

8-5-2018

Prior Day Effect in Forecasting Daily Natural Gas Flow from Monthly Data

Maral Fakoor
Marquette University

George F. Corliss
Marquette University, george.corliss@marquette.edu

Ronald H. Brown
Marquette University, ronald.brown@marquette.edu

Electrical and Computer Engineering Faculty Research and Publications/College of Engineering

This paper is NOT THE PUBLISHED VERSION; but the author’s final, peer-reviewed manuscript.

The published version may be accessed by following the link in the citation below.

2018 IEEE Power & Energy Society General Meeting (PESGM). [DOI](#). This article is © Institute of Electrical and Electronic Engineers (IEEE) and permission has been granted for this version to appear in [e-Publications@Marquette](#). IEEE does not grant permission for this article to be further copied/distributed or hosted elsewhere without the express permission from IEEE.

Contents

Abstract:.....	2
SECTION I. Introduction to disaggregation	2
SECTION II. Forecasting Daily flow from monthly data.....	3
A. General Models	3
B. General Models with Prior-Day Adjustment.....	4
C. Forecast Daily Data from Monthly Flow	5
SECTION III. Results.....	6
A. Prior-Day Adjusted Daily Models.....	6
B. Prior-Day Adjusted Disaggregation.....	7
C. Discussions.....	10
SECTION IV. Conclusions	10
References	10

Prior Day Effect in Forecasting Daily Natural Gas Flow from Monthly Data

Maral Fakoor

GasDay Project, Marquette University, Milwaukee, WI

George F. Corliss

GasDay Project, Marquette University, Milwaukee, WI

Ronald H. Brown

GasDay Project, Marquette University, Milwaukee, WI

Abstract:

Many needs exist in the energy industry where measurement is monthly yet daily values are required. The process of disaggregation of low frequency measurement to higher frequency values has been presented in this literature. Also, a novel method that accounts for prior-day weather impacts in the disaggregation process is presented, even though prior-day impacts are not directly recoverable from monthly data. Having initial daily weather and gas flow data, the weather and flow data are aggregated to generate simulated monthly weather and consumption data. Linear regression models can be powerful tools for parametrization of monthly/daily consumption models and will enable accurate disaggregation. Two-, three-, four-, and six-parameter linear regression models are built. RMSE and MAPE are used as means for assessing the performance of the proposed approach. Extensive comparisons between the monthly/daily gas consumption forecasts show higher accuracy of the results when the effect of prior-day weather inputs are considered.

SECTION I. Introduction to disaggregation

Natural gas local distribution companies (LDCs) need to purchase the amount of natural gas that their customers need every day. Hence, it is important for LDCs to know how much gas they need to buy. Purchasing more than needed amount causes additional cost such as storage or penalty fees, but not having enough requires purchasing gas on the spot market at a premium price. Therefore, it is necessary to have accurate daily forecasts to reduce the operational cost to LDCs and subsequently to reduce the price for the natural gas customers.

When an LDC provides services to a new geographic zone or if a new daily gas demand forecasting model is to be built, there are cases in which only monthly/billing-cycle consumption data are available and not daily consumption. If we want to train the model with monthly consumption data and then forecast daily gas usage based on that, this can cause large errors because non-daily data does not contain enough information for estimating daily consumption.¹

The disaggregation problem is to separate an aggregate into its component parts. Data can be information gathered from multiple streams or measures and reported as aggregated data or all the mass is measured together. Disaggregation decomposes the aggregate into smaller units to acquire information needed.²

Disaggregation has a wide range of applications including hydrology, energy efficiency for electricity and natural gas, and others. In hydrology, Lee and Jeong³ used a model that combines k-nearest neighbor resampling (KNNR) with a genetic algorithm to downscale and disaggregate daily to hourly precipitation. In other research,⁴ a stochastic auto-regressive model is employed on stream flow time series (volume of water that moves through a specific point in a stream during a given period) to estimate daily flow from monthly data.

