatively high loss. Coaxial cable signal loss is a function of its diameter, dielectric construction, temperature, and operating frequency. A ballpark figure is 1 dB of loss per 100 feet. Half-inch diameter aluminum cable has 1 dB of attenuation per 100 feet at 181 MHz; at one-inch diameter, the attenuation drops to 0.59 dB per 100 feet. The logarithm of the attenuation of cable (in dB) varies with the square root of the frequency. Thus, the attenuation at 216 MHz (within TV channel 13) is twice that of 54 MHz (within TV channel 2) since the frequency is four times as great. If channel 2 is attenuated 10 dB in 1,000 feet, channel 13 will be attenuated 20 dB…. Since attenuation varies with frequency, the spectrum in coaxial cable develops a slope. This is partially compensated with relatively simple equalization networks in the amplifier housings.

"The attenuation of the cable is a function of temperature and aging of components. These amplifiers use a pilot signal to control automatic-gain-control (AGC) circuits. A second pilot signal at a substantially different frequency than the first allows the slope of the attenuation characteristic to be monitored and compensation to be introduced with automatic slope control (ASC) circuits. Thus, long cascades of amplifiers can, once properly set up, maintain their performance over practical ranges of temperature and component aging."

In tree-and-branch topologies with long amplifier cascades, and long cable segments connecting them, losses from attenuation can vary greatly across the course of normal temperature swings in an annual weather cycle. As a result, sweeping these runs twice a year to adjust levels and gain controls in preparation for winter and summer temperatures was critical to ensuring service reliability.

Advancements in PNM Technology

PNM has been around for at least 10 years in mainstream cable operations and many new capabilities continue to be developed. While the existing DOCSIS 3.0 specifications may have reached maturity, new use cases and valuable extensions continue to evolve. Some examples include DOCSIS downstream blind equalization analysis and carrier to interference noise ratio (CNIR) for ingress noise detection. Without a doubt, more capabilities will continue to evolve; however, only adding limited incremental value to any existing PNM stack.

More so, the greatest opportunities ahead exist with artificial intelligence and machine learning to help elevate new use cases and drive operational value in ways that have not been previously conceived. Weather prediction and correlation, such as proposed within this paper, is one example of that.

Within the DOCSIS 3.1 specification, many capabilities remain untapped, largely because they are currently incomplete and unavailable. Even in the case of some functionalities, many operators have not even begun to capitalize on the capabilities that already exist.

Influence of PNM on Network Repair Prioritization

The evolution of intelligence tools, and more recently the proliferation of PNM tools, has had a significant impact on the prioritization of coaxial network repairs. It has been nearly 10 years since Comcast introduced its first PNM software program (the "Scout Flux") to Beta, and we continue to develop tools that offer greater visibility into the performance of its networks. Add to that the capability of nearly all CPE to report some level of intelligence back to the tool sets, and you have a comprehensive view of all corners of the network. No longer is plant performance informed by a small quantity of DOCSIS channels on a relatively small number of devices. The ability to construct, analyze and match full downstream signatures in multiple portions of each individual premise within a node, in addition to the traditional DOCSIS frequencies, means that an operator can develop an accurate map of all the impairments in the upstream and downstream at virtually every component level within that node. Individual premise issues can be isolated and identified, and with the integration of system design prints, network level impairments can be correlated to active or passive components with a fairly high degree of certainty. This allows for impact scaling, and prioritization of impairment resolution with greater precision, as well as improved task management. Since causality can be attributed to the component level with greater accuracy, dispatch of the proper fix agent is more effective and efficient.

A Glimpse into the Future

Technology consumers today are interacting with our networks and the MSOs themselves in ways that are very different than they were just a few short years ago. They converse digitally with each other, and with service providers, frequently, and across multiple platforms. These new and dynamic communications avenues are going to be critical to the feedback loop and data ingest of our learning machines and decision engines. Intelligent line-of-questions (LOQs) will understand network performance at every level, and inform digital solutions to a dynamic workforce. As a result of the proliferation of Artificial Intelligence and Network Monitoring programs, machines will become more adept at ingesting large quantities of disparate data and connecting symptoms to causality, and dispatching the correct fix agent to affect the repairs. Technology, at its core, should connect people to what's important to them, and deliver it wherever, whenever, and on whatever device they choose to consume it. Understanding the behaviors of our transmission medium in the context of annualized weather patterns, and building networks that are as impervious to localized weather as possible, will allow us to keep our customers connected to the communications products and services that they rely on us to provide.

Weather

Effects on Compromised Plant

Through systematic observation using PNM capabilities, certain examples of environmentally stimulated plant failures become obvious. One of the most typical signatures recognized by field technicians is referred to as a “suck out.” In actuality, an RF suckout is a half period of a standing wave. That’s to say, if a longer contiguous visible spectrum were available, it would be perceived as a periodic standing wave. Given the physical constraints of velocity of propogation and reflections, suckouts visible within 1GHz of occupied spectrum are limited to very short fault distances. Typically these will be high energy reflections within less than a few inches of reflective cavity. This is important because it makes them very easy to localize with traditional tree-and-branch analysis of the network topology.

