GasDay–GasHour Inequality

and Improvements to the GasDay Forecast*

Jackson Streeter

jackson.streeter@pop.belmont.edu - jstreeter8@ymail.com

Belmont University

Department of Mathematics and Computer Science

Marquette University

Department of Mathematics, Statistics and Computer Science

Computation Across the Disciplines: Research Experience for Undergraduates

Department of Electrical and Computer Engineering

GasDay Lab

2013-08-02

To Jeff, Jim, Nathan, Cameron, J.T., and Isak.

I have grown up with you and would not be the man I am today without you.

My friends, this research is done in dedication to you,

for the excellence you have instilled in me.

*This work was supported by the National Science Foundation (NSF) under grant CCF-1063041.

Abstract

GasDay provides natural gas demand forecasting software to Local Distribution Companies (LDCs) across the United States. The GasDay software suite supplies multiple time horizons of forecasting including: hour, day, month, and year. This research focuses on the hour and day features. Of particular interest is the inequality between the sum of hourly forecasts with a corresponding day forecast. This paper explores multiple methods of converging these forecasts and the effects those methods have on forecast accuracy.

1 Introduction

This paper was written to complete the expectations of participating in the Computation Across the Disciplines: Research Experience for Undergraduates (REU) hosted by Marquette University's (MU) Department of Mathematics, Statistics and Computer Science (MSCS). Although hosted by the MSCS Department, this research was conducted in the GasDay research lab of MU's Department of Electrical and Computer Engineering. The researchers in the GasDay Lab are the primary audience for this paper.

1.1 Contributions

The main goal of this research was to solve the GasDay–GasHour Inequality. This research has produced significant results:

1) Evidence for the Sum of GasHours being a competitive input for GasDay ensemble forecast

2) Solution for the GasDay–GasHour Inequality

This paper will begin with an introduction to the natural gas industry. It will provide the context for the research. Section 2 will review some mathematics concepts that will appear in the research. Section 3 discusses the early material of our research that was conducted to help prepare the researchers for the topics to come. Sections 4 and 5 contain the major pieces of research that this paper provides. Section 6 provides a conclusion and a review of the author's experience during the research.

1.2 Research Introduction

Natural gas is an important natural resource that is used all over the world for heating, cooking, drying, electricity generation, industrial processes, and industrial feed. Infrastructures have been built to facilitate the use of natural gas by individuals and businesses. Individuals and companies purchase natural gas from local utilities known as Local Distribution Companies (LDCs). The LDCs supply natural gas for their area by purchasing it from suppliers that might be hundreds or thousands of miles away. LDCs have two options for supplying their gas: purchasing reserves and purchasing on the spot market. Usually, it is more expensive to purchase gas on the spot market, so it is better for LDCs to purchase gas in advance. However, purchasing too much reserve is not good, as the suppliers of the reserve can charge a penalty for the gas that is not used. Alternatively, some LDCs have storage facilities of their own, and can use that space to ease the difference between forecasts and actual usage. Not all LDCs have this pleasure though so often the lowest cost for natural gas occurs when an LDC purchases the amount its customers will need.

Purchasing the right amount of natural gas requires a good forecast. Professional software and experts are often consulted for these forecasts. One example of forecasting software is GasDay. GasDay provides hourly, daily, monthly, and yearly consumption estimates for LDCs to help them decide how much natural gas to purchase for a given gas day. This paper reports research to enhance this product.

Figure 1 illustrates the infrastructure of the natural gas industry [3].

1.3 Problem Background

Predictions made by the GasDay and GasHour products follow different procedures. The GasDay forecast is an ensemble forecast composed of a Multiple Linear Regression (MLR) component and an Artificial Neural Network (ANN) component (as seen in Figure 2). The ensemble measures the error that the models have given over the past days and gives a greater weight to the model that has been performing better (with less error) lately. Then, a weighted measurement is produced, which is GasDay's prediction for the natural gas demand of the coming days. This process is independently run 8 times to produce forecasts for the next 8 days.

GasHour forecasts are created by multiple linear regression models. This process is run independently 106 times to produce forecasts for the next 106 hours. A daily demand can be calculated from the GasHour forecasts by summing the 24 forecasts that correspond to the same gas day. This

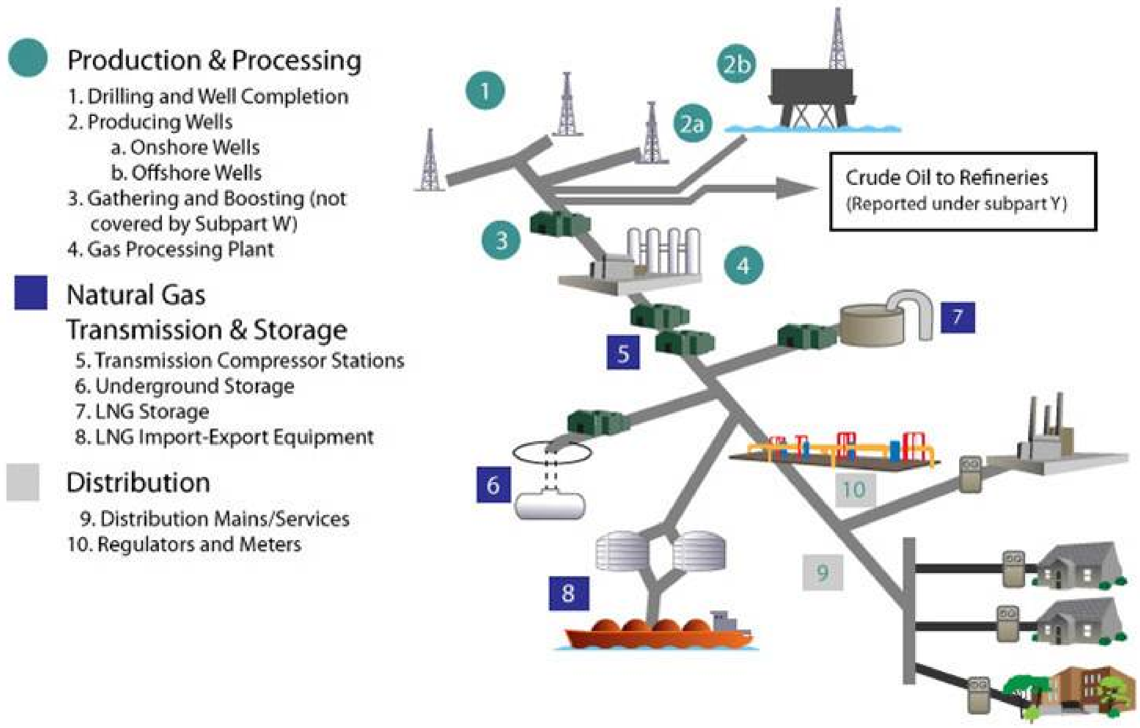


Figure 1: This flow chart represents how natural gas is distributed throughout the country Figure courtesy of [3]

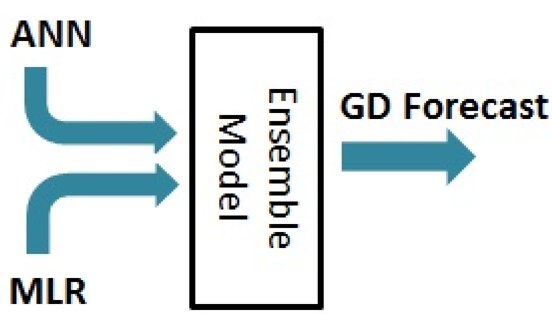


Figure 2: GasDay ensemble model

forecast will be referred to as the Sum of GasHours forecast throughout this paper. This forecast can be more accurate than the GasDay forecast as the aggregate errors tend to cancel each other out. On the other hand, the GasDay model is often more accurate than the Sum of GasHours forecast, as the GasDay forecast has two models to depend on for accuracy. It is in this very issue that the main topic of this paper is centered: The Sum of GasHours forecast for a given gas day does not equal the corresponding GasDay forecast. This will be referred to as the GasDay–GasHour Inequality throughout this paper.

The fact that this discrepancy exists is not surprising; rather it is expected. There are many contributing factors to this difference. One of the biggest factors for the discrepancy is data quality. Hourly and even daily flow data is often not measured with precision and is sometimes not measured at all. A major input for both of forecasts is recent data. When recent data is missing or inaccurate, future forecasts become inaccurate also.

Data quality is an even greater issue for hourly data (not more important, but more susceptible to error). For example, a reading taken five minutes late can contribute relatively large differences in both that hour and the following hour's flow measurement. Hourly readings can also be influenced by changes in pressure in the pipe line. If an LDC is expecting a large demand in gas for the morning hours tomorrow, the night before they might increase the pressure in its distribution pipes to meet the extra demand. The intended result is a flow through the pipe that exceeds actual demand for an hour or two. Later less gas will need to replaced in the pipeline. This action does help prepare LDCs for meeting customer demand, but it can skew the data we are trying to collect. During the first few hours of this process, the meters will read an increase in flow from the pipes (since extra gas is being pumped in). For the next couple of hours though, the meter under-reports demand because less gas is needed to replace the gas that is being taken out. Figure 3 illustrates these effects. These readings, though skewed, are not incorrect, but they are misleading. GasHour creates inaccurate models when it uses this noisy data.

GasDay tends to not have these particular issues, since the changes in pressure tend to even out

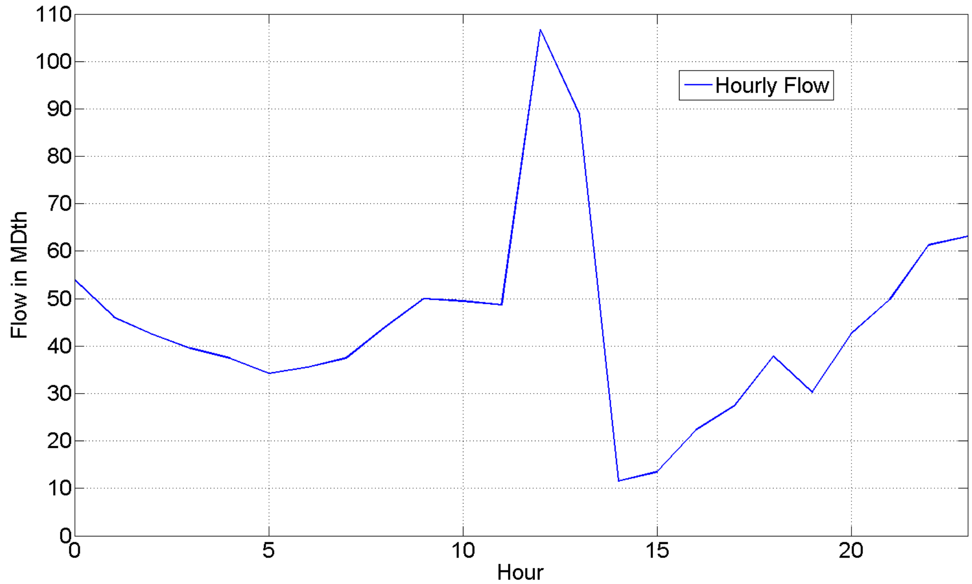


Figure 3: This is a sample model of how pressure changes may effect flow. Hours 11 and 12 have a sharp increase in flow not because the demand increased, but because the LDC pumped more gas into the pipeline to increase the pressure. The following hours report lower flows because less gas is needed to fill the pipeline.

through the day, and a measurement that is a few minutes late is not a significant amount of the day. However, data quality is as important to GasDay as it is to GasHour. Missing and inaccurate data can skew the GasDay forecast as much as the GasHour forecast. Data quality remains one of the most important factors in forecasting, and it is a primary contributor to the GasDay–GasHour Inequality.