Armel et al.⁵ did disaggregation by applying a set of statistical approaches for extracting appliance level data using electricity data characteristics. An artificial neural network (ANN) is applied to do load profile identification and disaggregation.⁶ A Time Series Reconstruction (TSR) algorithm that uses a regression model and correlated variables to construct an estimate of unobserved time series natural gas consumption data was implemented by Vitullo.⁷ Askari et al.⁸ found a method that handles multiple time series with variable time intervals and tested on a gas network using Lagrange Multipliers method.

Using information from days other than today helps to improve results of models. Using yesterday's inputs is discussed in Ishola,⁹ which showed prior day temperature has an important role in natural gas consumption in extreme cold events. Linear regression models which included prior day adjustment forecast daily gas flow more accurately than models in which just today's temperature is considered.¹⁰

In this paper, linear regression daily models are built, and the effects of including prior-day inputs to improve the accuracy of the models are discussed. Using findings from the mentioned models, methods to forecast daily flow from historical monthly consumption are studied.

SECTION II. Forecasting Daily flow from monthly data

A. General Models

The aim of the current work is to build a model to forecast daily gas demand from monthly/billing-cycle consumption. In the literature, several mathematical models are used to forecast daily flow from historical daily data. Linear regression models are widely used in forecasting. Vitullo's research shows that the coefficients of the linear regression model parameterized on daily inputs and linear regression model parameterized on monthly/billing-cycle inputs are close.^{10,11} For ease of use, the first model will be called a daily model and second one will be called monthly model. Given the appropriate weather variables, the monthly can predict the demand characteristics within reasonable error margins.

Linear regression daily models are to be built. The general n-parameter linear regression model is to forecast daily gas flow from historical daily data:

$$\hat{S}_k = \beta_0^D + \sum_{i=1}^{n-1} \beta_i^D x_{k,i} \quad (1)$$

where β_i^D is the i^{th} daily coefficient, k is the day index, and x are inputs. The superscript D stands for *daily*. In the simplest two-parameter daily model, the input is HDDW. Heating degree day (HDD) and wind adjusted heating degree day (HDDW) are

$$HDD_{ref} = \max(T_{ref} - T, 0) \quad (2)$$

and

$$HDDW = \begin{cases} HDD \frac{152 + ws}{160} & ws < 8 \\ HDD \frac{72 + ws}{80} & ws > 8. \end{cases} \quad (3)$$

In (2), reference temperature T_{ref} is set to 65 F . A two-parameter model models heatload when temperature is below 65 F and baseload when temperature is above that. To use more information from weather, other parameters such as HDDW55 can also be added. In a four-parameter model, cooling degree day (CDD) has been added to help model demands for temperatures above 65 F ,

$$CDD_{ref} = \max(T - T_{ref}, 0). \quad (4)$$

Linear trend, or Trend, is another term that can be studied for gas consumption, as it can affect baseload and heatload values. The effects are modeled by two parameters, one to reflect its effect in the baseload as Trend itself, and the other to model its effects for heatload values as MHDD * Trend, in which MHDD is the mean heating degree day, the average of HDD65 and HDD55:

$$\begin{aligned} \hat{S}_k = & \beta_0^D + \beta_1^D Trend + \beta_2^D MHDD * Trend \\ & + \beta_3^D HDDW65_k + \beta_4^D HDDW55_k \\ & + \beta_5^D CDD65_k \end{aligned} \quad (5)$$

B. General Models with Prior-Day Adjustment

In all of the previous models, information about the current day's weather is considered. However, it is known that the prior-day's weather has an important role in current day's gas consumption.^{1,9, and 12} However, prior-day weather effects cannot be extracted from monthly data. By using domain knowledge, we can account for typical prior-day weather effects in the monthly to daily demand disaggregation process. For example, when we use the two-parameter model for disaggregation, consider a three-parameter model with a change in HDDW from the previous day

$$\hat{S}_k = \beta_0^D + \beta_1^D HDDW_k + \beta_2^D \Delta HDDW_k \quad (6)$$

where $\Delta HDDW_k = HDDW_k - HDDW_{k-1}$. Rearranging (6) gives

$$\hat{S}_k = \beta_0^D + \beta_1^D \left(HDDW_k + \frac{\beta_2^D}{\beta_1^D} \Delta HDDW_k \right) \quad (7)$$

