Nonlinear and Spline Regression Models for Forecasting Gas Flow on Exits of Gas Transmission Networks

Radoslava Mirkov[1] , Herwig Friedl[2]

[1] Humboldt Universität zu Berlin, Department of Mathematics, Unter den Linden 6, 10999 Berlin, Germany, mirkov@math.hu-berlin.de
[2] Graz University of Technology, Institute of Statistics, Münzgrabenstraße 11, 8010 Graz, Austria

Abstract: The flow of natural gas within a gas transmission network is studied with the aim to predict gas loads for very low temperatures. Two models for describing dependence between the maximal daily gas flow and the temperature on network exits are presented. A Brain-Cousens regression model is chosen from the class of parametric models. As an alternative, a semi-parametric logistic regression based on penalized splines is considered. The comparison of prediction based on both models is included.

Keywords: nonlinear regression; penalized splines; gas flow, design temperature.

1 Introduction and Model Motivation

We study historical data of the flow of gas transported in networks in order to support a reliable and realistic prediction of the future gas flow. The forecast of gas loads at the so-called design temperature is of particular interest. The design temperature is the lowest temperature at which the gas operator is still obliged to supply gas without failure, and lies between $-12°$C and $-16°$C. Such low mean daily temperatures are very uncommon in Germany, and there is no gas flow data available at the design temperature. For this reason gas operators are forced to use the predicted gas loads at the design temperature, and we present here two models useful for the forecast.

Data is obtained from measuring stations within the German pipeline network operated by Open Grid Europe GmbH, one of the largest German gas transporters. It contains hourly gas flow for the period between January 2004 and June 2009, and the corresponding mean daily temperatures. We study the dependence of gas loads and air temperature on all exits along the pipelines. Typical exits in such networks are public utilities, industrial consumers and storages, as well as exits on border and regional crossings. Since we want to maximize the transportation capacity through the pipelines, we concentrate on the daily maximum flows y_i^{max}, $i = 1, \dots, n$ $(n = 2005)$,

at each exit, for every exit in the network. We consider the standardized daily maximum flows $y_i = y_i^{max}/\bar{y}$, where $\bar{y}$ denotes the empirical mean of all maximal daily gas flows at one specific measuring station.
The following model to describe the dependence of the standardized maximal daily gas loads y_i on temperature t_i is studied:

$$y_i = S(t_i) + \varepsilon_i, \tag{1}$$

where t_i stands for the weighted four-day-mean temperature with the weights $(0.53, 0.27, 0.13, 0.07)$, and $\varepsilon_i \overset{\text{iid}}{\sim} \mathcal{N}(0, \sigma^2)$ are error terms, for $i = 1, \ldots, n$, as suggested in Cerbe (2008).
Friedl et al. (2011) explore different modeling possibilities for this problem, and suggest several appropriate variants for the function $S(t_i)$. They also compare advantages and disadvantages of both approaches.

2 Nonlinear and P-Splines Regression Models

We fit a parametric as well as a semi-parametric nonlinear logistic regression model and analyze the properties of the gas flow through the pipelines in dependence of the temperature and the forecast based on these models. The so-called Brain-Cousens model (BC-model) is proposed by Ritz and Streibig (2008) for this kind of problems, while many authors propose some variant of spline regression, see e.g. Jones et al. (2009), Jarrow et al. (2004), Eilers and Marx (1996). A comparison of both approaches for a duck growth problem is presented in Vitezica et al. (2010).
In the class of parametric models, we consider the BC-model, which is defined by

$$S(t_i) = \theta_4 + \frac{\theta_1 + \theta_6\left(\frac{\theta_2}{t_i - 40°\text{C}} + d_i\theta_5\right) - \theta_4}{1 + \left(\frac{\theta_2}{t_i - 40°\text{C}} + d_i\theta_5\right)^{\theta_3}}, \tag{2}$$

where

$$d_i = \begin{cases} 1 & \text{if day } i \text{ is a working day,} \\ 0 & \text{if day } i \text{ is a holiday or at weekends,} \end{cases}$$

indicates whether the gas loads occurred on working days or on weekends and holidays.
The parameters in model (2) are used as follows: θ_1, θ_6 and θ_4 define the upper and lower asymptotes, θ_5 indicates the type of the day, while θ_2 and θ_3 describe the shape of the decrease of the curve. We use initial values provided in Friedl et al. (2011). The results of the evaluation are given in Table 1. Parameters θ_5 and θ_6 in the model are significant, implying that the that the expected gas loads differ during the week (W) and on weekends and holidays (H), and the upper asymptote is a line with slope θ_6. For low temperatures, the modified upper asymptote in the BC-model implies the