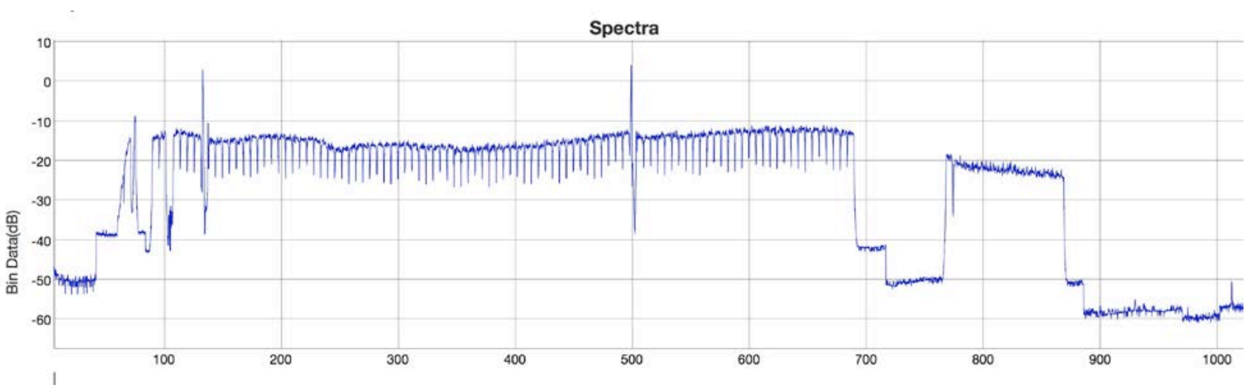

Figure 3 – Typial RF Spectrum

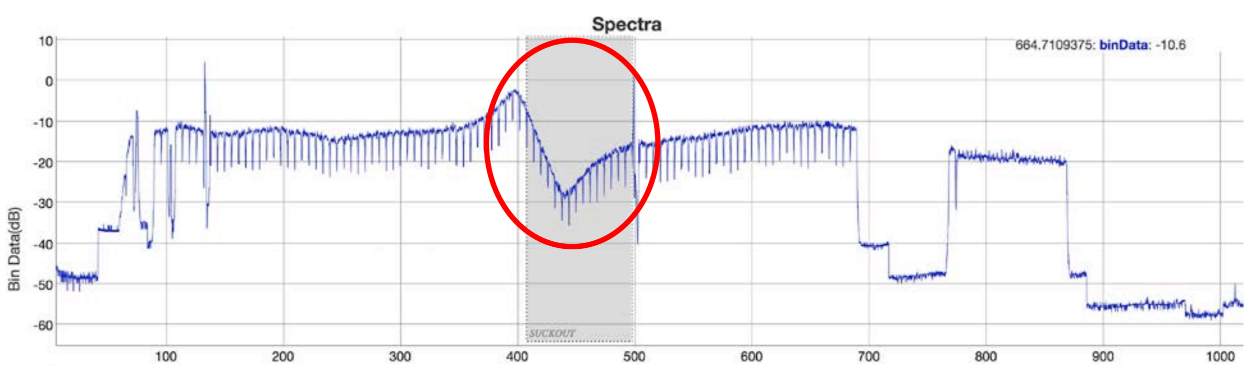

Figure 4 – Impaired RF Spectrum, Temperature Induced

Effects on Workforce

The introduction of a weather informed prioritization model to an already dynamic workforce will represent a significant disruption to any work distribution model. The additional layer of weather as prioritization suggests a time element that was not previously a component in the model. Since most weather predictions follow a 10 day model, there becomes a shifting 10 day period that predicts outcomes based on current data sets, and potential vulnerabilities in the node architecture. This level of variability is quite contrary to many current Demand and Preventive workforce distribution models. The reactive Demand Maintenance activities are already informed by plant failures and weather events, but only after the fact. Preventive Maintenance activities as mentioned previously, typically follow one of two paths: Scheduled or Intelligence tools informed. These models have been optimized by MSOs for many years now, and have provided admirable results thus far. A sliding predictive analysis model that implies prevention of uncertain future events will be difficult to elicit immediate adoption under the best circumstances. Line Maintenance Technicians, like most technicians, are creatures of habit, but tend to be longer tenured and entrenched in traditions and behaviors long earned. Introducing a change with uncertainty to this degree will be a significant challenge to adoption. Additionally, the current points or productivity models used to measure workforce efficiency may also have to undergo some enhancements as tasks are created for repairs prior to the predicted weather behaviors.

Weather Data

Weather data requirements

The weather data is needed for two tasks, each with their own set of requirements. It is assumed that each task will be iterative, with lessons learned being applied to modify the data requirements as needed.

Task 1 is to initially use temporally fine-grained (e.g. hourly) actual data to discover and quantify the strength of correlations between various weather events and plant impairments.

Task 2 is to initially use temporally coarser-grained (e.g. daily) forecast data to predict plant impairments based on predicted weather events, using the correlations established in Task 1. The predictions will in turn feed the plant maintenance ticket Reprioritization Engine (RE).

Geograpic scope

The geographic scope of Comcast's nationwide footprint was one of the few weather data requirements known at the beginning. Companies with a smaller footprint may possibly have a wider range of data vendors to choose from.

Task 1 – correlation establishment (pre-trial)

In order to test the viability of the concept we chose six trial markets with diverse climates, and collected weather data for each of them. The trial markets are:

- Albuquerque
- Denver
- Miami
- Minneapolis
- Philadelphia
- Seattle

The correlation establishment phase of the project requires fine geographic and temporal resolution of the actual weather data, in order to establish reliable correlation data between various weather events and various plant impairments. The assumption of the need for frequent updates was based on our assumption that tight temporal and geographical coupling between network impairment data and weather data could increase accuracy in determining the existence and strength of the correlations. Thus, the time scale would need to be finer than every 12 or 24 hours. The frequent data collection during the trial period will likely be enough to establish the correlations. Once the trial is over we can optionally reduce the collection frequency of actual weather data.

Note that weather forecast data is generally available on a 12 or 24-hour update cycle, not hourly or even finer, as the actual weather data is.

Initial assumptions

The details of the first iteration of requirements were a chicken and egg problem, as we didn't yet have any data to analyze in order to determine what weather data fields would be useful, and how geographically close the weather station needed to be to the network assets to generate reliable correlation data, and what geographic granularity would be desirable in order to determine correlations between specific weather data and specific plant impairment data. Once we collected several months' worth of data we would be able to analyze the correlations and then refine the data requirements based on what was proven to be useful or not.

Our initial guess regarding likely correlations of weather data vs. Plant impairment data included the following hypothesis:

Num	Description	Symptom	Comcast Data source	Notes
1	Cracked coax results in water incursion	High freq rolloff	SPECTRA	Can form a cavity. Water can close it up.
2	Wind results in intermittent RF dropouts	DOCSIS deregisters, video macroblocking	WOPR	Collected every 4 hours in WOPR. Daily snapshots in MELD.
3	Failing amp affects spectrum power	DS MER impairments, many symptoms depending on root cause and line gear type.		
4	Bad shielding leads to noise ingress	DOCSIS US and DS impairments		
5	Fast temp change causes shielding problems due to different metal expansion rates.	Noise incursion, CPD.		Metals plate out, form diodes, becomes a mixer.
6	High temp causes electronics failure (power supply, amp, node)	Low or missing RF		

Based on those initial guesses we required at least these fields in the weather data:
Temperature

Wind speed and direction
Precipitation

Any extra fields, such as humidity, barometric pressure, UV index may prove to be useful as we gain more experience establishing correlations, so will be retained if supplied by the vendor.

Task 2 – Ticket Reprioritization (trial phase)

Once the correlations have been established, it is anticipated there will not be a need for ongoing frequent data collection for that purpose once that phase is completed. Collecting data less frequently after the correlation establishment period will reduce the data cost and computational and storage resource requirements.

Selection of weather data vendor

Vendor selection required some investigation to find out which vendors had actual and forecast data with the desired data characteristics to meet our requirements. These fell into two broad categories: Government-supplied data and commercial data.

The government data lacked the geographic granularity needed, especially for rural areas. Often it is collected at airports, which we assumed would be fine for small or mid-size cities, but not very large cities or smaller towns that lack airports. We felt that the correlation establishment phase would require at least ZIP code level resolution.

Private data vendors offer much finer geographic granularity. We selected Weather Underground, which has some 250,000 weather stations online.

Collection and storage of weather data

Our trial consisted of six cities with diverse climates. For each city we selected one existing weather station from which to collect data, and correlated that to network elements within two miles of that weather station.

Weather Underground has an API that is used for data collection. For our trial we launched a curl command to each weather station in the trial as an hourly cron job.

The data is returned as a JSON file. This is then parsed and inserted into an SQL database by code that Vishnu wrote.

