Exhibit No. __ (BR-2)
Docket No. UG-17____
Witness: Brian Robertson

WASHINGTON UTILITIES AND TRANSPORTATION COMMISSION, Complainant, v. CASCADE NATURAL GAS CORPORATION, Respondent.	DOCKET UG-17_____

CASCADE NATURAL GAS CORPORATION

EXHIBIT OF BRIAN ROBERTSON

FORECAST MODEL

31-Aug-17

CASCADE NATURAL GAS CUSTOMER AND DEMAND FORECAST MODEL

DESIGN DOCUMENT
6/1/2017

This document contains the forecast methodology and supporting documentation for the 20-year customer and demand forecast results generated as part of the combined load study.

TABLE OF CONTENTS

I. Overview

The purpose of this document is to discuss the methodology used for the customer, demand, and peak day demand forecast models for Cascade Natural Gas Corporation ("Cascade" or "Company"). The Company will also describe the underlying data used for each model, the sources of the data, how the data was scrubbed, and how the data was formatted for modeling.

II. Data and Data Sources

a. Customer Data

Customer data was gathered through Cascade's Customer Care & Billing ("CC&B") System. The file reports the customers broken out by 7 categories: CIS_Division (State), Town, Year, Accounting Month, MR_CYC_CD, SIC_CD, and Tariff. The Accounting Month indicates which month the data was billed. The MR_CYC_CD code is the cycle code that explains which billing cycle the data was on for that month. The SIC_CD is the Standard Industrial Classification ("SIC") code that explains what type of customer is behind that data. For example, MR_CYC_CD with code CA01 represents the cycle dates of January 4th to February 3rd. If the billed date is February 4th for this example, the data will be represented with an Accounting Month of February. Cascade will later explain how the data is matched to a calendar month.

b. Demand Data

Demand usage data is gathered through the pipelines' Electronic Bulletin Board ("EBB") System. The three pipelines (NWP, GTN, and Enbridge) post daily usage data at the citygate level. The citygate is where Cascade takes ownership of the gas from the pipelines' distribution system into the Company's distribution system.

c. Weather Data

Cascade utilizes Schneider Electric to gather daily weather data information. This data is gathered at seven weather locations: four in Washington and three in Oregon. The four in Washington are Bellingham, Bremerton, Yakima, and Walla Walla. The three in Oregon are Baker City, Redmond, and Pendleton. Normal weather is defined as the average daily temperatures of the most recent 30 years of historical data which results in the average annual temperatures as well. The company uses a heating degree day ("HDD") as the unit of measure for temperature. HDD is calculated by taking the average temperature from a day and subtracting it from a reference temperature. If the reference temperature less HDD is negative, then the company gives that day a 0 for HDD. The company uses 60 °F as the reference temperature. For example, a 50 °F day will result in 10 HDDs (60-50).

d. Population and Employment Data

The Company uses Woods & Poole Economics, Inc. ("W&P") for annual Population and Employment actuals and projections. The data is listed at the county level for both Oregon and Washington.

III. Formatting

Cascade's data inputs for its customer and demand forecasts are in multiple different formats so the data must be converted into a usable and consistent format for use in Cascade's model. The CC&B data is at the town level, pipeline data is at the citygate level, weather data is at the weather location level, and the W&P data is at the county level. Since each of these inputs is broken out at different levels, Cascade must allocate and associate the data into a consistent format the Company can analyze in its model.

a. Formatting Customer Care & Billing data for the Customer Forecast

To perform the customer forecast, Cascade must match the town data from CC&B to the county data of W&P. Prior to the allocation, Cascade must convert the therms and customers from an accounting month into calendar months. To do this, Cascade uses the monthly data from the pipelines and matches this to the CC&B therm data. Then, using the cycle codes, the data is shifted until it matches as close to the pipeline data as possible. The Company found that shifting the first thirteen billing cycles matched the data with approximately a 5% error. After the Company matches the CC&B data to a calendar month, Cascade matches the town to the county it belongs to and allocates all of the customers to that county. Cascade uses the Company's tariff to assign the customers into four groups: Residential, Commercial, Industrial, and Interruptible. Since W&P is an annual number, the same number for population and employment is applied to each month for the county. Cascade also gives an indicator value of one for each month excluding January. Once Cascade has the data formatted, the Company removes outliers from the data. Below is an example of the formatting for the customer forecast in Table 1-1. Once the Company runs the forecast, the county data is allocated to the citygate level.

