Simple model for prediction of loads in district-heating systems

Erik Dotzauer*
Birka Heating, SE-115 77, Stockholm, Sweden

Received 1 August 2002; received in revised form 27 August 2002; accepted 31 August 2002

Abstract

In order to improve the operation of district-heating systems, it is necessary for the energy companies to have reliable optimization routines, both computerized and manual, implemented in their organizations. However, before a production plan for the heat-producing units can be constructed, a prediction of the heat demand first needs to be determined. The outdoor temperature, together with the social behaviour of the consumers, have the greatest influence on the demand. This is also the core of the load prediction model developed in this paper. Several methodologies have been proposed for heat-load forecasting, but due to lack in measured data and due to the uncertainties that are present in the weather forecasts, many of them will fail in practice. In such situations, a more simple model may give as good predictions as an advanced one. This is also the experience from the applications analyzed in this paper.

© 2002 Elsevier Science Ltd. All rights reserved.

Keywords: Heat-load forecasting; District heating; Linear least squares

1. Introduction

Due to the large operational costs involved, efficient operation of the producing units in a district heating system is desirable. However, before a production plan can be constructed, a prediction of the heat demand need be determined. Previous works on heat load forecasting, see e.g. [1], conclude that the outdoor temperature, together with the social behaviour of the consumers (referred to as the social component), have the greatest influence on the demand. Weather conditions like wind, global radiation (sunshine) and precipitation have less effect.

* Fax: +46-8-671-8283.
E-mail address: erik.dotzauer@birkaenergi.se (E. Dotzauer).
Other features than weather and social behaviour may affect the load. For example, in the district-heating systems analyzed in the paper, some of the consumers are located near a lake, which implies that for some wind-directions the current heat demand will vary depending on if there is ice or not on the lake. This kind of phenomenon, which is impossible to model explicitly with sufficient accuracy, is part of the stochastic component. Also shortages of measurements can be seen as a stochastic behaviour.

Several methods for load prediction have been suggested and implemented. These include time-series models of ARMA type [2,3], Kalman filters [1,4], and algorithms based on Artificial Neural Networks [5]. Most applications in the subject consider the prediction of electrical-power loads. The main difference between electrical-power and district-heating loads is that time delays, which may be significant in a district heating network, do not appear in a power grid. However, similarities between the two problem types indicate that the same type of algorithms may be used.

The basic idea in many applications is to construct one single demand-curve by combining a weather forecast with the historical load and weather curves. However, a disadvantage of modeling the heat demand as only one curve is that time delays, restrictions on distribution capacity in the network and fluctuations in the distributed water temperature cannot be considered. To master this, the heating network can be modeled as a network, in which the nodes and the arcs have associated variables for distributed water temperature and flow. Such a model was presented in [1].

To make full advantage of the network modelling, measurement points in the network for temperature, flow and pressure are needed. Often in practice, only the heat that leaves the production plants is measured, which implies that the single curve type of modelling indeed may be appropriate. In the case with only one curve, the network dynamics are part of the stochastic component.

Assuming a good modeling and reliable input data are available, the above-presented modelling techniques will in many cases generate good predictions of the district-heating demand. In practice, the situation may be different. Lack of measured data and uncertainties in the weather forecasts may make detailed modeling worthless. In such situations, a more simple model may give as good predictions as an advanced one. This motivates the relatively simple modelling presented in the present paper.

The model is based on the assumption that the heat demand can be described sufficiently well as a function of the outdoor temperature and the social component. In Section 2, an explicit expression for the temperature dependent part is developed. By using the expression and removing the temperature dependent part from the measured input data, the social component can be identified. This is discussed in Section 3. The forecasting algorithm is stated in Section 4. Section 5 presents some computational results, and finally, in Section 6, some conclusions are given.

2. The temperature function

The aim with the calculations in this section is to derive an explicit expression for the temperature-dependent part of the load. It is obvious that the temperature
dependence is non-linear. For relatively high outdoor temperatures, the temperature has less influence. For example, the load will almost be the same for 25 °C and 27 °C. A corresponding conclusion is also true for relatively low temperatures, e.g. whether the outdoor temperature is −28 °C or −30 °C does not matter, the production units will produce at their maximum rate anyway.

Measured data are used in the calculations. Let \( H \) be the number of hours of the horizon from which the historical data are taken. The historical data used are the measured outdoor temperatures \( t^i \), \( i = 1, \ldots, H \), and the measured heat loads \( q^i \), \( i = 1, \ldots, H \). The expression modelling the heat demand \( q_i \) in hour \( i \) is defined as

\[
q_i = f(t_i) + g(i),
\]

where the function \( f(t) \) models the temperature dependent part and \( g(i) \) models the remaining part.