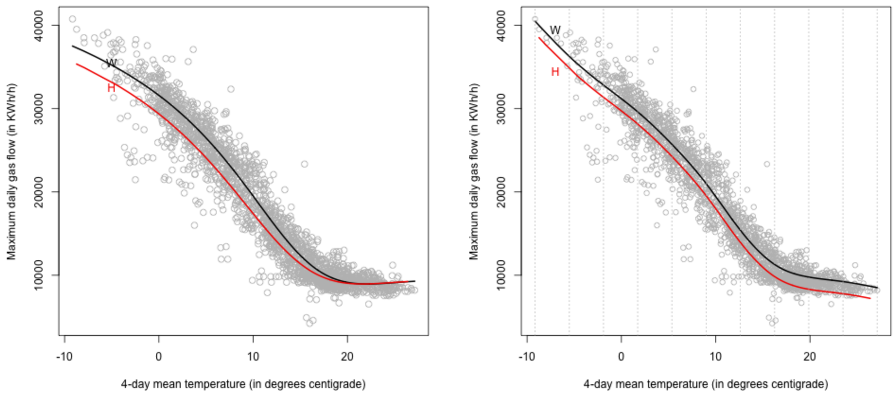

FIGURE 1. Fitted BC-model with indicator day (left) and penalized splines regression with indicator day (right) based on cubic B-splines on the mesh with 10 segments and the second order penalty $\lambda = 2.51$.

day-specific increase of the mean gas flow for approximately 2 times scaled $\bar{y}$ when the temperature decreases for 1°C. The graphical representation of model (2) is shown in Figure 1 (left).

TABLE 1. MLEs (std. errors) of the BC-model.

θ_1	θ_2	θ_3	θ_4	θ_5	θ_6
3.1805	−28.0417	6.5713	0.5229	−0.0447	−2.0807
(0.1062)	(0.7960)	(0.3431)	(0.0197)	(0.0031)	(0.2269)

Alternatively, Friedl et al. (2011) suggest the penalized splines (P-splines) approach, and assume that the function $S(t_i)$ is the linear combination of basis functions $B_j, j = 1, \ldots, m$, on the mesh Δ, given by

$$S(t_i) = \sum_{j=1}^{m} a_j B_j(t_i) + a_{m+1} d_i, \tag{3}$$

and B_j are basis functions of the B-spline of degree q, and the mesh Δ is an equidistant grid over $m - q$ segments, i.e. with $m - q + 1$ inner knots. The regression coefficients are obtained taking into account the smoothing penalty λ. We refer to Figure 1 (right) for a graphical representation of the fitted P-splines model and the position of the inner knots based on cubic B-splines on the mesh with 10 segments and the second order penalty $\lambda = 10^{0.4} = 2.51$.

3 Prediction

The models presented in Section 2 are now utilized for the prediction of gas loads at the design temperature. Recall, the design temperature lies outside of the domain of the predictor variable $t_i, i = 1, \ldots, 2005$. To this end, we replace the existing temperature t_i by a new predictor variable $\tilde{t}_k, k = 1, \ldots, \tilde{n}$, generated as an equidistant grid of temperatures, which includes low temperatures of interest. In particular, we generate $\tilde{t}_k$ starting from the lowest possible design temperature, i.e. $-16°$C, and go up to $35°$C with step size 1. Based on the new data and the fitted models (2), and (3), the predictions

$$\tilde{y}_k = S(\tilde{t}_k) + \varepsilon_k, \qquad \varepsilon_k \overset{\text{iid}}{\sim} \mathcal{N}(0, \sigma^2), \qquad k = 1, \ldots, 52,$$

are calculated.

The predicted values based on the BC-model are obtained using the `predict` method in R, as described in Ritz and Streibig (2008). P-splines allow straightforward smooth extrapolation, and we exploit this property to forecast gas loads at the design temperature. The second order penalty implies the extrapolation by a linear sequence, cf. Eilers and Marx (2010).

Figure 2 illustrates the prediction based on the BC-model and P-splines regression. At the design temperature of $-12°$C the predicted gas loads on working days based on models (2), and (3), are 38817, and 43048 KWh/h, respectively.

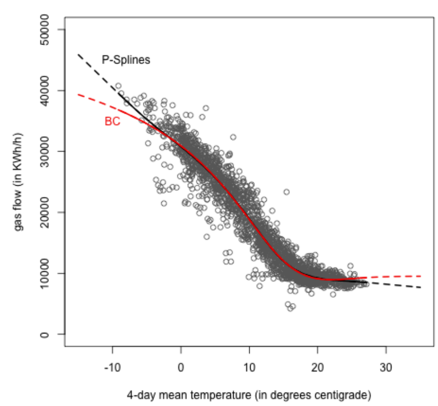

FIGURE 2. Prediction for working days based on the BC and P-splines model.