Table 1-1: Customer Forecast Format

County	Class	Year	Month	count	Population	Employment	Feb	Mar	Apr	May	Jun	Jul	Aug	Sep	Oct	Nov	Dec
Adams	Residential	2004	1	1097	16.664	8.57	0	0	0	0	0	0	0	0	0	0	0
Adams	Residential	2004	2	1092	16.664	8.57	1	0	0	0	0	0	0	0	0	0	0
Adams	Residential	2004	3	1093	16.664	8.57	0	1	0	0	0	0	0	0	0	0	0

b. Formatting Pipeline Data for Demand Forecast

Cascade must also convert the pipeline data into a usable format. The data that is pulled from the pipelines is the core and non-core usage data at the citygate daily level. Cascade has a report that tracks the daily non-core data at the citygate level so the Company uses this to back out the non-core numbers leaving the core citygate data points. Once Cascade has the core data, the next step is to match the CC&B data to the citygate data. The Company uses an allocation that was determined based on past data to allocate the town level to the citygate level. Allocating the CC&B data to the citygate level allows Cascade to build monthly allocation percentages for the four customer classes. Using these allocation percentages, Cascade can determine the amount of therms that belong to each customer class for each citygate. After these steps, Cascade will have daily therm core usage by customer class. The next step is for Cascade to associate the weather data to each citygate. Cascade uses proximity to citygates and geographic similarity to determine which citygates belong to a weather location. After the weather data is associated to each citygate, the actual and forecasted customers are allocated to the citygates. Once Cascade has usage and customers, the usage is divided by customers to come up with a use per customer ("upc"). Cascade gives an indicator of "1" to weekend days and "0" to weekdays. Cascade is also giving an indicator of "1" for the month the data is in and a "0" otherwise for all months excluding January. An example of the formatting of upc for analysis is given in Table 1-2.

Table 1-2: Demand Forecast Format

Citygate	Class	Year	Month	Day	Weekend	HDD	upc	Feb	Mar	Apr	May	Jun	Jul	Aug	Sep	Oct	Nov	Dec
ACME	Residential	2010	7	1	0	6	0.115493122	0	0	0	0	0	1	0	0	0	0	0
ACME	Residential	2010	7	2	0	3.5	0.092394497	0	0	0	0	0	1	0	0	0	0	0
ACME	Residential	2010	7	3	1	1	0.10394381	0	0	0	0	0	1	0	0	0	0	0
ACME	Residential	2010	7	4	1	2.5	0.127042435	0	0	0	0	0	1	0	0	0	0	0
ACME	Residential	2010	7	5	0	0	0.092394497	0	0	0	0	0	1	0	0	0	0	0
ACME	Residential	2010	7	6	0	1	0.046197249	0	0	0	0	0	1	0	0	0	0	0
ACME	Residential	2010	7	7	0	0	0.05774656	0	0	0	0	0	1	0	0	0	0	0
ACME	Residential	2010	7	8	0	0	0.046197249	0	0	0	0	0	1	0	0	0	0	0
ACME	Residential	2010	7	9	0	0	0.05774656	0	0	0	0	0	1	0	0	0	0	0
ACME	Residential	2010	7	10	1	0	0.05774656	0	0	0	0	0	1	0	0	0	0	0
ACME	Residential	2010	7	11	1	0	0.046197249	0	0	0	0	0	1	0	0	0	0	0
ACME	Residential	2010	7	12	0	1	0.069295873	0	0	0	0	0	1	0	0	0	0	0

IV. Customer Forecast

Cascade utilizes Autoregressive Integrated Moving Average ("ARIMA") models for the customer forecast as well as the demand forecast, which will be discussed in the next section. Below is the formula the Company uses to run the first regressions:

$$C_{Class}^{CG} = \alpha_0 + \alpha_1 Pop^{CG} + \alpha_2 Emp^{CG} + \alpha_m I_m + ARIMA\epsilon(p, d, q)$$

Model Notes:

- $C_{Class}^{CG} = Customers\ by\ Citygate\ by\ Class$
- $Pop^{CG} = Population\ by\ Citygate$
- $Emp^{CG} = Employment\ by\ Citygate$
- $m = month$
- $I = Indicator\ variable, where\ 1\ if\ the\ month\ indicated, 0\ otherwise. (Feb - Dec)$
- $ARIMA\epsilon(p, d, q) = Indicates\ that\ the\ model\ has\ p\ autoregressive\ terms, d\ difference\ terms, and\ q\ moving\ average\ terms.$