Other factors contributing to the GasDay–GasHour Inequality include that GasDay and GasHour are created using different models with different data sets, and that GasDay is a more complex forecast model, by virtue of more research by customer requests. These issues are not directly addressed in this paper (except the research aspect). Instead, this research focuses on adjusting the forecasts to force them be to equal and observing the effects of such an adjustment.

1.4 Relevance of Study

GasDay research is supported by the licensing fees that LDCs pay the lab to use the GasDay software. Since LDCs are GasDay's biggest financial supporter, their research requests are given high priority in the lab. The GasDay–GasHour Inequality is very important to the lab because it is a question that was posed to the GasDay Lab by a GasDay user.

This research has other potential benefits, such as improving the GasDay and GasHour models.

1.5 Acronyms and other Clarifications

Terms and acronyms that are important to understand for this paper are listed below.

GasDay	1) The GasDay Research Laboratory at Marquette University 2) The software package that LDCs use to forecast natural gas demand for different periods of time 3) A model within the GasDay software package that forecasts natural gas demand for the next 8 days
gas day	A period of 24 hours that the natural gas industry has defined and agreed to for the purpose of purchasing, measuring, and reporting natural gas. The gas day period goes from 9 A.M. - 9 A.M. Chicago time. For example, on January 28, 2013, the measured usage of natural gas for that day is the natural gas that is used from 9 A.M., January 28, 2013, to 9 A.M., January 29, 2013.

Top/Bottom of the hour	Refers to when an LDC reports their natural gas usage. A top of the hour LDC reports the natural gas usage for 4 P.M. to 5 P.M. as its 4 P.M. usage. A bottom of the hour LDC reports the natural gas use for 4 P.M. to 5 P.M. as its 5 P.M. use. This report assumes bottom of the hour reporting, implying that the natural gas usage for January 28, 2013 can be found be summing the reported use of each hour between and including 10 AM., January 28, 2013 through 9 A.M., January 29, 2013.
GasHour	A product within the GasDay software suite that forecasts natural gas demand for the next 106 hours
Sum of GasHours	A day's worth of natural gas demand, forecasted by the sum of the 24 GasHour forecasts corresponding to a given gas day
LDC	Local Distribution Company, LDCs are local utilities that provide natural gas for consumers
LR	Linear Regression
MLR	Multiple Linear Regression
ANN	Artificial Neural Network
GD	GasDay
GH	GasHour
Flow	This is the recorded actual natural gas usage for a given gas day
MDth	Thousands of Decatherms, 1 MDth is 10,000 therms, or approximately equivalent to 1,000,000 cubic feet of natural gas.
HDD	Heating Degree Day $\max(0, 65 - \text{Temperature})$

Design Day	An extreme day that incurrs an unusually high gas flow, generally corresponds with very cold temperatures
Defect	GasDay forecast - Sum of GasHours forecast
RMSE	Root Mean Squared Error, a measurement of error. $$\text{RMSE} = \sqrt{\frac{\sum_{k=1}^{N} (\widehat{S}_k - S_k)^2}{N}} \quad (1)$$ where N is the number of measurements, $\widehat{S}_k$ is the forecasted measurement, and S_k is the reported measurement.

2 Literature Review

GasDay's ensemble forecast is created by a combination of a MLR model and an ANN model. GasHour's forecasts are created by a MLR model. These models, along with other pertinent topics, will be discussed lightly below. For more details about these models, please refer to the texts and articles referenced.

2.1 Markov Process (Markov Chain)

A Markov process is a matrix that contains probabilities of an object (or in this case, a forecast) moving from state to state. These probabilities can come from observed transitions or from established probabilities. If the transition process remains the same, and if the sample data is representative of the population, then a Markov process can be used for forecasting [1]. Markov processes are useful because a forecaster does not need to know anything about the past states of the system, rather, they just need to know what the current state of the system is. Another useful property of Markov processes is that future states beyond the next state can be predicted by raising the matrix to the n^{th} power, where n is the number of states in the future. If the process for chaining states changes, then the Markov process becomes much less accurate as a forecast. In regard to this research, a Markov process was considered as an alternative to the current GasDay adaptive ensemble forecast, but a change in the forecast method may cause a constructed Markov process to lose accuracy.

For more details about Markov processes, please refer to Chapter 6 in [5].

2.2 Multiple Linear Regression (MLR)

MLR is a very common in predictive modeling. It can account for multiple input variables, all of which contribute toward a linear relationship. For example, suppose that for N days ($1 \leq \mathrm{k} \leq \mathrm{N}$), there are M ($1 \leq \mathrm{j} \leq \mathrm{M}$) independent variables ($x_{k,j}$) which contribute to customer demand S_k. Then the formula for S_k is

$$S_k \approx \widehat{S}_k = \beta_0 + \sum_{j=1}^{m} \beta_j x_{k,j}. \tag{2}$$

Each β_j represents a relationship factor between variable j and the forecasted flow. Each $x_{k,j}$ represents the value of the j^{th} variable on the k^{th} day. In the GasDay model, β_0 represents base load. Variable $j = 1$ could be any of a set of variables. Suppose $j = 1$ corresponds to Heating Degree Day. Then β_1 represents a relationship factor between Heating Degree Day and the forecasted flow, and $x_{k,1}$ represents the actual or forecast value of the Heating Degree Day on Day k [9]. As mentioned above, this model assumes that each input variable contributes to a linear relationship. If this is not the case, then the model does not forecast accurately what is trying to be modeled. Natural gas demand is a good candidate for a piecewise multiple linear regression model though, as it has definable linear relationships to temperature, wind speed, and other easily collectable data.

For more background on MLR models, please refer to Chapter 6, (Section 6.8 for Recursive Least Squares) in [2] or to Sections 6.7 and 6.8 in [4].

2.3 Artificial Neural Network (ANN)

In contrast to MLR, ANN's are good models of nonlinear systems. Consisting of inputs, input functions, activation functions, and outputs, ANN's are complex, matching the nature of the systems they are trying to replicate. To replicate nonlinear systems, ANN's first go through a training process [7], which learns the nonlinear relationships between input variables and output. These relationships can interpolate data that does not exist in the training data well, which makes this forecast a reliable forecast on typical days. On the other hand, ANN's are very poor extrapolators. On an atypical day (a day near or at design day conditions), the ANN forecast may make a poor forecast. If there are sufficient extreme weather cases in the training data, then it is possible that the ANN can produce an accurate forecast.

For more background on ANN's, please see Section 20.5 in [8].

2.4 Ensemble Forecasting

Ensemble forecasting is the practice of combining forecasts created by different methods to create a single, more accurate forecast [1]. The GasDay forecast is created by a combination of a MLR and an ANN model. Ensemble forecasts are often more accurate than the average of its components, and can sometimes be more accurate than its best component [1]. Ensemble forecasts can be considered flexible since their components can be added to, taken away from, or even replaced by different components. It is this attribute of ensemble forecasting that will be exploited in Section 4.

For more details about ensemble forecasting, please see [1].

2.5 Cubic Splines

Cubic splines are used to model the behavior of nonlinear systems by fitting cubic polynomials to a set of discrete points [6]. A cubic spline, represented by $s(x)$, is composed of several polynomial functions, $q_k(x)$. The polynomial $q_k(x)$ functions are "tied together" at points provided by the discrete point set, referred to as knots [6]. At each knot x_j, the spline is constructed such that $q_k(x_j)$ and $q_{k+1}(x_j)$ are cubic function with equal y-values, first derivatives, and second derivatives. The equal y-values make the spline continuous at the knots. The equal first derivatives will make the instantaneous slopes at each knot equal, and the equal second derivatives will make the concavity at each knot the same. The equal slopes and concavity at each knot make the transitions between functions appear smooth. Cubic splines are good models for approximating many data sets throughout science, and will be considered in this paper to distribute error.

For more details about Cubic Splines, please see Section 6.3 in the text [6].

3 Error Analysis

This section is a result of necessary preliminary studies of the GasDay model. These studies helped prepare us for developing a solution to the GasDay–GasHour Inequality.

3.1 Forecast Error

An understanding and analysis of forecast error can lead to an improvement of the GasDay forecast. First, we created a plot of forecast error, as seen in Figure 4. This figure suggests that errors are greater in the winter than in the summer and that the GasDay forecast probably is a better forecast than the Sum of GasHour Forecasts.

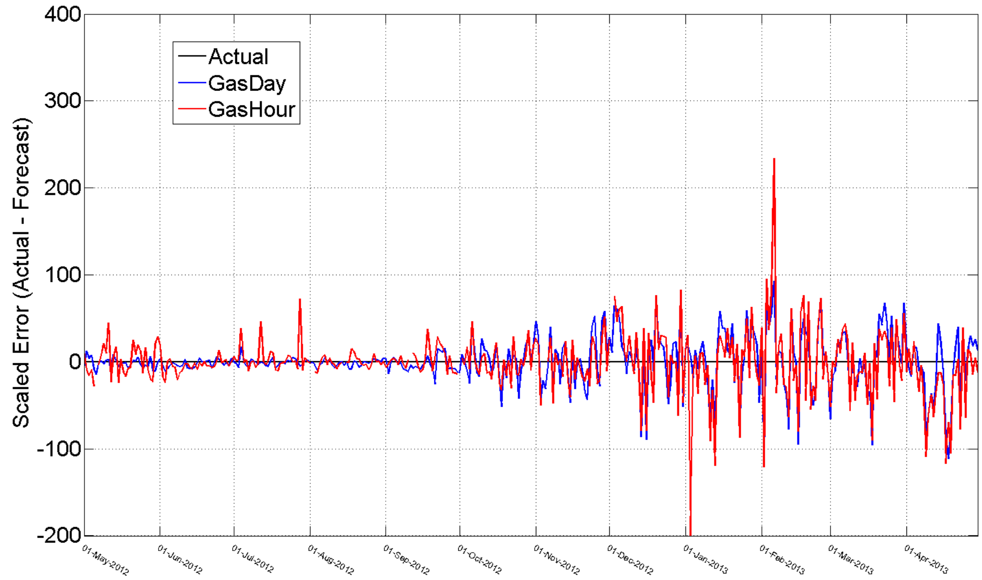
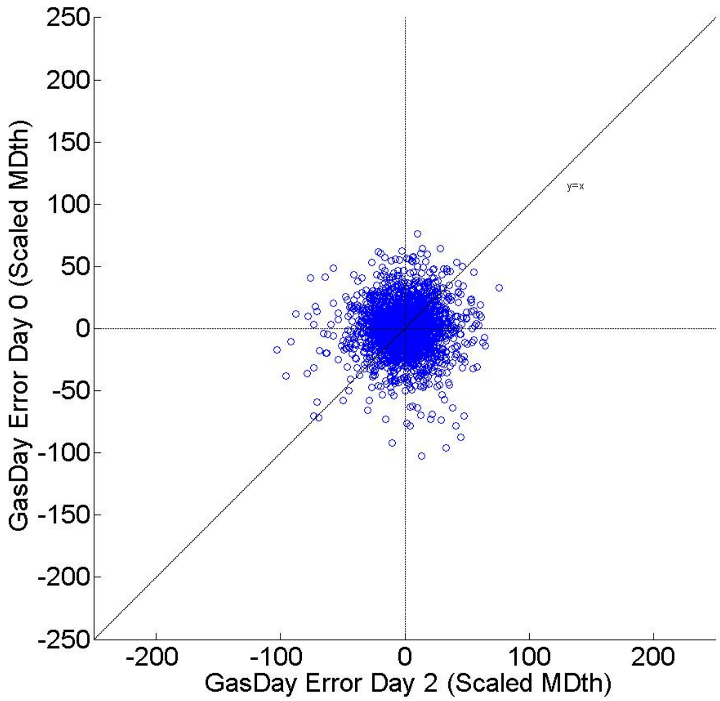