The temperature dependent part \( f(t) \) is assumed to vary as a piecewise linear function, see the illustrating example in Fig. 1. Here a function with five segments is used, but the number of segments can of course be chosen arbitrarily. Given the temperature levels \( \tau_j, j = 1, \ldots, 4 \), the function \( f(t) \) is described by nine parameters: the heat levels \( \beta_j, j = 1, \ldots, 4 \), and the slopes \( \gamma_j, j = 1, \ldots, 5 \). The remaining part \( g(i) \) is simply modelled with 168 parameters \( \alpha_j, j = 1, \ldots, 168 \), one parameter for each hour of the week. For each hour \( i \) of the historical horizon, the parameter \( \alpha_j \) that corresponds to the same hour of the week as \( i \) is used, or equivalently,

\[
g(i) = \alpha_j
\]

where \( i \) corresponds to the same hour of the week as \( j \).
The reason for using exactly 168 parameters is that it is possible to identify a possible pattern in \( g(i) \) with the weekly duration. A natural assumption here is that the social component may differ between working days and weekends.

A suitable choice for the parameters \( \tau_j, j = 1, \ldots, 4 \), is to distribute them equally with values at the same level as the measured temperatures. The remaining parameters describing the model, i.e. \( \alpha = (\alpha_1, \ldots, \alpha_{168}) \), \( \beta = (\beta_1, \ldots, \beta_4) \) and \( \gamma = (\gamma_1, \ldots, \gamma_5) \), are determined by solving the optimization problem,

\[
\min_{\alpha, \beta, \gamma} \left[ \sum_{i=1}^{H} (\hat{q}_i - f(\hat{t}_i) - g(i))^2 \right].
\] (3)

This is a linear least squares problem with linear constraints [which are implicitly included in the piecewise linear function \( f(t) \)].

The procedure for defining the function \( f(t) \) is summarized in the following algorithm:

**Step 1.** Choose the parameters \( \tau_j, j = 1, \ldots, 4 \), such that one fifth of the \( H \) measured temperatures \( \hat{t}_i \) are located in the intervals \(( -\infty, \tau_1], (\tau_1, \tau_2], (\tau_2, \tau_3], (\tau_3, \tau_4] \) and \((\tau_4, \infty) \), respectively.

**Step 2.** Solve problem. This will generate the model parameters, \( \alpha \), \( \beta \) and \( \gamma \) that define \( f(t) \) and \( g(i) \).

Even in the case when e.g. the wind and the global radiation have significant effect on the heat demand, the above modelling procedure may indeed be usable. In this case, the temperature \( \hat{t}_i \) can be seen as a temperature equivalent that is corrected for the current wind, etc.

3. Modelling the social component

The feature that has a major influence on the heat demand, other than the outdoor temperature, is the social behaviour of the consumers. An obvious insight is that this heat consumption may consist of both yearly, weekly and daily patterns. The yearly variation is considered by using historical data taken from the corresponding period in previous years. Alternatively, historical data from a period of a few weeks immediately before the forecast horizon can be used. The latter alternative is to prefer when the number or type of customers have changed significantly from previous years.

The weekly and daily patterns are modelled explicitly on the function \( w(i) \). In its most simple form, which shall be used when a variation with weekly duration can be identified, the function \( w(i) \) is chosen equal to \( g(i) \), i.e.

\[
w(i) = \alpha_j,
\] (4)

where \( i \) corresponds to the same hour of the week as \( j \).

If no weekly pattern can be identified, more robust results may be achieved by assuming that all the days of the week have identical profiles. In this case the function \( w(i) \) is defined as follows. First, remove the temperature dependent part from the load by computing the residuals \( r_i = \hat{q}_i - f(\hat{t}_i), i = 1, \ldots, H \). Let \( \delta_j, j = 1, \ldots, 24 \), be the
daily pattern, and compute each component $\delta_j$ as the average of the residuals $r_i$ whose index $i$ corresponds to the same hour of the day as $j$. Finally, the function $w(i)$ is defined as

$$w(i) = \delta_j,$$

where $i$ corresponds to the same hour of the day as $j$.

As an illustrative example, the measured load, $\hat{q}_i$, $i = 1,...,H$, the residuals $r_i$, $i = 1,...,H$, and the corresponding function $w(i)$, $i = 1,...,H$, defined from (5) are shown in Fig. 2. Six weeks of historical data are used, which means that data for 42 days are available. The upper solid line is the measured load and the lower solid line is the residual that remains when the temperature dependent part is removed. From the figure we see that despite the level of the original load curve changes over time, a social component with daily duration can easily be identified. The daily pattern is depicted with dots in Fig. 2.

The consumers of course do not act equally in all similar situations, which implies that the consumers also contribute to the stochastic component of the load.

