# TEMPERATURE/DEMAND METHODOLOGY FOR PLANNING PURPOSES THE TD76 CODE OF PRACTICE

PLANNING DEPARTMENT
HEADQUARTERS

November 1987

British Gas

BRITISH GAS plc

# TEMPERATURE/DEMAND METHODOLOGY
# FOR PLANNING PURPOSES
# THE TD76 CODE OF PRACTICE

**Planning Department Headquarters**

November 1987

Note: This document was reconstructed in May 2005 from a scanned copy of the original. As such it will not have been subjected to the same QA as the original and may therefore be subject to errors.

**Index**

**TEMPERATURE / DEMAND METHODOLOGY FOR PLANNING PURPOSES THE TD76 CODE OF PRACTICE**

# 1. INTRODUCTION

1.1. In 1980 a new code of practice concerning Regions' temperature/demand methodology was introduced in British Gas. This resulted from the work of a joint H.Q./Regional Steering Group and was contained in a document which has since become known by the shorthand title of TD76. The document was updated in 1982 but further changes to the methodology have been introduced from time to time since then. The purpose of this report is to consolidate the current code of practice into a single document. For ease of reference the format of the original TD76 document has been retained.

1.2. Most of the material in this report is the same as that in the original TD76. TD76, however, proposed many new ideas which have subsequently become accepted as common practice, and the wording in this report has been altered to reflect this where appropriate. The major differences of substance from the original report are the introduction of a section on chilled temperature (section 5.6); the expansion of sections 7 and 8.1 (in line with the Steering Group report of 1985); the change in the procedure for updating the historical temperature data series (para 4.4.1); the deletion of the section on references; the omission of much of the detailed material in Appendices 5,7,9, and 12, the updating of Appendix 13 in line with the 1988 ROP; and the deletion of Appendix 14.

1.3. This is a technical report intended for Operational Research (O.R.) and Corporate Planning practitioners. However, although a fair degree of background knowledge on the part of the reader is assumed much of the report is comprehensible to the non-statistician. The report is the definitive reference for questions concerned with temperature/demand relationships in the planning context, and with the calculation of peak day demands and load duration curves in particular.

1.4. The report draws a distinction between definitions and their interpretation on the one hand and the derivation and application of demand models on the other. In the first case the definitions (sections 3, 4, 5, and 8.1) are intended to be precise and unambiguous and all Regions should comply with them on the grounds of consistency (Appendix 2). In the second case a framework for deriving, fitting, and using models is recommended (sections 6, 7, and part of 8) which allows Regions considerable flexibility of approach. The report concentrates on recommendations and guidelines with supporting argument mainly in the appendices.

## 2. SCOPE AND PHILOSOPHY OF THE REPORT

### Scope

2.1. There are many different purposes for which demand/temperature relationships are used. Broadly these fall into three classes as follows:

- short term forecasting for grid control operational purposes
- temperature correction of actual sales or sendout for monitoring and control
- long term forecasting for planning purposes

2.2. Ideally a single temperature/demand methodology would cover all three purposes. In practice because of the wide variation in detail and accuracy of data, in time scales, and in the contexts in which results are used, a simple all-purpose methodology is inappropriate. This report is concerned with identifying good methodology for planning purposes although it may be that some of the code of practice is relevant for wider purposes as well

2.3. There are broadly three sorts of planning for which demand/temperature methodology is needed

- peak day planning e.g. for sizing of Regional transmission systems
- peak period planning e.g. for the LNG programme for peak shaving from the National Transmission System
- peak "within-day" planning e.g. for sizing Regional storage, and establishing policy over the use of linepack and LNG; and for network analysis.

Although these three are inter-related, the criteria appropriate for planning in the third area, within-day planning, are outside the scope of this report which therefore concentrates on the first two areas, peak day planning and peak period planning.

2.4. The most important information required by H.Q. from Regions for peak day and peak period planning is:

- Estimates of peak day demand in future winters on an average basis and a 1 in 20 basis.
- Estimates of load duration curves in future supply years on an average basis and a 1 in 50 basis.

The average estimates are required for several reasons including:

- to allow interpolation between average and severe conditions so as to enable H.Q. to understand better the possible range of outcomes in conditions less severe than 1 in 20 or 1 in 50

- to allow H.Q. to study the behaviour of the National Transmission System under a wide range of conditions. For example, the assessment of the ability of the NTS to transmit the annual contract quantity of gas from St. Fergus depends on estimates of average demand levels throughout the system during the summer months. The estimation of the number of days when seasonal storage can be refilled and the estimation of annual compressor fuel usage are also purposes for which the average load duration curves would be used.

2.5. British Gas at present plans to a 1 in 20 level for the peak day demand and a 1 in 50 load duration curve. The report gives guidance on how these levels should be determined but does not examine the question of whether the levels of 1 in 20 and 1 in 50 are the right levels for British Gas to be using. However, any change from these levels would not affect the structure or argument of the report or its recommendations. Thus anywhere in the report where the expression "1 in n" occurs for n equal to 20 or 50, a different value of n could be substituted instead. The effect would be that a different estimate from a probability distribution that was already being calculated would be selected, one corresponding to the (100 - 100/n) percentile rather than the 95 or 98 percentile as the case may be.

## Philosophy

2.6. There is a wide range of desirable features that a temperature/demand methodology should have. To be ideal it ought:

i) For Regions and HQ

- to give results as statistically accurate as possible e.g. not 1 in 40 or 1 in 60 if we want 1 in 50

- to be relevant i.e. to be necessary and sufficient for the estimation of peak day demands, and load duration curves

- to be practicable and easy to use by, and to be unambiguous to, the personnel using it

ii) For Regions

- to be not significantly inferior to other possible procedures

- to be sufficiently flexible that it can embrace a Region's particular circumstances

iii) For HQ

- to ensure consistency between Regions so that the same criteria are applied and the same standards adopted in all Regions, and the aggregation or comparison of Regional forecasts is thereby meaningful.

2.7. To some extent the objectives listed above are conflicting and it is a subjective judgement as to how to balance them. In making this judgement a distinction can be drawn between definitions and their interpretation on the one hand and the derivation and application of demand models on the other. In the first case the code of practice is intended to be precise and unambiguous and all Regions should adopt it on the grounds of consistency. This applies to sections 3, 4, 5, and 8.1 (except where some Regional latitude is explicitly allowed as stated in sections 4.1 and 5.1). In the second case a framework for deriving, fitting, and using models is recommended which allows Regions considerable flexibility of approach. This applies to sections 6, 7 and parts of 8.

2.8. A major plank in the argument for national guidelines applicable to all Regions is the need for a common approach on the grounds of consistency. The consistency argument underlies much of this report and it is so important that it is spelt out in some detail in Appendix 2. Briefly the argument is for equality of standards across Regions both with respect to each Region's investment and with respect to Headquarters' investment in the National Transmission System and allocation of gas between Regions.

# 3. THE DEFINITION OF PEAK DAY DEMAND AND LOAD DURATION CURVES

## 3.1. The Context of the Definition

3.1.1 This section is devoted to the detailed definitions of peak day demand and load duration curves. These definitions contain the term "connected load" which refers to the underlying demand for gas on the part of all types of gas consumer. Clearly connected load is changing through time as our customers and their use of appliances change; and some measure of it (such as annual demand under seasonal normal conditions) has to be estimated for past years and forecast for future years.

3.1.2 In the following definitions of the 1 in 20 peak day demand and the 1 in 50 load duration curve it is assumed that the levels of connected load are known, but of course they are only forecasts. Thus, the forecasts for a particular year of the 1 in 20 peak day demand and the 1 in 50 load duration curve are conditional on the forecast of the underlying levels of connected load in the year in question. The forecasting of these underlying levels for future years is a marketing and economic problem beyond the scope of this report. However the forecasts are made, the methodology of this report is intended to apply to central estimates of connected load in future years without the addition of a forecasting or any other sort of margin.

## 3.2. The 1 in 20 Peak Day Demand

3.2.1 The definition of 1 in 20 peak day demand is the demand that, in a long series of winters, with connected load held at the levels appropriate to the winter in question would be exceeded in one out of 20 winters, each winter being counted only once.

3.2.2 This definition differs from the best estimate of the demand at the 1 in 20 minimum winter effective temperature (see section 4.6) in that it reflects the need to take account of other sources of variation besides temperature. The other sources of variation which should be taken account of need to be made explicate so as to make the definition unambiguous. They are as follows:

(i) the residual error in the chosen model relating demand to temperature (and any other weather variables). This includes the error due to weather factors which are not identified in the above model

(ii) the effect arising from the level of connected load changing (even in a precisely known way) through the winter while the point in the winter when the peak day might occur can vary over a range of three months or so

(iii) the effect (which is independent of weather or growth factors) of differences in the average level of demand among the days of the week

(iv) the error that arises because (even allowing for effects (ii), and (iii)) there is a chance (related to the size of error (i)) that the second, or third etc. coldest effective temperature in the winter may give rise to a daily demand higher than the demand associated with the minimum effective temperature.

3.2.3 A methodology based on simulation should be used for deriving estimates of 1 in 20 peak day demand consistent with the definition in para. 3.2.1. An example is given in section 8.2 to explain what this means and how the sources of variation listed in para. 3.2.2 would be handled in a simulation model.

3.2.4 1 in 20 is the criterion currently used by the Industry for assessing peak day demand. It should be noted, however, that exactly the same definition would apply for 1 in n peak day demand only it would replace 20 (and the (100 - 100/n) percentile would replace the 95 percentile of a distribution). However, it should be noted that the average peak day demand is not the same as the 1 in 2 peak day demand. If a simulation model is used the average peak day demand is the simple average of all the peak day demands generated by the simulation model. The 1 in 2 peak day demand is the middle ranking (or median) of these demands. Because of the skewness of the distribution the 1 in 2 peak day demand is smaller than the average peak day demand.

## 3.3. The 1 in 50 Load Duration curve

3.3.1 The phrase "load duration curve" means a curve of load or demand on the vertical axis against days duration on the horizontal axis. The days duration correspond to the days in a supply year i.e. 365 days (or 366) in a consecutive period running from October 1st to September 30th. The 365 daily demands are ranked and the highest plotted at day 1, the nth highest at day it and the lowest at day 365. Thus any point on the curve corresponds to a demand level and a number of days on which that demand is equalled or exceeded in the supply year.

3.3.2 The definition of the 1 in 50 load duration curve is:

The 1 in 50 load duration curve is that curve which, in a long series of years, with connected load held at the levels appropriate to the year in question, would be such that the volume of demand above any given demand threshold (represented by the area under the curve and above the threshold) would be exceeded in one out of 50 years.

It is appreciated that this definition is very concise. To ease understanding, Appendix 1 provides a fuller explanation.

3.3.3 The above definition applies to a supply-year but analogous definitions should be used where appropriate to refer to a financial year or any other consecutive period of 12 months containing the complete winter period. The definition also applies to total load, but analogous definitions should be used where appropriate to refer to total firm load, or maximum potential load, or any other component of total load.

3.3.4 The definition reflects the need to take account of other sources of variation besides temperature. The sources of variation which should be taken account of are the same as those listed in para. 3.2.2. Thus the definition of the 1 in n peak day demand should be consistent with the definition of the 1 in n load duration curve. The 1 in 20 peak day demand should be taken as day 1 of the 1 in 20 load duration curve and the 1 in 50 peak day demand should be taken as day 1 of the 1 in 50 load duration curve, a demand by definition more severe than the 1 in 20 peak day demand.

3.3.5 A methodology based on simulation should be used for deriving estimates of the firm and maximum potential 1 in 50 load duration curves consistent with the above definition. This is discussed further in section 8.1.

3.3.6 The definition of para. 3.3.2 applies for the 1 in n load duration curve with it replacing 50. However, it should be noted that the average load duration curve or the load duration curve for the "average" winter is not the same as the 1 in 2 load duration curve. This is because for any winter daily demand level the volume of demand represented by the area under the curve and above the demand level has a skew distribution. The simple average of the volumes is larger than the middle ranking or median volume. Thus the average load duration curve represents a more severe winter than does the 1 in 2 load duration curve.

# 4. THE USE OF HISTORICAL TEMPERATURE DATA

## 4.1. Introduction

4.1.1 This section covers all aspects of temperature definitions and criteria and is divided into seven sections. Sections 4.2 and 4.3 are not mandatory in that the criteria of choice of temperature recording station and the definition of average and effective temperature may legitimately vary from those given where a Region has demonstrated an alternative that fits its data and circumstances better.

4.1.2 Sections 4.4 to 4.7 are mandatory except that a certain degree of latitude is contained in para. 4.5.2. Sections 4.6 and 4.7 are concerned with peak day temperatures and temperature duration curves. This information is required because there are some purposes for which simple analysis based on temperatures is an adequate proxy for a more complex analysis based on demands.

## 4.2. Temperature Recording Stations

4.2.1 Each Region should use as few temperature recording stations as is consistent with deriving an adequate demand/temperature relationship for the Region as a whole for planning purposes (i.e. calculation of load duration curves, peak day demand estimates, and temperature correction estimates). A single station should normally be sufficient. It is accepted that South Western, Southern, and North Eastern Regions use two.

4.2.2 The temperature recording station (s) used by the Region should be currently recording temperatures which allow the calculation of a daily average temperature (see para. 4.3.1 below) and should be reliable. In principle they should either be Met. Office stations or "approved" by the Met. Office.

## 4.3. Average and Effective Temperature Definitions

4.3.1 Each Region should define a daily average temperature as a function of temperatures within the day. The preferred definitions are ½ (max+min) over the 24-hour period or a simple average of 24 hourly or 12 two-hourly temperatures, coinciding with the gas supply day starting at 6am.

4.3.2 Each Region should define a daily effective temperature as a function of the average temperature on the day and proceeding days. The preferred definition is $E_i = \frac{1}{2} A_i + \frac{1}{2} E_{i-1}$

## 4.4. Length of the Temperature Data Series

4.4.1 The historical temperature data series to be used for analytical purposes should consist of daily average temperatures from October 1st 1928 to September 30th of the most recent supply year.

4.4.2 Any Region wishing to change its data series or to construct artificial data so as to comply with para. 4.4.1 should consult with H.Q.O.R. Department to reach an agreed series so as to minimise the differences between its own temperature data base and that used as part of the National Temperature Data Base.

4.4.3 The analyses based on the historical temperature data series should be updated simultaneously in all Regions at such time as agreed nationally and only at such time. This is to ensure that the inter-Regional consistency achieved by para. 4.4.1 is maintained into the future. The present practice is to update in October/November of each year adding one extra year to the series each time.

4.4.4 It should be assumed that there is no climatic trend through the years represented in the temperature data base and in subsequent and future years to the planning horizon (see Appendix 3)

### 4.5. Seasonal Normal Effective Temperature (SNET)

4.5.1 SNET should be calculated from the same historical temperature data series as referred to in para. 4.4.1. The present practice is to update at five year intervals starting in the Autumn of 1988.

4.5.2 For each day of the year an unsmoothed SNET can be calculated as a simple average of the effective temperatures on the same day in each year of the data base. The series of daily SNETs thus obtained should be smoothed so as to contain no day-to-day scatter. The method of smoothing may be by fitting a 2nd order Fourier series, or a centered moving average from which any odd anomalies are removed. The smoothed series should contain only two turning points, one in the winter and one in the summer because any more complication is unlikely to be of practical use and has no basis in meteorological theory. The SNET for any particular day or period should be read or calculated from the smoothed series.

### 4.6. The Probability Distribution of Winter Minimum Daily Effective Temperature

4.6.1 A minimum daily effective temperature should be calculated for each winter from the historical temperature data series referred to in para. 4.4.1.

4.6.2 The temperatures, one for each winter, obtained in para. 4.6.1 should be regarded as forming a random series through time.

4.6.3 The average winter minimum daily effective temperature should be defined as a simple average of the temperatures obtained in para. 4.6.1 above.

4.6.4 The ‘1 in 20’ winter minimum daily effective temperature (often just referred to as the 1 in 20 temperature) should be derived by fitting a Gumbel/Jenkinson probability distribution to the full set of coldest temperatures obtained in para. 4.6.1 above. The precise details of this procedure are described in Appendix 4.

4.6.5 The calculation of average and 1 in 20 winter minimum daily effective temperatures is one of the analyses referred to in para. 4.4.3.

## 4.7. The Probability Distribution of the Temperature Duration Curve

4.7.1 Temperature duration curves for each supply year in the temperature data base of para. 4.4.1 should be calculated by ranking days in the supply year according to daily effective temperature.

4.7.2 The series of degree-days per winter as defined by the area under each temperature duration curve and above some specified threshold temperature should be regarded as a random series through time, for all threshold temperatures.

4.7.3 The average temperature duration curve is defined as that curve which for any threshold temperature has the property that the area in degree-days below the temperature threshold is a simple average of the areas calculated individually for each of the supply years in the data base. Day 1 on the average temperature duration curve will generally be colder than the average winter minimum daily effective temperature as defined in para. 4.6.3. This point is discussed further in Appendix 6.

4.7.4 The 1 in 50 (and other "1 in n severe") temperature duration curves should be derived from an analysis for each temperature threshold of the degree-days below the temperature duration curve and above the threshold for all the supply years in the database. For each threshold the average of the degree-day values should be calculated (as in para. 4.7.3) and a cube-root normal distribution should be fitted to those values colder than the average. The 1 in 50 value should be calculated from the mean and standard deviation of the distribution. (The reasons for doing this and the precise method for doing it are described in Appendices 5 and 6). Day 1 on the 1 in 50 temperature duration curve should be defined as the 1 in 50 winter minimum daily effective temperature as defined in section 4.6.

4.7.5 The calculation of average and 1 in 50 temperature duration curves is one of the analyses referred to in para. 4.4.3.