Figure 4: Scaled Error for GasDay and Sum of GasHour Forecasts
Date Range: May 1, 2012 - April 30, 2013

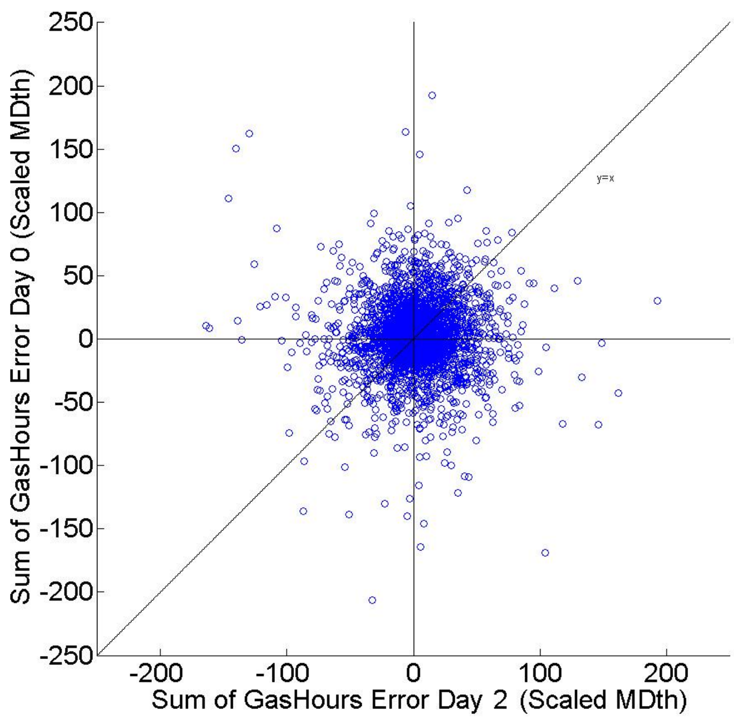
3.2 Two Days Ago Error

If the forecast error discussed in Section 3.1 were known, it could subtracted from the forecasts to give the actual natural gas demands. We decided to look for relationships from which the error can be predicted. We observed a relation between the error of the forecast for Day_{k-2} (two days ago) to the error for Day_k (today's forecast). It would seem that the error for Day_{k-1} would have a stronger relationship to the error for Day_k, but the forecast for Day_k is made during Day_{k-1}, so the LDC can prepare for Day_k. Since the forecast is made in Day_{k-1}, all of the data for Day_{k-1}, including total flow and forecast error, are unknown because the day has not finished yet. Therefore, Day_{k-2} error was observed.

Figures were generated for both the GasDay forecast and the Sum of GasHours forecast, as seen in Figure 5. If a linear relationship existed between these points, they would appear along a linear path. However, this is not the case as the points lie apparently randomly within a circle around the origin. This relationship cannot be modeled and was not explored further.



(a) Sum of GH Forecasts, Error of Day 0 vs. Error of Day 2



(b) GD Forecast, Error of Day 0 vs. Error of Day 2

Figure 5: Date Range: April 1, 2003 - April 30, 2013

3.3 Error Relation to Flow and Temperature

We decided to observe how the error of the GasDay and Sum of GasHours forecasts relate to temperature and flow. Based on Figure 4, it was expected that a greater flow would relate to a greater error because a larger flow introduces greater variability. Likewise, it was expected that colder temperatures would be related to greater errors, because lower temperature cause greater flows, and as stated before, a larger flow introduces greater variability. Figures 6a and 6b on plot the forecast error vs. the day's actual flow. The figures support our hypotheses about how actual flow relates to error. However, because the errors appear randomly placed, this relationship cannot be modeled, and it was not explored further.

Figures 7a and 7b plot the forecast error versus the day's actual temperature. The figures support our hypotheses about how actual temperature relates to error. However, because the errors appear randomly placed, this relationship cannot be modeled and was not explored further.

3.4 Day 0 Error

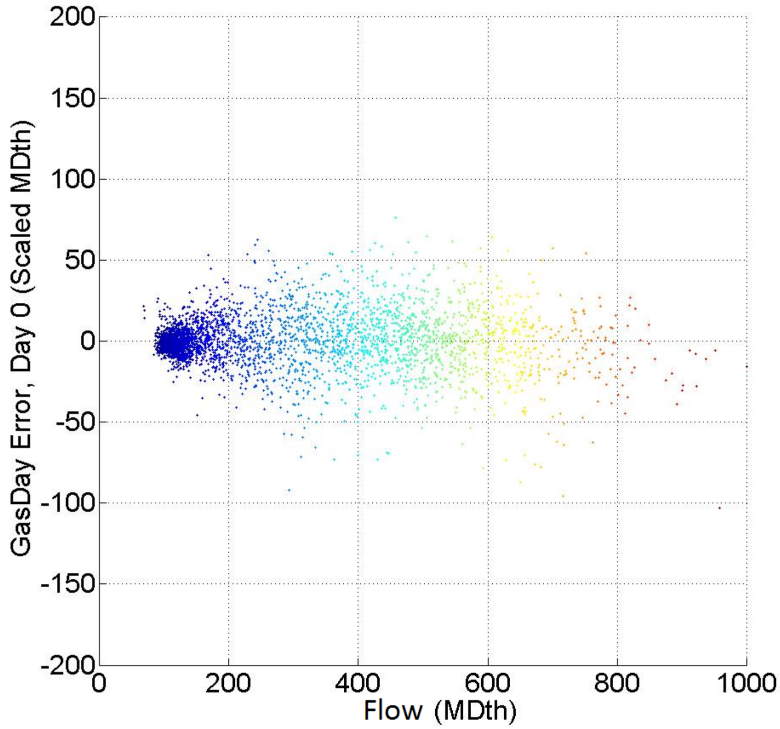
We observed the Day 0 GasDay forecast error vs. the Day 0 Sum of GasHours forecast error, as seen in Figure 8. In the graph, a general X shape can be observed. One of the lines is the $y = x$ line, and the other is the y-axis. There appears to be a general linear relationship along the $y = x$ line, but the `polyfit` function in MATLAB picked up on the points along y-axis. Due to time constraints, this model was not pursued, but it is revisited in the Future Work Section 6.2.1 .

3.5 Markov Processes

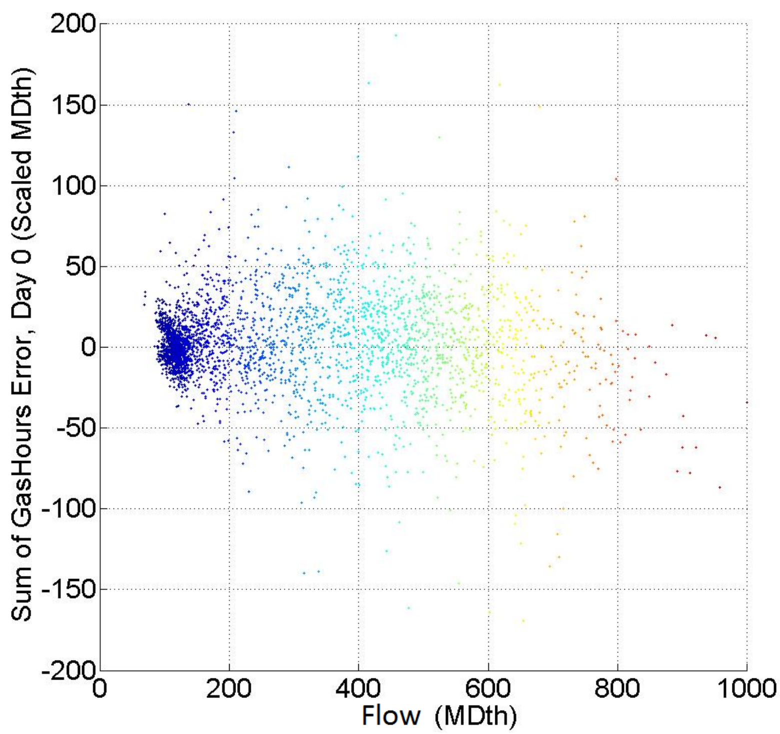
An analysis of Markov Processes can provide insight into how forecasts and actual flows relate to each other. First, we defined eight states that describe how three factors relate to each other:

(Reported Flow = A, GasDay Forecast = D, Sum of GasHours Forecast = H)

