Measurement of Domestic Hot Water Consumption in Dwellings

This report has been prepared by the Energy Monitoring Company in conjunction with and on behalf of the Energy Saving Trust with funding and support of the Sustainable Energy Policy Division of the Department for Environment, Food and Rural Affairs (Defra)

EXECUTIVE SUMMARY

This report describes the analysis of data on hot water consumption collected in approximately 120 houses. The project had four key goals: to measure volumetric consumption of DHW and the associated energy requirement; to identify DHW heating patterns in terms of both times and temperatures; to compare these results with the current BREDEM assumptions and finally to identify where in the dwelling water is being consumed. These goals have informed both the data collection and the analysis subsequently carried out on that data.

Hot water consumption, delivery temperature and incoming cold feed temperature were measured in all dwellings. In those with a system boiler an additional measurement was made of the primary pipework temperature, to enable the times of day at which water was heated to be identified. In a limited number of properties additional temperature measurements were made at each hot water outlet, allowing the destination of each run-off to be determined. Normally, data was accumulated at ten minute intervals. When a run-off was detected this was changed to five seconds for the duration of the run-off. This allowed accurate determination of the energy associated with the run-off and, where the additional information was available, the destination of that run-off.

The original sample consisted of 124 dwellings. Inevitably, some of these produced data which was not suitable for analysis. Initially data from five of these proved unusable for a variety of practical reasons. A further seven cases were dropped during the analysis process leaving an intial sample of 112. When the data was divided to examine the effect of boiler type a further five cases had no boiler at all or used a multipoint system, leaving 69 regular boilers and 38 combis. 23 dwellings featured the additional instrumentation described above, and of these 21 proved usable, consisting of 13 regular boilers and 8 combis.

Determining domestic hot water volume and energy consumption:

The mean household consumption has been found to be 122 litres/day, with a 95% confidence interval of ±18 litres/day. Statistical analysis of the flow data from each dwelling has considered the impact of geographical region, boiler type, number of occupants, and the number of those occupants who are children. It has revealed that the only one of these factors influencing consumption is the number of occupants. The mean energy content has been found to be 16.8 ± 2.2 MJ/day. Energy content of water delivered has been subjected to the same statistical analysis as the flow data, and has also been found to depend only on number of occupants.

Identifying DHW heating patterns in terms of time and temperature:

Across the whole sample delivery temperature has been found to be significantly below the widely assumed value of 60°C, with a mean value of 51.9°C estimated with a 95% confidence interval of ±1.3°C. The mean temperature among dwellings fitted with regular boilers is 52.9 ± 1.5°C. In houses with combi boilers it is 49.5 ± 2.0°C. This difference is highly statistically significant, and it is concluded that combi boiler owners routinely experience lower hot water delivery temperatures than regular boiler users.

In the case of regular boilers, data from the additional temperature measurement on the primary circuit has been used to determine the duration and times of hot water generation. This has allowed an estimate to be made of the times of day at which householders heat their water. The heating time has a mean of 2.60 hours/day, estimated with a 95% confidence interval of ± 0.35 hours/day. The overall schedule reveals that some households heat water as and when it is required, and the remainder generally heat between 8:00 and 10:00am, and again between 6:00 and 11:00pm.

Comparison between observed results and the current BREDEM assumptions:

Comparing the measured flow data with BREDEM reveals that the current model of consumption (based on the number of occupants in a dwelling) is appropriate. However, analysis of the average temperature rise of water as it passes through the heating system (derived from the initial cold feed temperature and the hot water delivery temperature) shows a value of 36.7°C, significantly lower value than the 50°C currently assumed in BREDEM. On this sample of dwellings BREDEM would over-predict energy consumption by approximately 35%. The current temperature difference is based on an assumed cold water inlet at 10°C and hot water delivery at 60°C. The discrepancy with the measured result is due partly to hot water temperatures lower than assumed, and partly to cold feed temperatures higher than assumed. Statistical analysis has demonstrated that cold water inlet temperature could be better estimated using a model that takes into account occupancy, boiler type and region. Hot water delivery temperature prediction could also be improved by using a simple model based only on boiler type.

Identifying where in a dwelling hot water is being consumed:

In dwellings where it was possible to monitor every hot water delivery point the measurement scheme has proved capable of identifying the destination location of almost all the hot water used. In other dwellings, where it was impractical to instrument one point the unallocated flow can therefore be attributed to that unmetered point with confidence. The relatively small number of dwellings which were equipped with this additional instrumentation means that it impossible to draw firm statistical conclusions. However, it can be seen that for locations such as the bathroom basin, bath and washing machine the difference in volume used is similar between regular and combi boilers. For the kitchen sink, however, the volume of water used by combi owners is significantly larger than in the dwellings with regular boilers. The most likely explanation for this is that users demand a higher temperature at the kitchen sink, this is harder to achieve with a combi boiler, and therefore more water is run off. At the other locations temperature is perhaps less critical, and combi systems can achieve this without the need for an extended run-off.