# 5. THE USE OF HISTORICAL WIND DATA

## 5.1. Introduction

5.1.1 The introduction of a wind variable into the basic gas demand model increases the complexity of analysis. However, the implications for the load duration curve are sufficiently important to make it essential for all Regions to include a wind variable in their analysis where it is found to be significant. Furthermore the variation in the load duration curves derived by using different definitions of the wind variable is sufficiently great to make it desirable for all Regions to adopt the same general form of wind variable on consistency grounds. (see Appendix 2)

5.1.2 Section 5 covers all the aspects of wind definitions and criteria and is divided into separate sections on roughly the same lines as section 4. Para. 5.3.6 and sections 5.4 and 5.5 are intended to be mandatory and Regions are strongly urged to follow paras. 5.3.4 and 5.3.5. The remainder of section 5 is not intended to be mandatory in that the criteria of choice of wind recording station and the definition of wind speed may legitimately vary from those given here for reasons of availability of data, or where a Region has demonstrated an alternative that fits its data and circumstances significantly better.

5.1.3 There is no useful analogue in the case of wind for sections 4.6 and 4.7 and probability distributions of peak day wind speed and wind duration curves are not needed. However, some Regions have defined a composite wind/temperature variable, “chilled temperature”, for which minimum values and duration curves can be calculated. This is discussed further in section 5.6.

## 5.2. Wind Recording Stations

5.2.1 Each Region should use as few wind recording stations as possible for the same reasons as given for temperature in para. 4.2.1. Wherever possible the same stations as are used for temperature data should be used.

5.2.2 The wind recording station(s) used by each Region should be currently recording wind speeds which allow the calculation of a daily measure of wind speed (see para. 5.3.1 below) and should be reliable. In principle they should either be Met. Office stations or “approved” by the Met. office.

## 5.3. The Definition of Wind speed and Chill Factor

5.3.1 Each Region should define a daily measure of wind speed as a function of wind speeds within the gas supply day. The preferred definition is average daily wind speed defined in as unbiased a way as possible e.g. the average of hourly readings over the 24 hour period starting at 6 a.m.

5.3.2 It is not recommended that Regions define a daily effective wind speed analogous to daily effective temperature.

5.3.3 Regions should not take account of wind direction (e.g. by calculating the northerly or north-easterly component of average daily wind speed) as there is no theoretical justification for such a refinement when temperature is already included in the demand mode. A refinement of this sort should only be used if additional historical data is available to support it and the Region can demonstrate that it results in a better fit of its demand model.

5.3.4 Each Region should define a chill factor variable i.e. a composite variable which is the product of wind speed and temperature terms. The preferred definition is:

$$C_i = \max\begin{pmatrix} W_i - \hat{W} \\ 0 \end{pmatrix} \times \mathrm{Max}\begin{pmatrix} \bar{A} - A_i \\ 0 \end{pmatrix}$$

where: $W_i$ = wind speed on day i as defined in para. 5.3.1

$\hat{W}$ = a threshold wind speed

$A_i$ = average temperature on day i as defined in para. 4.3.1

$\bar{A}$ = a threshold temperature

In this definition average temperature is recommended in preference to effective temperature on theoretical grounds and because it has been found usually to give a better fit in the demand model. Each Region should determine threshold levels for wind speed and temperature which result in the best fit of the demand model over a series of years. In using the chill factor the thresholds should not be allowed to vary from year to year. The temperature threshold should be fixed at a level so that the chill factor operates on a sufficient number of winter days that it can be estimated reliably but not on so many that it would usually operate over the period April to September.

5.3.5 Each Region should use the particular form of definition given in para. 5.3.4 on the grounds of consistency unless it can demonstrate to H.Q. that an alternative definition gives a consistently significantly better fit in its particular demand model.

5.3.6 All Regions are recommended to test for the significance of chill factor (see para. 5.3.4) in their demand model (along the lines discussed in section 6) over the period October 1st - March 31st. If it is significant it should be included in the model. If in any Region chill factor is not statistically significant over several past years' data then wind should be ignored and the remainder of section 5 is not applicable.

## 5.4. Length of the Wind speed Data Series

5.4.1 The historical wind speed data series to be used for analytical purposes should consist of daily measured wind speeds from October 1st 1928 to September 30th of the most recent supply year i.e. the same period as in section 4.4.1. However, it is not mandatory for Regions to acquire wind speed data for the period April 1st September 30th of each year.

5.4.2 Any Region wishing to change its series or to construct artificial data so as to comply with para. 5.4.1 should consult with H.Q. O.R. Department to reach an agreed series so as to minimise the differences between its own wind speed data base and that used as part of the National Wind speed Data Base. (See Appendix 7 for further details)

5.4.3 The analyses based on the historical wind speed data series should be updated at the same time as those based on the historical temperature data series and only at such time. (See para. 4.4.3).

5.4.4 It should be assumed that there is no climatic trend affecting wind speeds through the years represented in the wind speed data base and in subsequent and future years to the planning horizon.

## 5.5. Seasonal Normal Wind speed (SNW)

5.5.1 SNW should be calculated from the same historical wind speed data series as referred to in para. 5.4.1. The present practice is to update at five yearly intervals starting in the Autumn of 1988.(see also para. 4.5.1).

5.5.2 It is unnecessary to derive a daily SNW series along similar lines to those described in section 4.5 for seasonal normal effective temperature. It is considered sufficient for each Region to calculate SNW as a constant for the winter six months period (October 1st - March 31st) and use this as the standard to which demands are corrected when the demand model includes a chill factor term. The combined temperature and wind correction procedure is described in section 8.4.

## 5.6. Chilled Temperature

5.6.1 Some Regions have found it appropriate to combine the effective temperature and the chill factor variables in their models into a combined or "chilled temperature" variable. Using $C_i$ as defined in para 5.3.4 and $E_i$ as defined in para 4.3.2 then chilled temperature on day i, $X_i$ say, might be defined as:

$$X_i = E_i - k\,C_i$$

where k is a constant to be determined. A "standard" value for k used at H.Q. is 0.01. The value of k should be related to the ratio of the coefficients of $E_i$ and $C_i$ as derived from models fitted over a number of years in which these variables are separate (see section 6). If a

Region finds that this ratio is very unstable from year to year then the use of chilled temperature is not recommended.

5.6.2 If a Region using chilled temperature wishes also to define seasonal normal chilled temperature then the same procedure as described in section 4.5 should be followed substituting chilled temperature for effective temperature.

5.6.3 If a Region using chilled temperature wishes also to define the probability distribution of winter minimum daily chilled temperature or the chilled temperature duration curve it should follow the procedures described in sections 4.6 and 4.7 substituting chilled temperature for effective temperature.

# 6. THE FRAMEWORK FOR DERIVING A DEMAND MODEL

This section describes the framework within which each individual Region should derive the best demand model or models appropriate to its own circumstances and consistent with national requirements.

## 6.1. A General Form of Demand Model

6.1.1 In identifying an appropriate general form of demand model there is a great deal of prior knowledge based on the experience of the Industry over many years. Thus it is known that the level of connected load is affected by the economic environment, and that demand is affected by temperature and other weather variables, by the day of the week, and also by random fluctuations. Random error is sufficiently large, and data relating to extreme conditions sufficiently sparse, as to leave room for doubt about the best form of the underlying model.

6.1.2 A specific general model is not therefore recommended as superior to all others. However, the following fairly general linear additive model is thought to be appropriate as a starting point for most Regions:

$$D_i = a + bS_i + cT_i + dw_i + eg_i + \sum_{j=1}^{7} f_{ij} \{\alpha_j + \beta_j S_i + \gamma_j T_i + \delta_j W_i\} + u_i$$

where $u_i = \int u_{i-1} + \varepsilon_i$

and $f_{ij} = 1$ if $i \; \varepsilon$ class $j$ where $j = 1 =$ Fri; $j = 2 =$ Sat;…

$= 0$ if $i \notin$ class $j$ .... $j = 7 =$ Thur.

6.1.3 Although a more complex form of model is possible, the above is sufficiently general to illustrate the principles whereby a suitable model form should be derived. Each term is now discussed in turn:

- The dependent variable, $D_i$, is total daily sendout on day i, (translated from mcfd to m.ths using appropriate STP and CV assumptions) corrected if necessary by adding back an estimate of any interruption that has taken place on the day. In some Regions where reliable daily readings for interruptible load are available it might be appropriate to define $D_i$ as the total daily sendout excluding interruptible load.
- $S_i$ is a seasonal factor such as $SNET_i$, defined in section 4.5
- $T_i$ is a function of the effective temperature on day i as defined in para. 4.3.2 (Care must be taken to avoid multicollinearity between $T_i$ and $S_i$ by using a definition of the form $T_i = E_i - SNET_i$)
- $w_i$ is a measure of windspeed or chill factor on day i (see para. 5.3.4)
- $g_i$ is a measure of incremental connected load on day i (a growth term)
- $f_{ij}$ is a variable associated with the different pattern of demand on each day of the week. $\alpha$, $\beta$, $\gamma$ and $\delta$ reflect the fact that in principle

the demand pattern on different days of the week could interact with other variables.

- $ui_i$ is an autocorrelated residual error-term
- $\varepsilon_i$ is an independently and identically distributed error term

a,b,c,d,e, etc. are parameters to be estimated when fitting the model.

6.1.4 Regions may wish their model to include a measure of cold weather upturn (CWU) on day i (e.g. "night-override" effect or "cancellation of conservation" effect) related to consumer behaviour in extreme conditions. If Regions do this then the CWU variable must be included in the fitted model so that there is no double counting of the effect.

## 6.2. Deriving the Particular Model

6.2.1 In arriving at a particular form of demand model the starting point should be a fairly general model and good statistical procedures should be followed in making simplifications. The model described in section 6.1 can be used to illustrate this process as follows.

6.2.2 The first question is to determine the precise historical period or periods over which to fit the model. This is a major question and is left to section 6.3. Suppose, however, that a particular time period has been identified over which an appropriate model is to be fitted. The initial step would be to fit a fairly general form of model to the period in question by means of stepwise regression so as to decide on a set of variables which are statistically significant. This may reveal, for the sake of argument, that all the $\alpha_j$, $\beta_j$, $\gamma_j$ and $\delta_j$ are insignificant with the exception of $\alpha_1$, $\alpha_2$, and $\alpha_3$. In the example in Appendix 8 this is true for two out of three years and therefore the final model in this case would be:

$$D_i = a + bS_i + cT_i + dw_i + eg_i + \sum_{j=1}^{3} \alpha_j f_{ij} + u_i$$

6.2.3 At this stage it may be appropriate to experiment with different measures for some of the variables. For example a better fit over several years may be obtained from a different choice of threshold temperature in the definition of $w_i$ (see para. 5.3.4). It is recommended that the model is broken down step by step following sound statistical procedures until the gains by way of further simplification are out-weighed by the deterioration in the goodness-of-fit.

6.2.4 It is not considered appropriate to lay down hard and fast rules on the precise form of model that a Region should use. It is thought that a model similar to the one illustrated is likely to be appropriate in most Regions. In particular it is considered that whatever the precise form of the model, it should apply to all days of the week and that it should take account of autocorrelation in the residual error term if this is found to be significant.

6.2.5 Finally the model should be of a form that, when used in a predictive mode, it can generate estimates of peak day demand, and load duration curves, consistent with the criteria by which these are defined as laid down in section 3.

## 6.3. The Period Over Which The Model Should Be Fitted

6.3.1 The model is to be used primarily to estimate peak day demands and load duration curves. The latter requirement means that demand must be modelled over an entire supply year. It may be that the same model with the same parameters will be inappropriate at all times of the year. If this is so, a decision has to be made as to how to break the year up into separate periods for which separate models are then required.

6.3.2 The principles by which this decision should be made are similar to those described in section 6.2. The model should be fitted to periods of varying length and at different times of the year and if significantly different forms emerge these should be used. There are various statistical pitfalls (discussed in Appendices 8 and 9). Because of the critical importance of the mid-winter it is thought unlikely that less than two different models will be necessary, one to cover the winter six months and the other to cover the summer six months. Over and above this there may also be a cold period model. These models are likely to be of the same form although possibly the cold period model may drop the SNET term (due to the very small variation in SNET over the mid-winter period) and the summer model may drop the wind term (due perhaps to the lack of significance of wind at other than cold temperatures.)

6.3.3 Care must be taken to ensure that the models “blend in” with each other at points in the year when transfer from one model to another is taking place, in the above example at the beginning and end of the winter. In particular Regions are encouraged to take account of the switch-on/switch-off heating effect in Spring and Autumn. This is of special relevance in the context of temperature correction (see section 8.4).

6.3.4 The form of the model and the periods over which it is fitted should not be determined by a single year's data but should take account of a number of years' experience. Regions will need to ensure that their choices are reasonably robust by checking how the model would perform on several years' data and whether trends in the values of the model parameters from year to year are sensible.

6.3.5 Appendix 8 contains a more detailed exposition of the procedure for deriving and fitting a model.

# 7. FORECASTING MODEL PARAMETERS FOR FUTURE YEARS

Section 6 describes a framework within which each Region is recommended to derive the best models for total sendout for past years. This section describes the methods by which the parameters of the model might be estimated for future years so that the corresponding estimates of peak day demand and load duration curves can be derived.

## 7.1. General Principles

7.1.1 In order to estimate the parameters of the model for future years Regions are recommended to break down the total model for the past years into sub-models corresponding to each significantly different market sector, including unaccounted-for gas, in such a way that the sum of the sub-models is equal to the total model. Load growth can then be modelled separately for each market sector and sub-models for future years built up. Adding the sub-models for any forecast year will then give the required total model.

7.1.2 The precision with which the model for total sendout can be broken down by market sectors will vary from one Region to another depending on the availability of detailed daily information. It is recognised that in many cases some individual market sector models will inevitably be based on very limited and unreliable information. The purpose of constructing them is, however, primarily as a means for deriving a total model for future years in as consistent a way as possible. Each Region will have to judge for itself how detailed a breakdown to attempt bearing in mind that the individual sector models will need to seem intuitively reasonable and display consistent changes in parameter values over a series of past and future years.

7.1.3 A total model for future years could be obtained without breaking down the base year model. Net added load could be modelled by market sector and simply added to give a total added load model. This could then be added to the total model for the base year. However, the sub-model approach is preferred as it is seen to have the following advantages.

(i) Individual modelling of separate sectors provides a systematic way of taking account of differing trends in each sector and changes in the pattern of existing load.

(ii) Having explicit models of individual sectors facilitates the use of survey results in demand/temperature analysis.

(iii) It facilitates phasing sales forecasts by market sector (e.g. for monitoring).

(iv) It enables the identification of load factors for different kinds of business which is important for economic analyses of market development.

## 7.2. Past Years by Market Sector

7.2.1 In this section general features of the approach to identifying market sectors and breaking down the total model, for each period of the year, into sub-models for each sector, are discussed. Examples of three practical approaches are given in Appendix 10 to illustrate the wide range of ways in which the general principles can be interpreted.

7.2.2 It is recommended that the maximum level of detail to which Regions should consider breaking down the total model is to the following sectors:

- Domestic
- Commercial tariff
- Commercial firm contract
- Commercial interruptible contract
- Industrial tariff
- Industrial firm contract
- Industrial interruptible contract
- Gas used for own purposes
- Unaccounted for gas.

A minimum level of detail would be to model the domestic, commercial firm, industrial firm and the total interruptible sectors, together with gas used for own purposes and unaccounted for gas.

7.2.3 Regions should always derive a model for total sendout following the procedure recommended in section 6 and ensure that the sum of their sub-models is equal to the total model. The sub-models should therefore be of the same general form as the total model although it is not expected that all coefficients will be significant in all sub-models.

7.2.4 In estimating the coefficients of the sub-models Regions are recommended to adopt the procedure recommended for the total model to suit the level of detail and reliability of the information available for the market sector in question. As a general rule models should be estimated for those market sectors with the most reliable information first. These can then be subtracted from the total model and a sub-model for the remaining sectors obtained by differencing. Alternatively, if models have been estimated for all sectors then they should be scaled so that they sum to the total model.

7.2.5 The billing information for the non-domestic sectors is likely to be more detailed than that for the domestic sector. If a Region has daily meter readings or telemetered data for interruptible and large firm contracts it may be appropriate to follow the procedures outlined in section 6 to estimate sub-models for these sectors. If a Region has only monthly billing information for non-domestic sectors (tariff and contract) then a different approach to estimating the required sub-models should be adopted. If a Region has only a very few industrial contracts and these have very different characteristics it may be appropriate to treat these contracts on an individual basis.

7.2.6 For the domestic sector the billing information available is likely to be of limited value due to the preponderance of estimated reads. Amendments for unread gas also introduce error. However a total domestic model can be obtained from the analysis of weekly monitor data. This is preferable to a model obtained by subtracting the non-domestic sub-models from the total model. Use should also be made of the results of peak load surveys.

7.2.7 The set of sub-models obtained for the most recent year should be checked for compatibility with the sub-models used for previous years. Changes in the coefficients of a sub-model between successive years should be intuitively reasonable in the light of real changes experienced in the market sector which the sub-model represents.

7.2.8 Sendout and market sector models should be fitted to a number of past years' data so that a well-founded model can be derived for projection. In general Regions should exclude past years where the characteristics of markets are considered to be completely at variance with current circumstances. However at least the three most recent past years should be examined and Regions should always retain the most recent cold winter in their analysis until another cold winter occurs.

7.2.9 The following are some of the techniques for estimating parameters for the base model:

- Simple extrapolation of a trend between parameters and (say) sendout established for a number of past years
- Calculation of an average value of a parameter for which, typically, no clear trend can be established
- Calculation of a weighted average of values of a parameter, the weights chosen to reflect proximity of the year fitted and/or severity of the year fitted
- Selection of a preferred value of a parameter, for instance from a past severe winter
- Selection of a ratio between parameters which is to be preserved.