State 1: D > H > A

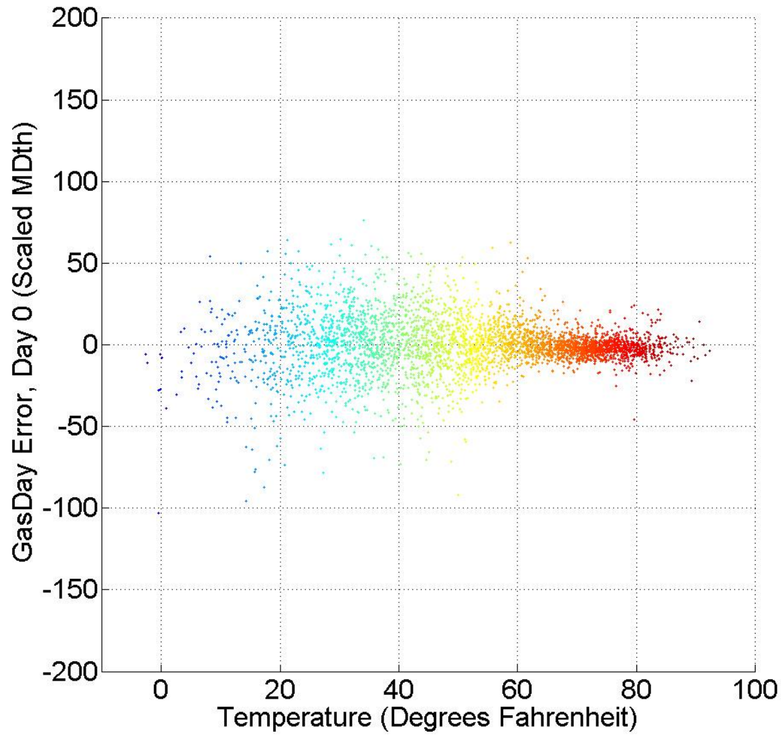


(a) Relation of GasDay Error on Day 0 to Day 0 Reported Flow

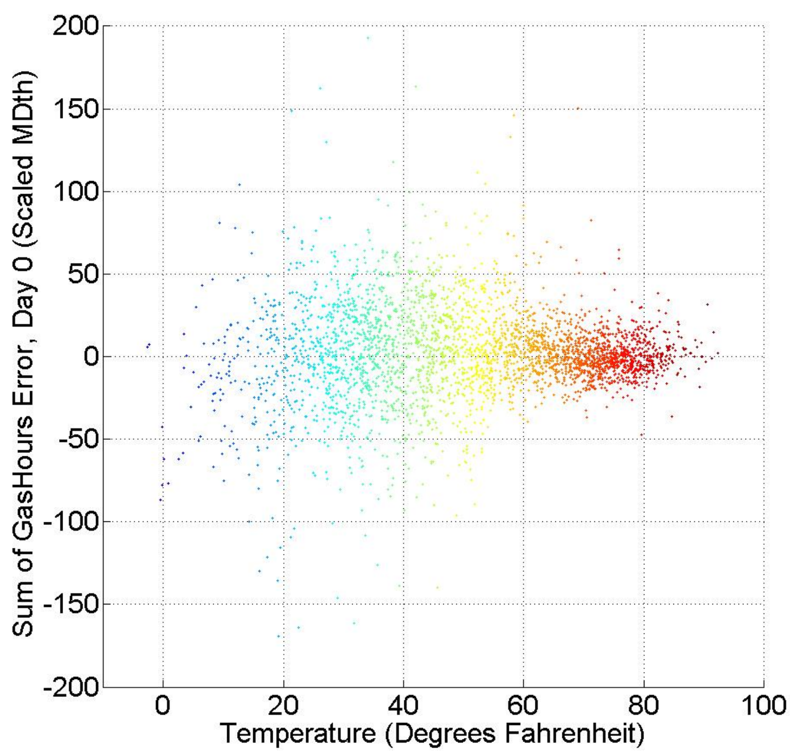


(b) Relation of Sum of GasHours Error on Day 0 to Day 0 Reported Flow

Figure 6: Date Range: April 1, 2003 - April 30, 2013



(a) Relation of GasDay Error on Day 0 to Day 0 Temperature



(b) Relation of Sum of GasHours Error on Day 0 to Day 0 Temperature

Figure 7: Date Range: April 1, 2003 - April 30, 2013

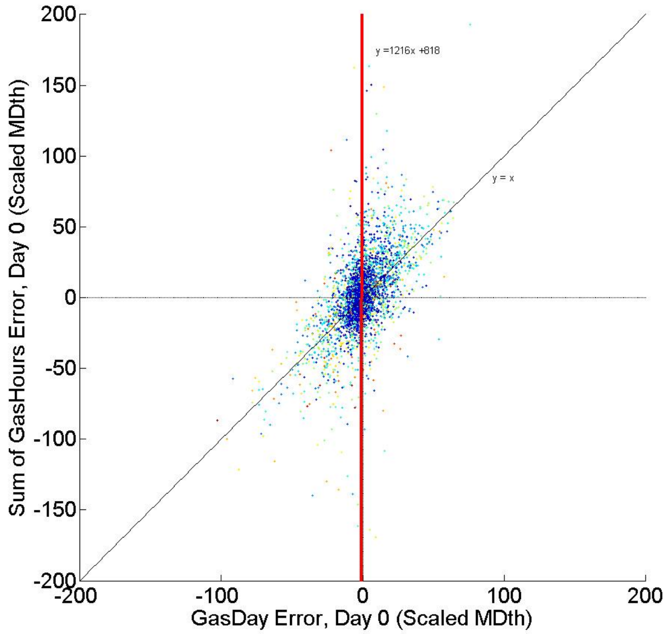


Figure 8: Error of Day 0 for GD and Sum of GH's
Date Range: April 1, 2003 - April 30, 2013

State 2: D > A > H, $|D - A| > |A - H|$

State 3: D > A > H, $|A - H| > |D - A|$

State 4: A > D > H

State 5: H > D > A

State 6: H > A > D, $|D - A| > |A - H|$

State 7: H > A > D, $|A - H| > |D - A|$

State 8: A > H > D

Table 2 shows the number of times a state transitioned to another state. Table 3 describes state transition probabilities. That is, from one day to the next, how likely are we to change from one state to another. This table can be taken further. If the Day_{k-1} error state is known, then only one row needs to be considered when looking at the probabilities. Table 4 tabulates these conditional probabilities. The problem with Table 4 is that often we do not know the error state on Day_{k-1} because the actual flow is not yet known. However, the actual flow from Day_{k-2} is known. In a

Table 2: The vertical column on the left is the error state from the preceeding day (Day_k), and the horizontal row on top is the error state for the proceeding day (Day_{k+1}). This table was created by comparing the error state of Day_k to the error state of Day_{k+1}. After the errors state were calculated, a 1 was added to the corresponding row and column. (Date Range: April 1, 2003 - April 30, 2013)

State	1	2	3	4	5	6	7	8	Row Sum
1	81	25	32	77	68	11	16	25	335
2	16	13	19	26	32	6	9	8	129
3	26	6	42	47	55	12	19	27	234
4	52	23	43	239	142	49	75	104	727
5	100	40	59	130	237	29	91	65	751
6	11	5	9	48	40	24	40	24	201
7	30	11	19	85	93	29	85	44	396
8	19	5	14	81	82	40	58	94	393
Column Sum	335	128	237	733	749	200	393	391	3166
	Number of Times State Transition Occurred								Total Sum

Table 3: This table was created by dividing each element in Table 2 by the Total Sum. This table says that on any given day within the date range, the probability of transition from State r to State c, can be found by looking at row r and column c. (Date Range: April 1, 2003 - April 30, 2013)

State	1	2	3	4	5	6	7	8
1	0.0256	0.0079	0.0101	0.0243	0.0215	0.0035	0.0051	0.0079
2	0.0051	0.0041	0.0600	0.0082	0.0101	0.0155	0.0028	0.0025
3	0.0082	0.0019	0.0133	0.0148	0.0174	0.0092	0.0060	0.0085
4	0.0164	0.0073	0.0136	0.0755	0.0449	0.0076	0.0237	0.0328
5	0.0316	0.0126	0.0186	0.0411	0.0749	0.0092	0.0287	0.0205
6	0.0035	0.0016	0.0028	0.0152	0.0126	0.0076	0.0126	0.0076
7	0.0095	0.0035	0.0060	0.0268	0.0294	0.0092	0.0268	0.0139
8	0.0060	0.0016	0.0044	0.0256	0.0259	0.0126	0.0183	0.0297
	State Transition Probabilities							

Markov process, raising the matrix to the n^{th} power will provide the likelihood of being in a state on the n^{th} day. Hence, squaring Table 4 gives us the probability of being in an error state 2 days later. Table 5 shows the results for this calculation.

Table 2 shows that States 4 and 5 are the most likely states to change to. This suggests that the GasDay forecast is more likely to provided a smaller absolute error on a given day than the Sum of GasHours forecast. This corresponds with our earlier assessment that the GasDay forecast provides a more accurate forecast for daily gas use than the Sum of GasHours does. On the other hand, we expected to see the diagonal, from top left to bottom right, contain the largest probabilities. If this were the case, the state the forecast was in two days ago most likely would be the state of today's

Table 4: This table shows a Markov process that relates the likelihood of being in a defined state, given the state 1 day ago. This table was created by dividing each element in Table 2 by its corresponding Row Sum. (Date Range: April 1, 2003 - April 30, 2013)

State	1	2	3	4	5	6	7	8
1	0.2418	0.0746	0.0955	0.2299	0.2030	0.0328	0.0478	0.0746
2	0.1240	0.1008	0.1473	0.2016	0.2481	0.0465	0.0698	0.0620
3	0.1111	0.0256	0.1795	0.2009	0.2350	0.0513	0.0812	0.1154
4	0.0715	0.0316	0.0591	0.3287	0.1953	0.0674	0.1032	0.1431
5	0.1332	0.0533	0.0786	0.1731	0.3156	0.0386	0.1212	0.0866
6	0.0547	0.0249	0.0448	0.2388	0.1990	0.1194	0.1990	0.1194
7	0.0758	0.0278	0.0480	0.2146	0.2348	0.0732	0.2146	0.1111
8	0.0483	0.0127	0.0356	0.2061	0.2087	0.1018	0.1476	0.2392
	State Transition Probabilities							

Table 5: Markov Process that relates the likelihood of being in a defined state, given the state 2 days ago. This table was created by squaring Table 4. (Date Range: April 1, 2003 - April 30, 2013)

State	1	2	3	4	5	6	7	8
1	0.1308	0.0492	0.0872	0.2340	0.2323	0.0547	0.1006	0.1112
2	0.1171	0.0467	0.0922	0.2265	0.2410	0.0565	0.1091	0.1109
3	0.1102	0.0394	0.0873	0.2269	0.2379	0.0605	0.1155	0.1224
4	0.0957	0.0372	0.0700	0.2443	0.2296	0.0667	0.1247	0.1318
5	0.1174	0.0450	0.0803	0.2217	0.2463	0.0570	0.1190	0.1132
6	0.0923	0.0359	0.0658	0.2354	0.2327	0.0700	0.1408	0.1271
7	0.0994	0.0382	0.0686	0.2303	0.2378	0.0655	0.1371	0.1231
8	0.0881	0.0331	0.0616	0.2300	0.2323	0.0733	0.1399	0.1418
	State Transition Probabilities							

forecast.

We stopped analysis of Markov processes at this point due to time constraints and other pressing topics. Section 6.2.2 discusses potential future work for this topic.

3.6 Data Collection Error

Data collection was discussed in Section 1.3 as a contributing factor to the GasDay–GasHour Inequality. Data quality is very important for producing accurate forecasts. It is very important for the lab to have good data, so much so that there is active research in data cleaning for this lab. In this section, we explore an example of data quality that emphasizes the importance of collecting good data.