In the case of nonlinear regression models, the standard error estimates do not change substantially when we leave the domain of the predictor variable. Figure 3 (left) represents the predicted values for working days based on the BC-model (2), and the corresponding standard error bands. The naïve method based on the assumptions of the normality of error terms and the variance homogeneity is emplyed to determine standard errors of parameters. Some other methods for constructing prediction intervals for nonlinear regression can be found in Gauchi et al. (2010), and Ritz and Streibig (2008).

It is well known that extrapolation in the case of splines can be unsafe for the prediction, although the model provides a good fit for gas loads. This fact is reflected in the shape of the error bands for the P-splines model (3). Due to the local smoothing, the fit is very good with a small error band width within the domain of the predictor variable, while the increase in the width of error bands is large as soon as we extrapolate. The Bayesian estimate of the standard error bands for the fitted P-splines model for working days are shown in Figure 3 (right).

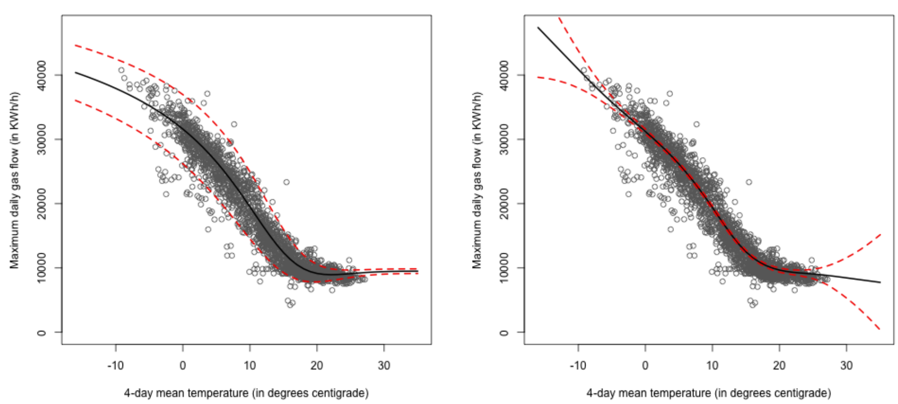

FIGURE 3. Prediction for working days based on the BC model (left) and P-splines (right) with the corresponding standard error bands.

4 Conclusions

We investigate prediction based on the nonlinear BC-model and on the semi-parametric P-splines regression. Both the BC-model and the P-splines reflect the behavior of gas flow for low temperatures in a realistic way. We note that the nonlinear regression models are generally more difficult to handle than the local smoothers like the P-splines, because of their numerical properties. Contrary to them, the P-splines methodology is a very flexible simpler alternative, but it does not support the multiple regression techniques, and one cannot exploit the desirable flexible temperature effects. The forecast of gas loads based on the BC-model is safer than the one relying on the P-splines, due to the numerical construction of models.

References

Cerbe, G. (2008). *Grundlagen der Gastechnik*. Hanser Verlag.

Eilers, H.C., and Marx, B.D. (1996). Flexible Smoothing with B-splines and Penalties. *Statistical Science*, **11(2)**, 89-121.

Eilers, H.C., and Marx, B.D. (2010). Splines, Knots, and Penalties. *Wiley Interdisciplinary Reviews: Computational Statistics*, **2(6)**, 637-653.

Gauchi, J-P., Vila, J-P., and Coroller, L. (2010). New Prediction Interval and Band in the Nonlinear Regression Model: Application to Predictive Modeling in Foods. *Communications in Statistics - Simulation and Computation*, **39(2)**, 322-334.

Jarrow, R., Ruppert, D., and Yu, Y. (2004). Estimating the Interest Rate Term Structure of Corporate Debt With a Semiparametric Penalized Spline Model. *Journal of the American Statistical Association*, **99(465)**, 5766.

Jones, G., Leung, Y., and Robertson, H. (2009). A Mixed Model for Investigating a Population of Asymptotic Growth Curves Using Restricted B-Splines. *Journal of Agricultural, Biological, and Environmental Statistics*, **14(1)**, 66-78.

Friedl, H., Mirkov, R., and Steinkamp, A. (2011). Modeling Gas Flow on Exits of Gas Transmission Networks. *Submitted to a Special Issue of the International Statistical Review on Energy Statistics*.

Ritz, C., and Streibig, J.C. (2008). *Nonlinear Regression with R*. Springer.

Vitezica, Z.G., Marie-Etancelin, C., Bernadet, M.D., Fernandez, X., and Robert-Granie, C. (2010). Comparison of nonlinear and spline regression models for describing mule duck growth curves. *Poultry Science*, **89**, 1778-1784.