Validation of weather forecast data accuracy

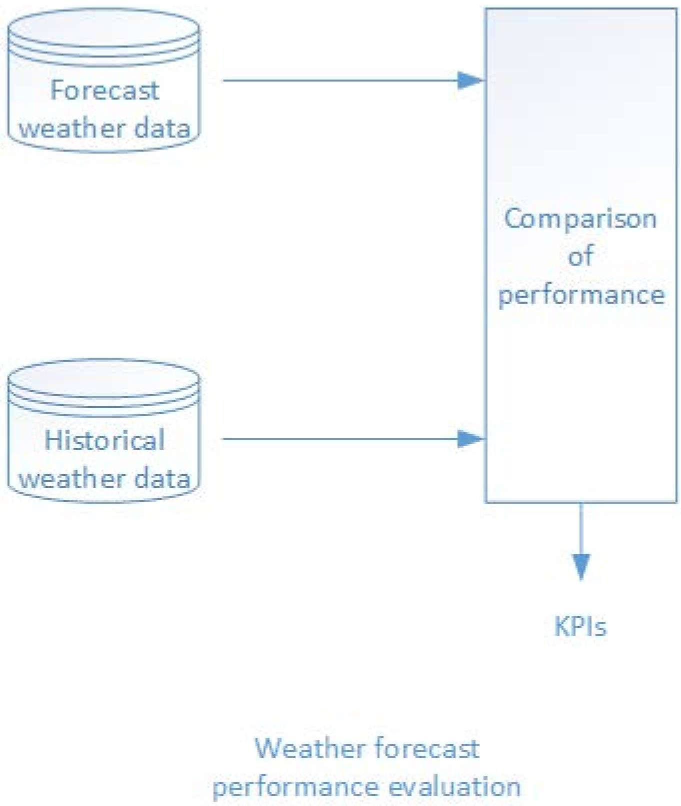

Figure 5 – Performance Evaluation

The point of this project is to have enough confidence in weather forecast data to use it to reprioritize plant repair tickets based on predicted impending weather events. In order to have that confidence, we must analyze the expected errors in the forecast data.

Weather Underground provides a 10 day forecast for various regions, which is updated every day. Note that the forecast region data covers a far broader area than a particular weather station. For example, there is one regional forecast for the Denver metro area, which may contain hundreds of individual weather stations.

In order to analyze the accuracy of those forecasts for each region, we start with the actual data for individual weather stations of interest within that region, for a particular date, and then compare the forecasted values for each field (temp high and low, wind speed, and precipitation) for each of the 10 days. For example, if the actual data is available for the 11th of the month, then we will look at forecast data from the 1st in order to get the 10-day forecast value, forecast data from the 2nd in order to get the 9-day forecast value, etc. The error for the 10-day, 9-day, etc. is then calculated and charted.

An example is shown in Figure 6. This data is from the Denver area trial weather station, with the actual weather temperature data for June 29, 2018. This data was very surprising. We expected the 10-day forecast to have the maximum error, with the error diminishing as the lead time diminished, probably in a highly nonlinear way (much worse at 10 days than at 1 day). Instead, the low temp forecast error decreased in an approximately linear way, but the high temp forecast error actually got worse as the time date got nearer, with the prior day's forecast missing by nearly 10 degrees!

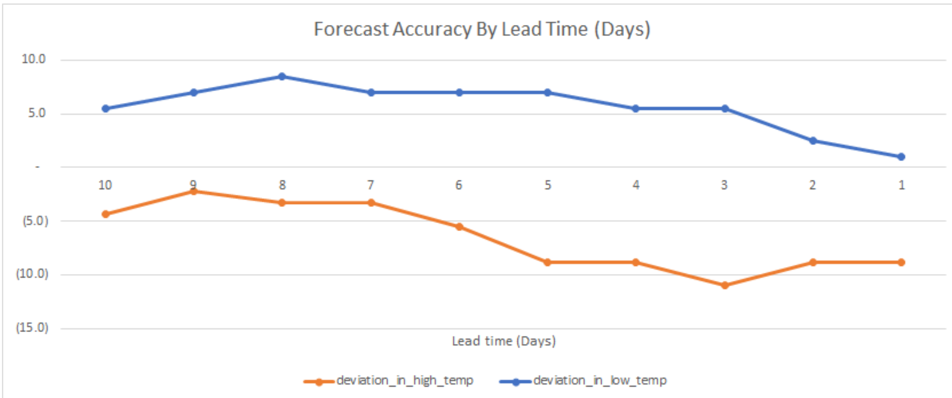

Figure 6 – Forcast Accuracy

Our beginning assumption is that different markets will have different error trend patterns depending on season. For example, Denver is fairly predictable in the summer and winter, but spring and fall are very unstable, given the city's proximity to the Rocky Mountains. As we collect and analyze more data over the span of months and years we will find out if that's true or not for each market.

Formatting and integration of weather data

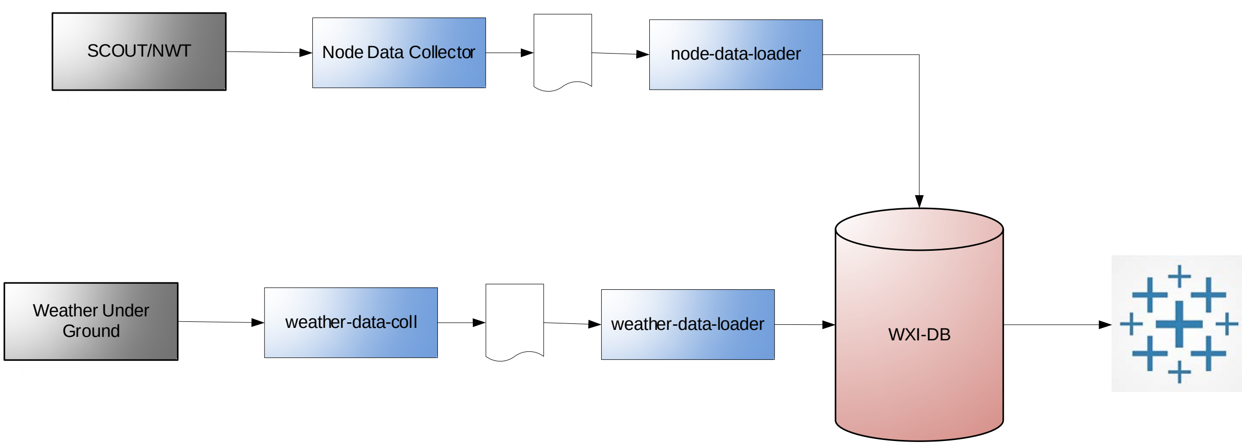

Figure 7 – Data Process Model

Understanding forecasted and the actual data was crucial for determining the viability of the concept, and as part of this trial we started collecting the data from Weather underground in January of 2018. The next step was to compare the forecasted data to the actual forecast for the day, and to determine the deviation between the forecasted data and the actual.

We received a 10 day forecast and wanted to determine which day out of the 10 day's forecast we can rely on the most. Finding deviations between the forecasted data and actual gave us that idea, which we could later use in our prediction algorithm using ML & AI.