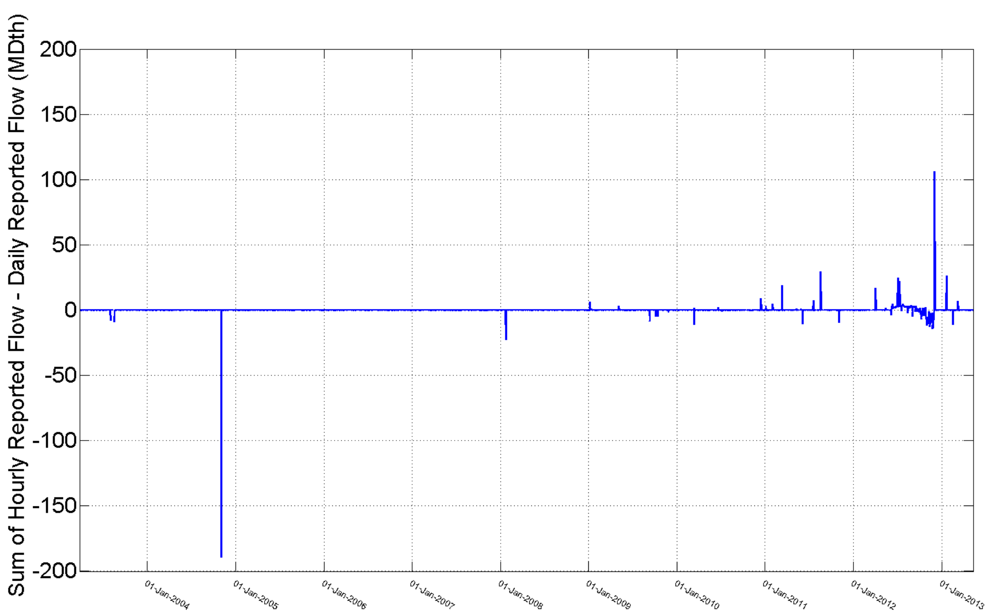


Figure 9: Difference Between Sum of Reported Hourly Flows and Reported Daily Flow
There is a significant difference between the forecasts in 2012
Date Range: March 26, 2003 - May 11, 2013

In Figure 9, the difference between the sum of reported hourly flows and the reported daily flows is plotted. This plot should be a line that lies on the x-axis; however, the plot has many jumps, and significantly differs for an extended period during 2012. Because differences exist between these datasets, that means that the GasDay forecast and GasHour forecast will be created using different historical data. This is a factor that leads to differing GasDay and GasHour forecasts. Part of the issue for improving these forecasts is ensuring that they train on quality data. Although this research is not focused on improving data quality, it is reemphasized here for the GasDay Lab.

4 GasHour as an Input to the GasDay Forecast

This portion of the research was a hinge point between preliminary analysis and developing a solution to the GasDay–GasHour Inequality. The prior research was all in preparation for developing a solution for the GasDay–GasHour Inequality. The following research is the development of said solution.

Ultimately, the plan for the GasDay–GasHour Inequality is to set the sum of the GasHours equal to the GasDay forecast. GasHour forecasts are going to be shifted because, as discussed in Section 1.3, the GasDay forecast is generally a more accurate forecast. That is in general though. GasHour has the potential of being a good predictor for daily gas usage. Since GasDay is an ensemble of its MLR and ANN components, GasHour is really in competition with the MLR and ANN forecasts. However, the proposed solution will put the GasHour forecasts at the mercy of the MLR and ANN components. If the MLR and ANN perform poorly over a period time, the GasHour forecasts will then suffer too, (having been forced to equal the ensemble of those two). If GasHour forecasts were shifted to equal the GasDay forecast in the current state, then good GasHour forecasts could be shifted toward bad GasDay forecasts for the sake of having equal forecasts. We used the Sum of GasHours as an input of the ensemble that creates the GasDay forecast, so that the good GasHour forecasts are no longer at the mercy of the other components but are still influenced by them.

Table 6: RMSE for GasDay, MLR Component, ANN Component, and Sum of GasHour Forecasts
Date Range: April 1, 2003 - April 30, 2013

	Day 0	Day 1	Day 2	Day 3
GasDay	71.56	96.79	115.10	137.09
ANN Component	72.83	234.12	340.04	385.74
MLR Component	79.47	234.38	337.38	378.75
Sum of GasHours	125.99	225.42	320.29	370.64

4.1 Root Mean Squared Error of GasDay and GasHour Forecasts.

Analysis of the Markov Processes in Section 3.5 suggested that GasDay was more often a better forecast for daily natural gas demand than GasHour. Analysis of the Root Mean Squared Error (RMSE) of the forecasts agreed with this assessment. Over a 10 year period, the RMSE for both models and the component models of the GasDay forecast are calculated in Table 6.

As expected, the GasDay forecasts performed much better than the Sum of GasHours. However, Days 1, 2, and 3 are interesting. In each of these days, the Sum of GasHours performed slightly better than the component models of the GasDay forecast. This suggests that the Sum of GasHours forecast is a competitive forecasting model ANN and the MLR for each day after Day 0. With this assessment, we altered the components of the GasDay forecast.

We would have added the Sum of GasHours as a co-component of the ensemble GasDay forecast; however, time constraints did not allow us to implement such a change. Within our allotted time, we were able to replace the MLR component of the GasDay forecast with other values. The MLR was replaced because the ANN is performed better on Day 0, as seen in Table 6. Though the MLR appears to perform better on later days, Day 0 was the most important factor to this research, so the MLR was replaced. For multiple time periods, the following values were placed in the GasDay Forecast: Actual Flow, Actual Flow - 35, Actual Flow + 35, the LR model, and Day 0 Sum of GasHours forecasts. GasDay forecast predictions were calculated using those inputs.

Figure 10 is a plot of the weights associated with the GasDay forecasts that used the Sum of GasHours and the ANN as components. Remember that in ensemble forecasting, a forecast that has been preforming better than its competitor forecasts is weighted more by the ensemble. Forecasts

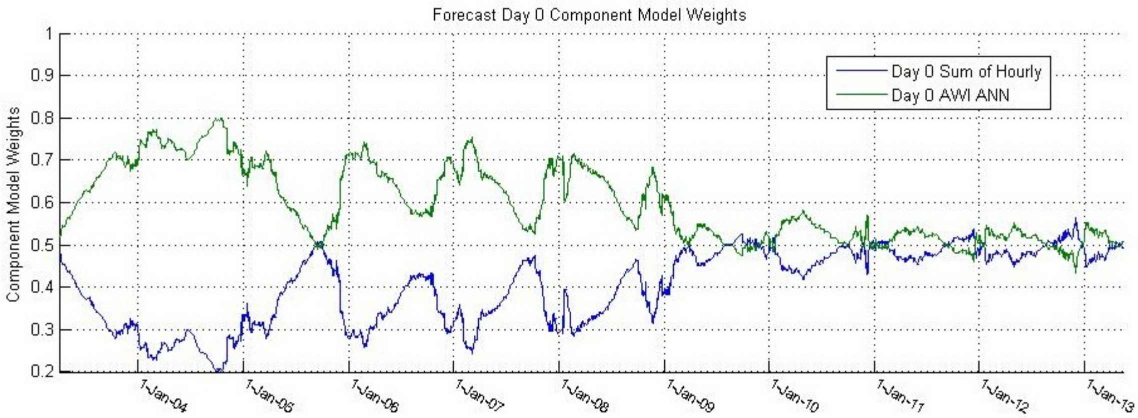


Figure 10: GasDay Ensemble Model Day 0 Component Weights
Date Range: April 1, 2003 - April 30, 2013

that have been preforming worse than its competitor forecasts are it weighted less by the ensemble. In the beginning years, the Sum of GasHours was not performing well, and was weighted less by the ensemble model. It is during these times when adjusting the GasHour forecasts to reflect the GasDay forecast probably improves the GasHour forecasts. However, starting in January 2009, the GasHour model started to perform much better, retaining significant weight through the remainder of the period, even gaining a greater weight at times than the ANN. During the times when the Sum of GasHours is a better forecast than the components of the GasDay ensemble, the good GasHour forecasts would be forced to resemble the less accurate GasDay forecasts, worsening the GasHour forecasts. By including the Sum of GasHours as a component of the GasDay model, the ensemble model works out these issues and allows both models to reflect which has been performing better lately.

Though including the Sum of GasHours as a component of the GasDay forecast improves both forecasts in theory, we wanted to see how doing this actually affected the GasDay model. A RMSE for these each forecasts was calculated. Table 7 summarizes these results.

One might expect the actual flows to provide the lowest RMSE, but because the GasDay forecast does not train on data with low errors, it handles the actual flow data differently than the rest of the

Table 7: Root Mean Squared Error (RMSE) calculations of the Day 0 GasDay forecast with different components used in place of the MLR component. Note that the forecasts for 5/10/12 - 4/30/13 were produced with a dataset with a date range of 5/1/12 - 4/30/13.

	5/10/12 - 4/30/13	3/26/03 - 5/11/13
Actual	55.42	34.95
Actual - 35	32.73	32.11
Actual + 35	29.53	30.16
MLR	103.77	71.83
Sum of GasHours	109.22	79.81

data. As for the Sum of GasHours, it provided a competitive GasDay forecast to the MLR GasDay forecast during the time periods. This finding justifies exploring how adding of the Sum of GasHours as a component effects the GasDay forecast. This topic is discussed further in Section 6.2.3.

5 GasDay–GasHour Inequality

To reiterate the problem, the sum of 24 GasHour forecasts corresponding to a single gas day, does not equal the GasDay forecast for the same gas day. Two potential solution are discussed in this paper. First, adjust the GasDay forecast to equal the GasHour forecast. Second, adjust the GasHour forecasts so that their sum equals the GasDay forecast. The first solution is not desirable, as the GasDay forecast is often the better forecast (As discussed in Sections 1.3, 3.5, and 4). This section will discuss adjusting GasHour forecasts so that their sum equals the GasDay model.

In this section, three adjustments to the GasHour forecasts are proposed: the Naive Method, the Cubic Spline Method, and the Piecewise Linear Method. These are methods of deciding how to spread the error. These methods are not themselves creating forecasts. Rather, they are functions developed to store how much each GasHour forecast should be adjusted. Within this context, a new definition of the problem can be created: The sum of the hourly adjustments should equal the difference between the GasDay model and the sum of the GasHours (this will be called the defect).

When attempting to solve this problem, it became apparent that each day's inequality should not be solved individually. When doing so, issues of continuity can become a problem when forecasting shifts from day to day, as seen in Figure 11a. This problem is dealing with discrete data points in hourly forecasts, so continuity is technically a non-issue in these forecasts. However, consider that natural gas flow is a continuous phenomenon. The GasHour model, although discrete, is trying to model a continuous phenomenon, so adjustments to the model should reflect the continuous nature of the phenomenon it is trying to model. The Naive Method is an example of solving the problem for each day individually, and the continuity issue is explored further in this section.

5.0.1 `fmincon`

`fmincon` is a function in MATLAB that minimizes a function while meeting given constraints. This function accepts a cost function (the function to be minimized), linear constraints, upper and lower bounds for the solution, and a function for nonlinear constraints. `fmincon` completes the minimization by iteratively selecting values of the vector of independent variables until all constraints are met within a tolerance level. This function is used in the Cubic Spline Method and in the Piecewise Linear Function.

5.1 Naive Method

The Naive Method is very simple. First, divide the defect by 24. Then this number is added to each hour's forecast during the day. Because this method does not consider the next day's defect, the graph of the adjustment to the GasHour forecast contains jump discontinuity, seen in Figure 11a. The naive solution solves the GasDay–GasHour Inequality. However, it fails to reflect the continuous nature of gas flow.