The prior day weather sensitivity coefficient $\frac{\beta_2}{\beta_1}$ represents the dependence of day-to-day gas consumption when the temperature changes. This impact is a result of the thermodynamics of the buildings, which cause lags when conveying heat and other factors.¹² The value for $\alpha = \frac{\beta_2}{\beta_1}$ is typically in the range $[-0.3, -0.15]$. Hence, for the two-parameter disaggregation model, we assume a value for $\frac{\beta_2}{\beta_1}$ and replace $HDDW_k$ with

$$\begin{aligned} & HDDW_k + \frac{\beta_2^D}{\beta_1^D} \Delta HDDW_k \\ &= \left(1 + \frac{\beta_2^D}{\beta_1^D} \right) HDDW_k \\ &\quad - \frac{\beta_2^D}{\beta_1^D} HDDW_{k-1} \end{aligned} \quad (8)$$

to include the impact for prior-day weather changes. The same analogy can be made for higher order linear regression models.

C. Forecast Daily Data from Monthly Flow

A similar procedure can be applied to build linear regression models of monthly/billing-cycle data, or *aggregated data*. The general linear regression aggregation model can be formulated as

$$\sum_{k=1}^N \hat{S}_k = N * \beta_0^M + \sum_{k=1}^N \sum_{i=1}^{n-1} \beta_0^M x_{k,i} \quad (9)$$

with aggregation interval N (such number of days in a month for the problem of gas demand forecast), where superscript M stands for *monthly*. For a dataset containing more than one month, (9) is repeated for each month. As Vitullo showed, the coefficients of the two daily and

monthly linear regression models are interchangeable, given the appropriate weather inputs for temperature-sensitive operational areas. This has been verified experimentally.^{10,11} Therefore, it is assumed that having a linear regression model of the monthly/billing-cycle data built from training inputs, the same coefficients can be used to forecast the daily flow for test set using daily weather inputs with high degree of accuracy.

Linear regression can only account for the overall shape of the gas demand curves, as the process of aggregation is not reversible. However, we aim to infer the fluctuations in the daily demand by considering the prior-day's weather parameters. By taking advantage of daily models trained on a monthly dataset and considering its coefficients, we incorporate the prior-day's inputs by introducing the prior-day adjusted formulation, similar to (8) into the monthly model using daily weather inputs to forecast daily flows. The next section contains the results of various daily/monthly linear regression models and forecasting with and without considering the effects of prior-day adjustment.

SECTION III. Results

A. Prior-Day Adjusted Daily Models

To see the effects of prior-day adjustment on forecasting the daily data, daily inputs such as temperature, wind speed, and actual daily flow are used. For the purposes of training and testing, eight and two years of daily weather/flow data are used, respectively. For each experiment, data from 10 different operational areas around the US are collected. The daily flows are scaled between zero and 1000, and $N = 30$ is considered.

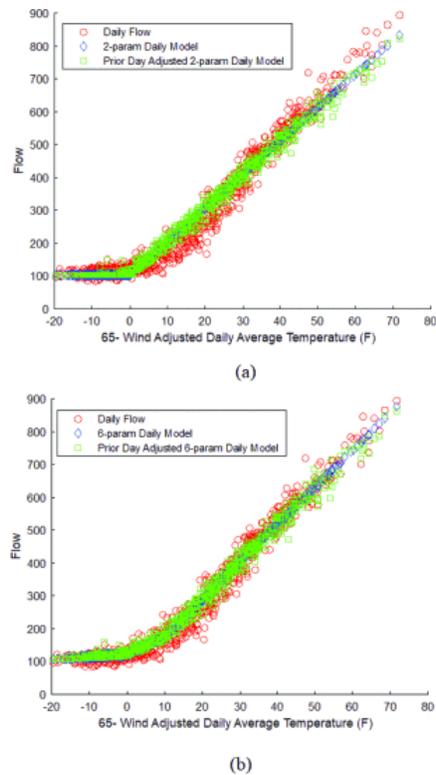


Figure 1. Comparison between the actual daily flow values (red circles), the daily models without the prior-day adjustment (blue diamonds) and with prior-day adjustment (green squares), for 2-parameter (a) and 6-parameter (b) linear regression models.