All these techniques (and others) are acceptable in principle provided that they are appropriate to the Region's experience. It is likely that Regions will wish to give more weight to models fitted to more recent years and to severe winters but these should not be considered in isolation from data from other years. Whichever techniques are used Regions must take care to ensure that the parameters of the base model are consistent. For example if the values of particular parameters are constrained then the values of other parameters should be re-estimated taking the constrained value(s) into account.

7.2.10 The general guidelines can be interpreted in many different ways (see Appendix 10) and within each approach there will be differences due to the level of detail of information available and its reliability. Each Region has to judge for itself how to make the best use of its past data, and how best to break down its total sendout model. The main concern throughout should be with obtaining sub-models which give reasonable estimates of demand over a period of past years and which can be summed to the total model. These sub-models then provide the springboard from which the total model for future years can be built up.

## 7.3. Future Years by Market Sector

7.3.1 In this section the general principles of the approach to obtaining sector models for future years are outlined and in Appendix 10 practical interpretations of these principles are given.

7.3.2 The total demand model for each of the future years covered by the ROP should be built up from the corresponding market sector models. The market sector models for future years should be obtained by projecting forward the models for past years which have been derived along the lines suggested in section 7.2. As stated in para. 3.1.2, the projections for future years should in all cases be central estimates. These will generally be based on Marketing forecasts of load growth and changes in market structure, and will therefore be very dependent on the quality of such forecasts. Nevertheless, it is important that the resulting total model for each future year should appear reasonably consistent with that for other future years and past years.

7.3.3 The changing nature of market sector temperature/demand relationships should be taken into account when parameters for future years are estimated. Regions should make specific allowance for changes in the characteristics of the base load, and the characteristics of load which is expected to be added and load which is expected to be lost. The Region's forecasting system should also have some method of handling the possibilities that the changes may be effective at identifiable points within the year or spread throughout a year or spread throughout a portion of a year.

7.3.4 For non-domestic sectors Marketing forecasts of load growth and associated load factor should be used to build up the sub-models for future years. It is recommended that large loads are separately identified and the remaining unidentified load subdivided between a numbers of uses which may have significantly different load characteristics.

7.3.5 For domestic sectors changes in the sub-models should be based upon forecasts of appliance sales, the existing appliance population, estimates of load factors for existing sales (which may vary in the future as insulation levels increase), forecasts of load factors for added sales, appliance consumption forecasts, etc. Relationships between load factors and the coefficients of the sub-models should be established in order to obtain coefficients for future years.

7.3.6 A model based on historical sendout data would normally be expected to be more accurate than individual sector models. When sendout models are disaggregated to sector models the sector models should therefore be constrained so that the sums of the parameters of the sector models equal the parameters of the sendout model. When sector models for future-years are combined into total firm and maximum potential demand models this should be done by simple aggregation.

## 7.4. Summary

7.4.1 Section 7 describes the principles by which the Steering Group recommends Regions to approach the problem of forecasting the parameters in a fairly complex total daily sendout model for future years. Appendix 10 describes three approaches to the practical problem of applying these principles, the so-called "load factor" approach, the "sector modelling" approach, and the "integrated model" approach.

7.4.2 There are many other related approaches and it is up to each Region to decide the approach which suits it best bearing in mind the quality of data it has available. The object of the exercise is to obtain the best forecast of the total daily sendout model possible for the years of the ROP consistent with the ROP assumptions and Marketing forecasts, as a means to forecasting peak day demands and load duration curves. This is best achieved by breaking down the models into sub-models, projecting these sub-models forward, and re-aggregating them, each stage being undertaken in a consistent manner and care being taken to ensure reasonableness and consistency between the parameters in the models derived for successive years, both past and future.

# 8. USING THE MODEL(S)

The primary purpose of the demand model or models derived in section 7 is to provide the mechanism for forecasting peak day demands and load duration curves for future years and for providing the means of correcting current demand for temperature and wind deviations from normal. In the case of peak day demands and load duration curves the model needs to provide forecasts consistent with the strict definitions of these terms as laid down in section 3, and the criteria for treating variations in temperature and wind conditions as laid down in sections 4 and 5.

## 8.1. Load Duration Curves

8.1.1 Regions are required to make returns to H.Q of average and 1 in 50 load duration curves for each year covered by the R.O.P. A possible way of producing these curves is via temperature duration curves (derived along the lines of section 4.7) and a simple demand/temperature relationship whereby a temperature duration curve of given probability is translated into a load duration curve of the same probability. This approach is considered to be inadequate in that it fails to produce load duration curve's consistent with the required definition, as given in section 3, and neither can it take adequate advantage of the richer form of demand model described in section 6.

8.1.2 Load duration curves of the required form should be produced via a simulation model. A description of a suitable simulation model is given in Appendix 11. Such a model has been successfully run by all Regions to produce average and 1 in 50 load duration curves since the early 1980s.

8.1.3 If for any reason the simulation approach is not used then steps must be taken to ensure that the resulting load duration curves are consistent with the definitions and criteria of sections 3, 4, and 5. It may be possible to derive decision rules whereby load duration curves derived by other means can be amended so as to be similar to those which would have been derived by simulation. However, in view of'the directness and simplicity of the simulation approach it is thought that such a procedure is likely to introduce self-defeating complexities.

8.1.4 The total firm load curve and the maximum potential load curve (i.e. total firm plus maximum potential interruptible curve) are the two most important curves for Company planning. They should therefore be prepared in a standard way between one Region and another. They should be obtained by deriving equations which fully represent the market in question (i.e. total firm or maximum potential) and these equations should be used separately in a simulation model with no subsequent additions to the resulting curves. Provision for cold weather upturn, cancellation of conservation, scheduled loads etc. should be made in the basic equations and should not be made through additions to the simulated curves.

8.1.5 A particular problem arises in modelling the minimum essential interruptible curve, which differs from the total firm and maximum potential curves in that the load which it represents is expected to be set to zero for a part of the year. The minimum essential interruptible load curve represents the demand from interruptible customers after maximum contractual interruptions, and taking into account loss of some interruptible capacity due to operational factors. (Regions should choose to which part of the load curve to assign such lost interruptible capacity in as realistic a way as possible).

8.1.6 Regions usually assume that the minimum essential curve will initially be set to zero for the first n days, where n is the shortest number of days for which any of the interruptible customers may be interrupted. The curve steps up from zero on day n to a positive level on day n+1 and similar increases occur at other durations of interruption until all interruptible durations in the contracts are assumed to be exhausted. This approach implies that all available interruption has been successfully scheduled into the coldest days of a 1 in 50 winter and that the full effective contractual volume of interruption has been achieved.

8.1.7 A practical problem arising from the procedure outlined above is the treatment of the various "saw-teeth" which occur when the minimum essential interruptible curve is added to the total firm curve. Regions may either convert the "saw-teeth" into horizontal sections of load curve or reorder the day numbers of the minimum essential curve so as to produce a curve which joins the total firm and maximum potential curves with a smoothly increasing curve. In either case the volume below the minimum essential curve and the volume between the total firm and the maximum potential curve above the minimum essential curve are to be preserved.

8.1.8 The decision as to which of the above two approaches to use is left with Regions because it must be taken against the background of Regional knowledge about flexibility for rescheduling interruption in cold weather. Simple reordering of the minimum essential curve makes no allowance for rescheduling of interruption however and obscures the representation of the various lengths of interruptible contract. Regions should therefore adjust each sawtooth individually to ensure that the resulting firm plus minimum essential curve is either horizontal or monotonic decreasing at all points.

8.1.9 Regions should take account of the fact that the effectiveness of interruption is likely to vary with weather conditions, for instance between average and severe years, and with duration of pipeline allocation. For example one Region has calculated effectiveness ranging from 86% to 92% for pipeline levels varying between "normal" and about 10% lower than normal. Furthermore at "normal" pipeline levels some interruption capacity is likely to be "sterilised" in the sense that contracts will be of longer duration than is implied by the pipeline allocation.

8.1.10 Bearing these problems in mind Regions should continue to calculate both theoretical and effective volumes of interruption and should take into account the "normal" level of pipeline allocation when calculating effective capacity. Effective interruption is defined to be the theoretical volume of interruption above a given threshold less the expected (or actual) supply to interruptible customers planned to be (but not in fact) interrupted at or above the same threshold. Regions should adjust their firm plus minimum essential curves to reflect the forecast effective levels of interruption (see para. 8.1.5 above). No deduction should however be made for sterilised interruption.

8.1.11 Direct simulation of the interruptible load duration curves is not recommended. The total firm and maximum potential curves should be obtained by simulation, and the maximum potential interruptible curve should be obtained by subtraction of these two curves on a daily basis. It is recognised that the resulting curve cannot be considered to be a genuine 1 in 50 curve. However, the interruptible curve is of lesser importance than the total firm and maximum potential curves which are both genuine 1 in 50 curves (see para. 8.1.4).

8.1.12 Load duration curves for the domestic, commercial, and industrial firm market sectors are required to add up to the total curve. The 1 in 50 sector curves will not therefore be genuine 1 in 50 curves in the sense that the total firm and maximum potential 1 in 50 curves are genuine. One possible approach is to simulate load duration curves for each sector using the sector models and compare these with the simulated total firm curve. The sector curves derived in this way would need to be adjusted so that they summed to the total firm curve. An alternative approach is to estimate demands for each sector at a number of temperatures using the sector models and express them as proportions of total demand. The proportions could then be used to split total demand between sectors over the relevant section of the load duration curve. Other approaches are also acceptable provided that they reasonably reflect the information contained in the sector models.

## 8.2. Peak Day Demand

8.2.1 The definition of the load duration curve is such that there is a direct link with the definition of peak day demand. For all n, day 1 on the 1 in n load duration curve is defined as the 1 in n peak day demand. Thus the 1 in 20 peak day demand is by definition day 1 on the 1 in 20 load duration curve.

8.2.2 The simulation model used for deriving load duration curves should also be used to derive peak day demands. The maximum daily demand occurring in each simulated year should be noted and the same procedures as described in section 4.6 for minimum effective temperature should be followed.

8.2.3 Para. 3.2.2 noted the sources of variation other than temperature that should be taken account of when estimating peak day demand. To demonstrate how this can be done in a simulation model the following illustrative example is given:

$$D_i = a + bT_i + c_i + \sum_{j=1}^{3} \alpha_j f_{ij} + \varepsilon_I$$

Where $D_i$ is the demand on day i, $T_i$ the temperature on day 1, and

$f_{i1}$ = 1 if day i is a Friday and 0 otherwise
$f_{i2}$ = 1 if day i is a Saturday and 0 otherwise
$f_{i3}$ = 1 if day i is a Sunday and 0 otherwise

and the $\varepsilon_i$ are independently normally distributed with variance $\sigma^2$
On fitting the model, estimates of a, b, c, $\alpha_j$ and $\sigma$ are found

8.2.4 For the temperatures corresponding to each of the N winters in the database, daily demand is simulated using a random number stream to generate values of $\varepsilon_i$. From the resulting N peak day volumes, one for each winter, average, 1 in 20 and 1 in 50 peak day values are derived. This procedure is repeated to give a total of 7r estimates of the peak day statistics (in Appendix 11, r is 4). In r of the simulation runs the winter begins on a Monday, in the next r it begins on a Tuesday etc. Finally the 7r peak day statistics are averaged in each case to get the desired average, 1 in 20 and 1 in 50 peak day estimates.

8.2.5 Through this process the sources of variation listed in para. 3.2.2 have each been taken account of as follows:

- the residual error by explicitly simulating it and the error due to other weather factors by assuming it is contained in the residual error

- the error due to growth in connected load through the winter and to differences in the days of the week through their representation in the model (the c and f terms) and hence the simulation

- the effect of the second, third etc. coldest days through the simulation of demand through the whole winter period.

8.2.6 The above description is for illustrative purposes only. It is not suggested that this particular model is the best one. The question of the most appropriate form of model is discussed in section 6.

## 8.3. Other Uses of the Demand Model in Combination with a Simulation Model

8.3.1 The simulation model, described in Appendix 11, has mainly been developed for the estimation of peak day demand and load duration curves. However, a realistic daily demand model used in combination with a simulation model can have much wider application. It is potentially a very powerful tool for examining all aspects of demand profiles. In any situation where one wishes to estimate average or extreme values for the demand over a specified period of time a simulation model is the most reliable way of doing this. Examples of two possible further uses of the simulation model are suggested in Appendix 13 where the H.Q. information requirements for the supplementary statements are described. They are discussed below for illustrative purposes.

8.3.2 Regions are required to submit details of their half-yearly natural gas requirements in the average year on statement CPD2 (see para. A13.7). To ensure that this information is consistent with the load duration curve for the average year, the half-yearly requirements can also be obtained from the simulation model. For each half-year the simulation model will generate a total demand corresponding to each of the N winters in the database. This will simply be the average of the 28 simulations for any year. Averaging these N values will then provide the half-yearly demand in the average year required on statement CPD2.

8.3.3 Weekly demand profiles for both an average and a cold year are required on statement PSI (see para. A13.13). The simulation model can also be used to derive this information. For each week of the supply year the model will generate an average and maximum daily demand corresponding to each of the N years' temperature and windspeed data. The average demand for each week is simply the average of the N average demands for that week. The cold weather demand in each week can be obtained by fitting a probability distribution of the kind described in section 4.6 to the simulated peak demand in each month, estimating the monthly 1 in 20 values and interpolating between these to get the weekly values.

## 8.4. Weather Correction

8.4.1 This section of the report reproduces the material, with minor amendments, from section 8.4 of the original TD76 report. However. the subject of weather correction was not treated in any great detail in TD76 and this section should therefore be treated as an introduction to the subject only. A further more detailed report, TD119, was written in 1984, and subsequently a new working party was convened in 1987 to review the methodology again.

8.4.2 The procedures described in section 6 yield models of sendout and sales which cover the whole year and these models should also be used for weather correction purposes. The correction is for weather rather than temperature to cover the situations where the models include a wind variable. The correction process should attempt to adjust actual demand to the level which would have resulted had temperatures (and winds) been normal.

8.4.3 Thus following the recommendations of previous sections, the demand/temperature model will provide an estimate of demand on day i as a function of effective temperature, a chill variable involving both wind speed and actual temperature, and other variables such as seasonality, day of week, etc., i.e.

$$\check{D}_i = f(E_i, w_i, A_i) + x_i$$

Where $D_i$ is estimated demand
$E_i$ and $A_i$ are defined in section 4.3
$w_i$ is wind speed as defined in para. 5.3.1
$x_i$ is the sum of all the other terms in the model, excluding the residual.

8.4.4 An equivalent demand at standard weather conditions on day i is given by $D_{is} = f(SNET_i, SNW_i, SNAT_i) + x_i$

where SNET is defined in section 4.5
SNW is defined in section 5.5
SNAT is the seasonal normal actual temperature, which for most Regions may be taken as equal to SNET

8.4.5 The weather correction on day i is then

$\check{D}_i$-$\check{D}_{is} = f(E_i, w_i, A_i) - f(SNET_i, SNW_i, SNAT_i)$ and the "weather corrected" demand is taken as

$\check{D}_{ic} = D_i - (\check{D}_i - \check{D}_{is})$

where $D_i$ is the actual demand on day i.
$\check{D}_{ic}$ differs from $\check{D}_{is}$ by the addition of the actual residual error on day 1.

8.4.6 If chill factor does not occur in the model (as will be the case over the summer six months) the terms in $w_i$ and $SNW_i$ disappear. In either case the correction is not affected by the presence of an autocorrelated residual error structure in the model.

8.4.7 Where no interruption occurs, the total demand is equal to sendout. If there is interruption, an estimate of the demand lost should be added to the sendout before carrying out the weather correction. In estimating the lost demand, attention should be paid to the weather sensitivity of the interrupted load so that the estimated total actual demand is a true reflection of what the demand would have been if there had been no supply constraint. No account should be taken of any expected interruption in the average year.

8.4.8 The weather correction of sendout should be carried out on a daily basis and the corrections aggregated for longer periods. Sales by market sectors are normally only available on a period basis, quarterly or possible monthly, so it is recommended that the weather correction for sales should be obtained from the total weather correction of sendout over the same period. The sendout correction should be allocated to market sectors in proportion to the values of the coefficients of the temperature terms in the sector temperature/demand models, as outlined in section 7, taking account of any assumed temperature sensitivity of unaccounted-for gas. Although this procedure ignores possible variations in the effect of wind it should not result in any serious error.

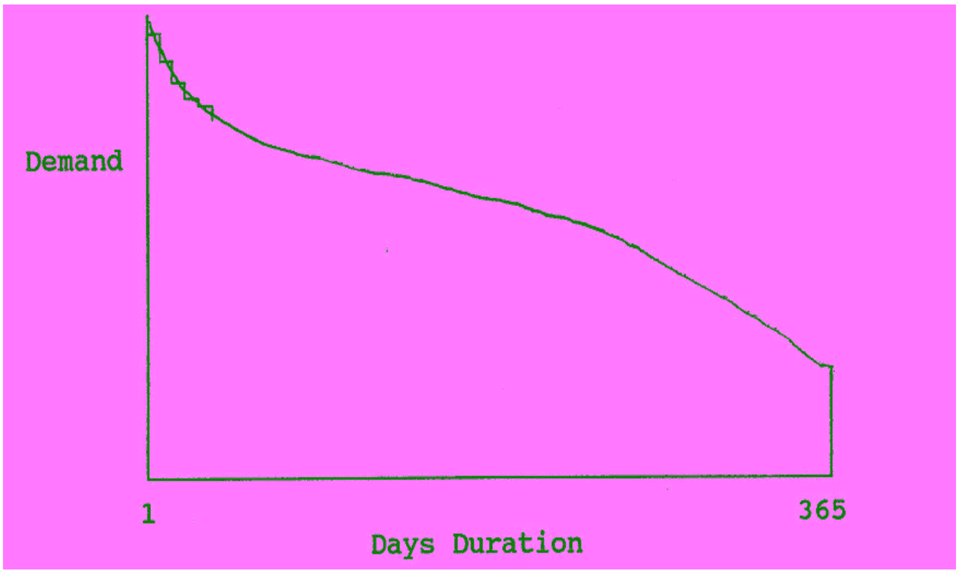
8.4.9 As discussed in section 6 the procedure for parameter estimation in demand models fitted to different periods of the year should specifically identify lower weather sensitivity over the summer period where found to be significant. This should be taken account of in the temperature correction procedure. In particular, there should be an appropriate adjustment, by means of a temperature cut-off at very warm temperatures, to the procedure. In addition Regions should develop methods for detecting switch-on and switch-off of seasonal loads and use the resulting decision rules in adjusting the parameters of the demand model over the periods concerned.