For the trial we continued collecting using the shell script and a cron job that ran every hour to pull both forecast data as well as the actual data. Forecasted data consisted of forecast for the next 10 days and the actual data was hourly reading of that day’s forecast.
Forecasted data was also collected hourly, and the average of high and low temperature taken, to compare against the Max and Min value of different weather points from the actual data to determine the deviation.

Actual Weather Data Example:

```
{
 "response": {
 "version":"0.1",
 "termsofService":"http://www.collectionweatherdata.com/weather/api/d/terms.html",
 "features": {
 "conditions": 1
 }
        }
 ,      "current_observation": {
                "display_location": {
                "full":"Littleton, CO",
                "city":"Littleton",
                "state":"CO",
                "state_name":"Colorado",
                "country":"US",
                "country_iso3166":"US",
                "zip":"80122",
                "magic":"1",
                "wmo":"99999",
                "latitude":"39.590687",
                "longitude":"-104.947212",
                "elevation":"1707.2"
                },
                "observation_location": {
                "full":"Denver Centennial, Colorado",
                "city":"Denver Centennial",
                "state":"Colorado",
                "country":"US",
                "country_iso3166":"US",
                "latitude":"39.59",
                "longitude":"-104.95",
                "elevation":"5620 ft"
                },
                "estimated": {
```


},
"station_id":"KCOLITTL344",
"observation_time":"Last Updated on July 27, 5:00 AM MDT",
"observation_time_rfc822":"Fri, 27 Jul 2018 05:00:52 -0600",
"observation_epoch":"1532689252",
"local_time_rfc822":"Fri, 27 Jul 2018 05:01:03 -0600",
"local_epoch":"1532689263",
"local_tz_short":"MDT",
"local_tz_long":"America/Denver",
"local_tz_offset":"-0600",
"weather":"Overcast",
"temperature_string":"61.3 F (16.3 C)",
"temp_f":61.3,
"temp_c":16.3,
"relative_humidity":"86%",
"wind_string":"Calm",
"wind_dir":"West",
"wind_degrees":264,
"wind_mph":0,
"wind_gust_mph":0,
"wind_kph":0,
"wind_gust_kph":0,
"pressure_mb":"1019",
"pressure_in":"30.11",
"pressure_trend":"+",
"dewpoint_string":"57 F (14 C)",
"dewpoint_f":57,
"dewpoint_c":14,
"heat_index_string":"NA",
"heat_index_f":"NA",
"heat_index_c":"NA",
"windchill_string":"NA",
"windchill_f":"NA",
"windchill_c":"NA",
"feelslike_string":"61.3 F (16.3 C)",
"feelslike_f":"61.3",
"feelslike_c":"16.3",
"visibility_mi":"10.0",
"visibility_km":"16.1",
"solarradiation":"--",
"UV":"0","precip_1hr_string":"0.00 in (0 mm)",
"precip_1hr_in":"0.00",
"precip_1hr_metric":" 0",
"precip_today_string":"0.00 in (0 mm)",
"precip_today_in":"0.00",

"precip_today_metric":"0"}
}

Forecasted Weather Data Example:

"simpleforecast": {
"forecastday": [
{"date":{
"epoch":"1532739600",
"pretty":"7:00 PM MDT on July 27, 2018",
"day":27,
"month":7,
"year":2018,
"yday":207,
"hour":19,
"min":"00",
"sec":0,
"isdst":"1",
"monthname":"July",
"monthname_short":"Jul",
"weekday_short":"Fri",
"weekday":"Friday",
"ampm":"PM",
"tz_short":"MDT",
"tz_long":"America/Denver"
},
"period":1,
"high": {
"fahrenheit":"86",
"celsius":"30"
},
"low": {
"fahrenheit":"58",
"celsius":"14"
},
"conditions":"Partly Cloudy",
"icon":"partlycloudy",
"icon_url":"http://icons.wxug.com/i/c/k/partlycloudy.gif",
"skyicon":"",
"pop":20,
"qpf_allday": {
"in": 0.00,
"mm": 0
},
"qpf_day": {

"in": 0.00,
"mm": 0
},
"qpf_night": {
"in": 0.00,
"mm": 0
},
"snow_allday": {
"in": 0.0,
"cm": 0.0
},
"snow_day": {
"in": 0.0,
"cm": 0.0
},
"snow_night": {
"in": 0.0,
"cm": 0.0
},
"maxwind": {
"mph": 10,
"kph": 16,
"dir": "WNW",
"degrees": 282
},
"avewind": {
"mph": 9,
"kph": 14,
"dir": "WNW",
"degrees": 282
},
"avehumidity": 51,
"maxhumidity": 0,
"minhumidity": 0
}
]
}

Loading of Actual and forecast data was done in Scala and using spark. The flexibility of Scala and spark combined made the coding effort small and this data was loaded into MySQL to analyze and create Tableau dashboards. This gave us the easiness to select date or date ranges to see the deviations in different weather points.

Tableau Dashboard:

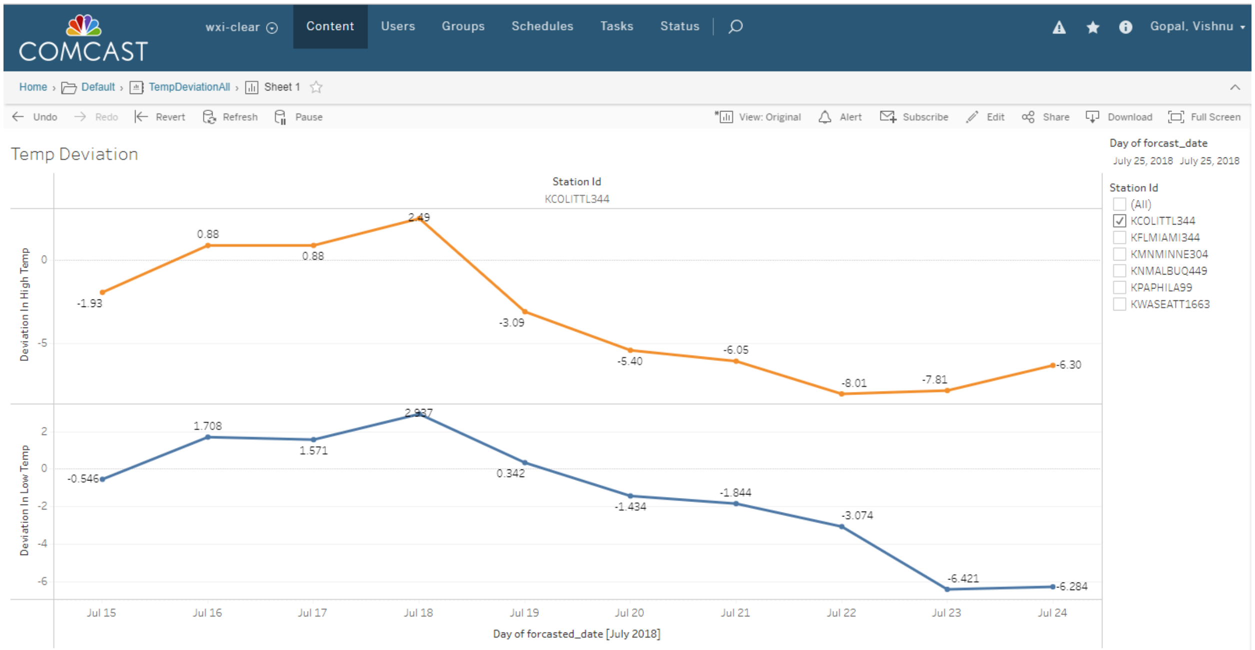

Figure 8 – Temperature Over Time

Integration of weather data and plant data

Analysis of Weather History vs. Plant History

As of July 2018, we have collected about six months of actual weather data for one station in each of the six trial markets, along with about two months of regional weather forecast data for the six trial markets.