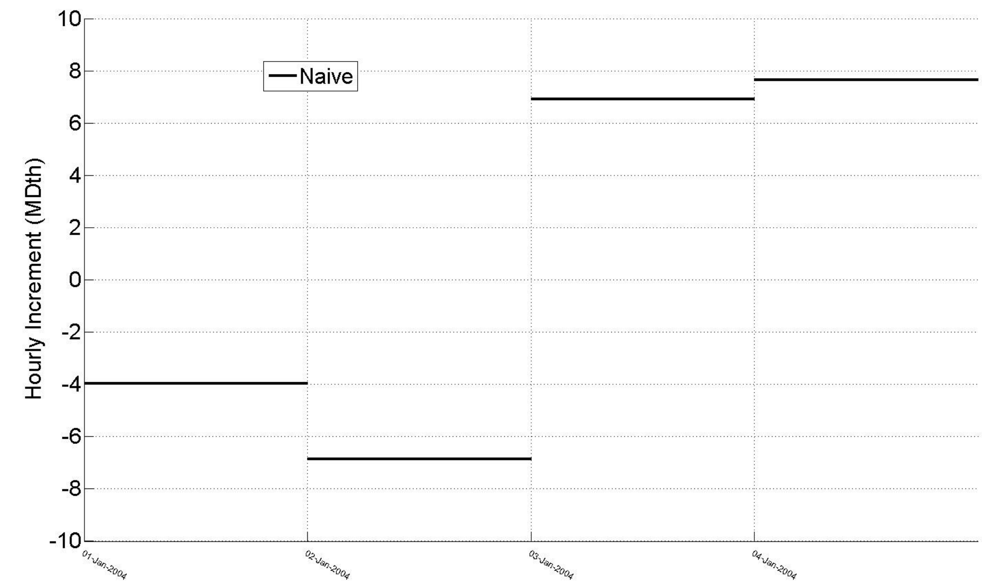
Consider a forecast day when the sum of GasHour predictions are 96 MDth greater than the GasDay forecast. By the naive method, we simply subtract 4 MDth to each hours' forecast, and the adjustment is finished. This day is fixed, and, on its own, this solves the GasDay–GasHour Inequality. However, suppose that on the next day GasHour predictions are 96 MDth less than the

GasDay forecast. The naive solution suggests we add 4 MDth to each hour. Again, this day is fixed, and the GasDay–GasHour Inequality is solved. Suppose also that the forecast for hour 23 of day 0 is 63 MDth and the forecast for hour 0 of day 1 is 65 MDth. Before the adjustment, it is clear that there is an increase in flow from hour 23 of day 0 and hour 0 of day 1. After the adjustment, however, the new flows are 67 MDth and 61 MDth respectively. The post-adjustment forecast suggests that there is a decrease in flow, rather than the increase in flow suggested by the original forecast. This "drop" in flow can be misleading and only occurs due to the adjustment. Scenarios like this should be avoided. A continuous set of adjustments will not necessarily prevent similar scenarios; however, it would be more unlikely with a continuous set.

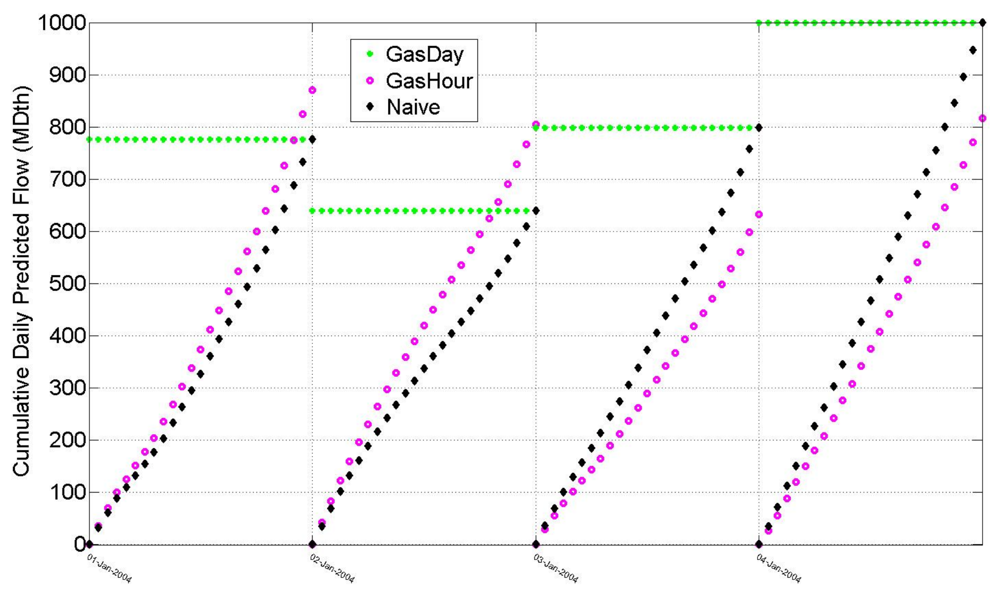
5.2 Cubic Spline Method

The Cubic Spline Method recognizes that even though GasHour forecasts are discrete points, the gas flow is actually continuous. MATLAB was used to construct these splines. Two MATLAB functions were critical to this process: `spline` and `fmincon`. `fmincon` was used to select y-values for knots to be placed every 4 hours. `spline` was used to construct the cubic splines through the knots that `fmincon` selected. `fmincon` was constrained in its selection by the following parameters. The first and most important constraint is that for the created splines, the sum of the adjustments for each day must equal the defect. The next constraint is that there should be no negative flows, that is, the cubic spline should not adjust an hourly flow to be less than 0. If this were to happen, GasHour would be telling the LDC that for that hour, they are going to have negative demand. (This was primarily a concern for the Cubic Spline Method, because negative extremes in the adjustments produced by this method occasionally produced negative forecasts.) Lastly, the function that `fmincon` was minimizing was the 2-norm for the vector containing all of the adjustments (effectively minimizing the adjustments for each hour).

The result of this process is seen in Figure 12. The Cubic Spline Method provides an oscillatory, continuous adjustment. The continuous part is what was desired. However, the oscillatory increment



(a) Hourly Increment Graph
There is jump discontinuity in this graph. The adjustment for each hour is the point on the line that corresponds to each hour. At the discontinuous points, the adjustment for that hour correspond to the adjustment from the left.



(b) GasDay and Sum of Hourly Forecasts
The adjustments force the Sum of GasHours to equal the GasDay forecast.

Figure 11: Naive Method
Date Range: January 1, 2004 - January 4, 2004

adds varying adjustments throughout the day. This oscillating adjustment is not desirable. Hours at the ends of each day often contain extremes in adjustment, so we turned to a new method.

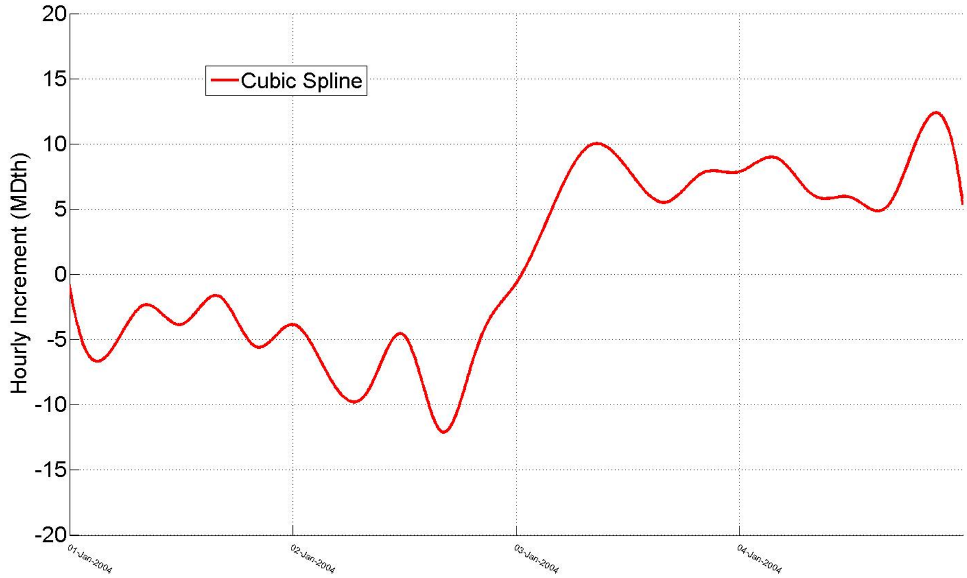
5.3 Piecewise Linear Method

The problem for the Naive Method is that it produces discontinuous forecast adjustments. A problem that both the Naive and Cubic Spline Methods have is that they adjusted the first forecast hours as much or sometimes more than the rest of the hours in the Day 0 forecast. Since the first the hours are the ones that GasHour can be expected to get most nearly correct, adjustment to these hours should be avoided, and passed on to later hours. These are issues that the Piecewise Linear Method addresses.

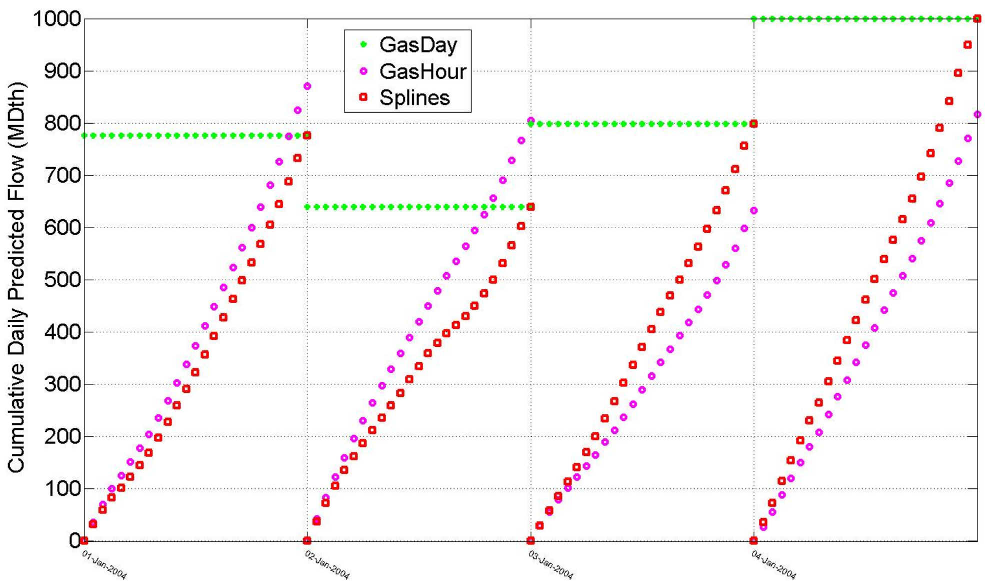
Given the Naive Method adjustments, suppose we adjust only 2 hours at each end of the day. The goal of adjusting these sets of 4 hours is to line them up such that they fit a linear path, but still contribute the necessary adjustment to make the Sum of GasHours equal the GasDay forecast. Imagine taking the last two adjustments for the Naive Method on a given day and the first two adjustments for the Naive Method from the next day, and bending them so that the adjustments for each day are connected. If this were the only adjustment, then the sum of the adjustments would not equal the defect for the day. So again, `fmincon` is used.

`fmincon` was run to select the adjustment values for the 20 hours throughout each day that were going to be adjusted equally. Between the points, MATLAB's `polyfit` function was used to construct a linear function between these values. Day 0 was a little different. The hour 0 adjustment was fixed to 0, and a linear function was constructed to meet the selected y-value at hour 6. This function was again held under the constraint that the sum of the adjustments must equal the defect for the day. Lastly, `fmincon` was trying to minimize the slopes between each adjustment change.