## APPENDIX 1 - THE DEFINITION OF THE LOAD DURATION CURVE

A1.1 In section 3.3 the 1 in 50 load duration curve is defined in a concise way which is familiar to operational researchers and corporate planners in British Gas but less so to others. The purpose of this Appendix is to elaborate in simpler terms on the meaning and purpose of the load duration curve.

A1.2 The term "load duration curve" is not strictly accurate because what is meant is a histogram of 365 steps each of width one day. Because there are so many steps the histogram is usually approximated as a smooth curve. For most purposes this does not matter but for some purposes, especially those concerned with accurate estimation of the area under the top part of the curve, it is important that the analyst keeps this point in mind.

A1.3 A typical load duration curve is shown below. The top part of the curve, which represents demands on the coldest days of the year, is shown in more detail than the rest, both as a daily step function and as a smooth curve drawn through the steps. It can be seen that a smooth curve drawn through the steps intersects the axis at a higher point than does the step function. However, it is the top step of the step function which gives the correct measure of the peak day demand.




A1.4 A load duration curve is constructed by ranking in order of size the daily demands occurring in a 365 day period. It gives no information at all about the pattern of demand within any 24 hour period. Neither does it give any information at all about the relationship between successive days' demand through time. As such it is a very limited representation of demand over a period of a year. However, for many purposes, particularly the determination of LNG storage requirements on the national system it is ideal because it captures in a simple form the essential information from a very complex picture.

A1.5 The simulation model described in section 8 of this report, simulates daily demands corresponding to the weather experience of any 365 day period, usually a supply year. From this simulated demand pattern over a year many analyses can be conducted but so far as load duration curves are concerned, the daily demands are ranked in order of size and graphed or tabulated to produce a load duration curve corresponding to the weather experience of that particular year. In this way load duration curves can be produced corresponding to any actual year. If these curves were all drawn on the same graph many of them would cross over each other many times. In such cases it would not be possible to say that one curve was more or less severe than another. In particular, one winter may give rise to the highest peak day, another winter may be the worst at the 21 day duration, another at the 63 day duration and so on. A method is needed for artificially creating a “1 in 50” load duration curve, an average curve, or indeed any “1 in n” curve, which has the required properties throughout its length.

A1.6 The appropriate method depends on the purpose to which the resulting “1 in n” curves are to be put. Because our predominant interest is in accurate forecasts of volumes of gas above particular demand thresholds ( such as the pipeline level) the method chosen is one that ensures that whatever the demand threshold is, the volume of gas represented by the area under the curve and above the threshold is the best forecast we can make e.g. if it is a 1 in 50 load duration curve then this volume of gas will be our best estimate of the volume which will occur in a 1 in 50 year. This is a more roundabout way saying what is said in para 3.3.2.

A1.7 In practical terms, we take our N actual simulated load duration curves, pick a demand threshold, and calculate the N volumes of gas, represented by the area underneath each curve and above the demand threshold. We then fit a probability distribution to these N values and from this calculates a 1 in 50 value. We repeat this process for a wide range of other demand thresholds. We then draw the 1 in 50 load duration curve in such a way that the area beneath it and above each demand threshold is equal to the value we have calculated. In this way an artificial curve is constructed which has the desired property that the area beneath it and above any demand threshold represents a 1 in 50 volume of gas.

## APPENDIX 2 - THE CONSISTENCY ARGUMENT

A2.1 There are many aspects of forecasting, particularly in the area of temperature/demand relationships and planning criteria, in which assumptions have to be made of a very uncertain, and even arbitrary, sort. In some cases, particularly those to do with day-to-day operational practice in Regions, it is appropriate that these assumptions are made independently in each Region and it does not matter much to the Company as a whole whether they are made quite differently in different places. In other cases, the assumptions are important because they have far-reaching effects on the allocation of resources within the Company. In such cases there is a strong argument in favour of consistent assumptions across all the Regions. It is contended that the forecasting of 1 in 20 or 1 in 50 demands in future winters is of the latter sort.

A2.2 The argument arises from the observation that what any particular Region forecasts as a 1 in 20 peak day or 1 in 50 load duration curve is not 1 in 20 or 1 in 50 in any absolute sense but only relative to a framework of underlying assumptions which it tries to ensure are as near the truth as possible. These assumptions cannot be avoided and, although as much science may be brought to bear as possible, many of them are simply untestable and therefore largely arbitrary. This realisation undermines any argument that any one particular set of assumptions is right or best in any absolute sense.

A2.3 Thus we might not be at all confident, for example, that a load duration curve is truly 1 in 50 rather than 1 in x but we can, through a more consistent treatment of temperature statistics and demand/temperature methodology, ensure that all Regions are each planning roughly to 1 in x, even though we might never know whether x is even close to 50. It is contended that this is a better situation than one where there is a wide range of assumptions across the Regions and consequently a wide, though unintentional, difference in their risk levels.

A2.4 It might be argued that the aggregate of twelve Regional estimates is more likely to be closer to the true national 1 in 50 estimate if as much diversity as possible is built into the assumptions behind the individual Regional estimates i.e. if we deliberately seek a wider range of approaches. This is a statistical argument based on the so-called “law of large numbers” which would imply that the effect of under-estimates in some Regions would tend to be cancelled out by over-estimates in other Regions. It is considered however, that this argument is misplaced in this context because it is felt that the individual Regional estimates are just as important as their total, both to the Regions themselves, and for purposes such as the design of the National Transmission System as a whole, and the allocation of LNG between Regions. Whilst the simple aggregate of the Regional estimates is also important there is no guarantee that, with only 12 figures to aggregate, a wide range of approaches would give rise to a better national total.

A2.5 A distinction can be drawn between definitions and their interpretation on the one hand and the methods used in deriving and applying a demand model on the other. In the first case, (sections 3, 4, 5 and 8.1), the consistency argument should take precedence over arguments based on local preferences or circumstances. In the second case, (sections 6, 7 and parts of 8), there may be good reasons for a variety of practice across Regions. So long as each Region's methods are internally consistent and based on sound principles they should be judged on their own merits by their performance in forecasting.

## APPENDIX 3 - THE SPAN OF YEARS FORMING THE TEMPERATURE DATA BASE

A3.1 In section 4.4 it is stated that all Regions should use the period from October 1st 1928 to September 30th of the most recent supply year as the span of years forming the temperature data base. In this Appendix the justification for this is outlined. The text of sections A3.2-A3.12 is essentially the same as that contained in Appendix 3 of TD76 which was written when the most recent supply year was that ending on September 30th 1979. Thus no reference to subsequent winters is made until Section A3.13.

A3.2 The steps in the argument are as follows:

a. It is first necessary to establish that key results of temperature analysis differ significantly according to the particular span of years that is used as the database.

b. It is next necessary to examine climatological and statistical arguments for favouring one particular span of years rather than another.

c. From this it is concluded, on the basis of the consistency argument in Appendix 2, that it is desirable for all Regions to use the same span of years and that this span should run to the present from an agreed date starting in 1928 or very shortly before, or else starting at some date between about 1870 and 1890.

d. It is then necessary to establish which, if any, particular spans of years within the above limits are practically feasible for all Regions to use.

e. From this it is concluded that a daily temperature series covering the span of years beginning in 1928 can be constructed for each Region and is the best one for all Regions to use.

A3.3 Table A3.1 is based on a weighted average of effective temperatures for each Region and is derived from the H.Q. National Temperature Data Base.

Table A3.1 – The Effect of Different Spans on Severe Winter Estimates

| Span of years | Length | 1 in 20 minimum effective temp $^{o}$ F | 1 in 50 degree-days below 32 $^{o}$ F | 37 $^{o}$ F |
|---|---|---|---|---|
| 1. 1947/8 – 1976/7 | 30 | 25.9 | 82 | 277 |
| 2. 1946/7 – 1975/6 | 30 | 25.5 | 106 | 331 |
| 3. 1937/8 – 1966/7 | 30 | 24.4 | 143 | 385 |
| 4. 1917/8 – 1976/7 | 60 | 25.1 | 105 | 324 |
| 5. 1962/3 – 1976/7 | 15 | 25.3 | 88 | 338 |
| 6. 1961/2 – 1976/7 omitting 1962/3 | 15 | 26.9 | 39 | 216 |
| 7. 1928/9 – 1977/8 | 50 | 24.8 | 124 | 349 |
| 8. 1929/30 – 1978/9 | 50 | 25.2 | 118 | 349 |
| 9. 1928/9 – 1978/9 | 51 | 24.8 | 126 | 360 |

A3.4 The above table shows the sensitivity of 1 in 20 and 1 in 50 estimates to the inclusion or not of particular severe winters in a small span of years. For example line 1 excludes the 1946/7 winter and the estimates are low, whereas line 3 includes the 1939/40 winter and the estimates are high. The sampling error of 1 in 20 and 1 in 50 estimates reduces as the size of the sample grows and on these grounds the length of span to be used in estimating 1 in 50 values should be as long as possible and at least 50 years. On the strength of the above table any of lines 4,7,8, or 9 might be a suitable one to choose.

A3.5 However, ideally the sample size should be much larger and we should therefore look at the possibility of using the longest span of years for which data is available. Table A3.2 shows an analysis based on the longest series of daily temperatures in existence which has been compiled by the Meteorological Office as representative of Central England. This series runs from 1826 to date.

Table A3.2 – Estimates Based on the Central England Series

| Span of years | Length | 1 in 20 minimum effective temp $^{o}$ C | 1 in 50 degree-days below -1 $^{o}$ C |
|---|---|---|---|
| 1828/9 – 1977/8 | 150 | - 6.5 | 65 |
| 1878/9 – 1977/8 | 100 | - 5.8 | 64 |
| 1828/9 – 1877/8 | 50 | - 7.5 | 70 |
| 1878/9 – 1927/8 | 50 | - 6.1 | 75 |
| 1928/9 – 1977/8 | 50 | - 5.6 | 58 |
| 1928/9 – 1978/9 | 51 | - 5.7 | 58 |

A3.6 We can infer from Table A3.2 that the nineteenth century had colder winters than this century and it can be seen that the spans of years resulting in the mildest 1 in 20 and 1 in 50 estimates are the ones relating to the periods starting in 1928/9. Table A3.3 shows all the winters in the three 50-year periods whose degree days below $-1^{o}$C exceed 30. (Incidentally neither 1826/7, 1827/8 nor 1978/9 comes into this category.)

Table A3.3 – Central England Cold Winters since 1827

| 50-year Period | Winter | Minimum Effective Temperature $^{o}$ C | Degree-days below -1 $^{o}$ C | Rank |
|---|---|---|---|---|
| 1828/9-1877/8 | 1829/30 | - 6.6 | 47 | 7= |
| | 1837/38 | - 9.6 | 71 | 2 |
| | 1840/41 | - 7.9 | 47 | 7= |
| | 1844/45 | - 5.6 | 37 | 11= |
| | 1854/55 | - 7.0 | 56 | 5 |
| | 1859/60 | - 7.6 | 32 | 16= |
| | 1860/61 | - 4.1 | 35 | 15 |
| | 1866/67 | - 7.8 | 36 | 13= |
| | 1870/71 | - 6.3 | 42 | 10 |
| 1878/9-1927/8 | 1878/79 | - 5.9 | 49 | 6 |
| | 1880/81 | - 7.8 | 76 | 1 |
| | 1890/91 | - 4.7 | 37 | 11= |
| | 1892/93 | - 4.5 | 32 | 16= |
| | 1894/95 | - 7.4 | 70 | 3= |
| 1928/9-1977/8 | 1928/29 | - 6.9 | 32 | 16= |
| | 1939/40 | - 6.4 | 36 | 13= |
| | 1946/47 | - 5.3 | 46 | 9 |
| | 1962/63 | - 7.4 | 70 | 3= |

A3.7 It can be seen from Table A3.3 that nine of the worst 18 winters of the last 150 years occurred in the first 50-year period, five in the second period and four in the third period. On the basis of degree-days below –1$^{o}$C the winter of 1962/3 ranked only 3rd, and that of 1946/7 only 9th. Also, there is a very long gap between 1894/5 and 1928/9 when no very severe winter occurred. Closer inspection of the records of individual winters reveals that 1916/7 was the only even moderately severe winter during this period. These results are consistent with the findings of climatologists whose general view is that the early part of this century was exceptionally warm when seen against long-term trends in the climate.

A3.8 Although long-term climatic trends certainly exist it has been decided that in the context of using temperature data from the last 50-150 years as the basis for forecasts over the next 5-10 years they are too small and unpredictable to justify making explicit corrections to our daily temperature series. (See Appendix 3 of the original TD76). However the span of years should be chosen judiciously to lessen the chance that it inadvertently covers a period which is either too warm or too cold.

A3.9 The above argument and the results of Tables A3.2 and A3.3 lead us to conclude that the most appropriate span of years from those illustrated in Table A3.3 is that given by line 9, the 51 year period from 1928/9 - 1978/7. This is because the spans represented by lines 4,7 and 8 all look mild in the context of Table A3.2. This is clear if one compares line 7 of Table A3.1 with line 5 of Table A3.2 both of which refer to identical periods. We would therefore conclude that we should either use the period 1928/9 - 1978/9 or else a much longer period, perhaps from some time in the 1870s or 1880s. To use a period starting after the severe winter of 1894/5 and before the mid 1920's would be erring on the mild side.

A3.10 A further observation about sampling error is relevant in reinforcing the conclusion, from the above arguments, that all Regions should use the same span of years and that this should be a span beginning no later than 1928. This is that over the British Isles as a whole the severity of particular winters is very well correlated from place to place e.g. 1962/3 was amongst the most severe winters in all Regions and 1974/5 amongst the warmest. Thus in so far as the sample of winters represented by a particular span of years is, due to chance, an imperfect sample from some underlying unknown distribution the imperfection shows itself roughly in the same way in all Regions. So if the same span of years is used in all Regions a further benefit is that differences between Regions' results arising from sampling error will be minimised.

A3.11 We have now reached point d) in the argument. The argument for recommending 1928/9 as the starting winter of the data base is strengthened by the limited availability of historical series of daily temperature across the country and the problem of generating artificial data from those Regions whose series of daily temperatures does not stretch back this far.

A3.12 In consultation with the Met. Office, at the time of the original TD76 report, H.Q. O.R. Department looked at all the long series of temperature data available and the way they might be used to generate artificial data. A factor which was prominent in alighting on 1928/9 as the starting point was the absence of a continuous long term series for the London area which goes back beyond this year. Furthermore if we start earlier than 1924 a complication arises with East Midlands because of the lack of Nottingham data, and if we start earlier than 1915 a similar problem arises with Eastern Region due to the lack of Rothamstead data. As we move further back towards l900 the problems become more numerous. To move back to a starting date in the 1870's or 1880's would involve all Regions in the construction of an extensive amount of artificial data with the exception of Southern Region. Although on the grounds of sample size this is attractive it is judged that the disadvantage of the extra amounts of pseudo-data required by all other Regions swings the argument in favour of a 1928/9 start date.

A3.13 Since 1981 all Regions have used a historical database which runs from 1st October 1928. In the first instance this covered the 51 year period 1928/9 - 1978/9. This was updated in May 1985 to cover the 56 year period 1928/9 - 1983/4. In July 1987 Corporate Planners Committee received a paper (CPC 87/17) which recommended the adoption of the rules contained in paragraphs 4.4.1 and 4.5.1 of this report. These were accepted and subsequently endorsed at Matching Panel. CPC 87/17 contains the detailed arguments behind this decision which are not repeated here. As at October 1987 the data series was of length 59, a significant improvement in terms of sample size over the original 51.

## APPENDIX 4 THE PROBABILITY DISTRIBUTION OF PEAK DAY TEMPERATURES AND DEMANDS

A4.1 This Appendix is written in terms of peak day demand but it applies equally to winter minimum daily effective temperatures as defined in section 4.6. However, in the latter case, the sign of each temperature would have to be reversed so that the most severe value was the largest rather than the smallest value.

A4.2 The following description of the procedure is based on the theory of extreme values described in the book by Gumbel and the paper by Jenkinson (see the original TD76 for the full references). In Gumbel's book, three types of distribution known as types 1, 2, and 3 are discussed. The distinction between the three types can be seen if the ranked extreme values are plotted against the corresponding Gumbel probabilities (i.e. $-\ln(-\ln((i - ½)/n))$). If the values have a type 1 distribution the plot will be a straight line. A type 2 distribution will give a curve of increasing gradient and a type 3 distribution will give a curve of decreasing gradient, with an upper bound or asymptote. Many Regions have found, and this was confirmed by research by the Steering Group in 1979, that distributions of types 1 or 3 give good fits to their demand data.

A4.3 In the case of a type 1 distribution the parameters of the underlying straight line distribution can easily be obtained by regression methods. However, in the case of a type 3 distribution, the parameters of the underlying distribution can be estimated by regression methods only if the value of the asymptote is chosen arbitrarily. The resulting estimates are sensitive to the choice of bound and this is seen as a disadvantage of using this method where the aim is to achieve consistent estimates across all Regions.

A4.4 In the paper by Jenkinson, however, a method of fitting a Gumbel type 3 distribution is derived which avoids the need for any subjective definition of the asymptote. Jenkinson defines the probability that the peak day demand, D, is less than some given value d as:

$$\exp \{ - (1 - (d - d_o) /a)^{1/k}\}$$

where $d_o$, a and k are parameters to be estimated. (Jenkinson redefines the origin for d as $d_o$ so that in his paper $d - d_o = X$).