As discussed before, two-, three-, four-, and six-parameter linear regression models are generated from training daily weather and flow data. It is a common practice to show the flow data with respect to the relative wind adjusted daily average temperature, TempW,

$$TempW = \begin{cases} 65 - HDDW65 & HDDW65 > 0 \\ Temp & HDDW65 = 0. \end{cases} \quad (10)$$

Figs. 1 (a) and (b) show the forecast results for test set of two- and six-parameter daily linear regression models parametrized on daily data, with and without the prior-day adjustment with respect to TempW. The prior-day weather sensitivity coefficient is set to $\alpha = -0.25$, based on the average of minimum errors calculated from 10 different training datasets used for this study. For each figure, the red circles show the actual daily flow values, while the blue diamonds are the two- and six-parameter linear regression models' flow forecasts. Green squares, on the other hand, are drawn by including the prior-day adjustment component into the linear models. While the linear models are not able to represent the flow fluctuations of the actual data, prior-day adjusted models show more variations, and therefore enable more accurate modeling. This can be seen from Table I, where the RMSE and MAPE errors of linear regressions models of order two, three, four and six, with and without prior-day adjustment are presented. Table I shows that the four-parameter daily model with a small difference from six- and three-parameter daily model can better forecast daily flow and also that the prior day adjustment improves forecasting regardless of degree of the model.

TABLE I. Comparison Between Daily Models With And Without Prior-Day Adjustment (PDA) With The Ground Truth Daily Flow. MAPE And RMSE Are Average Values Over 10 Different Operational Areas Across US.

	MAPE		RMSE	
	<i>Daily Model</i>	<i>PDA Daily Model</i>	<i>Daily Model</i>	<i>PDA Daily Model</i>
2-Param	15.89	14.75	40.53	36.61
3-Param	14.70	13.56	36.40	31.45
4-Param	14.33	13.24	35.70	30.87
6-Param	14.35	13.28	36.44	31.76

B. Prior-Day Adjusted Disaggregation

For the disaggregation problem, where all daily weather inputs are available, but flow values are the available monthly or by billing cycles, the coefficients from daily model trained on monthly inputs are similar to the coefficients of the daily model trained on daily inputs, as discussed in the work of Vitullo et al.,^{10,11} To quantify forecast errors, 10 operational areas with known daily flow are considered, similar to the previous experiment for prior-day adjusted daily models. In each case, eight years' of data are used for training, weather and flow data are aggregated to generate the monthly data.

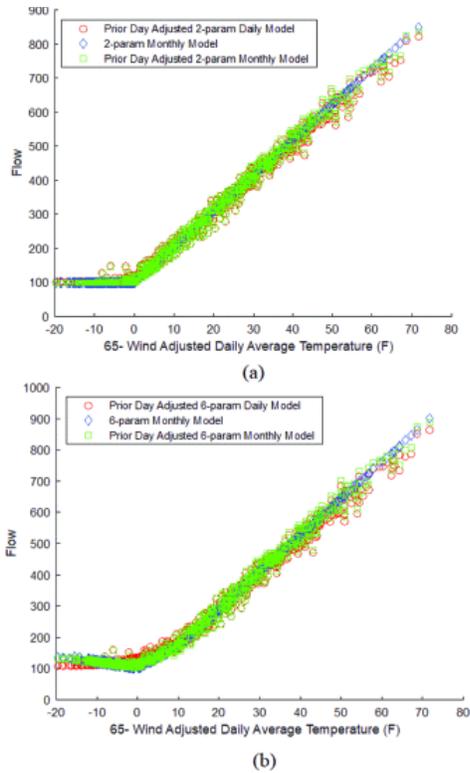


Figure 2. Comparison between prior-day adjusted daily models (red circles), monthly models without prior-day adjustment (blue diamonds) and with prior-day adjustment (green squares) for 2-parameter (a) and 6-parameter (b) linear regression models.