The results of this adjustment can be seen in Figure 13. This method addresses the continuity issue that the Naive Method presents. This method connects the naive adjustments, after small perturbation. This method also addresses the issue of the oscillatory adjustments presented by the

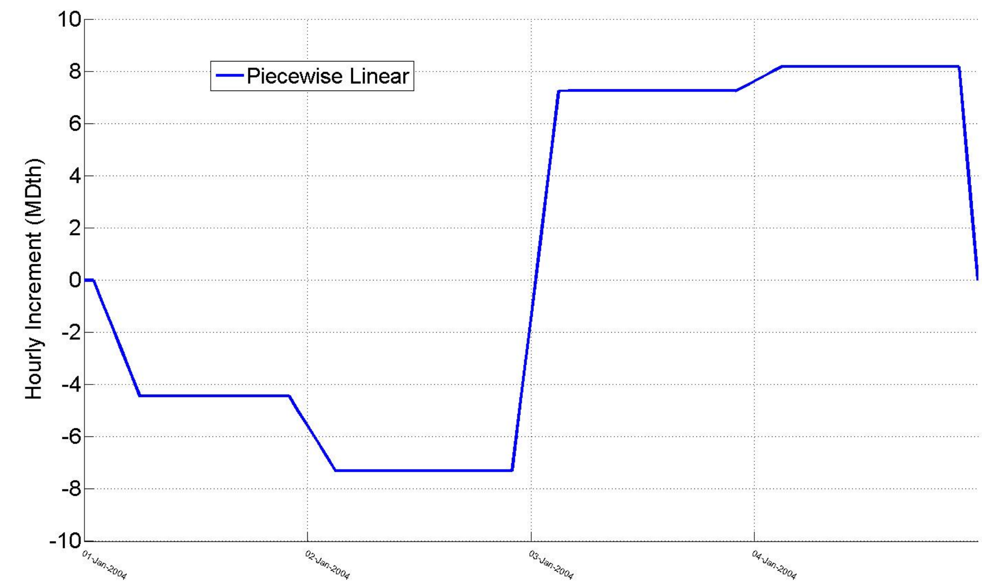


(a) Hourly Increment Graph
There is no discontinuity, but these adjustments are oscillatory.

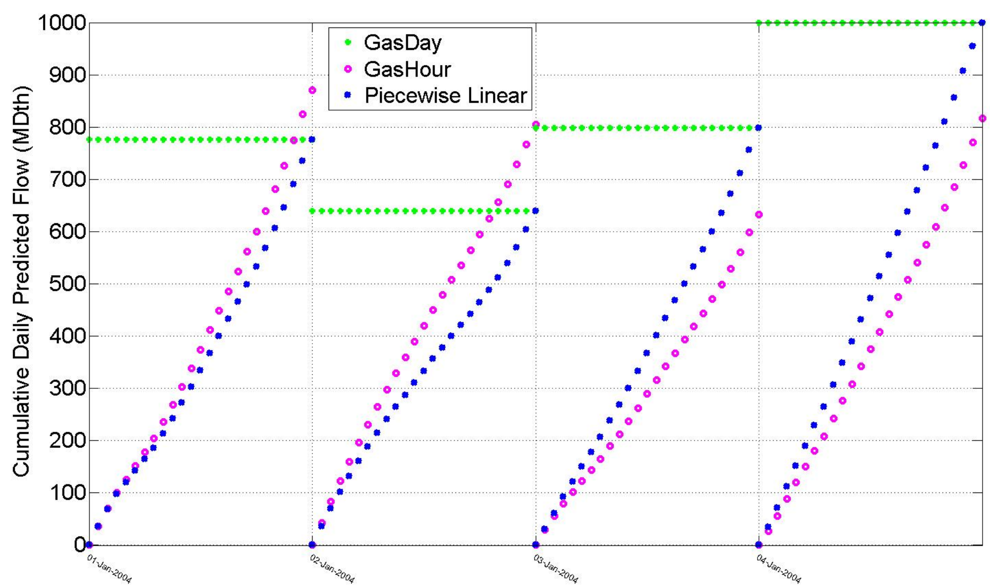


(b) GasDay and Sum of Hourly Forecasts
The adjustments force the Sum of GasHours to equal the GasDay forecast, however there is a small knot on Day 1. This knot corresponds to an increase of the decrement during the same time. Oscillating adjustments can cause these knots. This is not desirable, as it changes the shape of the GasHour forecasts.

Figure 12: Cubic Spline Method
Date Range: January 1, 2004 - January 4, 2004



(a) Hourly Increment Graph
There is no discontinuity, and the adjustments do not change much within the same gas day.



(b) GasDay and Sum of Hourly Forecasts
The adjustments force the Sum of GasHours to equal the GasDay forecast. There is no knot on Day 1.

Figure 13: Piecewise Linear Method
Date Range: January 1, 2004 - January 4, 2004

Cubic Spline Method. This method applies the same adjustment to most of the hours in each day and transitional adjustments to hours between days. The Piecewise Linear Method is the adjustment method we propose as the solution to the GasDay–GasHour Inequality.

5.3.1 Longterm Analysis of Piecewise Linear Method

The Piecewise Linear Method is our proposed solution, and it certainly solves the GasDay–GasHour Inequality. However, an analysis of how the adjustments effect the forecast is necessary before the solution is implemented. This analysis was performed by measuring the RMSE for each hour's forecast (between 0 and 95) over a 10 year period. The adjusted and original forecast RMSEs were calculated and can be seen in Figure 14.

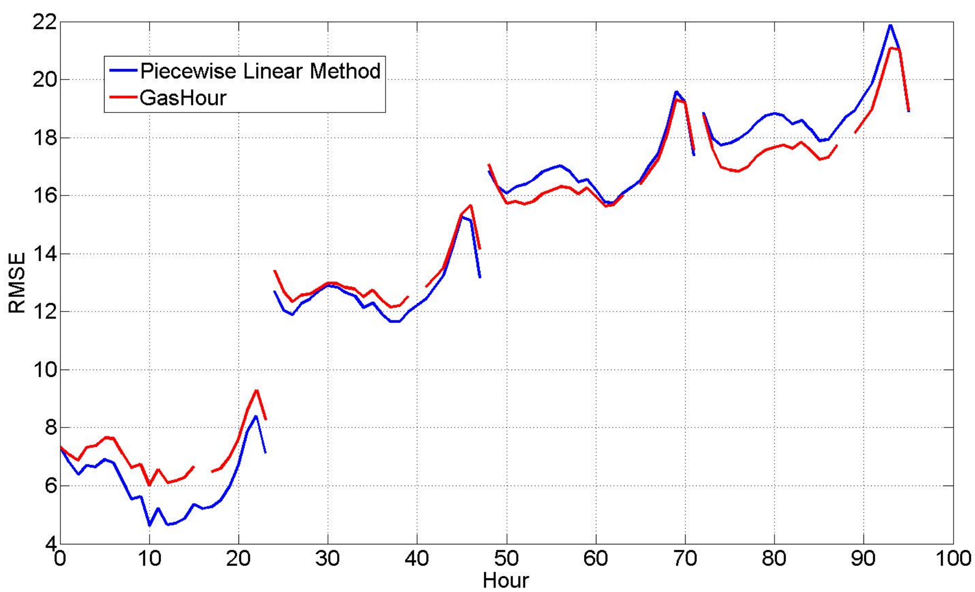


Figure 14: RMSE of GasHour and Adjusted GasHour Forecasts
Date Range: April 1, 2003 - April 30, 2013

It is clear from Figure 14 that the Day 0 forecasts were improved by the adjustment. However, it is unclear if there is improvement for Days 1 and 2, and the Day 3 forecasts were arguably worsened. This outcome is interesting, as these measurements reflect the results observed in Table 6. However, it does not reflect the results of the one might expect from adjusting toward the GasDay forecasts, but instead the results reflect the nature we would expect from adjusting the forecasts toward the components. In Table 6, the Day 0 component forecasts were considerably better, but the Day 1, 2, and 3 forecasts were slightly worse than the Sum of GasHours forecast. It would seem that adjusting the GasHour forecasts toward a GasDay forecast should cause it to improve. More investigation is needed to understand this phenomenon.

Given the analysis in its current state, the lab and its customers are presented with the question of importance. **Is it more important to have agreeing or accurate forecasts?** It is the author's opinion that a more accurate forecast is more desirable than agreeing forecasts. Also, if the Sum of GasHours is implemented as a component of the GasDay ensemble, then it could be observed how well GasHour has been performing lately by observing how much it has been weighted as a component of the GasDay model, as seen in Figure 10. For Days 1, 2, and 3, it is the author's recommendation that the adjusted be omitted. However, forcing the Day 0 forecasts to agree increases the accuracy of the GasHour forecasts. Therefore, we recommend that the adjustment be applied for Day 0.

6 Review

The research in this paper is now finished. Further exploration was not completed due to time constraints. This section reviews our work.

6.1 Conclusions

Error modeling is common practice in forecasting. In this paper, we have explored different plots of error to see if there were definable relationships that had not been explored previously. We observed that the forecast error from 2 days ago, the relationship of error to temperature, and the relationship of error to actual flow were not good predictors for the error of today's model. We observed that there may be a linear relationship between the errors of day 0 forecasts for the GD and GH forecasts. With some data cleaning, a relationship potentially could be defined and used to forecast the error for each model.

Ensemble forecasts generally improve in accuracy when more components are added. In this paper, we have replaced the MLR component of the GasDay ensemble with the Sum of GasHours. The RMSE of the new forecasts were competitive enough to suggest that the Sum of GasHours could be a potential input for the GasDay model. More importantly, if a method were implemented to adjust the GasHour forecast to equal the GasDay forecast, GasHour should be implemented as a component of the GasDay forecast, so that the GasHour forecasts are not at the mercy of the ANN and MLR models.

Three solutions were developed to solve the GasDay–GasHour Inequality. The Naive Method was rejected because it provided discontinuous adjustments. The Cubic Spline Method was rejected because its adjustments were oscillatory. The Piecewise Linear Method was chosen as the best solution. It provides continuous adjustments and is not subject to the variability of adjustments of the Cubic Spline Method. Over a period of 10 years, the Piecewise Linear Adjustments were shown to have considerably improved the hourly forecasts for Day 0. Adjustments for Days 1, 2, and 3 were inconclusive or made the forecasts worse. We recommend this solution be implemented for the Day 0 forecasts and used at discretion for the Day 1, 2, and 3 forecasts.

6.2 Future Work

There are many items of research and implementation that stem from this work.

6.2.1 Day 0 Error

The Day 0 error, as modeled in Figure 8 and discussed in Section 3.4, has potential for a good model of error. If some data cleaning were performed to rid the dataset of the points along the y-axis that do not generally fit in with the $y = x$ line, an equation may be able to be produced to model this error. The data along the y-axis is not bad data. Rather, these points are created when the Sum of GasHours has a significant error, and the GasDay forecast is accurate. This may occur on days with weather extremes or other anomalies. It could be said that the GasDay model performed *too well* on these days. If these days could be identified and removed, then there would be a set of data where the error of the Sum of GasHours and the GasDay act similarly. These errors potentially could be fit with a linear model.

