Interim report

Modelling the demand for gas by industry
Data Science (with Statistics) MSc

Daniel James Dennis
160413256
June 11, 2021
Supervisor: Sarah Heaps

Introduction

As the climate crisis looms, we must not only innovate, but adapt to be more efficient in our energy consumption. Gas is a vital part of the energy sector in the UK, with 57% of industrial and commercial energy needs being met by gas in 2020 (*Enabling the gas markets plan*, 2019). As the UK moves towards its target of net-zero emissions by 2050, satisfying the changing demand for gas will be integral in this fight. The National Grid has submitted the Gas Markets Plan (GMaP), in which one of the three focus areas for 2020 is "balancing", describing the process of trying to maintain equal levels of gas brought on and off the network. Herein lies an issue – it is essential to be informed of how much gas will be required in one day's time, one month's time or one year's time.

This research will use daily data provided by Northern Gas Networks for two regions, known as local distribution zones (LDZ's), in the North and North East of England. This spans just over a decade and features daily metered demand from large industrial users[1], a composite weather variable (CWV) and calendar information (day of the week and date). This will be combined with statistical methods, namely time series analysis, while model-fitting will be carried out in the Bayesian framework.

There is a diverse literature regarding modelling the demand for gas, highlighting the sectors need for such research and the many angles the problem can be tackled from. Studies modelling the demand for gas are rarely focused solely on industrial consumers, usually focusing on residential and commercial consumers, or an aggregate of all three, generally providing short-term to medium-term forecasts. These often involve a model based on weather-related and calendar-related factors. Weather-related factors largely influence gas consumption due to consumers changing their heating demands, while calendar-related factors provide a platform to model the differences in gas consumption on weekdays, weekends, and public holidays, caused by changes to working patterns. A number of other factors may be included, such as oil prices, GDP, and population growth.

The work of Lyness (1984), recognised the presence of the main patterns of demand on daily, weekly, and yearly scales, while Vitullo et al. (2009), and Soldo (2012), provide comprehensive overviews of the statistical and computer science-driven models that have been developed in this area. Sánchez-Úbeda & Berzosa (2007), present a novel approach combining quantitative and qualitative forecasts for industrial users in Spain. Huntington (2007), and Zhu et al. (2014), include several socioeconomic variables to their models, which both feature some autoregressive component. Heaps et al. (2020), examine the demand for gas in the same two regions featured in this paper, however their work investigates the effect of public holidays on residential and commercial consumers' daily demand for gas using a four-state, non-homogeneous hidden Markov model.

Aim and objectives

The importance of understanding future demands for gas shapes the motivation for this paper; to model and forecast the daily gas demand from industrial consumers in two regions of England over a medium-term horizon (up to about 5 years ahead). To achieve this, four statistical goals need to be completed. First, the exploratory data analysis (EDA) and literature will be used to derive appropriate statistical models for data from the two LDZ's. Second, a Markov chain Monte Carlo algorithm will be constructed to fit both models. Third, both models will be fitted to the data and the results interpreted and analysed. Fourth and finally, the models will be combined, and inferencing procedures will be undertaken to produce a joint model so that the shared structure in the data can be exploited. The project is on track to be completed by its submission date, with the Gantt chart displaying a structured plan to do so.

Overview of progress

The project began by loading the data into RStudio, isolating the necessary columns from the data frame, removing any null entries, then joining a data frame to it which contained the bank holidays from each year.

[1] Large industrial premises which have their meters read daily, *Gas demand forecasting methodology* (2016)

Exploratory data analysis was undertaken, providing a thorough visual overview of the data. This sought to provide insight on how the demand varied on weekdays, weekends, and bank holidays, how the demand varied throughout the year, and how the demand varied with the CWV. Firstly, the EDA gave evidence that weekdays see higher demand than weekends, and non-bank holidays see higher demand than bank holidays. It is assumed the reason behind this is that industrial consumers are likely to reduce capacity or stop operations on weekends and bank holidays, thus lowering their demand for gas. Secondly, the EDA showed signs that the demand for gas peaked in winter and troughed in summer, which could be due to industrial consumers requiring less heating at their sites through summer, thus decreasing demand for gas. Thirdly, the EDA displayed a weak negative correlation between the CWV and the daily metered demand. This further suggests some link between the weather and the demand for gas, in that as the weather improves (i.e., temperature increases, wind decreases), demand decreases, suggesting some link to heating requirements. There may, however, be some less obvious reason causing the demand in summer/warmer periods to drop. The North East daily metered demand data displays a time series where observations alternate between four well-defined levels, while the data for the North exhibits periods of linear growth with some step changes to the gradient throughout.

