Sigmoid Models Utilized in Optimization of Gas Transportation Networks

Radoslava Mirkov [1], Herwig Friedl [2], Hernan Leövey [1], Werner Römisch [1] and Isabel Wegner-Specht [1]

[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 an aim to apply these results in order to optimize such networks. The analysis of real data provides a deeper insight into the behavior of gas in- and outflow. A sigmoid regression model is chosen to describe dependence between the maximal daily gas flow and the temperature on network exits.

Keywords: Sigmoid Regression; Gas Transport; Optimization.

1 Introduction

Transportation and supply of natural gas is a an important topic, and we study the flow of gas transported in networks in the past in order to support the optimization of such networks and thus improve the supply of gas. To do so we fit a nonlinear regression model and analyze the properties of the gas flow through the pipelines in dependence of the temperature.
Data is obtained from measuring stations within the German pipeline network, and contains hourly gas flow for the period of the last five years. Mean daily temperatures are also provided. We study the dependence between gas consumption and air temperature on all exits along the pipelines. Since we want to maximize the transportation capacity through the pipelines, we concentrate on the daily maximum flows y_i, $i = 1, \ldots, n$, at each exit, for every exit in the network.

2 Sigmoid Regression

In what follows, we concentrate on the data observed at one specific station. Based on the Cooperation Agreement (2008) between gas companies, we choose the following sigmoidal growth model to describe the dependence of gas consumption on temperature:

$$y_i = \mu \times S(\theta|t_i) + \varepsilon_i, \tag{1}$$

where μ denotes the overall mean of all maximal daily gas flows at one specific measuring station, t_i stands for the weighted four-day-mean temperature with weights $(0.5333, 0.2667, 0.1333, 0.0667)$, the sigmoid function S with parameter $\theta = (\theta_1, \theta_2, \theta_3, \theta_4)$ is given by

$$S(\theta|t_i) = \theta_4 + \frac{\theta_1 - \theta_4}{1 + \left(\dfrac{\theta_2}{t_i - 40°C}\right)^{\theta_3}}, \tag{2}$$

and $\varepsilon_i \overset{\text{iid}}{\sim} \mathcal{N}(0, \sigma^2)$ are error terms.
The choice of this model is based on physical properties of gas, see e.g. Cerbe (2008). The four parameters in model (2) have the following meaning: θ_1 and θ_4 are the upper and lower asymptotes, and the other two parameters describe the shape of the decrease of the (logistic like) curve. More precisely, θ_2 is the inflection point of the curve, and θ_3 is proportional to the slope at $t = \theta_2$ (cf. Ritz and Streibig, 2008).
The starting values for the iteration necessary to calculate the maximum likelihood estimates, as provided in the Cooperation Agreement (2008), are given in the Table 1. Alternatively, according to Seber and Wild (2003), we obtain the crude initial estimates of θ_1 and θ_4 from the scatterplot, while θ_2 and θ_3 can be obtained using the linearization

$$y^* = \theta_3 \log(-\theta_2) - \theta_3 \log(-t + 40°C).$$

Note that the starting values, which have been estimated from data are much closer to the fitted parameters than the values given in the Cooperation Agreement (2008), whereas both sets of starting values yield the same model parameters. The model specified by (1) and (2) is fitted in R using

TABLE 1. Starting Values of the Regression Parameters.

	$\theta_1 - \theta_4$	θ_2	θ_3	θ_4
Agreement	2.3878	−34.7213	5.8164	0.1208
Estimated	1.4737	−32.0284	6.0055	0.4375

the function `nls()`. We refer to Table 2 for the estimated parameters in the fitted model.

TABLE 2. Maximum Likelihood Estimates of the Regression Parameters.

$\theta_1 - \theta_4$	θ_2	θ_3	θ_4
1.5862	−32.6474	6.6644	0.4468

In Figure 1(left) we show the model fitted to data describing the typical gas outflow for public utilities, as well as the curves corresponding to both sets of starting values given in Table 1. This figure suggests that the sigmoid model reproduces the gross characteristics of the gas flow well, though it obviously underestimates the mean responses for low temperatures.