Figs. 2 (a) and (b) show the result for test set's forecasts from two- and six-parameter daily linear regression models trained on monthly inputs, with and without the prior day adjustment with respect to TempW. For all, the prior-day weather sensitivity coefficient is set to $\alpha = -0.25$ as in the previous experiment. Here, unlike in Figs. 1 (a) and (b), the red circles represent flow from prior-day adjusted daily models and not the actual daily values. This is to show how well the model trained on monthly data forecasts, as ideally we can approach the forecast made with historical daily data. Table II shows the RMSE and MAPE when comparing the forecast results from the monthly models, with and without the prior-day adjustment. As a result, the two-parameter daily model trained on monthly data forecasts daily flows which are closer to the forecasts made from two-parameter daily data trained on daily flows. The effect of the prior day adjustment is considerable compared to the case in which prior day adjustment was not used.

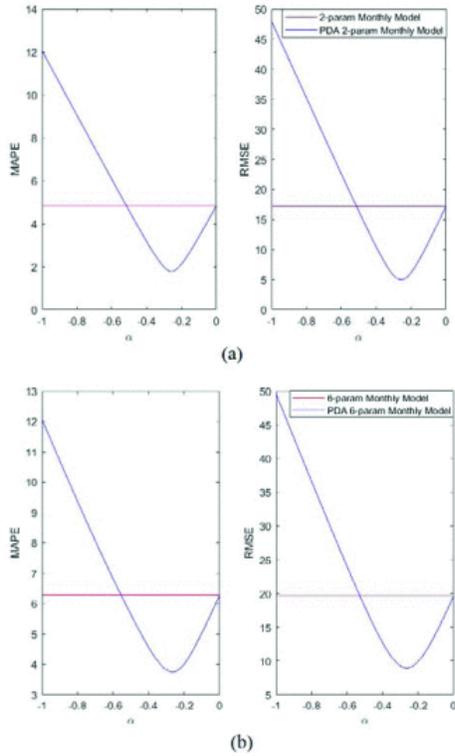


Figure 3. Comparison between the error metrics MAPE and RMSE for 2- (a) and 6-parameter (b) models, with (blue curves) and without (red lines) the effects of prior-day adjustment. The horizontal axis shows the changes in α value in its range $[-1,0]$.

To see the effect of changing the prior-day weather sensitivity coefficient α on the accuracy of the models, its values is changed over its range $[-1,0]$, and the RMSE and MAPE are computed for the monthly models and prior-day adjusted daily models. Fig. 3 (a) and (b) show the results for two- and six-parameter linear regression models, respectively. $\alpha = 0$ shows the case when only today's weather is considered, and $\alpha = -1$ means yesterday's flow was the only input. In each figure, the straight red line represents the error metrics without any prior-day adjustment, while the blue line shows the change in the error metric when the α value is changed in its range. The presented results are for one of the ten operational areas' test data. The results from other areas was qualitatively similar.

As can be seen, the average α does not result in the least possible error for this specific operational area. However, the error here is computed with the knowledge of actual daily flow, as prior day adjusted two- and six-parameter daily model is built from historical daily data, which is not the case for the problem of disaggregation. However, more optimized α values can be computed if the similarity between the weather variables of the operational areas used for computation are taken into consideration. Exploring the effects of such optimization is left to future research.

TABLE II. Comparison Between Disaggregation Results Computed By Monthly Models With And Without Prior-Day Adjustment (PDA) With The Prior-Day Adjusted Daily Models. Mape And Rmse Are Average Values Over 10 Different Operational Areas Across US.

	MAPE		RMSE	
	<i>Monthly Model</i>	<i>PDA Monthly Model</i>	<i>Monthly Model</i>	<i>PDA Monthly Model</i>
2-Param	6.67	3.73	20.32	9.50
3-Param	6.34	3.92	24.28	13.52
4-Param	7.90	5.43	25.53	15.14
6-Param	7.88	5.41	25.47	15.14

C. Discussions

Our experiments show that considering the effects of prior-day weather inputs to forecast the daily gas consumption yields reduced error metrics and higher accuracy, regardless of whether daily models are considered or disaggregation results. However, inclusion of more weather parameters does not always result in lower errors. For disaggregation, inclusion of more parameters consistently results in higher error compared to the same order of parameter daily model characterized on daily inputs, probably the result of over-parametrization of the linear regression modeling. In one hand, the dynamic behavior of the gas consumption can change over the observation period. On the other hand, the nonlinearities associated with the system may not be fully accounted for by linear regression modeling.¹³ More rigorous analysis of the sensitivity of the models to individual parameters is out of scope of the current work and is left to future research.