Based on the reversible steps commented upon in the EDA, a hidden Markov model seems appropriate for the North East data (see Frühwirth-Schnatter (2006)). The model follows, where the state equation is $Pr(S_t = k|S_{t-1} = j) = \lambda_{jk}$ and the observation equation is $Y_t|S_t = k \sim N(\mu_k, \sigma_k^2)$ for $k = 1, ..., K$. In order to learn about the parameters of the hidden Markov model, a Gibbs sampler with data augmentation will be used. This alternates between drawing samples of the hidden states given the parameters and the parameters given the hidden states. To carry out the latter steps, the full conditional distributions of the model parameters need to be computed. The full conditional distributions for $\mu_1, ..., \mu_k$ and $\sigma_1^2, ..., \sigma_k^2$ have been calculated from the priors $\mu_1, ..., \mu_k \sim N(m, v)$ and $\sigma_1^2, ..., \sigma_k^2 \sim invgamma(\alpha, \beta)$, where

$$\mu_k|\boldsymbol{y}, \boldsymbol{s}, \sigma_k^2, \xi \sim N\left(\frac{vN_k\bar{y}_k + \sigma_k^2 m}{vN_k + \sigma_k^2}, \frac{\sigma_k^2 v}{vN_k + \sigma_k^2}\right)$$