A4.5 In order to estimate the parameter k Jenkinson uses the relationship:

$$2^k = \delta_1 / \delta_2$$

where $\delta_1$ is the standard deviation of the peak day demand and $\delta_2$ is the standard deviation of the maximum of any pair of the peak day demands. $\delta_1$ is estimated from the original sample but to estimate $\delta_2$ a second sample has to be defined in which the ith ranked value is repeated (2i-1) times, this being the frequency with which any value is the maximum of any pair in which it occurs. If the original sample contains 59 values the second sample will contain $(59)^2$ values.

A4.6 Having obtained the value of k the other parameters are obtained as:

$$a = \frac{1}{\sqrt{(2k)! - (k!)^2}}$$

$$d_0 = \check{D} - a\,(1 - k!)$$

where $\check{D}$ is the mean of the peak day demands
The 1 in n peak day demand $D_n$ is then given by

$$D_n = d_o + a\,(1-(-\log_e(l- 1/n))^{\,k})$$

A4.7 This procedure was found to work well on all the data to which it was applied and it has the advantage of being entirely objective and hence is more likely to give rise to consistent results across all Regions if used universally.

A4.8 In theory the type 1 distribution is the limiting value of the type 3 distribution as $k \rightarrow 0$. However, in practice Jenkinson's procedure could not be used if $k = 0$. A very small value of k would indicate that the values exhibited only a slight departure from linearity while a high value of k would indicate a very definite curvature.

A4.9 Jenkinson's procedure for fitting the type 3 distribution was considered to be to be the most flexible and objective procedure for obtaining the distribution of peak day demands and it is therefore the method that has been adopted (see paragraph 4.6.4),

A4.10 The procedure described here is included in the simulation model described in Appendix 11. If, in the simulation procedure, k is calculated to be negative or smaller than 0.005 then it is set equal to 0.005 and this results in a curve very close to a straight line being fitted.

## APPENDIX 5- THE PROCEDURE FOR ESTIMATING THE 1 IN 50 LOAD DURATION CURVE

A5.1 This Appendix is derived from Appendix 5 of the original TD76 which referred throughout to a data series of length 51. In this report this is changed to 59 (the appropriate length as at October 1987) but this will change in future in line with paragraph 4.4.1. This does not affect the argument of the Appendix.

A5.2 The most sensitive aspect of the derivation of the 1 in 50 load duration curve is the method by which a probability distribution is fitted to the volumes above each threshold demand. A great deal of research was conducted into this question and Appendix 5 of the original TD76 described the conclusions and some of the reasoning behind them. The proposed procedure is only summarised here. It has been incorporated in the simulation model described in Appendix 11.

A5.3 The starting point for the procedure is a 59 year series of accumulated volumes above a series of demand thresholds. At each threshold the volumes are censored at the average volume (see paragraphs A5.6 - A5.8 below) and a cube root normal distribution fitted to the ranked volumes above the average. The distribution is fitted by regressing the cube roots of the ranked volumes against the corresponding normal order statistics. The resulting parameters are averaged over 28 different simulations. The series of the means and standard deviations obtained from each threshold in turn is smoothed using a 3 point moving average and an estimate of the 1 in 50 volume at each threshold is calculated from the smoothed values. This gives the integrated 1 in 50 load duration curve which is then differentiated (see para. A5.4 below). At a certain point towards the top of the curve, this procedure breaks down as there are an insufficient number of winters with non-zero volumes to which the distribution can be fitted. Above this “transition” point the curve is derived by an algorithm which is described in detail in Appendix 6.

A5.4 The smoothed estimates of 1 in 50 volumes provide the integrated load duration curve. The load duration curve is derived from this as follows. If $V_i$ and $V_{i+1}$ are the volumes above two consecutive thresholds $D_i$ and $D_{i+1}$, then the duration at the intermediate threshold

$$\frac{D_i + D_{i+1}}{2} \text{ is simply given by } \frac{V_{i+1} - V_i}{D_i - D_{i+1}} \text{ days.}$$

The demand on a particular day number is simply obtained by interpolating between successive durations. As the load duration curve is required in histogram form the demand for day i is taken as that for day (i – ½ ) on the continuous curve.

A5.5 The procedure is easily generalised to 1 in n curves for other values of n besides 50. However, because it is based on only those winters colder than average from a sample of size only 59 it should not be used for values of n outside the range $3 < n < 100$.

A5.6 Although the procedure was developed for the generation of load duration curves, it is intended to apply also to accumulate degree-days below a series of temperature thresholds and the estimation of the 1 in 50 temperature duration curve.

A5.7 It was argued in Appendix 5 of the original TD76 report that as our prime concern is with the average year and colder than average years, in particular the 1 in 50 value, the fitting of the distribution should not be allowed to be distorted by those volumes associated with milder than average winters. However, it was not obvious from the data how heavily it should be censored before fitting the cube root normal distribution, so several alternative decision rules were tested as follows:

i) censoring zero volumes only
ii) censoring all volumes less than half the average volume
iii) censoring all volumes less than the average volume
iv) using only the largest five volumes.

A5.8 It was concluded that decision rules i) and ii) produced unacceptable estimates of 1 in 50 volumes because of the adverse influence of small volumes on estimates of interest. However, decision rule iii) appeared to censor the data sufficiently to give reasonable estimates of 1 in 50 volumes as far as could be judged from inspection of the data. Decision rule iv), the "top 5" rule, also gave reasonable estimates of 1 in 50 values as these always tended to lie within the range of the largest two volumes in the sample.

A5.9 It was felt on balance that as the idea had become established that the 1 in 50 estimate should be derived from a distribution fitted to a major portion of the data, rule iii) would be more acceptable than rule iv). It was also felt that rule iv) was too dependent on the particular experience of the severe winters of 1962/3 (and 1946/7) and although these winters were the coldest everywhere across Great Britain the implication under rule iv) that their conditions were of almost the same probability everywhere would not be acceptable. Rule iii) it was felt, would temper the particular effect of these two winters through taking account of all colder than average winters in the sample.

# APPENDIX 6 - THE METHOD FOR OBTAINING THE TOP OF THE TEMPERATURE DURATION OR LOAD DURATION CURVE

A6.1 This Appendix is written entirely in terms of load duration curves but it applies equally to temperature duration curves as described in section 4.7.

A6.2 The 1 in 50 load duration curve can be derived satisfactorily over most of its range from an analysis of the accumulated annual volumes above a series of demand thresholds. However, as the demand threshold gets higher and higher the estimates of 1 in 50 volumes become more and more difficult to make as more and more of the winters in the sample give rise to zero volumes. This Appendix describes a method of deriving the demands corresponding to the top few days of the 1 in 50 load duration curve such that the curve is consistent with the definition in section 3.3.

A6.3 The method basically involves choosing a point on the load duration curve above which the volume analysis described in Appendix 5 breaks down, and completing the curve by fitting a cubic equation subject to four constraints:

(1) Day 1 should be the 1 in 50 peak day demand estimated from an analysis of maximum daily demands.

(2) The demand given by the fitted curve must equal the demand given by the directly obtained curve at the point where the two curves meet.

(3) The area under the fitted curve and above the demand threshold corresponding to the point where the two curves meet must be equal to the 1 in 50 volume derived from the volume analysis for this threshold.

(4) There should be a smooth transition from the fitted curve to the directly obtained curve i.e. where the two curves meet their slopes should be the same.

Assuming the fitted curve is a cubic, conditions (1) - (4) give rise to four equations in the coefficients of the cubic which can be solved to define the fitted curve precisely.

A6.4 There is a problem in deciding at what threshold the volume analysis breaks down and hence the point at which the fitted curve should take over. Experience has shown that a good rule is to take the first point, moving up the curve, at which 5 or fewer values above the average remain. (However, if this demand threshold has a duration of less than 8 days, the first point with a duration of more than 8 days should be used as the curve fitting routine can only reasonably be used over a period of about 8 days in order to satisfy the four conditions).

A6.5 This procedure is illustrated by the following example:

The results of a simulation are shown in the table below:

| Threshold<br>k.th | 1 in 50 value *<br>k.th | Duration +<br>days |
|---|---|---|
| 9000 | 1981 | |
| 8900 | | 8.02 |
| 8800 | 3585 | |
| 8700 | | 12.24 |
| 8600 | 6032 | |
| 8500 | | 18.16 |
| 8400 | 9754 | |
| 8300 | | 23.77 |
| 8200 | 14507 | |

* The 1 in 50 value obtained by the recommended method described in Appendix 3 i.e. a cube-root normal distribution fitted to values censored at the average value.

+ The duration is obtained by the simple difference method.

The first threshold with at least 5 values above the average is 8800.

The 1 in 50 peak day demand is 9715

The duration of demand above 8800 = $\dfrac{12.24 + 8.02}{2}$ = 10.13 days

The slope of the curve at 8800 = $\dfrac{8900 - 8700}{8.02 - 12.24}$ = 47.39 k.th/day

Let the fitted curve be

$D(t) = a + bt + ct^2 + dt^3$

<u>Condition 1</u> concerns day 1 and as we are approximating a histogram by a continuous curve this is assumed to be t = ½

(1) $a + .5b + .25c + .125d = 9715$

<u>Condition 2</u> is given by D(10.13) = 8800

(2) $a + 10.13b + 102.6169c + 103.951d = 8800$

<u>Condition 3</u> is given by the integral of D(t). The volume up to day t is

$at + bt^2/2 + ct^3/3 + dt^4/4$

Note that this is the total volume in the first t days and not the volume above a threshold equivalent to day t. The volume above the threshold 8800 is 3585, so the total area below the curve must equal 3585 + 8800 x 10.13 = 92729

(3) $10.13a + 51.31b + 346.50c + 2632.557b = 92729$

Condition 4 is given by the derivative of D(t)

$$D'(t) = b + 2ct + 3dt^2$$

(4) $b + 20.26c + 307.85d = -47.39$

Solving equations (1) - (4) for a, b, c, d gives

$$D(t) = 9828.4 - 237.26t + 21.457t^2 - .79537t^3$$

A6.6 This procedure provides an objective and unambiguous estimate of the top of the load duration curve which is consistent with conditions (1) to (4). Furthermore, it can be generalised for values of n other than 50. However, as the volume analysis is based only on those winters colder than the average it should only be used for values of n in the range $3<n<100$.

A6.7 This procedure is not applicable to the average load duration curve because condition (1) is not consistent with the definition of the average load duration curve in para. 3.3.6. In case of the average load duration curve the volume analysis breaks down for day 1. The area under the continuous curve and above a demand threshold has a skew distribution. Therefore if the demand at day $^1/_2$ on the continuous curve is taken to represent day 1 on the histogram, the area under the histogram will not be consistent with the volume analysis. The following very simple procedure can be used to overcome this problem. From the volume analysis two consecutive demand thresholds can be identified which give rise to durations just above and just below 1 day on the continuous curve. Interpolating between the values provides a day 1 demand for the histogram that is consistent with the volume analysis.

A6.8 The above procedures should be used by all Regions. The procedures are embedded in the simulation model described in Appendix 11.

# APPENDIX 7 - ISSUES CONCERNING THE INCLUSION OF WIND IN THE DEMAND MODEL

A7.1 In section 5 the definitions and criteria concerning the use of a windspeed variable are described. In this Appendix some of the thinking behind these recommendations is discussed. Interest in the effect of wind on gas demand became widespread following the exceptionally high demands experienced on a few very windy days in February 1979. Subsequently all Regions investigated the significance of windspeed and now include it in some form in their demand models.

A7.2 In practice very few of the Met. Office temperature stations used by Regions also record windspeed and there are only 12 Met. Office Stations in Great Britain with long term wind records. It is the Met. Office view that variations in windspeed across the country are less important than those in temperatures and therefore it is less important for a Region to have windspeed measurements from a location close to its main demand centre. Each Region is therefore recommended to use one of the long series of wind data even if it is not within the Region's geographical area. The creation of artificial wind data for a station with a short series in a better location is a less favoured alternative.

<u>Definition of the Wind Variable</u>

A7.3 There are basically two issues concerning the definition of the wind variable, assuming that several alternative definitions are all found to be significant. The first concerns the goodness-of-fit that the demand model including the wind variable has to the data, and the second concerns the effect that this same demand model has on the resulting estimate of load duration curves. This second point was researched in some depth at H.Q. at the time of the original TD76 report. it was found that different definitions, each fitting the data equally well, can result in significantly different 1 in 50 load duration curves, (see para. A7.9 below). On the grounds of consistency all Regions therefore are recommended to use the same form of definition.

A7.4 The particular definition, as stated in para. 5.3.4 is recommended as a result of considering both theoretical arguments as to the appropriate form of variable, and the empirical results of fitting various forms to different data. On statistical grounds, and also on the basis of models of the physics of heat transfer, a composite variable is preferable to a variable that consists of windspeed only. The composite variable is usually referred to as chill factor which is the product of a wind term and a temperature term. For the purposes of building design the Met. Office suggests a chill factor variable in which the square root of windspeed is taken, e.g. $W_i^{1/2}(65°F - T_i)$, but consider a more general term of the form recommended in para. 5.3.4 also satisfactory. It is also worth pointing out that the experience of other countries with colder climates than the U.K. also leads to the best measure of the effect of wind being of the chill factor sort.

A7.5 Empirical studies have been undertaken of several alternative definitions of the wind variable, viz, W, $W^{1/2}(\check{T} - T)$, $W(\check{T} - T)$, in the terminology of para.5.3.4, where W was measured from a variety of different thresholds, and where T and Ť were interpreted as effective temperatures in some cases and actual temperatures in others. The overall results of these investigations were that there was generally little to choose in terms of goodness-of-fit between the various alternatives (see Appendix 8), but average temperature was better than effective temperature. The inability to discriminate between many of the models was on account of the very limited amounts of data and the confidence intervals around the estimates of residual error being wide and therefore overlapping. Nearly all forms of variable were, however, highly significant.

Fitting the Model Including a Chill Factor

A7.6 The chill factor should only be included in models, which are fitted to periods of the year when the temperature is expected to be cold enough for it to operate on a reasonable number of days. If a Region were using the regime of fitting periods suggested in Appendix 8, it would include the chill factor in the winter and cold period models. It would then only be necessary to collect windspeed data during the October to March period.

A7.7 The chill factor should be included in the winter and cold period models for all years over which the models are fitted, although it may not be significant in those years when there were few "windy" days. When smoothing the coefficients over consecutive years (on the lines recommended in Appendix 8) a Region should be strongly influenced by the coefficients obtained in the "windy" years.

A7.8 The fact that the inclusion of windspeed or chill factor in the demand model covering the summer period (April 1$^{st}$ - September 30$^{th}$) is not recommended is not because it is believed that wind has no effect on demand over this period. Rather, it is thought that the degree of improvement in goodness-of-fit of the demand model and the significance of the results on the load duration curve are likely to be too small to warrant the effort involved in data collection and analysis.

The Effect of Wind on the Load Duration Curve

A7.9 The simulation model can be used to estimate the effect on the load duration curve arising from the inclusion of a wind variable in the demand model, and to quantify how this effect varies according to the particular definition of wind variable employed. The results of some such investigations are described in background papers to the original TD76 report.

# APPENDIX 8 – COMPARATIVE PERFORMANCE OF DIFFERENT MODELS

## A8.1 Introduction

The general form of the model in Section 6.1 is

$$D_i = a + bS_i + cT_i + dw_i + e\,g_i$$
$$+ \sum_{j=1}^{7} f_{i\,j}\ \{\alpha_j + \beta_j S_i + \gamma_j T_i + \delta_j w_i\} + u_i$$

where $u_i = \rho\, u_{i-1} + \varepsilon_i$

and $f_{i\,j} = 1$ if $i \in$ class $j$ where $j = 1$ = Fri; $j = 2$ = Sat;…

$= 0$ if $i \notin$ class $j$ .... $j = 7$ = Thur.

In developing a model each Region should start from a fairly general form and then look for simplifications, following good statistical procedures. (Alternatively, the Region could start with a simple model and build up to the general model, again following good statistical procedures). The final form of the model that will be arrived at by applying this procedure may vary from Region to Region. The purpose of this Appendix is to illustrate how a Region might go about identifying the best model for its own circumstances starting with the general form. In Tables A8.1, A8.2 and A8.3 particular models are quoted where they serve to illustrate the points discussed in this Appendix. The work described was undertaken in 1980 but remains relevant today.

## A8.2 The Seasonal Term

The most usual seasonal term is SNET, although other measures have been suggested. Unless an unusual temperature statistic is found which itself allows for seasonal variation, the seasonal term is normally significant in models fitted to three months data or more. For models fitted over a period as short as a-month the term may not be important, but a significant seasonal term has been found in models fitted to six weeks' data, mid-January to end-February (as in the model shown in Table A8.1).

## A8.3 Temperature Deviation

If a seasonal term is included in the model the temperature is usually taken as the deviation from SNET. The general form of the model given above includes a single coefficient c, i.e. it assumes a constant temperature response over the period of the model. This assumption should be checked if the model is fitted over a number of months. In particular it is expected that there will be a significant difference in the coefficient between summer and winter.

## A8.4 Other Weather Variables

Other weather variables should be investigated, particularly the chill factor variable defined in para. 5.3.4. Models (1) - (4) in Table A8.3 show the effect of including different definitions of a wind variable in models fitted to the winter of 1978/79. It can be seen that the residual standard error is reduced by any of the wind variables but the greatest reduction is found with a chill factor of the recommended form.