Although the current sample of houses with the additional instrumentation required to determine the destination of hot water run-offs is small, it has served to prove that the measurement technique is effective. To obtain more robust information, however, it would have to be carried out in a substantially larger number of dwellings.

CONTENTS

1 Introduction

This report describes the analysis of data gathered to understand the use of hot water in domestic dwellings.

The goal of this document is twofold. It describes the cleaning, checking and analysis of the data gathered, and presents the results obtained. Equally importantly it contains the full 'audit trail' for that analysis, providing enough information for a third party to check the analysis, or even to recreate it in its entirety.

After a brief review of the project goals and the data collection scheme the process of cleaning the data prior to analysis is described in detail. Subsequent analysis begins by characterising volumetric hot water consumption, both as a daily average quantity and as an hourly profile. The energy delivered as hot water is also explored. Next, hot water heating patterns, both in terms of hot water delivery temperatures and, where applicable, times of heating are examined. Measured consumption and temperatures are compared with the assumptions currently built into BREDEM.

Finally the analysis moves to a subset of the sample which was equipped with additional instrumentation to allow the location of hot water use to be identified.

2 Key project goals

At the beginning of the project, four key goals were identified. These informed the physical monitoring strategy, and have subsequently determined the analysis carried out on that data:

1. Measure DHW consumption and energy delivered;
2. identify DHW heating patterns – time and temperature;
3. Compare the observed results with the current BREDEM assumptions;
4. Identify where in the dwelling water is being consumed.

3 Outline of data collection scheme

In all cases volumetric water consumption was measured to a resolution of 0.1 litre. In order to establish the energy requirement associated with this delivery the cold inlet temperature (either to the hot water storage tank or to the combi boiler) was measured, together with the delivery temperature.

Figure 3.1 shows a typical installation on a regular boiler, and Figure 3.2 shows how Combi boilers were monitored.