Cascade runs this model for each of the 55 citygate and citygate loops by class where applicable. A citygate may only feed one or two classes. First, the Company checks for stationarity. If the data is non-stationary Cascade would difference the data, repeating the step until the data is stationary. Most times, the Company will not difference the data or difference it only once. Once the differencing is determined, Cascade runs the regression and checks for autocorrelation. Cascade uses the Autocorrelation Function ("ACF") and Partial Autocorrelation Function ("PACF") to determine moving average or autoregressive terms for the model. Cascade would then remove non-significant variables. Typically, the model would only choose one of the two between population by citygate and employment by citygate. The Company noticed that if a non-significant monthly indicator variable was removed, the model would provide less robust results, therefore, some monthly indicator variables were left in even when non-significant. Cascade uses Akaike Information Criterion ("AIC") and Mean Absolute Percentage Error ("MAPE"), along with other statistics, in determining which model to use.

V. Demand Forecast

As previously mentioned, Cascade utilizes ARIMA models for the demand forecast as well. Below is the model used for the demand forecast:

$$\frac{Therms}{C_{Class}^{CG}} = \alpha_0 + \alpha_1 HDD^{CG} + \alpha_m I_m + \alpha_w I_w + ARIMA\epsilon(p,d,q)$$

Model Notes:

- $C_{Class}^{CG} = Customers\ by\ Citygate\ by\ Class.$
- $HDD^{CG} = Heating\ Degree\ Days\ assigned\ to\ Citygate\ from\ Weather\ Location.$
- $m = month$
- $w = weekend$
- $I = Indicator\ variable, where\ 1\ if\ the\ month\ indicated, 0\ otherwise. (Feb - Dec)$
- $ARIMA\epsilon(p,d,q) = Indicates\ that\ the\ model\ has\ p\ autoregressive\ terms, d\ difference\ terms, and\ q\ moving\ average\ terms.$

Cascade runs this model for each of the 55 citygate and citygate loops by class where applicable. Cascade next runs the regression and check for autocorrelation. Cascade uses the ACF and PACF to determine moving average or autoregressive terms for the model. Cascade then removes non-significant variables. As with the customer forecast, Cascade uses AIC and MAPE, among other statistics, in determining which model to use.

VI. Peak Day Forecast

To forecast peak day usage, the Company parses the data and uses the 3rd quartile of coldest days. Cascade removes the effects of warm weather on usage. After parsing the data, Cascade runs linear regressions on the data with monthly indicators. Cascade uses the following formula for peak day forecasting:

$$\frac{Therms}{C_{Class}^{CG}} = \alpha_0 + \alpha_1 HDD^{CG} + \alpha_m I_m$$

Model Notes:

- $C_{Class}^{CG} = Customers\ by\ Citygate\ by\ Class.$
- $HDD^{CG} = Heating\ Degree\ Days\ assigned\ to\ Citygate\ from\ Weather\ Location.$
- $m = month$
- $I = Indicator\ variable, where\ 1\ if\ the\ month\ indicated, 0\ otherwise. (Feb - Dec)$

Cascade runs this model for each of the 55 citygate and citygate loops by class where applicable. The Company runs the model and remove non-significant variables. Similar to the customer and demand forecast, Cascade uses AIC and MAPE, among other statistics, in determining which model to use. Once the models are finalized, Cascade analyzes peak day using three different HDD scenarios; Average, System max, and Citygate max. The average peak day uses the average HDD from the coldest days in each of the past 30 years as an HDD for each weather location. System max peak day uses the coldest system wide peak day in the past 30 years, which was December 21, 1990. Citygate max finds the coldest day in the past 30 years and creates a hypothetical day where all weather locations experience the coldest day HDDs in the same day.

VII. Final Product

After running each forecast, the Resource Planning analysts run a backcast to test for quality assurance. If any issues are identified, the Company will re-run those models. The Company produces a monthly customer and demand forecast and an annual peak day forecast for each of the 55 citygate and citygate loops. The forecasts are broken out by class: Residential, Commercial, Industrial, or Interruptible. In addition to using the forecast model for weather normalization in Cascade's rate case, Cascade uses the forecast model for Cascade's SENDOUT optimization model, citygate study, 5-year revenue plan, Purchase Gas Adjustment, and the Northwest Gas Association 10-year outlook report.