A8.5 Weekend Effects

The models only included four daily variables and have allowed $\alpha$, $\beta$ and $\gamma$ to vary for weekdays (Monday - Thursday), Friday, Saturday and Sunday. A significant difference in $\alpha$ is usually found for each of the three weekend days. Occasionally $\beta$ is significant and rarely $\gamma$. Fitting constant $\beta$ and $\gamma$ has made little difference to the residual error as indicated (in the case of $\gamma_2$) in models 5 and 6 in Table A8.3. Models for two previous years (not shown) had contained only significant $\alpha_1$, $\alpha_2$ and $\alpha_3$ for weekends whereas for the third year model 5 shows $\gamma_2$ also to be significant. To reach a common form of model for all three years model 6 was used with $\gamma_2$ forced out resulting in only a slight increase in residual standard error over model 5, from 192 to 194.

A8.6 Growth

A simple linear growth term was included in the models. For some sets of data this term may not be significant. This is so for short periods (e.g. two months) and in some cases for even six month periods, as in the case of the winter six months for 1977/78 in model 7 in Table A8.3. However, the growth term for the following winter six months was significant as in model 8. To reach a common form of model over a period of years the growth term would only be included if it was found to be significant in the majority of years provided there were no step problems from one year to another.

A8.7 Length of Fitting Period.

As is usual in statistics, the choice of fitting period is a matter of balancing conflicting factors. A model over a short period is likely to fit better and may be simpler but the parameters will be estimated on relatively few points and so may be unstable. The optimum length of period during the winter was investigated by fitting the general model using stepwise regression to five different time periods as follows:

- January and February (beginning well after the New year holiday)
- December t o February (missing out the Christmas/New year period)
- November to the coldest day (missing out the Christmas/New year period)
- November to the peak day (missing out the Christmas/New year period)
- October t o March (missing out the Christmas/New year period)

The goodness of fit of a model is usually judged in the first instance on its residual standard error and on this criterion the fit for the longest time period was not much worse than for the shortest. However, the most important purpose of the winter demand model is to simulate the period around the peak day. It was thought that the long period model might fit well on average but not fit well enough over the main period of interest. To test this, the residual standard error, the mean of the residuals over the January - February period, and the projected demand at the 1 in 20 temperature were calculated for the model fitted to each of the five periods in turn. Further criteria examined were the standard error of the estimates at the 1 in 20 temperature. For each of the five time periods the performance of the model against the criteria listed above can be compared. The results in Table A 8.1 relate to 1977/78 data and indicate that the six-month model fitted rather less well for the period of interest than the shortest period model. There was no consistent difference in the extrapolation to cold temperatures and the parameters of the six - month model were estimated with much less error. The choice between accuracy and stability in these cases appeared finely balanced. Each Region should make its own judgment.

One interesting point that emerged is that the model fitted from 1st November to the peak day gave a very poor fit to the data for January and February.

For the summer months there is no period of particular interest and a model fitted to the six months April to September may be sufficient. However, attention should be paid to the temperature sensitivities in the model as it may be found that there is a lower response in the period June to August. Attention should also be paid to the increased incidence of holidays in the summer period and their resultant effect of demand. Attention should also be paid to the timing of the switch-on/switch-off heating effect. Ideally the transition from a winter to a summer model (and vice-versa) should coincide with this (see also section 8.4). One possible regime would therefore be to use three models:

(a) a summer model fitted and used for the period April to September
(b) a winter model fitted to the period October to March and used f or this period except for a selected cold period e.g. January (after the New Year) to February.
(c) a cold period model fitted to the coldest period of the winter, which may vary from year to year, but used for a selected mid-winter period as indicated in (b).

When more than one model is used in this way through the year care must be taken that there is no major discontinuity when changing from one model to another. Both models should be evaluated for an overlapping period around the point of change. If the discrepancy appears too great for a change at a given date the models should be merged, for example by forming a weighted average with the weights changing from one model to the other.

The simulation model described in Appendix 11 allows load duration curves to be generated via demand models which divide the year into up to 10 separate periods (and in principle it could allow any number). If, however, three basic periods are used as suggested above there is plenty of scope within the simulation model as presently formulated to handle the changeover from one model to another.

## A8.8 Autocorrelated Residual Structure

For ordinary least squares fitting of the models, it is assumed that the residual errors are independent random variables. This assumption should be tested by examining a plot of the autocorrelation function (acf) or through the use of the Durbin - Watson statistic although this only tests for first order autocorrelation. In practice very few data sets examined so far have produced acfs with significant second (or higher) order autocorrelation. In all Regions the residuals have been found to be correlated, and a suitable modelling routine to take account of this should be used. The inclusion of autocorrelated residual structures invariably gives a marked reduction in the residual error no matter how many explanatory variables are included in the original model. The improvement in fit has usually been as great or greater than that obtained by adding additional variables to a basic model i.e. a correct model for the residuals is as important in improving the fit as are weekend effects, wind effects, growth and so on.

Comparisons on two separate sets of data summarised in Table A8.2 indicate that the improvement in fit by adding .an autocorrelated residual structure to the basic model is greater than that obtained by adding wind and growth terms. When such a structure was added to the model as well as wind and growth terms, the fit was further improved. Further details of the results for the second set of data are shown in Table A8.3 where models 9 - 12 have no autocorrelated residual structure and models 13 -16 include a first order autocorrelated residual structure (AR 1 ). For the results in Table A8.2 examination of the autocorrelation function indicated significant autocorrelation at lag 1 and so an AR 1 model was used.

Estimates of the autocorrelation parameter ( ʃ) have been found to be reasonably stable within data sets, for a given model structure. For one set of data examined, values of ʃ between 0.4 and 0.6 were associated with a change in residual standard error from the optimum (at ʃ = 0.5) of 164 to 166. There was also very little change in the estimates of the other parameters.

The autocorrelation parameter is modeling something, which obviously occurs in the real world - a persistence of the demand pattern not accounted for by the weather variables included in the model. As such, it should be included in any simulation using the fitted model.

The simulation model described in Appendix 11 calculates daily demands through the year and accumulates them to form load duration curves. In doing this it deliberately discards information on the pattern of demand through the winter. Such information may however be very valuable for purposes beyond those covered in this report and the simulation model is consequently a potentially much more powerful tool for examining demand profiles. However, this potential cannot be realized unless the underlying demand model used in the simulation is also realistic with respect to autocorrelation.

## A8.9 Consistency Between Years

In determining the structure of the model to be used, data from several past years should be examined (five years appears reasonable). It is suggested that the models, periods of fitting, and so on should first be investigated for each year in isolation and the structures then compared. If a variable is found to be significant in four out of five years it should be included for all years. Similarly, a variable significant in only one out of five years may make little difference if it is excluded from all years. Other more borderline cases should be examined to determine the overall effect of including or excluding variables in order to arrive at a common structure for the model to be applied to all years.

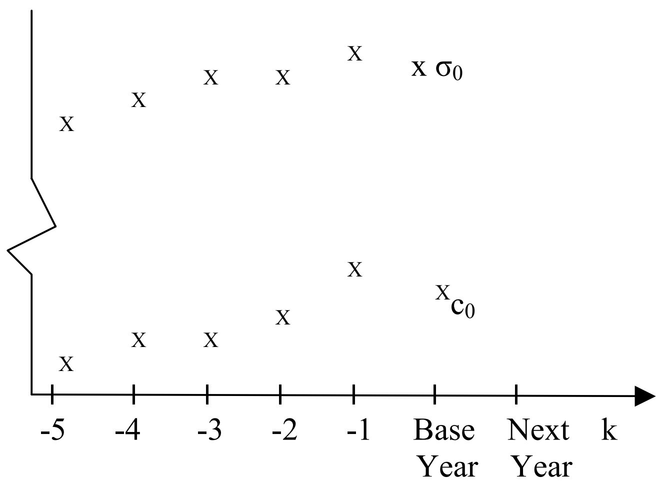
In this way, a common form of model(s) is derived for each of the last five years. Suppose for the sake of argument that the cold period models areas follows:

$$D_{ik} = a_k + b_k SNET_i + c_k (E_i - SNET_i) + d_k w_i + \sum_{j=1}^{3} \alpha_{jk} f_{ijk} + u_{ik}$$

where $u_{ik} = 0.5\, u_{i-1,k} + \varepsilon_{ik}$ and $\varepsilon_{ik}$ is $N(0, \sigma_k^2)$

and k = 1 to 5 refers to the last 5 winters in turn .

One way of smoothing parameters $a_k$, $b_k$, $c_k$, $d_k$, $\alpha j_k$ and $\sigma_k$ for the purposes of projection to future years is to plot them on a graph of the sort shown below in the case of $c_k$ and $\sigma_k$




Suppose, for the sake of illustration, that the base year (k=0) was particularly mild but that the previous winter (k= -1) was severe and experienced a wide range of temperatures, and that earlier winters (k= -2 to k= -5) were all fairly average. In this case, it would be recommended that in fitting a smooth curve through the ck and $\delta_k$ other things being equal, the curve should pass closer to the point corresponding to k= -1 than to the point corresponding to k=0. This is because of the argument of Appendix 9 and the argument that the experience of cold weather is of greater relevance than that of mild weather, even though it may be less recent.

The precise way in which smooth curves are obtained as the basis for projection forward to future years must be a matter for individual Regional judgement. It is not suggested that the projections would necessarily be simple extrapolations; they would depend on marketing and economic forecasts and be derived along the lines described in section 7 and Appendix 10.

Table A8.1

Models Fitted To Different Periods Of The Winter

The form of the model is:

$$D_i = a + bS_i + c(T_i - S_i) + e\,G_i + \sum_{j=1}^{3} \alpha_j F_{ij} + u_i$$

| Fitting Period | Parameter Estimates | | | Residuals for Jan/Feb. | | Estimated demand at 1 in 20 temp. |
|---|---|---|---|---|---|---|
| | a | b | c | Mean/Bias | St.Dev. | |
| Jan-Feb | 3685(40) | -142(15) | -106(5) | 0 | 62 | 4271(81) |
| Dec-Feb | 3468(62) | -184(12) | -103(4) | -1 | 68 | 4236(82) |
| Nov-Feb (coldest day) | 3077(88) | -98(11) | -92(3) | -20 | 113 | 4068(99) |
| Nov-Feb (peak day) | 3084(81) | -98(10) | -90(4) | -21 | 106 | 4053(89) |
| Oct-March | 3405(36) | -135(4) | -100(3) | 29 | 77 | 4147(90) |

(Figures in brackets are standard errors of the estimates)

The Effect On The Residual Standard Error of Including an Autocorrelation Term

The models are:

1. $D_i = a + bS_i + C\,(T_i - S_i) + \sum_{j=1}^{3} \alpha_j F_{ij} + u_i$

2. $D_i =$ ″ ″ ″ ″ $+dW_i$

3. $D_i =$ ″ ″ ″ ″ $+eG_i$

4. $D_i =$ ″ ″ ″ ″ $+dW_i + eG_i$

models 5 - 8 are of the same form as models 1 - 4 but have an
autocorrelated residual error structure.

| Model | Winter WM Data 1978/79 | | | Winter SC Data 1978/79 | | |
|---|---|---|---|---|---|---|
| | **Residual Standard Error *** | **Durbin - Watson** | $\rho$ | **Residual Standard Error *** | **Durbin - Watson** | $\rho$ |
| 1 | 240 | 0.8 | | 195 | 0.6 | |
| 2 | 203 | 1.0 | | 178 | 0.7 | |
| 3 | - | - | | 165 | 0.8 | |
| 4 | 194 | 1.1 | | 142 | 1.1 | |
| 5 | 192 | 1.7 | 0.6 | 134 | 2.1 | 0.7 |
| 6 | 173 | 1.8 | 0.6 | 132 | 2.2 | 0.7 |
| 7 | - | - | - | 130 | 2.0 | 0.6 |
| 8 | 171 | 1.8 | 0.5 | 125 | 2.0 | 0.5 |

* Residual Standard Error = $\hat{U}_i$ for model 1 - 4
″ ″ ″ = $\hat{E}_i$ for models 5 – 8

Table A8.3

Some Models which Illustrate Certain Aspects of The Procedure Described In Appendix 8

The parameters in the table relate to the following models fitted to the period October - March 1978/9 (except model 7 which was fitted to 1977/8)

1. $D_i = a + bS_i + c(T_i - S_i) \qquad + \sum_{j=1}^{3} \alpha_j F_{ij} + u_i$

2. $D_i = \;'' \quad '' \quad + dW_i \quad + '' \quad '' \quad ''$

3. $D_i = \;'' \quad '' \quad + dW_i(\check{T}-T_i) + '' \quad '' \quad ''$

4. $D_i = \;'' \quad '' \quad + dW_i(\bar{A}-A_i) + '' \quad '' \quad ''$

5. $D_i = a + bS_i + c(T_i - S_i) + dW_i(1+\check{T}-T_i) + e\,G_i$

$$+ \sum_{j=1}^{3} F_{ij}(\alpha_j + \beta_j S_i + \gamma_j(T_i - S_i)) + u_i$$

6. $D_i = a + bS_i + c(T_i - S_i) + dW_i(1+\check{T}-T_i) + e\,G_i + \sum_{j=1}^{3} F_{ij}\alpha_j + u_i$

7. $D_i = \;'' \quad '' \quad '' \quad + \sum_{j=1}^{3} F_{ij}\alpha_j + u_i$

8. $D_i = \;'' \quad '' \quad '' \quad + e\,G_i \quad '' \quad ''$

9 - 16 are models 1-8 on Table A8.2.

| | a | b | c | d | e | $\alpha_1$ | $\alpha_2$ | $\gamma_2$ | $\alpha_3$ | $\int$ | $\sigma_e$ |
|---|---|---|---|---|---|---|---|---|---|---|---|
| 1 | 8255 | -368 | -275 | | | -207 | -976 | | -956 | | 241 |
| 2 | 7922 | -354 | -283 | 26.0 | | -192 | -193 | | -913 | | 216 |
| 3 | 7862 | -338 | -252 | 3.5 | | -211 | -948 | | -908 | | 203 |
| 4 | 7857 | -339 | -247 | 3.6 | | -200 | -930 | | -886 | | 190 |
| 5 | 7024 | -319 | -245 | 3.0 | 1.7 | -211 | -985 | -29 | -913 | | 192 |
| 6 | 7018 | -319 | -249 | 3.0 | 1.7 | -212 | -955 | | -913 | | 194 |
| 7 | 7284 | -302 | -204 | 3.1 | | -273 | -1000 | | -877 | | 160 |
| 8 | 7018 | -319 | -249 | 3.0 | 1.7 | -212 | -955 | | -913 | | 194 |
| 9 | 4362 | -191 | -142 | - | - | -128 | -475 | | -328 | | 195 |
| 10 | 4291 | -186 | -138 | 4.6 | - | -138 | -474 | | -337 | | 178 |
| 11 | 3996 | -164 | -123 | - | 2.6 | -115 | -147 | | -318 | | 165 |
| 12 | 3917 | -159 | -118 | 4.7 | 2.6 | -125 | -477 | | -329 | | 142 |
| 13 | 4335 | -185 | -143 | - | - | -134 | -470 | | -296 | 0.7 | 134 |
| 14 | 4314 | -185 | -141 | 1.7 | - | -134 | -465 | | -299 | 0.7 | 132 |
| 15 | 4055 | -165 | -134 | - | 1.9 | -127 | -473 | | -296 | 0.6 | 130 |
| 16 | 3976 | -161 | -128 | 2.6 | 2.3 | -121 | -467 | | -302 | 0.5 | 125 |

In the models described in Tables A 8.1, A8.2 and A8.3:

$S_i$ is $SNET_i$ (see section 4.5)
$T_i$ is $E_i$ (see para.4.3.2)
$G_i$ is the day number

$W_i$ is average daily windspeed (see para.5.3.1) and the $F_{ij}$ are dummy variables for Friday, Saturday and Sunday.

## APPENDIX 9 ERRORS IN TEMPERATURES USED IN GAS DEMAND MODELS

A9.1 In principle a problem can occur in estimating the parameters of the demand model where there are errors in the temperature variable. This situation might arise where the temperatures a Region is using in its temperature/demand model are not the 'true, temperatures influencing demand. This might be due to the wide geographical spread of a Region and the consequent difficulty in obtaining a representative temperature reading.

A9.2 It can be shown that errors in the independent variable will lead to a downward bias in the parameters of a model estimated by least squares regression. In Appendix 9 of the original TD76 report this problem was examined in detail and only the conclusions of the analysis are given in this report. Briefly, both least squares and instrumental variable methods were used to estimate the parameters of a simple model of the $D = a + bT$ form fitted to a limited period of the winter for four successive winters. The main conclusion was that although the least squares estimate of b was biased downwards the bias was very small in a cold winter where a wide range of temperatures had been experienced. In the light of this result, it was recommended that when smoothing parameters over a number of years (on the lines described in Appendix 8) more weight should be given to those years where a wide spread of temperatures occurred than to other years (even if the other years are more recent)

A9.3 On balance it was concluded that the difficulties in applying the instrumental variable technique within the framework of section 6 and Appendix 8 outweighed the benefits in terms of the improvement in the estimate obtained, as any underestimate in peak day demand due to errors in temperatures is likely to be small. Regions should, however, be aware of the possible effects of errors in measurement of the independent variables in the demand model, and test for these if there is reason to suspect they may be significant.

# APPENDIX 10 – FORECASTING MODEL PARAMETERS FOR FUTURE YEARS

A10.1 Introduction

Section 7.2 describes the general features of the approach to breaking down the total sendout model into market sector sub-models and section 7.3 outlines the general way in which the coefficients of these market sector models could be projected forward for future years and the models summed to give a total model for future years. In this Appendix three alternative practical interpretations of the general principles are described to illustrate the variety of ways in which the procedure can be implemented. These approaches have been termed the "load factor" approach, the "sector modeling" approach, and the "integrated model" approach and they are assumed to apply in Regions A , B, and C respectively.

Past Years by Market Sector

A10.2 The "Load Factor" Approach

The "load factor" approach illustrates that a total daily sendout model can be broken down into a reasonable set of sub-models even when information on individual market sectors is limited and of poor quality.