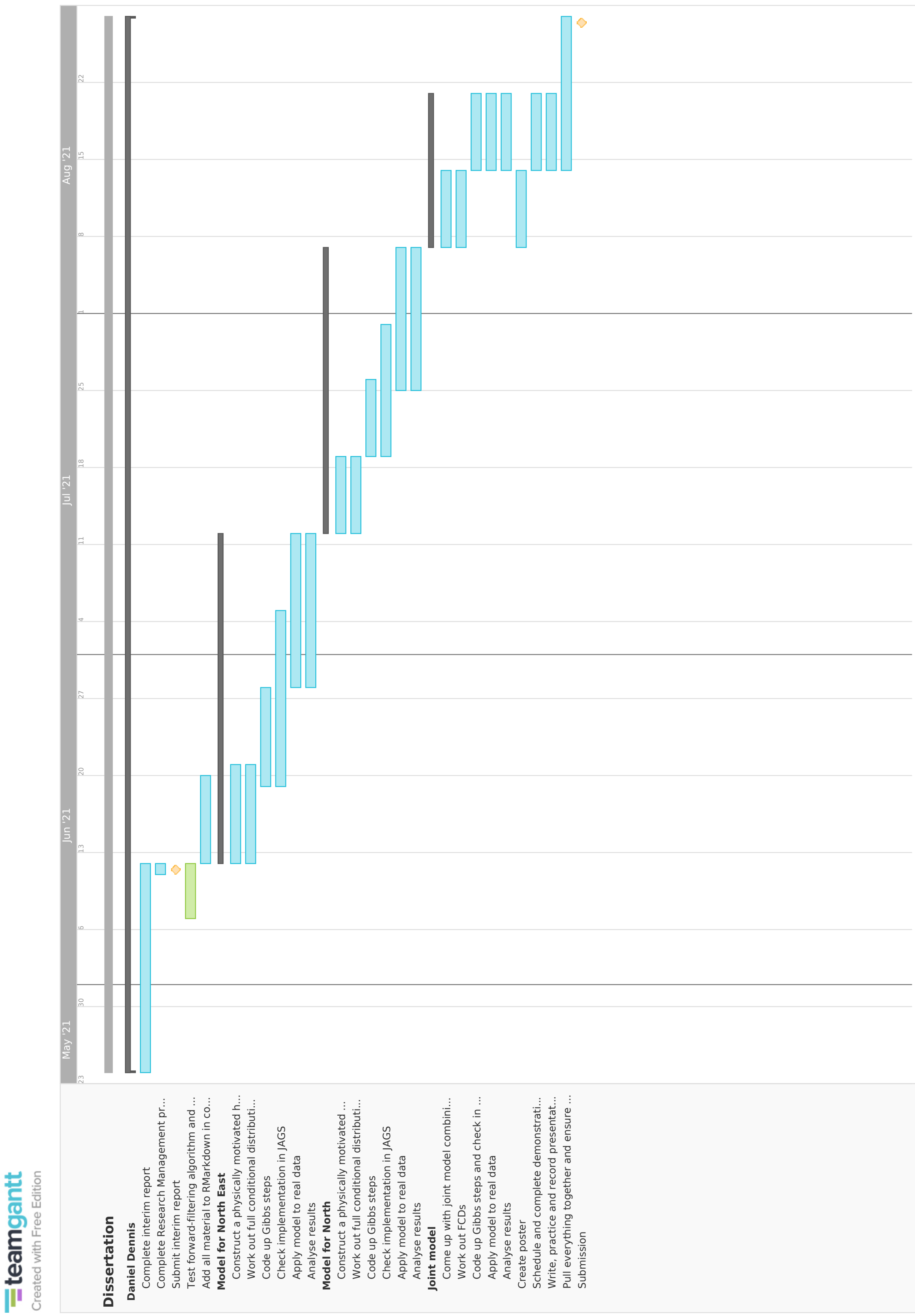
and

$$\sigma_k^2|\boldsymbol{y}, \boldsymbol{s}, \mu_k, \xi \sim invgamma\left(\alpha + \frac{N_k}{2}, \beta + \sum_{t:s_t=k} \frac{(y_t - \mu_k)^2}{2}\right)$$

where N_k is the number of observations in state k and $\bar{y}_k = \frac{1}{N_k}\sum_{t:s_t=k} y_t$ for $k = 1, ..., K$.

The next step will be to complete the forwards-filtering-backwards-smoothing algorithm to sample the hidden states, before creating a physically motivated hidden Markov model for the North Eastern data adapting the Gibbs sampler to fit the model. Later, a dynamic linear model will be derived for the Northern data. If time permits, a joint model will be created which combines the two aforementioned models.

Project plan



References

Anon (2019) *Enabling the gas markets plan.* National Grid.

Anon (2016) *Gas demand forecasting methodology.* National Grid.

Frühwirth-Schnatter, S. (2006) *Finite mixture and markov switching models.*

Heaps, S.E., Farrow, M. & Wilson, K.J. (2020) Identifying the effect of public holidays on daily demand for gas. *Journal of the Royal Statistical Society: Series A (Statistics in Society).* 183 (2), 471–492.

Huntington, H.G. (2007) Industrial natural gas consumption in the united states: An empirical model for evaluating future trends. *Energy Economics.* 29 (4), 743–759.

Lyness, F.K. (1984) Gas demand forecasting. *Journal of the Royal Statistical Society: Series D (The Statistician).* 33 (1), 9–21.

Sánchez-Úbeda, E.Fco. & Berzosa, A. (2007) Modeling and forecasting industrial end-use natural gas consumption. *Energy Economics.* 29 (4), 710–742.

Soldo, B. (2012) Forecasting natural gas consumption. *Applied Energy.* 92 (C), 26–37.

Vitullo, S.R., Brown, R.H., Corliss, G. & Marx, B.M. (2009) Mathematical models for natural gas forecasting. *Can. Appl. Math. Q.* 17 (4), 807–827.

Zhu, Q., Lu, Q., Zhou, X. & Lai, K.K. (2014) A driving force analysis and forecast for gas consumption demand in china. *Mathematical Problems in Engineering.* 2014 (1), 1–11.