We also have a long history of plant data, which has been automatically collected for years, as part of our normal network operations.

With data sets for actual weather data and actual plant data for the six trial markets, we are now ready to look for correlations. The general plan is to feed these data sets into three AI analysis tool instances, each using a different algorithm to find correlations between actual weather data and plant impairment data. The algorithms and configuration parameters will be tweaked to find the strongest and most reliable correlations between the data sets.

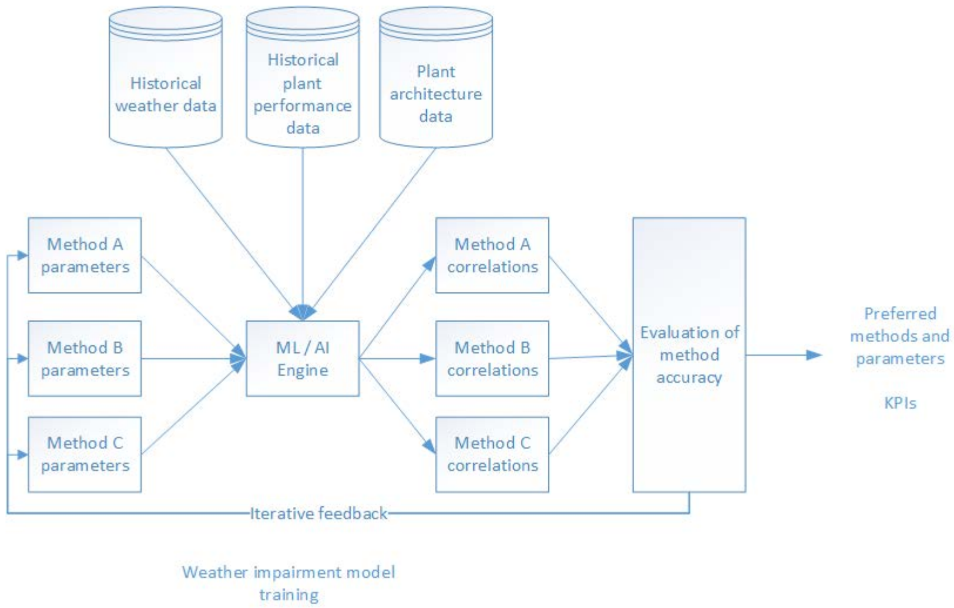

Figure 9 – ML / AI Flow Diagram

Once the machine learning technique is optimized, it can be used to reprioritize the existing plant maintenance ticket backlog, in order to avert customer-impacting problems triggered by impending weather events. The effect of this reprioritization on customer experience can be compared to the BAU process, in order to evaluate the performance of the system. The flow is illustrated in Figure 10, below.

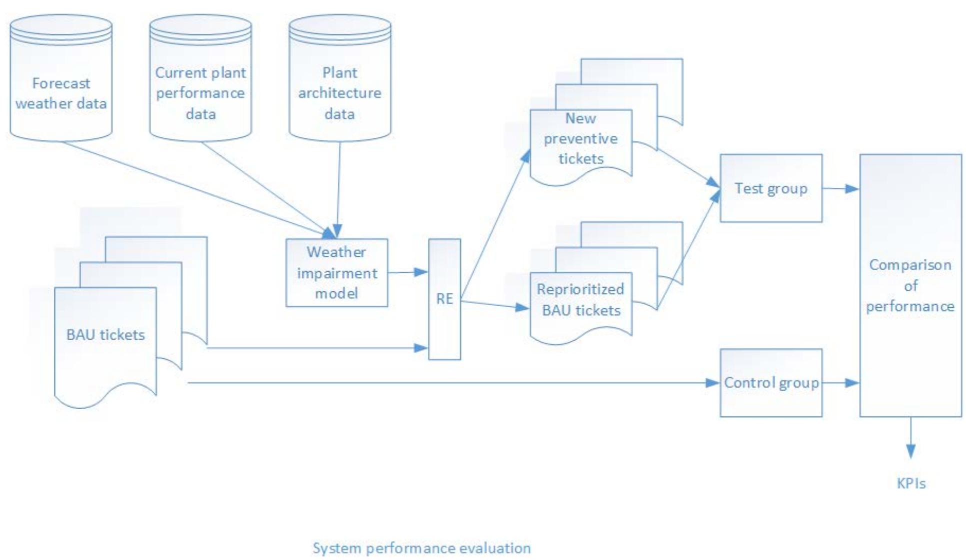

Figure 10 – Analysis Flow Diagram

Characterization of Plant Assets

Previously we discussed the evolution of network architecture, from long tree-and branch-architectures, to the more modern, Node + "x" architectures. As described, the network topology of the cable network is defined by the needs of the local subscriber base. As such, there are many varied and diverse architecture styles still being utilized by MSOs. Within the same service area, operators can have nodes with over 100 amplifiers, and 10 amplifier cascades, as well as nodes with only subscriber taps attached (Node + 0). Each of these architecture styles are matched to their service delivery areas for maximum efficiency, however they could be affected differently by localized weather patterns. For the purposes of determining the impact of weather on coaxial cable networks, it is important that we develop some categories or plant groupings by which we can more clearly understand those impacts. For the purposes of this paper, we have broken the nodes in this study into groups according to several attributes: Aerial cable placements versus underground cable placements, total number of amplifiers in a node, and the maximum number of amplifiers in cascade within a node.

Artificial Intelligence, Machine Learning, and the Future of PNM

With the ever-widening adoption of machine learning into the enterprise, there are many substantial improvements that can be made regarding PNM. Vastly different areas of technology and computing have been affected positively by the research and implementation of these techniques. Take text and video, for example: The relatively new fields of NLP (natural language processing) and CV (computer vision) have made large impacts to those fields. It has also allowed for previously near-impossible tasks to suddenly become possible to give us object recognition in images and more accurate and natural speech to text. As long as the data and telemetry exist, there are many opportunities to enhance current PNM methods as well as enhance predictions beyond the tools currently in use. Additionally, as more telemetry is collected, more often it will become increasingly more difficult to analyze and efficiently gain knowledge from this wealth of new data without the use of machine learning.