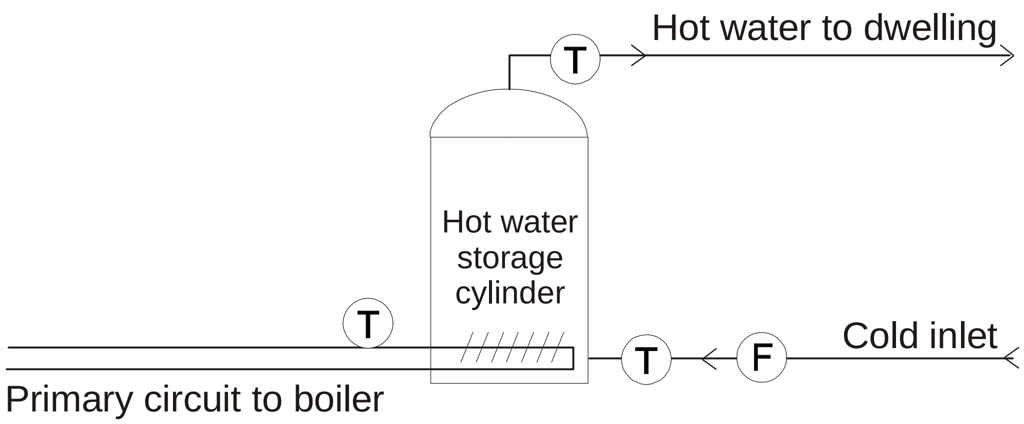

Figure 3.1: Monitoring configuration for regular boiler

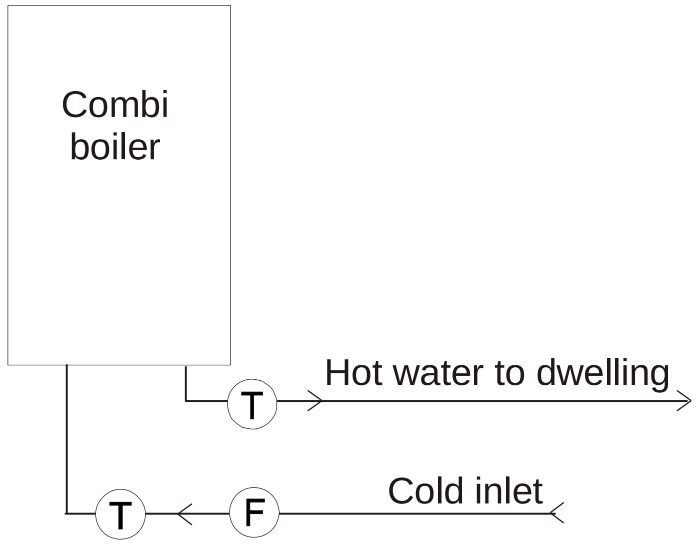

Figure 3.2: Monitoring configuration for a combi boiler

To determine the point within the dwelling where water was consumed a sub-sample of dwellings was equipped with additional instrumentation, which used a simple strap-on sensor to measure the temperature of the hot water pipe to each location or appliance. By determining the temperature rise seen at each location each time a run off was observed the likely destination of that run off can be determined, and both water and energy consumption allocated to that particular appliance. Figure 3.3 shows a typical installation.

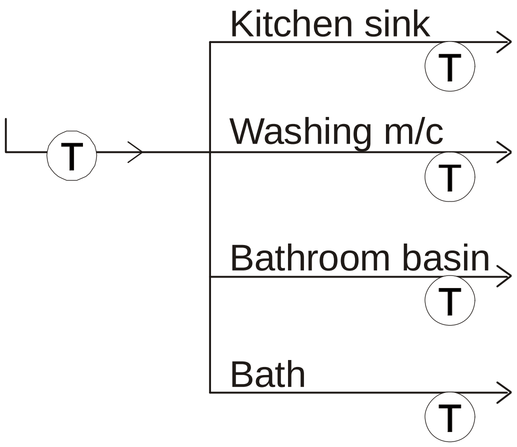

Figure 3.3: Example layout of additional instrumentation

The temperature sensors used for this location monitoring were of poorer accuracy than those used for the main feed and delivery temperature measurements. They were mounted simply by strapping them to the outside of the pipe, rather than being immersed in the flow. Data from them should therefore be considered only indicative. In particular they do not allow the estimation of pipe losses by comparison with the temperature of the water leaving the regular boiler hot water cylinder or the combi outlet.

The data collection equipment was configured to measure all the sensors and make a record of them every ten minutes. As soon as a run off was detected the sampling rate was increased to once every five seconds. This higher speed data collection lasted until the run off ceased.

The original sample consisted of 124 dwellings. Inevitably, some of these produced data which was not suitable for analysis. Table 3.1 summarises the size of sample used at each stage of the analysis. The data from five of these proved unusable for a variety of reasons, leaving a sample of 119 homes. A further seven cases were dropped during the analysis process.

	Size of sample	
Initial sample	124	
Data rejected at preliminary cleaning stage: 5 cases	119	
Data rejected after preliminary inspection: 7 cases	112	
Data rejected due to boiler types neither regular nor combi: 5 cases	Regular boilers	Combi boilers
	68	39

Table 3.1: Successive reduction in sample size during data inspection and analysis

A total of 23 dwellings were equipped with the additional instrumentation which allowed the destination of each hot water draw off to be identified. Two of these datasets proved unusable, leaving 21 cases for analysis. Of these, 13 were equipped with regular boilers and 8 with combis.

Details of the other sample characteristics have already been documented elsewhere.

4 Outline of the data analysis process

The software developed for analysis operates in two sections:

- file amalgamation: the multiple data files supplied by the monitoring contractor are joined together to produce a single data file for each dwelling. All subsequent analysis is then carried out from these 'master' files. Spurious blank header lines are removed as part of this process. The start time of each file is checked against the finish time of the preceeding file, and if data has appeared twice this is corrected. This process and other data cleaning measures are described in detail in Appendix 1.

- data reduction: subsequent analysis programs work from the master files described above. Although the data amalgamation process has removed some errors there may still be others, such as out of range readings. The analysis software carries out a series of further checks on each data record before using it. Depending on the information required from a particular analysis this process falls into one of two distinct categories:

    - record at a time calculations: these include accumulating total dwelling volume and energy consumption, and determining heating regular operation schedule;

    - block calculations: a new block of data is generated each time a run-off begins. This allows calculations to be made which involve many data records simultaneously. A block includes the zero flow readings taken immediately after a run off. This is necessary when allocating where water has gone to in dwellings when we need to find the maximum increase in outstation temperature is required, and due to thermal inertia effects the temperature does not reach its maximum value until after the run off has stopped. When compiling profiles account must be taken of the fact that a block may cross a profile time boundary.