In Region A, for example, the only daily information available is that for total sendout. The procedure of section 6 is followed to obtain a total model covering the whole year, for several past years. In this case, there are 3 versions of the model for each year, one covering the summer six months, one the winter six months and another overriding the winter model for January and February, the so-called cold period model. Quarterly and annual sales information is available in Region A for eight market sectors as follows: Domestic, Industrial Tariff, Commercial Tariff, Industrial Firm Contract, Commercial Firm Contract, Industrial Interruptible Contract, Commercial Interruptible Contract and Own Use. The sum of these eight sectors equals the total sendout after allowing for unaccounted for.

It is considered most important to obtain a reasonable breakdown of the cold period model for each year, so for each sector a load factor is estimated. The load factors are really no more than educated guesses but are implicitly based on survey analyses and individual contract information. As the load factor gives a relationship between the demand on the average day and the "peak day", it can be used with annual sales information to provide a rough estimate of demand on the "peak day", in each market sector. For the purposes of this analysis the demand on the "peak day", is interpreted as the demand at the 1 in 20 temperature. The sum of these individual sector demands is compared with the total demand at the 1 in 20 temperature as estimated from the total sendout model for the Region in each year. If these totals are very different the load factors are adjusted and the procedure repeated until a set of reasonably consistent load factors is obtained in each year. The load factors may also be adjusted to ensure that the progression from year to year is reasonable in the light of known changes in the composition of the load in each market sector.

Having arrived at a final best estimate of the load factor for each sector, for each year, simple demand/temperature models of the D = a+bT form are derived by solving the two simultaneous equations relating to average and 1 in 20 temperature conditions. The parameters of the total model are then broken down and allocated to each market sector model pro rata with the relationships between the values and across the simple sector models. This procedure is illustrated in the following example:

e.g. Suppose the total model for the most recent year is

$$D_i = a + bS_i + cT_i + dW_i + eG_i + \sum_{j=1}^{3} \gamma_j F_{ij} + u_i$$

and we have derived eight sector models plus an unaccounted-for model of the form

$$Dk_i = \alpha_k + \beta_k T_i \qquad\qquad k = 1 \ldots\ldots. 9$$

for the cold period and $\sum\alpha_k = \alpha$ and $\sum\beta_k = \beta$
Then the simplified form of the total sendout model is

$$D_i = \alpha + \beta T_i$$

An estimate of the complex version of each individual sector model can be derived by making simple assumptions such as:

$$\beta_k / \beta = c_k / c = d_k / d$$
$$\alpha_k / \alpha = a_k / a = e_k / e = \gamma_{jk} / \gamma_j$$
for each k.

Market sector models for the summer and winter periods can be obtained by breaking down the appropriate total sendout model assuming that the parameters are allocated between market sectors in the same proportions as in the cold period models.

A10.3 The "Sector Modeling" Approach

The "load factor" approach described above is essentially a "top down" type of approach adopted by Region A because data on individual market sectors is limited. In contrast, the "sector modeling" approach is a "bottom up" type of approach, which is used in a Region which has enough detailed information to allow modeling of individual market sectors.

Region B, like Region A, has daily information for total sendout and follows the procedure of section 6 to derive total sendout models for a series of years. Region B also has 3 versions of the model covering the same period of the year as Region A. However, in contrast to Region A, Region B has daily information on most contracts from telemetry and loggers and has supplemented this by a survey to give full coverage of the contract market. Region B can therefore derive models following the procedure described in section 6 for any combination of contracts for the same time periods used for its total model. A few contracts do not lend themselves to this approach, because of their peculiar day of the week or holiday effects or uneven growth, and are therefore treated differently. For these a daily profile of seasonal normal demand and a temperature sensitivity are derived, giving a model of the form shown below for the kth contract.

$D_{ki} = X_{ki} + a_k (S_i - T_i)$

$D_{ki}$ = demand of contract k on day i

$X_{ki}$ = seasonal normal demand of contract k on day i

$a_k$ = temperature sensitivity of contract k

Region B obtains a daily model for the tariff sectors by subtracting the daily contract models from the corresponding total model. As the tariff market is mainly domestic the breakdown between Domestic, Industrial and Commercial components can be achieved for any period of the year by making reasonable assumptions. However, Region B also derives a completely independent model of the Domestic market using the results of the Annual Peak Load Surveys and/or Domestic monitors. The daily meter readings from the surveys are used to derive models of the general form for specific appliance groups following the framework suggested in section 6. The parameters of these models are checked for consistency over a period of years. Apart from providing a completely independent check on the model derived by differencing, the very detailed breakdown into appliance groups provides a good basis for forecasting.

A10.4 The "Integrated model" Approach

The "load factor" approach of Region A and the "sector modelling", approach of Region B both involve deriving market sector models to correspond with a total sendout model for chosen periods of the year in several past years. The "integrated model", approach adopted by Region C is quite different in that five years of data on both total sendout and demand in individual market sectors is analysed together to derive a single model for each market sector. Nevertheless, this approach is thought to be within the spirit of the general principles set out in section 7.2.

In common with Region A, the only daily information available in Region C is that for total sendout. A model of the general form recommended in section 6 is derived for each of the past five years. One model is fitted to a full years data. The particular form of model for each year is:

$$D_i = \sum_{k=1}^{12} a_k x_{ik} + c (T_i - S_i) + dW_i + \sum_{j=1}^{3} \gamma_j F_{ij} + u_i$$

The $x_{ik}$ are weights associated with a particular day of the year which allow interpolation between the coefficients, $a_k$ which can be interpreted as mid-month seasonal sendout for each month of the year. The other terms are as shown in section 6.

Region C identifies the same eight market sectors as Region A but before attempting to break the model down by market sectors makes certain assumptions about the forms of the sub-models for particular sectors. The day of the week effects (estimated from a survey of selected customers) are assumed to be applicable only to non-domestic models while the wind chill effect is only included in the domestic sub-model.

Monthly billed information is available for the six non-domestic sectors and this is analysed for the whole five year period. Briefly the analysis for each sector involves correcting to a standard number of days in each month, temperature correcting these sales using an assumed temperature sensitivity and then calculating a twelve month centred moving average (i.e. a trend). The seasonal pattern of variation about the trend in each of the five years is expressed as a monthly factor. These factors are averaged over the five years to provide a single set of monthly seasonal factors for inclusion in the sub-model. Applying these factors to the mid-month trend for the five years provides a series of smoothed mid-month sales for each of the non-domestic markets. These are compared with the corresponding actuals and adjustments made to the parameters if necessary.

Having derived an acceptable set of non-domestic sub-models, the total smoothed non-domestic mid-month sales are subtracted from the total mid-month seasonal sendouts after allowing for unaccounted-for, to provide a series of mid-month seasonal demands for the domestic market. To derive a model for the domestic sector the concept of an equivalent fire as measure of connected load is introduced. This relates all appliances to a common standard, which is chosen to be a main living room fire. The domestic mid-month seasonal demands are plotted against the establishment of equivalent fires separately for each month. A line is fitted by eye to the five points for each month. The intercept of the line is interpreted as the domestic base load and the slope as the heating load per equivalent fire. The temperature sensitivity is analyzed in a similar way. Separate models are then derived for domestic base load and heating load. The base load model is of a very simple form having a mid-month seasonal demand per customer and temperature sensitivity. The heating load model is similar except that demand is related to equivalent fire establishment and the wind chill effect is included.

The single set of sub-models obtained from the five-year analysis is used to generate total sendouts for the most recent year. If the correspondence with actual values is not satisfactory, particularly in the winter, adjustments are made to the parameters.

Future Years by Market Sector

A10.5 The ”Load Factor” Approach

In the case of Region A the breakdown of the total model by market sectors concentrated on the cold period and was based on identifying annual sales and load factors for each of eight market sectors. To project forward the cold period models for each future year it is simply necessary to forecast growth in annual sales and associated load factor, taking account of any changes in the structure of the existing load, and hence derive appropriate values - of $\alpha_k$ and $\beta_k$. Adding the sector models gives the total model for the cold period for each forecast year which is then compared for consistency with the total models in each of the past years. Forecast models for other periods of the year are obtained by scaling the past years' models in proportion to the changes in the cold period models.

Region A then has a set of total demand models for each forecast year which can be input to the simulation model described in Appendix 11 to produce the load duration curves and peak day demand forecasts for the years of the ROP.

A10.6 The "Sector Modelling" Approach

Region B was fortunate in having good daily information on which to base its contract sector model for all periods of the year in past years. These models are projected forward on an individual contract basis where possible and in the case of other forecast sales by identifying likely industrial categories and the associated pattern of demand. For the domestic sector, Region B has used the peak load curves to obtain models of the general form suggested in section 6, for specific appliance groups for the mid-winter period. This detailed breakdown means that models for future years for the domestic sector can be obtained very easily by scaling the coefficients of the appliance group models in proportion to the expected net growth in the number of appliances, and summing the group models. Any expected changes in demand pattern of particular appliance categories can be allowed for by adjusting the appropriate coefficients. The resulting total domestic model can be compared with that of past years and other future years and adjusted to be consistent from one year to the next as appropriate. Region B can then follow the same procedure as Region A to obtain total models for all periods of the year, for each forecast year, which can be input to the simulation model to produce the required load duration curves and peak day demand forecasts.

A10.7 The "Integrated Model" Approach

Region C used a different approach for obtaining market sector models for past years and its approach to building up forecasts for future years is also different although it makes use of the same basic information. One of the differences between the "load factor" and "sector modelling" approaches and the "integrated model" approach to the breakdown of the total model is that the latter provides one model applicable over the 5 year period analysed while the former approaches provide separate models for each year. This distinction also applies in using the sector models for forecasting. In the cases of Regions A and B a set of models has been derived for each forecast year but in the case of Region C the set of models derived for the past years is used for the future years as well and the appropriate level of connected load is input as a variable in order to derive demands for a particular year. The forecasts required by Region C are therefore the annual rate of sales for each month of the non-domestic sectors and the establishment of equivalent fires for the domestic sector. These forecasts are built up from Marketing forecasts of expected load growth. Any changes in the relative consumption of appliances are taken account of in forecasting the number of equivalent fires.

The "integrated model" approach provides a total demand model for each forecast year, which has a form very similar to that recommended in section 6. It will produce a daily seasonal normal demand profile for each forecast year for total sendout and a sensitivity to temperature and other weather factors which can be input to the simulation model.

# APPENDIX 11 - DESCRIPTION OF THE SIMULATION MODEL

A11.1 Introduction

In this Appendix a simulation model for the generation of load duration curves is described. The description is very similar but not identical to that given in the original TD76. A summary of the model is first given, followed by sections covering the input to the programme, the method of simulation, and the output from the program. A more detailed exposition of the latest version of the model can be obtained from H.Q. OR Department. Throughout this Appendix the number 59 is used (being the number of years in the database as at October 1987). This will change in future in accordance with section 4.4.

A11.2 Summary

The model produces the average and 1 in 50 load duration curves and the 1 in 20 peak day demand for any year, given a relationship between demand and a set of independent variables (principally temperature) on each day in the year and a 59 year series of daily values of the independent variables.

For each day of the first year (say 1/10/28 - 30/9/29) a random error is generated and added to the demand, to create a daily demand profile corresponding to the year's temperatures. At the same time a second demand profile for the year is created using the antithetic random number stream. This procedure is repeated for 14 independent random number streams to give 28 simulations of the year. The procedure is then repeated for each of the 59 years.

For each simulation, for each of the 59 years the maximum daily demand and the accumulated demand above each of a series of threshold levels of demand are recorded. The 59 values are than ranked and used in the estimation of average and extreme values in each of the 28 simulations.

“Average” peak day demand is simply the average of the 59 maximum daily demands. The 1 in 20 and 1 in 50 peak day demands are estimated after fitting a probability distribution as described in Appendix 4.

For each threshold level of demand the average volume above that threshold level of demand is simply the arithmetic average of the 59 values from the simulation. The 1 in 50 load duration curve is obtained following the procedures described in Appendix 5 and Appendix 6.

A simplified illustration of the operation of the simulation model is shown in Figure A11.1

A11.3 The Input to the Model

In order to simulate daily demand and hence load duration curves the model requires as input an equation of the form $D_i = f(T_i \text{ etc.})$ for each i = 1 …. 365(366). In principle any function could be used and a different one for each day of the year could be used and a different one for each day of the year could be input. However, the program is currently written so that up to 10 different demand models can be input corresponding to up to 10 divisions of the supply year. These demand models are derived quite separately from the simulation model by the procedures described in Sections 6 and 7.

The particular demand model used in the program, for all time periods, has the following form: -

$D_i = C_1 + C_2\,SNET_i + C_3\,(E_i - SNET_i\,) + C_4W_i + C_k + C_{14}\,g_i + HOL_i + u_i$

Where: $k = 5, 6, 7, 8, 9, 10$, or 11 depending on
the day of the week of day 1, $k = 5$ being Monday etc.

and $u_i = C_{12}\,u_{i-1} + C_{13}\,\varepsilon_i$

The independent variables are:

$SNET_i$ = Seasonal Normal Effective Temperature on day i as defined in section 4.5.

$E_i$ = Effective Temperature on day i defined as: $E_i = 0.5\,E_{i-1} + 0.5\,A_i$
where $A_i$ = Average temperature on day i

$g_i$ = Growth variable for day i. The day number counting from October 1st has been used in runs of the model to date.

$u_i$ = auto-correlated residual error term

$W_i$ = chill factor for day i as defined in para.5.3.4.

$\varepsilon_i$ = a standardized random normal variate.

$HOL_i$ = the demand that is "closed-down" due to holidays on day i

The parameters are the $C_j$ ($j = 1$ …..14) which are input to the program separately for each of up to 10 demand models covering the supply-year. The program allows for adjustments to the constant term ($C_1$) for each day of the week ($C_5$ for Monday, $C_6$ for Tuesday etc.) but in practice in fitting the demand model (an activity which precedes and is quite separate from the simulation model) only Friday, Saturday and Sunday have significant coefficients, and $C_5$, $C_6$, $C_7$, and $C_8$ have been input as zero. $C_{12}$ is the coefficient of autocorrelation of the residual error term estimated when fitting the model. If the model has been fitted using ordinary least squares $C_{12}$ is input as zero. The standard error of the demand model is input as $C_{13}$. If a wind term is used its coefficient is $C_4$; otherwise $C_4$ is set to zero.

The model requires daily values of SNET and the growth term for a complete year starting on October 1st. A daily value for the estimated amount of load closed down for holidays can also be input for each day of the year. Daily average temperature and windspeed data are required for the 59 year period.

## A11.4 The Simulation of Daily Demand

The program takes the daily values for the temperature (and windspeed) variables one year at a time and calculates a demand profile for a complete supply year using the appropriate demand model for each period of the year. The random error term included is based on a series of standard normal random variates. For each day in the year the antithetic demand is also calculated. The technique of antithetic variates has been found to reduce the variance of the mean of the simulated results (this is discussed in the next section of this Appendix). The basic idea is that if one simulation by chance gives rise to a daily demand profile whose mean level is above the true mean level, then a better estimate of the true mean level will be obtained by combining the results of this simulation with those from a simulation which has a mean below the true mean - the antithetic simulation. Such pairs of simulations are generated by using random variates in the second simulation that are perfectly negatively correlated with those used in the first.

In the program, if the direct demand on day i is given as:

$$D_i = f\,(T_i \text{ etc.}) + u_i$$

then the antithetic demand will be:

$$D'_i = f\,(T_i \text{ etc.}) - u_i$$

so that: $D'_i = D_i - 2u_i$

We refer to pairs of simulations conducted in this way as “antithetic pairs”.

The maximum daily demand and the accumulated demand (or volume) above a series of threshold demand levels are recorded for each simulation. The program uses 28 threshold levels, which are automatically calculated in such a way that the intervals between thresholds are smaller at high demand levels.

The procedure is followed for 14 antithetic pairs of random error streams to give up to 28 simulated demand profiles. The reason for 14 is that it is a multiple of the number of days in the week. For each year of daily temperature and windspeed values 7 weather series are generated by shifting the weather data by -3,-2,-1,-0,1,2, and 3 days. The reason for doing this is to eliminate the day of the week effect cause by October 1st falling on one day of the week rather than another in any year.

Two pairs of simulations are then performed with each of the 7 weather series for each year making 14 pairs of simulations in all. The 28 values of the maximum daily demand and the volume above each threshold are calculated and it is these values (one set for each of the 59 years) that are used in estimating peak day demands and load duration curves.

## A11.5 The Optimum Number of Simulations

The choice of the number of simulations to carry out is bound to be somewhat subjective, as the benefits of additional accuracy from a greater number of simulations must be weighed against the time and cost involved in doing them. A detailed description of the way in which this choice was made can be obtained from HQ O.R. Department.

It is clear that the larger the number, n say, of Monte-Carlo replications of a particular winter the smaller will be the variance associated with the average of the results. One way of judging what is an acceptable value for n is to compare for each winter in the database the volume above a particular threshold given by a deterministic simulation, $V_1$ say, with that given by the corresponding Monte-Carlo simulation, $V_2$ say. The larger n the smaller the variance of $V_2$.

In order to minimize the combined effects of the day of the week and random errors, while keeping the number of simulations as low as possible, it was decided that antithetic pairs should be used and the number of pairs should be a multiple of 7. With fourteen pairs of simulations on test data $V_2$ was always greater than $V_1$ and in all but three years (out of 51) $V_1$ was outside the 95% confidence limits for $V_2$. Consequently 14 antithetic pairs of simulations were judged to be sufficient.

## A11.6 Estimating Peak Day Demand

The "average" peak day demand is simply the arithmetic average of the maximum daily demand for each of the 59 simulated years.

1 in n peak day demands, particularly 1 in 20 and 1 in 50 values, are estimated by fitting a Gumbel-Jenkinson probability distribution to the ranked maximum daily demands, as described in Appendix 4, to each of the 28 simulation runs. The desired 1 in n level is the average of the resulting 1 in n values in each simulation run.