The History of Machine Learning And Artificial Intelligence in Comcast Operations

For years, most PNM tools and methods have used conditional logic to implement well known thresholds and alerting from existing device telemetry. These methods have been created and implemented by some of the leading experts in the field and have been instrumental in the shift to proactivity. However, some of these methods can only do so much.

A number of efforts have been researched and worked on. Some of these include predicting when cable modems are active and thus scheduling maintenance activity during periods of predicted low usage, so as to minimize customer impact. Another example is enhanced matching of RF signatures of devices with known issues, so that devices which may have been on the edge of the thresholds are correctly included in an impairment grouping.

Moving forward, many more efforts, aside from this weather correlation study, are underway as well, such as using machine learning and graph theory to better isolate problem RF active and passive elements in the field. Also a number of efforts are analyzing customer impact correlations to certain telemetry.

The Commoditization of ML / AI

With the growth and availability of machine learning software frameworks and tools, it has never been easier to begin leveraging sophisticated techniques that can learn from data and in some cases reduce development time and effort. That said, however, time and investment still needs to be spent on research, development and implementation of these new techniques.

Software

Newer software widely in use today such as scikit-learn python library, or Google's TensorFlow library and framework, lower the barrier of entry and democratize machine learning for a wider audience. In certain circumstances, custom algorithms or implementations still need to be developed, depending on the need of the PNM system or tool -- but mostly these tools should be able to integrate. There are many more fantastic libraries and tools being developed that aren't listed here. Even with all of these, a certain set of skills is required to successfully integrate them into the tool or system.

Skills

There is an oft-cited Data Science Venn Diagram by Drew Conway (reference 3) which portrays the required combination of skills to be successful in implementing machine learning methods. Understanding the difference between classification and regression are good, but to fully grasp why a certain model is producing a specific output it takes some additional knowledge. This isn't to say that every team needs a data scientist, but to dig down into the reasoning behind certain outputs it does help. As the libraries and frameworks move faster forward, the algorithms will continue to improve and become more explainable which will further drive adoption and reduce cost.

Cost / investment

Investment is always an important factor when determining if an existing PNM tool or process should be enhanced, or, in some cases, replaced by a machine learning technique. Usually the highest chunk of the cost is research, followed by development and implementation. The main factor for the increased research cost is simply time. During the research phase there are many items which need to be performed by the researchers and engineers. These are problem definition, data gathering, data cleaning and manipulation, data standardization and normalization, machine learning model development and evaluation and finally implementation.

The Convergence of PNM and ML / AI

Given the fact that machine learning libraries and frameworks are more accessible than ever, greater convergence with PNM will continue. Some of the examples are tool enhancements and new processes, but the time will come when entire systems are replaced with advanced and complementary learning systems. The promise of some of these emerging technologies is that simply based on data, the system can learn optimal paths, unnecessary actions, recommendations and more.

The Composition of an AI Enhanced PNM Program

Hardware & Software

This is very dependent on specific implementations. Our study used weather data from a variety of hardware platforms. From our point of view, the weather station hardware was irrelevant; what mattered was the data was in a standardized format. Computing was done on commodity PC and cloud assets.

Data and Modeling

As previously discussed in the architecture overviews, today's networks are widely varied in their construction and cable placement. For the purposes of this trial, we selected node areas that were comprised of entirely aerial cable placements, or entirely underground cable placements. That singular delineation greatly reduced the volume of nodes considered for evaluation, but those nodes would be more directly informative of the impact of weather on the performance and symptomology being measured. From within those node populations, the next determining factor was cascade length, or depth. There is significant variance in cascade depth among the node populations under review, which ultimately should allow for an analogous population consistent with many architecture styles employed across the country. Understanding that each node could be comprised of both aerial and underground segments, the learnings from this trial, and as the data models evolve, we can take the learnings from the various cable placements, and construct a more hybrid approach to the impact on weather on mixed placement architectures.
Profiles were established for candidate nodes, consisting of of the following features:

- Node size: the number of cable modems being serviced by the node, bucketized into: Node +0 (0), Small (1-19), Medium (20-49), Large (50-99), and Extra-Large (100+)
- Maximum reach: the maximum number of amplifiers between the node and a cable modem, bucketized into: Node +0 (0), Short (1-2), Medium (3-5), Long (6-8), and Extra Long (9+)
- Plant distribution: All Underground, and All Aerial. Nodes with mixed distribution (underground and aerial) were excluded
- Distance from weather station
- Total cable length

The complete feature set used in the machine learning model include:

- Node size: 0 (Node +0), 1 (Small), 2 (Medium), 3 (Large), 4 (Extra Large)
- Maximum reach: 0 (Node +0), 1 (Short), 2 (Medium), 3 (Long), 4 (Extra Long)
- Plant distribution: 0 (underground), 1 (aerial)
- Weather feature
- Impairment feature

Possible weather features include:

- Temperature: degrees farenheit
- Relative Humidity: percent
- One hour temperature change

- Average temperature across X hours

Possible impairment features include:

- Noise: numeric measurement or bucketization by severity:
  0 (None), 1 (Non-Severe: < 70), 2 (Severe: > 70)
- Suckout: numeric measurement or bucketization by severity:
  0 (None), 1 (Non-Severe: < 70), 2 (Severe): > 70)
- Power: 0 (Normal), 1 (Non-Severe), 2 (Severe)
- Forward error correction (FEC)
- Signal to noise ratio (SNR)

Execution

A KPI of the project is the net change in CX brought about by use of the system in response to forecasted weather events. The initial trial will do this via a post-event analysis. This will be done by executing it against the ticket queue of a trial market that subsequently suffered weather damage to the HFC plant which is within the trial radius of the weather station. After the weather event we will have the BAU result, because that's what the market will have executed. By having the Reprioritization Engine (RE) run against the queue prior to the weather event, and analyzing the difference between the actual BAU customer impact scores, versus the calculated customer impact scores had certain tickets been prioritized, we will be able to determine the relative effectiveness of the system.

For example, say a severe rain storm hits a trial market. The BAU result was that a node went down due to a water-induced failure, causing a degraded signal for several hours for 200 customers. This happened because the repair ticket for that node was not high enough in the queue to be repaired before the storm hit. Had the RE been engaged, it would have prioritized that ticket, eliminating the degradation because proactive network maintenance, targeted at the node, suddenly, due to the weather forecast, became a much higher risk asset. Some other enqueued ticket(s), not at increased risk because of rain, got deprioritized. The effect of that will be weighed against the RE's improvement from keeping the node nominal, yielding a net change in customer impact score. That net change is a KPI.

Addressing Common Barriers to Adoption

There are several common barriers to adoption when introducing change into any organization, and particularly where the change is requiring the frontline technical workforce to execute or support the change. The first is the Operational Benefit to the Organization, or the Return on Investment (ROI), and the second is the "What's in it for Me" (WIIFM), as it pertains to the frontline technical workforce (adoption and execution). ROI as it pertains to prevention in the CATV model is generally outlined in a couple of key metric reductions, i.e. outages or trouble calls. As with any preventive program, it is often difficult to measure something that doesn't happen in a limited time period. Year over Year (YoY) performance comparisons in any given metric category can provide early anecdotal data to inform the potential performance gains in a trial or small market rollout. These market trials have historically been the basis upon which programs have been measured with respect to the ROI, and evaluated by the organization as to the worthiness to initiate the change to a larger, if not institutional scale.