If a linear model were found, a model such as

$$\begin{aligned}
C &= \alpha * (\text{GH Error}) + \beta * (\text{GD Error}) \\
C &= \alpha * (\text{Actual} - \text{GH}) + \beta * (\text{Actual} - \text{GD}) \\
C &= \alpha * \text{Actual} - \alpha * \text{GH} + \beta * \text{Actual} - \beta * \text{GD} \\
C + \alpha * \text{GH} + \beta * \text{GD} &= (\alpha + \beta) * \text{Actual} \\
\frac{C + \alpha * \text{GH} + \beta * \text{GD}}{\alpha + \beta} &= \text{Actual},
\end{aligned}$$

might be effective. This area has potential for improving the GasDay and GasHour Forecast.

6.2.2 Markov Processes

Markov processes have potential for improving GasDay forecasts, as discussed in Section 3.5. If the error state of the forecast were known, then the GasHour and/or GasDay forecast could be adjusted accordingly. For example, if it is known that today's forecast is in State 1, then it would be known that both forecasts are over-forecasting for the day. In this scenario, both forecasts could be adjusted downward to produce more accurate forecasts.

Conditional probabilities also were not carried as far as they could have been. The error state for Day 0 will not be known. However, it will be known how D relates to H (if D is greater than or less than H). With this known, a further conditional probability can be produced, separating the rows into columns 1-4 and 5-8. These tables were not produced due to time constraints; however it would be an easy process to complete.

Bayesian techniques are not used in any GasDay forecasting model. Markov Processes have a potential to be the first Bayesian technique to be added, but more detailed analysis needs to be conducted.

6.2.3 GasHour as a Component to GasDay

The research completed for this paper replaced the MLR forecast with the Sum of GasHours as a component of the GasDay adaptive ensemble forecast. A future researcher could consider adding the Sum of GasHours as a co-component to the GasDay forecast. Ensemble forecasts normally increase in accuracy when more components are added. The analysis performed in Section 4 suggests that the Sum of GasHours may be a good candidate for a new component. As seen in Table 6, the Sum of GasHours forecasts were very competitive with the ANN and MLR components of GasDay model, particularly on Days 1, 2, and 3. Adding the Sum of GasHours as a component can potentially improve the GasDay forecast on those days. It also will address the issue of conforming good GasHour forecasts to a bad GasDay forecast.

6.2.4 Sine Transformation and other Adjustments to the Piecewise Linear Method

Figure 13a plots the Piecewise Linear adjustments for each hour. When the plot changes from the standard error distribution to the sloped or transition period, the graph is non-differentiable. There is nothing "wrong" with this event; however a more gradual transition may be more desirable. One solution would be to replace the linear transition between adjustments with half a period of a sin wave. This adjustment would force the adjustments during the transition period to take on more of the error dispersement, removing some of the weight from the rest of the hours. This type of transition potentially can improve the GasHour forecast even further.

The improvement for the first day was least apparent during the first few hours (See Figure 14). This is the period where the Piecewise Linear Method implemented an increasing dispersion of error, starting with 0 for hour 0. During those early hours, it was also assumed that the GasHour would have the smallest error during those hours. This is not the case though, as seen in Figure 14. The first few hours of both the original and adjusted forecasts saw a decrease in the RMSE during Day 0, after the first few hours. Therefore, our assumption that GasHour would have the smallest error during the first few hours may not be correct, and should be revisited when improving this method. Different forms of dispersion for the first hours should be considered when improving the adjustment method.

6.2.5 Active Adjustments

All the forecasts that have been discussed in this paper have been for tomorrow's forecast and the days after. GasHour is often run throughout the day to help LDCs actively adjust to changing variables throughout the day. Research needs to be done to observe how adjustments to today's forecasts can be made to equal to today's GasDay forecast. It will probably be similar to the research presented in this paper. However, these forecasts will be different. There will be some hours where reported flows are available. These hours will not be adjusted, and that may have some adverse effects on the adjustment for the rest of the hours, particularly at the end of the gas day. Consider if the hourly flow for the first few hours is lower than expected. If the rest of the hourly forecasts are forced to take on the difference for the rest of the day, then those hours may take on a larger than necessary adjustment. This issue should be considered when implementing hourly adjustments.

6.2.6 Implementation

Implementation of the solution needs to be discussed. Consumers of GasDay have requested that this problem be solved, and this research has recommended a solution. As discussed in Section 5.3.1, these adjustments improve the Day 0 forecast, but not necessarily the Day 1, 2, and 3 forecasts. Consumers should be consulted on whether it is more important to have accurate or agreeing forecasts for the Day 1, 2, and 3 forecasts.

It is important to note that most of the Day 0 Hourly forecasts were improved, and this solution should been given serious consideration for implementation in the GasHour software.

6.3 Lessons Learned

I this section, I am going to review my time spent as a part of the REU and GasDay Lab. This section is less professional in the research paper-esque fashion, but I hope it provides desired details for those overseeing my time here.

6.3.1 REU Experience

While applying for REUs, this was the last one I applied for because it had the latest application closing date (March 1). This was my only application for the week, but was of course put off until February 28th because of other academic obligations. As I filled out the application though, I became more and more fond of what this REU had to offer. Mentorship. Weekly meetings with my peers. Largest paycheck. Interesting Projects. Great Location. Everything seemed so awesome here, and this REU became my number 1 choice. In the end, it became my only choice, as it was the only one I was accepted to. I am very thankful for that, and I hope that my work has lead the program to not regret doing so.

As appreciative as I am, I believe that my review would be incomplete if I did not provide my critiques of the program. Whether by my fault or the program's, I do not know my REU peers (besides my roommates) half as well as my GasDay peers. There are many reasons for this. It could be because I spent all of my working hours in the GasDay Lab. From what I have heard, even if I were to work in the REU lab, many of my REU peers did not choose to work there anyway. It could be that I do not drink, and many of the organized group activities that I heard planned were about visiting local bars, or including alcohol in other ways (which, the video game bar sounded awesome, but I would rather play video games at home with some Mountain Dew). It could be that I have an introverted personality and tend to avoid socializing. It could be that Bad Sci-Fi nights did not occur as planned. Whatever the case may be, I did not get to know my REU peers very well. That's not to say that the REU program did a poor job, that they could have done better, or that it is even the program's responsibility that I know my peers well, but it is something I am stating.

Overall, I believe the REU achieved its goal with me. I gained valuable experience in research. Although I am not any more sure if I want to attend graduate school, I think that has more to do with my desire to change majors. I am very appreciative of the culture and nature of research that I was able to experience, and will continually consider my experiences here as I go through my future.

6.3.2 GasDay Experience

When I look back at this experience, the REU tends to be a footnote to my experience at GasDay. From day one, I was thrown into the GasDay Lab with all of its students, faculty, projects, and culture, and I loved every second of it. These are my kind of people. Cookouts every week. Weekly meetings to discuss our research. Meetings between mentors and researchers happening all of the time. Customer interaction everywhere. This lab met all of my expectations and more. I will hold every future workplace environment I have to the standard that was set by this lab. I had a fantastic time this summer, and I am truly grateful to this lab and everyone in it.

6.3.3 Personal Experience

I had an awesome summer. I traveled to a part of the country I had never been before. This was my first experience living far from home, making my own meals, and truly learning how to take care of myself (At Belmont, I had a 16 meal a week meal plan and live an hour's drive from home). My roommates were great guys who taught me a lot about cooking, shopping, and meal planning, and they were really great guys to get to know over these ten weeks. I had a lot of fun playing frisbee, going to game nights, going to the mall, and finding various locations throughout Milwaukee.

I also had a really good job this summer. My past job experiences have included working at a Sonic Drive In, mowing lawns, tutoring, and being a full-time student. This job has truly been my favorite. I was not exposed to hot grease and freezing ice. I was not working in the sun all day (Although this is enjoyable, it can be tiring). I was not challenged with someone else's education. And I was certainly not required to do homework! (Well, I certainly did a lot of work, and some at my apartment, but it never had the essence of a ”homework” assignment.) Of course, I found joy in all of my previous experiences, but this experience was by far the best. I was placed in a lab where I was encouraged to work hard. I was given mentor guidance. I was consistently conversing and listening with colleagues on my and their projects. In the end, I was able to leave the lab with work that has great potential for improving the work they do there. I am very satisfied with the help and respect I was given from day one and the continued support I received everyday I was here in Milwaukee. This was an awesome experience, and I am so grateful that I was able to be a part of this.

Acknowledgements

First and foremost, I wish to thank Marquette University, Dr. Dennis Brylow, and Dr. Kim Factor for hosting this REU program. I have gained valuable experiences this summer that I will carry for the rest of my life.

I wish to thank the National Science Foundation for funding this program. This work was supported by the National Science Foundation under grant CCF-1063041.

I also thank my mentor for this project, Dr. George Corliss. Thank you so much for your dedication to excellence and your continuing love and support for high quality research. I am very grateful for the time, knowledge, and patience you invested in me.

I wish to thank the GasDay Lab and its many students and professors. I have never been in a research environment before, so you have set my expectations high for research, business, and workplace excellence. It was truly awesome to get to spend my summer with you. Thank you so much for welcoming me and supporting me this summer.

I wish to thank my parents, who for all of my life have supported my many endeavors.

I wish to thank my roommates, Nate and Bill. It was so great to have some guys to bounce ideas off of, as well as having someone available to help me cook, socialize, and truly be an adult for the first time in my life.

I wish to thank my Belmont academic advisors, Dr. Mike Pinter and Dr. Daniel Biles. They have fully supported all of my academic endeavors since coming to college and have continually helped guide me toward success in my undergraduate career.

References

[1] J. Scott Armstrong, editor. *Principles of Forecasting: A Handbook for Researhers and Practitioners*. Kluwer Academic Publishers, Norwell, Massachusetts, 4th edition, 2004.

[2] William L. Brogan. *Modern Control Theory*. Prentice Hall, Englewood Cliffs, New Jersey, 3rd edition, 1991.

[3] EPA. Subpart w basic information: Emissions from the petroleum and natural gas industry. Last Access: July 31, 2013. `www.epa.gov/ghgreporting/reporters/subpart/w-basicinfo.html`.

[4] Robert V. Hogg and Elliot A. Tanis. *Probability and Statistical Inference*. Prentice Hall, Upper Saddle River, New Jersey, eighth edition, 2010.

[5] Scott M. Lynch. *Introduction to Applied Bayesian Statistics and Estimation for Social Scientists*. Springer Science+Business Media, New York, New York, 2007.

[6] Melvin J. Maron and Robert J. Lopez. *Numerical Analysis*. Wadsworth Publishing Company, Belmont, California, third edition, 1991.

[7] David E. Remelfart and James L McClelland. *Parallel Distributed Processing: Explorations in the Microstructure of Cognition*, volume 1. MIT Press, Cambridge, MA, 1986.

[8] Stuart Russell and Peter Norvig. *Artificial Intelligence: A Modern Approach*. Prentice Hall, Upper Saddle River, New Jersey, second edition, 2003.

[9] Steven R. Vitullo, Ronald H. Brown, George F. Corliss, and Brian M. Marx. Mathematical models for natural gas forecasting. *Canadian Applied Mathematics Quarterly*, 17(7):807–827, Jan. 2009.