SECTION IV. Conclusions

We have investigated, the effect of prior-day adjustment of weather parameters in disaggregation of monthly/billing-cycle data. Given domain knowledge, linear regression models are considered for representing the relationship between the different weather parameters and the daily/monthly gas consumption and the trained models are used for forecasting. For the experiments, daily flow of 10 different operational areas from across US are considered as ground truth and two-, three-, four-, and six-parameter linear regression models are computed. At first, the effect of prior-day adjustment for accurate modeling of the original daily flow values is considered. Considering the effects of prior-day weather improves the accuracy of the model for all of the operational areas, regardless of the number of parameters. The same effects can be seen for the disaggregation problem. For this, the original daily flow values are aggregated to resemble the monthly/billing cycle data usually acquired by LDCs. Incorporating the prior-day adjustment into the monthly models reduces RMSE and MAPE errors in comparison to their not-adjusted counterparts. Overall, our experiments show the importance of including prior-day weather variables in daily flow forecasting and in disaggregation of monthly/billing-cycle flow data to daily gas consumption/demand.

References

1. S. R. Vitullo, R. H. Brown, G. F. Corliss, B. M. Marx, "Mathematical models for natural gas forecasting", *Canadian Applied Mathematics Quarterly*, vol. 17, no. 4, pp. 807-827, 2009.
2. S. Abbott, *Hidden curriculum*, Aug. 2014, [online] Available: <http://edglossary.org/hidden-curriculum>.

3. T. Lee, C. Jeong, "Nonparametric statistical temporal downscaling of daily precipitation to hourly precipitation and implications for climate change scenarios", *Journal of Hydrology*, vol. 510, pp. 182-196, 2014.
4. N. Rebora, F. Silvestro, R. Rudari, C. Herold, L. Ferraris, "Downscaling stream flow time series from monthly to daily scales using an auto-regressive stochastic algorithm: StreamFARM", *Journal of Hydrology*, vol. 537, pp. 297-310, 2016.
5. K. C. Armel, A. Gupta, G. Shrimali, A. Albert, "Is disaggregation the holy grail of energy efficiency? The case of electricity", *Energy Policy*, vol. 52, pp. 213-234, 2013.
6. Y. Xu, J. V. Milanović, "Artificial-intelligence-based methodology for load disaggregation at bulk supply point", *IEEE Trans. Power Syst.*, vol. 30, pp. 795-803, 2015.
7. S. R. Vitullo, G. F. Corliss, M. Adya, F. Nourzad, R. H. Brown, "Disaggregation of energy consumption data using correlated variables", *Canadian Applied Mathematics Quarterly*, vol. 21, no. 3, 2013.
8. S. Askari, N. Montazerin, M. F. Zarandi, "High-frequency modeling of natural gas networks from low-frequency nodal meter readings using time-series disaggregation", *IEEE Transactions on Industrial Informatics*, vol. 12, pp. 136-147, 2016.
9. B. I. Ishola, R. J. Povinelli, G. F. Corliss, R. H. Brown, "Identifying extreme cold events using phase space reconstruction", *International Journal of Applied Pattern Recognition*, vol. 3, no. 3, pp. 259-275, 2016.
10. R. H. Brown, P. E. Kaefer, C. R. Jay, S. R. Vitullo, "Forecasting natural gas design day demand from historical monthly data", *PSIG Annual Meeting. Pipeline Simulation Interest Group*, 2014.
11. S. R. Vitullo, "Disaggregating time series data for energy consumption by aggregate and individual customer", *PhD. Dissertation*, 2011.
12. P. E. Kaefer, "Transforming analogous time series data to improve natural gas demand forecast accuracy", *Master thesis*, 2015.
13. M. B. Beck, G. van Straten, *Uncertainty and forecasting of water quality*, Springer Science & Business Media, 2012.