Further, results of any level of adoption can be colored by the willingness of the frontline workforce to accept the change. If the initiative is overly complex, and doesn't result in a financial benefit for the frontline users, it may be difficult to institute, to prove out the benefit or ROI for the organization. When dealing with any change to the processes or procedures affecting frontline employees, training hours must be budgeted to outline the change, and fully define the WIIFM associated with the proposed change.

Technology and Operational Discontinuity

It is well established that technology and operations can be successfully integrated, but only through a dynamic integration model that is capable of connecting thought leaders and technologists with execution leaders in operations. Discontinuity, by definition, infers a break in something. Depending on perspective, discontinuity can be perceived as something is broken, like a process or program, or it could be described as a necessary disconnect while the organization course corrects in a more favorable direction. The evolution of network architecture has lent itself to successive changes in construction, design and plant management practices. Any operational or engineering changes suggested by the concepts or ideas within this document, would by definition require a temporary course correction, or discontinuity in the current operations model to provide early estimates on an ROI, and inform decisions on wider scale deployment.

New Prioritization Around 10 Day Forecasts

Solving Priority Conflicts

There are several challenges with any preventative activity or program that have to be overcome, or at the very least, acknowledged in order to evolve a model. First is the understanding that any plant management model is truly preventative. Second is quantifying a return on investment (ROI) calculation, attempting to monetize the impact of an event that was predicted to occur, but did not. Every organization has business imperatives that must be achieved in order to meet obligations to it's customers, employees or stakeholders. Those must be balanced against the operational imperative of protecting, improving performance of, and further monetizing their architectural assets. In order for the operational model to evolve to include a weather informed prioritization, it must first be determined to provide tangible benefit to the business imperative. By continuously engaging a weather analytic engine to a local geography, and analyzing key performance indicators against weather forecast data, one could argue that the prioritization model moves closer to achieving prevention in a meaningful way.

Existing Priority System

Early models of network maintenance prioritization relied on call center data to report areas of the tree-and-branch architectures to the line maintenance teams for investigation. In forward-only, broadcast video delivery systems, there were no other telemetry data sets to which line tech teams could respond. All maintenance was scheduled according to an annual visit cycle, or as a result of customer-reported trouble. As network topologies have gotten smaller, the individual node counts have risen significantly, and knowing the lines of demarcation between service areas had gotten more difficult. As technology and intelligence tools have evolved, so too have the potential data sets available to operators to build node performance metrics, and prioritization methodologies. According to Comcast XOC Engineer Larry Scott, Watchtower and Equilibrium are two of the internal toolsets that allow Comcast to performance-rate and prioritize its hundreds of thousands of nodes.

Watchtower is the primary and first level monitoring tool. It ingests information from multiple intelligence programs, and assigns performance values to each node according to the telemetry data from all of the DOCSIS-based CPE within that particular node. Comcast uses a device poller that calls to each device, retrieves the RF and impairment metrics for that device, then aggregates that data into the Watchtower interface. Each device location is then color-coded relative to its performance data for quick visual analysis, as well as correlated to adjacent devices in close proximity. If multiple devices in the same geographic area are determined to share a similar impairment, then Watchtower declares that those devices are part of an Event. Events are ascribed a score according to multiple logic points built into a scoring algorithm. Each node is then able to be assigned a “node score”. In this model, the higher the node score, the more impaired devices there are within the node. A node score of zero would imply that the devices within that node are operating in an environment that is free from impairments.

Nodes can then be preliminarily prioritized based on their score within a functional team area. Some of the customer premise device attributes that are measured for the purpose of ascribing node score include: Receive and transmit levels, Signal to Noise Ratio (SNR), Modulation Error Ratio (MER), and Codeword Errors (CER). Additional attributes are sent to Watchtower from other toolsets that analyze PNM data points, such as Comcast’s Spectra tool. Spectra analyzes the DOCSIS devices that are capable of providing full band capture information, and matches those signatures, based on impairment logic in the tool, to inform Watchtower when there are groups of devices within a node suffering from suckouts, waves, or other downstream signature impairments.

Another Comcast tool that informs Watchtower device performance is SASQWatch (Service Affecting Stream Quality Watch). SASQWatch looks at node metrics from the CMTS port perspective, as opposed to the aggregate DOCSIS device perspective, to inform data packet loss affecting the upstream CMTS port. Each CMTS port is analyzed at regular intervals, and if a threshold is breached, a message is sent to Watchtower to create an event, increase the score, and by extension, the priority. As Larry Scott explains, not all events are created or measured equally. Certain event types have a base score multiplier consistent with their severity. Like the SASQWatch events just described, a Severe SASQWatch event has a 35 point value multiplier, to increase the score more quickly. For example, if a CMTS port experiences severe packet loss, the amount of customers associated with that event will be multiplied by the time they were impaired, then multiplied by the event multiplier, in this case 35, to derive a total event score. All of the individual events are then summed, to equal the total node score. Node scores are characterized further by the aggregate node scores across incremental time periods (i.e. Current score, One Day score, and Three Day score). This can provide context to the duration of the impaired Customer Experience, as well as which nodes are currently impaired, and are immediately actionable by available line technician resources.

In addition to the device performance metrics, Watchtower also monitors all the customer premise equipment for an offline state, indicative of an outage. These events are “soaked” for a period of time for validity, then issued immediately to the available technicians on duty for assignment and correction via its embedded communications engine. Individual users subscribe to automated messages via e-mail, text message or both. The interface also houses the technicians' schedules and individual roles. Techs can be assigned to particular maintenance roles, such as Demand Maintenance (Outages, Severely Degraded Service events and Single Customer escalations), or Preventive Maintenance, (Sweep, Optimization and Leakage). Because Watchtower understands the number of techs on duty, and their individual roles, it routes the appropriate jobs to the tech best equipped to handle that particular task.

Node scores are prepared by Watchtower about every four hours, and sent for final prioritization in Comcast’s prioritization engine, called Equilibrium. EQ, as it’s known, is the intermediate funnel between Watchtower’s analysis and scoring engine, and its workforce management interface. EQ also ingests the local maintenance technicians' schedules and roles, and has settings to manage the number of node repair jobs that get issued into the maintenance team's work queue. Teams can adjust their job volume in different ways: Either schedule-based, according to the number of techs on active duty, or on a static basis. EQ administrators can set the job volume to one job per person or more, based on system needs, or as a static volume of a fixed number of jobs at any given time. In either case, as soon as a technician completes one job, EQ automatically issues the next and highest priority job into the queue. One job closes, the next automatically appears. Equilibrium's primary function, however, is to allow for customized prioritization at the local team level, while still maintaining focus on the primary customer impacting metrics. For instance, two nodes may have the same score, however one node may have a higher customer trouble call count than the other. EQ will issue the node with the highest trouble call count first. It also maintains score history, for the purpose of prioritization. If a node is partially impaired for 30 days, EQ will add score to the node in the background to increase its priority. Equilibrium can also be programmed to execute a function called "Compare and Replace". If a node has been issued to the work queue, but hasn't been issued to a technician yet, EQ may replace the node, with one of a higher priority. It is constantly analyzing every tech team's work queue, and searching the total node population for the most impactful repair to be repaired.