## A11.7 Estimating "average" and "1 in n" load duration curves

For each threshold level the "average" volume is calculated simply as the arithmetic average of the 59 simulated values.

In order to estimate the 1 in 50 volume above each threshold the program fits a cube root normal distribution to the ranked volumes after censoring those volumes less than the average volume. The parameters of the distribution are estimated by least squares regression. This is done separately for each simulation run of 59 years. The resulting parameters for 28 simulations are then averaged.

The mean and standard deviation at successive thresholds are smoothed using a three point moving average (allowing for unequal thresholds as described in Appendix 5 of the original TD76). The 1 in 50 values are then calculated from the smoothed parameters.

The integrated forms of the average and 1 in 50 load duration curves are differentiated. The top parts of the load duration curves are estimated following the procedure described in Appendix 6.

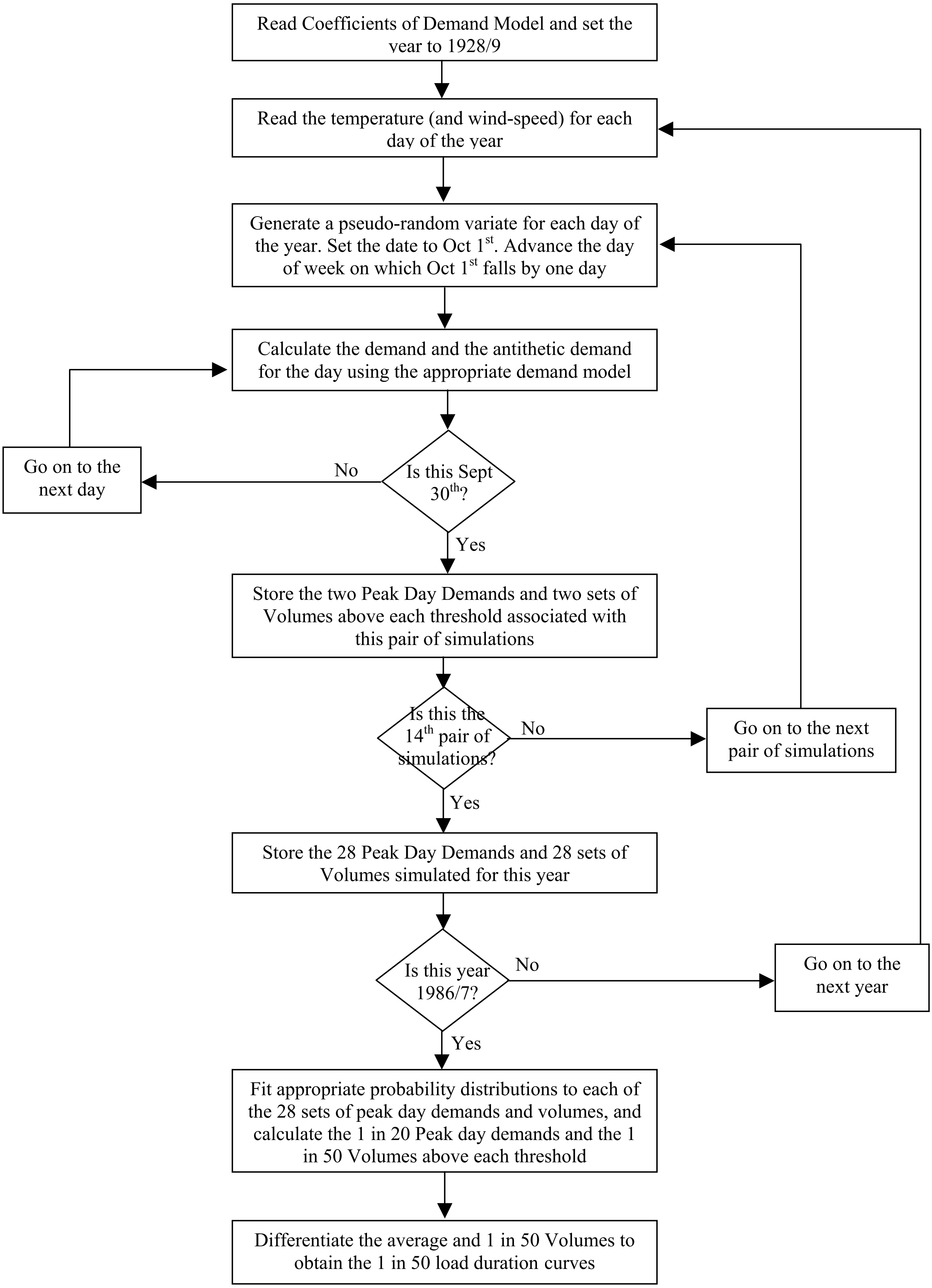
## A11.8 The output from the Simulation Program

The output from the program is as follows:

- The coefficients of the demand models used.

- The peak day demand for each year.
- The volume above each threshold for each year.
- The peak day statistics - mean, standard deviation, 1 in 20 and 1 in 50 values estimated from a Gumbel-Jenkinson distribution.
- The unsmoothed parameters of the cube-root normal distribution fitted to the sorted volumes at each threshold.
- The smoothed parameters for each threshold together with the average volume and sorted volumes above the threshold.
- The threshold at which the volume analysis ceases.
- The 1 in 50 and the average load duration curves day by day.

Simplified Illustration of the Operation of the Simulation Model - FIGURE A11.1



# APPENDIX 12 - THE OVERALL EFFECT OF THE TD76 CODE OF PRACTICE

## A12.1 Introduction

Appendix 12 of the original TD76 report considered the effect of all the changes introduced by the Steering Group in 1980. Most of the material in that Appendix is no longer relevant today and so is omitted here. One of the major changes, however, was the introduction of Monte-Carlo simulation as a replacement for the simpler approach of translating temperature duration curves. As temperature duration curves are still used for many broad-brush exercises it is felt worthwhile to repeat in this report the results of the research undertaken by the Steering Group on the effects of this particular change:

## A12.2 Monte-Carlo Simulation

Prior to 1980 most Regions used some form of temperature duration curve approach for deriving their load duration curves. Since 1981 all Regions have used Monte-Carlo simulation in order to produce a load duration curve that is consistent with the definition in section 3.3.

The effect of introducing simulation in a Region can be considered in two stages, that due to day-by-day simulation and that due to simulating the random error term on each.

There are many different variants of the temperature duration curve method. However, a reasonable approach where one has a model of the general form given in section 6, would be to arrange the days on the 1 in 50 temperature duration curve in the same order as the days in the SNET profile to provide a hypothetical (although quite artificial) 1 in 50 temperature profile. The demand model for any year can then be used with these temperatures for a full supply year assuming the day-of-the-week pattern associated with the year in question. The resulting demand profile can then be ranked to give a 1 in 50 load duration curve.

This method was used with a model fitted to West Midlands data for 1978/9 and the resulting load duration curve compared with that obtained from using the simulation model, described in Appendix 11, in a deterministic mode. The temperature duration curve approach was found to imply a peak shaving volume about 15% higher than the deterministic simulation, and at around 60 days the temperature duration curve approach continued to give slightly higher load duration curve, with a 2% greater volume, than the simulation model.

However, a load duration curve consistent with the definition in section 3.3.2 requires t h e day-by-day simulation of the random error in the demand model. On the test data, the load duration curve derived by Monte-Carlo simulation implied a peak shaving volume about 17% higher than that derived from a deterministic simulation. At the 60-day threshold there was less difference between the two curves, the Monte Carlo simulation implying a volume only about 2% larger than that derived from a corresponding deterministic simulation.

On balance, in this particular example, it appeared that the temperature duration curve approach produced a load duration curve quite similar to that produced by Monte-

Carlo simulation and that there would have been only about 2% difference in the resulting peak shaving volume. There is no theoretical reason why the results should have been so close but the effect would seem to depend on the particular duration curve method a Region might have been using.

# APPENDIX 13 - SUPPLEMENTARY TEMPERATURE / DEMAND INFORMATION REQUIRED BY H.Q.

## A13.1 Introduction

Apart from statements outlining gas sales forecasts in the ROP and Preliminary Marketing Statements, Regions are required to submit supplementary statements to Corporate Planning and to Production and Supply Division at H.Q. which are specifically concerned with temperature/demand relationships. The information required is defined in the "Notes for Guidance" issued with the statements each year.

This Appendix sets out in one place the information requirements of both Corporate Planning and Production and Supply Division. Its purpose is to provide clear definitions of all items of temperature/demand information required by H.Q. which are both internally consistent and consistent with the definitions elsewhere in this report.

This Appendix does not give detailed guidance on the precise format of the statements. This will continue to be given in the "Notes for Guidance" accompanying the statements each year.

## A13.2 The List of Statements

The following are the statements at present required by H.Q. Corporate Planning each year in June as a supplement to the Preliminary Marketing Forecasts, and in December as a supplement to the ROP. They are based on forecast levels of connected load consistent with the appropriate Marketing Assumptions:

| | |
|---|---|
| CPD1 (A) | Load duration data for the average year. |
| CPD1 (B) | Load duration data for the severe year. |
| CPD2 | Half yearly natural gas requirements in the average year. |
| CPD3 | Peak day demand for natural gas |
| CPD4 | Temperature/wind/demand model. |
| CPD5 | Average and severe year interruption capacity data. |
| CPD6 | Conservation in the domestic market. |
| CPD7 (A) | The average year temperature duration curve. |
| CPD7 (B) | The severe year temperature duration curve. |
| CPD8 | Peak day and degree-day temperature statistics. |

Production and Supply Division requires:

PS1 (A)    Weekly demand profile for a cold year.

PS1 (B)    Weekly demand profile for an average year.

PS2    Storage availability and requirements in the severe year.

PS3    Daily demand and diurnal variation by offtake for selected days.

In total the PS statements are only required once a year but at different times for the individual statements.

A13.3    CPDl (A) and CPDl (B)

CPDl (A) and CPDl (B) are the statements on which Regions return their load duration data to H.Q. for each ROP.
Data is required for several categories of sendout for average and severe (currently 1 in 50) conditions for six supply years ahead and for one past year (for the 1988 ROP the years required will be 1986/7 to 1992/3 inclusive).
The load duration curve is described in section 3.3 and Appendix 1, and Regions should ensure that their load duration data on statements CPDl (A) and CPDl (B) is consistent with the definition given in para.3.3.2.

Load duration data is required for the following categories of sendout:

i) Domestic

ii) Commercial Firm (Tariff & Contract)

iii) Industrial Firm (Tariff & Contract)

iv) Net Supplies to other Regions

v) Minimum Essential Interruptible and Scheduled

vi) Total Firm plus Minimum Essential Interruptible and Scheduled

vii) Maximum Potential Interruptible and Scheduled

viii) Cold Weather Upturn

The load duration curve is defined as a histogram for each of 365 days.
Regions should also provide the total volume under the load duration curve and identify the total amount of unaccounted for included in this.

In addition to day 1 on the average curve, Regions should provide a breakdown by demand categories of the average peak day demand on CPD1 (A) and, in addition to day 1 on the 1 in 50 curve, the 1 in 20 peak day demand should be given on CPD1 (B).

A13.4 Load Duration Data for Firm Sendout

For Supply/Demand matching at H.Q. the most important load duration curve is that for total firm plus minimum essential interruptible and scheduled (vi). It is expected that Regions will use a simulation model as described in Appendix 11 to derive a total firm load duration curve (which is not returned to H.Q.) and then break this down between the three firm market sectors (perhaps by using the individual sector models derived following the guidelines in section 7).

The load duration curves for each market sector should include the unaccounted-for gas associated with that market sector. In particular, contract load should only include that element of unaccounted-for that is directly attributable to these sales i.e. meter error and leakage from service mains but not leakage from the distribution system. The balance of unaccounted-for gas should then be attributed to tariff loads and this will tend to mean the balance of unaccounted-for being included in the Domestic load duration curve.

Regions should identify the net supplies to other Regions (iv). Supplies to other Regions should be positive and supplies from other Regions negative. Daily direct supplies of natural gas should be shown; and gas from other sources should not be shown.

A13.5 Load Duration Data for interruptible Sendout

The load duration curve for minimum essential interruptible and scheduled sendout (v) should relate only to that demand that has to be met after completely exhausting the maximum interruption periods in the contracts, making an appropriate assumption about the effectiveness of interruption. Sellers' Option loads should not be included. In contrast the load duration curve for maximum potential sendout to interruptible and scheduled customers (vii) should be derived on the basis that interruptible customers are not interrupted at all and that scheduled customers are interrupted for their scheduled period. Again Sellers' Option loads should be excluded. The difference between curves (v) and (vii) should therefore be the maximum interruption capacity on each day of the load duration curve assuming full effective interruption.

A13.6 The Cold Weather Upturn

As discussed in para.6.1.4 Regions may include a measure of cold weather upturn in their demand model to take account of changes in consumer behavior in extreme conditions, such as the cancellation of conservation or the leaving on all night of central heating. Any cold weather upturn must be included in the fitted model so that there will be no double counting of this effect.

If a Region includes a cold weather upturn of this sort in deriving its load duration data it should identify the amount of demand attributable to it in (viii) and be able to provide H.Q. with supporting details of the calculation. This should be the difference between the result using a linear model and that using a model with a CWU term (which may be non-linear). Ideally this would entail fitting models both with and without the CWU term and simulating both to arrive at a difference. An acceptable approximation would be the difference between the simulation result using a model fitted with a CWU term, and the simulation result using the same fitted model but replacing the CWU term by zero.

A13.7 CPD2

Regions' half-yearly natural gas requirements for natural gas in the average year should be shown on statement CPD2 for the six years of the ROP and for one past year. The breakdown by market sector is similar to that for the load duration statements and Regions should ensure that CPD2 is consistent with CPDl (A) in each year.

A13.8 CPD3

1 in 20 peak day demand for various categories of demand should be shown on statement CPD3, for the six years of the ROP (and should be consistent with CPDl (B)). A revised estimate for the most recent winter should also be shown. The peak day demand is defined in section 3.2. The simulation program described in Appendix 11 provides an estimate of the 1 in 20 peak day demand that is consistent with the load duration curve.

A13.9 CPD4

Regions are asked to provide, on Statement CPD4, a description of the temperature/wind demand model used in the preparation of the appropriate set of demand forecasts. Coefficients are requested for the first year ahead being forecast i.e. 1987/8 for the 1988 ROP.

A13.10 CPD5

Information is required, on CPD5, on the contractual and the effective annual volume of interruption capacity available in average and severe years in each of the six years of the ROP. The effective capacity in a severe year should be the difference between the annual totals for fields (v) and (vii) on Statement CPDl (B). Regions should also provide estimates of the expected volume of interruption in an average year.

A13.11 CPD6

Information is required on CPD6 on the reduction in annual domestic demand because of the effects of “good housekeeping” and “permanent measures” conservation. Information is also required on the peak equivalent of the annual reduction and on whether the peak equivalent is cancelled wholly or in part.

A13.12 CPD7 (A), CPD7 (B) and CPD8

On the CPD7 statements, Regions should supply average and severe year temperature duration curves based on their current database. For the 1988 ROP these will be based on the 59 year series 1928/9 – 1986/7, but this will change each year as stated in paragraph 4.4.1.

On CPD8 Regions should supply details of minimum effective temperatures and the cumulative numbers of degree-days below a threshold of $0^{o}$ C. The average and severe (1 in 50) temperature duration curves will change each year in line with the decision to increase the size of database by one year each Autumn. A seasonal normal temperature duration curve should also be supplied on each occasion when seasonal normal temperatures are revised. It is currently planned that they should be revised every five years starting in the Autumn of 1988, as stated in paragraph 4.5.1.

A13.13 PS1 (A) and PS1 (B)

These statements are required by Central Control for planning plant availability throughout the year, and are only required for the first year of the ROP (i.e. for the 1988 ROP this would be the supply year 1987/8). A demand profile for a cold year is required on PS1 (A) and for an average year on PS1 (B). PS1 (A) should assume no interruption while PS1 (B) should assume planned interruption.

The cold weather demand in each of the four week periods in the supply year is defined to be the daily demand that, in a long series of years, with connected load held at the level appropriate to the year in question, would be exceeded in that period in only one year out of twenty, each year being counted only once. This definition was specified in a memorandum from Central Control to Regional Grid Controllers dated 7th April 1982. The average (or seasonal normal) demand in each week of the supply year is defined to be the average daily demand occurring in that week over a long run of years, with connected load held at the appropriate level for that year.

As these definitions take account of variations other than temperature the procedure for calculating the profiles cannot be based simply on calculating demand at particular temperatures. It will therefore be desirable for Regions to use Monte Carlo simulation for calculating the weekly profiles as well as the load duration curves required by Corporate Planning. For average demand, the simulation model described in Appendix 11 can be used to generate the weekly profiles in the following way. For each week of the supply year the model will generate an average daily demand corresponding to each of the 59-year temperature and windspeed data. The average demand for each week is calculated as the average of the 59 demands for that week.

The simulation procedure described in Appendix 11 can also be used, with minor modification, to provide 1 in 20 four-weekly demands. The computer program can be made to store, in addition to the maximum daily demand in each year, the maximum demand in each of the twelve four-week periods. The required 1 in 20 values can then be estimated in precisely the same way as is done for the peak day demand. This can be done for both firm only and firm plus interruptible demand.

A13.14 PS2 and PS3

These Statements are required for the last six years of he ROP by Engineering Planning department at H.Q. Regions are asked to provide information that will assist H.Q. in both designing and planning the operation of the National Transmission System to meet demand in both average and severe conditions.

On statement PS2 Regions should give details of the storage available and required for selected days on the severe load duration curve as given on statement CPDl (B).

On statement PS3 Regions should provide a breakdown by offtake of the demand on certain specified days. These days are selected from those quoted on other statements, in particular CPDl (B), CPD3 and PS1 (A) and (B).