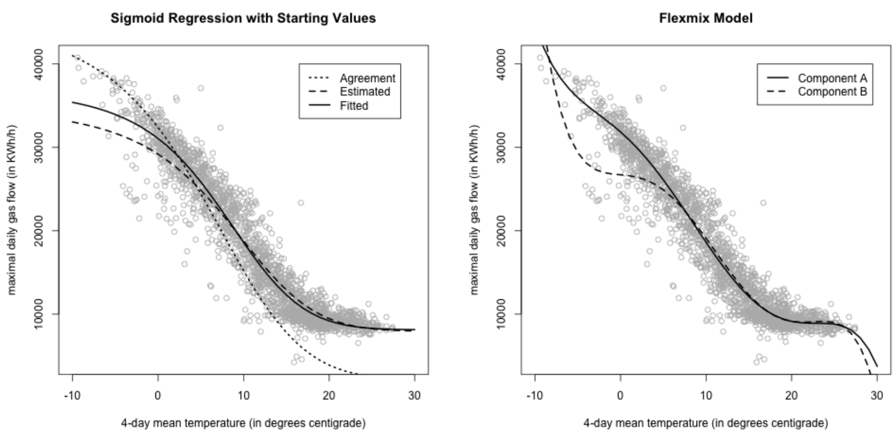

FIGURE 1. Fitted sigmoid regression (left) and flexmix model (right).

3 Mixture of Polynomial Regression Models

The nuances missed by the sigmoid model motivates the use of the flexmix approach introduced in Leisch (2004), which offers a framework for the flexible fitting of finite mixtures of regression models in the R environment. Since this framework mostly covers generalized linear models, we approximate the nonlinear sigmoid function S in 2 by a polynomial of the fifth grade, and fit a two-component mixture model

$$\begin{aligned} \text{Class A:} \qquad & y_i = \beta_0^A + \sum_{j=1}^{5} \beta_j^A t_{ij} + \varepsilon_i \\ \text{Class B:} \qquad & y_i = \beta_0^B + \sum_{j=1}^{5} \beta_j^B t_{ij} + \varepsilon_i. \end{aligned} \tag{3}$$

The model given by (3) is fitted in R using the function `flexmix()`, the parameters of the fitted model are given in Table 3.
Table 4 shows the estimated prior probabilities, the number of observations assigned to the corresponding clusters, the number of observations where the posterior probabilities are greater than 0.0001 and the ratio of the

TABLE 3. Mixed Regression Parameters.

Component	β_0	β_1	β_2	β_3	β_4	β_5
A	31841	-901.81	-42.43	-3.17	0.40	-0.0084
B	26712	-893.59	-7.16	-13.87	0.96	-0.0172

TABLE 4. Mixed Regression Properties.

Component	Prior Prob.	Cluster Size	Posterior Prob.	Ratio
A	0.779	1814	1972	0.9199
B	0.221	191	2005	0.0953

latter two numbers. The ratio of the second component is approximately 0.1, indicating the overlap of the classes for the large proportion of data. Figure 1(right) shows the fitted model obtained from the flexmix procedure. The fit for low temperatures differs clearly for both model components, and describes the gas flow in a more appropriate way.

4 Conclusions

Preliminary results show that a simple sigmoid model enables a good starting approach to the observed problem. The shape of the sigmoid function we used to model the dependence of the maximum gas flow from temperature is suitable, but there is room for improvement. We suggest the flexible fitting of mixture models. Since the approximation by polynomials is unstable, the flexmix approach would offer even better results, if one could generalize this method to mixtures of sigmoid models.

References

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

Cooperation Agreement (2008). *Vereinbarung über die Kooperation gemäß*§20 *Absatz* 1 *b) EnWG zwischen den Betreibern von in Deutschland gelegenen Gasversorgungsnetzen*. BMJ Deutschland.

Leisch, F. (2004). FlexMix: A General Framework for Finite Mixture Models and Latent Class Regression in R. *Journal of Statistical Software*, **11**, 1-18.

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

Seber, G.A.F., and Wild, C.J. (2003). *Nonlinear Regression*. New Jersey: John Wiley & Sons.