4. Load forecasting

A prediction of the heat demand is constructed by combining a weather forecast with the modelling presented in Sections 2 and 3. Let $I$ be the number of hours of the

![Fig. 2. Measured load and social component. Legend: upper solid line: measured load, lower solid line: remaining part left when the temperature-dependent part is removed, dots: day profile.](image-url)
forecast horizon. The temperature forecast is denoted by $\tilde{t}_i$, $i = 1,...,I$, and the heat demand forecast that will be computed is denoted by $\tilde{q}_i = 1,...,I$.

The heat demand forecast is constructed as follows:-

**Step 1.** Define the function $f(t)$ by performing the algorithm presented in Section 2.

**Step 2.** Define the function $w(i)$ from either (4) or (5).

**Step 3.** Predict the load $\tilde{q}_i$, $i = 1,...,I$, by combining the temperature forecast $\tilde{t}_i$, $i = 1,...,I$, with the load prediction model $\tilde{q}_i = f(\tilde{t}_i) + w(i)$, $i = 1,...,I$.

5. Theoretical and practical experience

This section presents the computational results using the algorithm developed in Section 4. Measured data from two district heating systems in the region of Stockholm, Sweden, are used in the analysis. The larger system is referred to as ‘System 1’, and the smaller as ‘System 2’. System 1 has a typical day load (winter day) of about 700 MW, and System 2 of about 300 MW. In Fig. 3 measured and predicted loads for 1 week of February 2001 are presented.

Table 1 presents results using measured data from year 2001. For each prediction computed, the time horizon from which the measured data are taken is 6 weeks long, i.e. $H = 1008$, and is chosen as the period immediately before the forecast horizon. The latter is 1 week, i.e. $I = 168$. To give a fair evaluation of the model, measured temperatures are used instead of forecasted ones. This implies that the result will not
be influenced by the uncertainties that are normally present in the weather forecasts. The algorithm version that uses Eq. (5) to compute $w(i)$ is used. The relative and absolute errors for each hour of the second day of the forecast horizon are computed. The reason for analyzing the data from the second day is because trading in the Scandinavian electricity-spot market is performed 1 day in advance, which implies that the predictions for the second day are the most important. In the table, on lines one and two, the average errors for each quarter of the year (Q1, Q2, Q3 and Q4) are shown. Line three gives the number of hours (in percentages) where the relative errors are less than five percent.

From the results, we conclude that the relative errors are lower during winter (Q1 and Q4) than during summer (Q2 and Q3). However, the absolute errors for quarter are quite low (9.33 and 4.98 MW). We also see that during winter (Q1 and Q4), about half of the predicted values have a relative error that is less than five percent. The corresponding number for the summer is twenty-five percent.

A deeper analysis of the results shows that in several of the predictions that were quite bad, the correct shape of the load curves were found, but vertically displaced to another level. This phenomenon may be a consequence of the relatively little amount of input data that were used. When the temperature forecast contains values on a level that is not present in the historical temperatures, the algorithm may fail to make a sufficiently good model of the temperature dependence. The remedy may be to use more measured data, which include temperatures on the current level.

During spring 2002, the model supported the production planning group at an energy company with heat-load predictions. Commercial software for the same purpose, Aiolos [6], which is a model of the ARMA type, was used in parallel. A practical conclusion is that the load curves suggested by the two systems are of about the same quality. Another obvious insight is that good temperature-forecasts are crucial for making relevant heat-demand predictions. This is also understood by analyzing the data from the two systems considered. In Systems 1 and 2, a change in the outdoor temperature by one degree Celsius implies a change in the corresponding heat demand by about 28 and 12 MW, respectively. These numbers should be compared with the absolute errors presented in Table 1. A concluding remark is that the study on making better load predictions should focus on improving the quality of the weather forecasts, than on developing advanced load prediction algorithms.

### Table 1
Relative and absolute errors for each quarter of year 2001

<table>
<thead>
<tr>
<th></th>
<th>System 1</th>
<th></th>
<th>System 2</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Q1</td>
<td>Q2</td>
<td>Q3</td>
<td>Q4</td>
</tr>
<tr>
<td>Relative (%)</td>
<td>6.67</td>
<td>13.42</td>
<td>15.22</td>
<td>8.29</td>
</tr>
<tr>
<td>Absolute (MW)</td>
<td>36.64</td>
<td>23.48</td>
<td>9.33</td>
<td>32.71</td>
</tr>
<tr>
<td>#Relative &lt; 5% (%)</td>
<td>45.88</td>
<td>26.23</td>
<td>21.97</td>
<td>41.92</td>
</tr>
</tbody>
</table>
6. Conclusions

The paper presented a method for predicting the heat demand in a district heating system. The modelling is based on the insight that the load is mainly affected by the outdoor temperature and the social behaviour of the consumers. The model has a very simple construction, yet predictions from it were comparable with those for other more sophisticated methods. Both theoretical and practical results are presented.

References