Proposed Priority System

As mentioned previously, Comcast uses a comprehensive prioritization engine, with multiple layers of priority "accelerators". It is likely that all cable operators today have their own versions of prioritization, predicated on weighting symptomology that is closely aligned to the customer experience. They may also layer in trouble call and repair call volumes. It is worthwhile mentioning that none of these engines create additional work. Rather, their intent is to weigh current symptomology with predictive data to manage their human resources, and apply them to the customer population that needs relief most immediately. These models, however, likely do not layer in architectural references in an informative fashion. They may consider node size, but only in so much as it informs the volume of devices attached to the network or node area.

This study is intended to add an additional predictive data point, which the operators can add as a more dynamic "last layer" of prioritization. For instance: In periods of relative stability in the weather forecast (based on forecast accuracy), the prioritization model would continue to operate with the original business rules applied to the program. In this model, the PNM predictive indicators, in conjunction with architecture qualifiers, will lend additional "weight" to the health of the node area for prioritization. As such, an operator would then, by necessity, have to examine its plant assets with a more myopic lens, taking into consideration regional weather patterns, as well as any architecture data available indicating cable placement and cascade depth.

Adapting This Into Existing Workforce Scheduling

Enabling a prioritization model based on the traditional 10-day weather forecasting scheme would require several things. First, a high degree of confidence in the predictive model would have to be established, as well as an agile workforce distribution model. The workforce management model would need to shift a portion of a limited resource pool from Demand activities, to Preventive activities in a dynamic fashion. Lastly the operator would need to add an additional layer of prioritization into its work management system based on regional weather and plant architecture design.

Operational Culture and Workforce Training Considerations

We have already referenced the tribal knowledge, or operational culture of a system or geography, as it has been shaped over the years by operation and business philosophies, as well as weather. What we have yet to really dive into, is the impact of enhanced intelligence tools, or PNM data to the operational training and change communication strategy. What is often overlooked in the proliferation of intelligence that defines symptomology is the repair stratagem of the line maintenance teams themselves, with respect to the actual repairs of the hardline plant. As mentioned earlier, the physical components comprising the Hybrid Fiber Coax (HFC) network have changed remarkably little in the previous six decades of cable television architecture. As such, the coaxial plant management practices themselves have changed remarkably little. Of primary import, are two very simple goals, which, because of scale, are remarkably hard to achieve. The goal of any plant management plan can be boiled down to: Nominal 75 ohm impedance, and shielding integrity. These two properties are the end goal of any plant repair effort. Sweep and Balance programs, leakage rideout programs, and noise mitigation can be argued as the primary daily activities of most line maintenance technicians, and all of those hours of efforts can be attributed to the same end result: 75 ohm impedance, and a closed system. This is not an indictment of PNM data or other intelligence tools. Quite the contrary, in fact. Those intelligence tools have provided unprecedented visibility and localization of impairments, and have allowed operators to more properly scale repair efforts, and ensure plant integrity post repairs. We mention the simplicity of the plant remediation efforts in the context of training, and the impact to the operational culture. With this in mind, and in the context of this trial, training then can be focused on the model used to prioritize plant repairs, rather than introducing a new metric or symptom into the mix. Because the plant remediation methods won't necessarily change (only the order in which they're applied to the node population), training the workforce should not add significant technical complexity to the change management process. Finally, the operational culture should likewise not suffer significant impact, as the remeditation processes employed by the technical workforce will also remain largely intact.

Organizational Alignment and Common Understanding

When large scale initiatives and large organizations collide, there is typically a wide range of adoption levels and results, and not all of them positive. It's similar to the old telephone game, where the message delivered by leadership, doesn't always resemble the message received by the end user. Further, different parts of the organization may have different goals, which may not be totally aligned. When introducing any change, it's incumbent on the owners of the initiative to understand the deltas in the ability of their regional teams to be able to ingest, and institute change. Particularly where preventive maintenance initiatives are concerned. Common understanding is an end state that ultimately determines the engines that any changes agency uses. Things to consider in determining the change management engine could include: Current state of the region/system, communications and training structure, or regional/local engineering educational structure.

Conclusion

It is well known experientially and anecdotally that weather behaviors can have significant impact on the performance of the outside plant networks, and the operator's ability to ensure trouble-free service delivery to their customers. The body of this work suggests that weather forecast data with greater than 90% accuracy in key weather metric categories, when leveraged against PNM datasets, can inform the operator of potential failures within their networks. This presents a window of opportunity to address the vulnerabilites, and minimize the potential impacts of that weather event to the customer experience.

Abbreviations

AI	artificial intelligence
AGC	automatic gain control
ASC	automatic slope control
ANN	artificial neural network
BAU	business as usual
CATV	cable television
CER	codeword error ratio
CINR	carrier-to-interference-plus-noise ratio
CLEAR	comcast leadership, engineering and relationships
CMTS	cable modem termination system
CPD	common path distortion
CPE	customer premises equipment
CV	computer vision
CX	customer experience
DevOps	development and operations
DOCSIS ®	data over cable service interface specification
DS	downstream
EQ	equalization, or equilibrium
ESPN	entertainment and sports programming network
FEC	corward error correction
FM	frequency modulation
FTTP	fiber to the premises
HBO	home box office
HFC	hybrid fiber-coax
Hz	hertz
ICFR	in-channel frequency response
ISBE	International Society of Broadband Experts
JSON	javascript object notation
KPI	key performance indicator
LOQ	line of question
MER	modulation error ratio
MSO	multiple-system operator
NLP	natural language processing
PON	passive optical network
PC	personal computer
PM	preventive maintenance

PNM	proactive network maintenance
pre-EQ	adaptive pre-equalization
RE	reprioritization engine
RF	radio frequency
RoI	return on investment
SA	spectrum analyzer
SASQwatch	Service Affecting Stream Quality Watch
SCTE	society of cable telecommunications engineers
SID	spectral impairment detector
SNR	signal-to-noise ratio
SQL	structured query language
TBS	turner broadcasting system
US	upstream
UV	ultraviolet
VHF	very high frequency
VoIP	voice over interner protocol
WG	working group
WIIFM	what's in it for me
YoY	year over year

Bibliography & References

[1] CableLabs DOCSIS® Best Practices and Guidelines - Proactive Network Maintenance Using Pre-equalization. 2012

[2] Expo2016, A Comprehensive Case Study of Proactive Network Maintenance – Larry Wolcott

[3] The Data Science Venn Diagram – Drew Conway, http://drewconway.com/zia/2013/3/26/the-data-science-venn-diagram