The analysis described in this report began during the first month of data collection. This allowed the bulk of the necessary software to be developed in advance, and also allowed a number of minor problems to be identified, which were subsequently resolved by the data collection team.

5 Data cleaning

At the end of the data collection phase of the project a DVD containing all of the data was submitted. This DVD is titled 'DHW data (final set) 4/9/07'. The data manipulation described in the remainder of this report is all based on this DVD.

5.1 Preliminary data cleaning

During data collection most of the folders which contained data had been renamed to include a reference to the householder's name. The first operation was therefore to rename all folders so that their name consisted only of the dwelling serial number. This allows the data analysis software to find each folder, and also provides householder anonymity.

The files to which data are written on the data collection PC have a name derived from the date of the download, and the serial number of the data logger used in each installation. This approach produced two problems:

- in two cases loggers were manufactured with duplicate serial numbers. The data from each logger therefore had to be kept separate, and renamed when data collection was complete;

- in one case a logger failed late in the project, and was replaced with a unit taken from a dwelling where data collection had finished. In order to maintain consistent file naming the data taken at the second dwelling had to be renamed.

A number of flow meters developed faults which resulted in occasional spurious pulse generation. All data was passed through a cleaning program by the monitoring contractor before being submitted for analysis. A minor problem with that program was that it sometimes inserted extra blank lines in the file header. Initially these were edited out by hand, but when it was realised the occurred quite a large number of times (approximately 75) the analysis software was modified to remove these lines automatically.

Finally, there were five dwellings which had suffered from problems throughout the data collection period and, after discussion with the data collection team these five cases were removed from the database.

Whenever a dwelling (or, in subsequent sections, a single file) is described as being removed from the dataset the data is not actually deleted. Instead removal is accomplished by renaming the file so that it no longer conforms with the naming system described in Appendix 1. In this way the data is ignored by the data analysis software, but can still be examined if required.

Some files contained extra channels which had been switched on in error when the data logger was installed. In all cases the error was detected by the analysis software, which checks that each record in each datafile contains the appropriate number of fields. In some dwellings the problem was subsequently fixed by turning the channel back off. In these cases the relevant files have been edited to remove the additional information. In other cases the additional channels remained in place throughout the data collection period, and simple null entry made in the file that contains the channel definitions for each dwelling. Again, these adjustments are all described in detail in Appendix 1.

When monthly consumptions were tabulated, a number of properties showed anomalous behaviour at either the start or finish of the data collection period. In some cases this may have been due to instrumentation problems either being fixed at the start of data acquisition, and in other cases it may be that faults developed so near the end of data acquisition that they were not fixed. Appendix 1 again contains full details of this process.

5.2 Timing checks

The program which builds the master file checks the time only at the start and finish of each data file, in order to ensure that, if two files overlap, there is no duplication of data. At the analysis stage the time of each record is checked. This process serves two purposes. If there were occasions inside the file when data was duplicated it would be identified (in practice there have turned out to be no such occasions). Finally, if there are periods of downtime when data was not gathered this too can be identified.

5.3 Channel checks

In a number of cases the cold and hot water temperatures were swapped over when they were installed. The analysis software holds a list of the dwellings in which this occurred, and the two channels are swapped back before analysis starts. Appendix 1 lists the dwellings which have this problem.

The next step in the checking process examines each record to see if it is in range. If a temperature is outside the range 0 to 80°C or a flow is outside the range 0 to 100 litres/minute then the record is considered invalid and is not included in the data analysis procedure. In the analysis of flow data only the hot and cold channels are checked. In the additional analysis of flow destinations all remote temperatures are checked. This distinction is important, as it allows a record with errors on a location sensor still to be used in the analysis of the total hot water flow and energy, even though it cannot be used for analysing the destination of the flow.

5.4 Flow checks

In some dwellings flow meters became unreliable towards the end of the data collection period. This unreliability manifested itself in the production of large numbers of spurious pulses. These cases were isolated by generating a monthly consumption figure over the whole monitoring period. Where this indicated that consumption had increased sharply the raw data was inspected to see whether there were still outlet temperature rises corresponding to each run off. In cases where there were not the dataset was truncated appropriately before use.

With all of these data cleaning measures in place, approximately 30 million run-off records were left for analysis.

6 Analysis of flow data

This section describes the analysis of the water volume and the associated flow, return and primary temperature data.

6.1 Hot water consumption

For each dwelling the total hot water consumption has been determined. In order to facilitate comparisons between dwellings monitored over different periods the flows are normalised to a per day basis. As shown on Figure 6.1 there is wide variation across the whole sample.

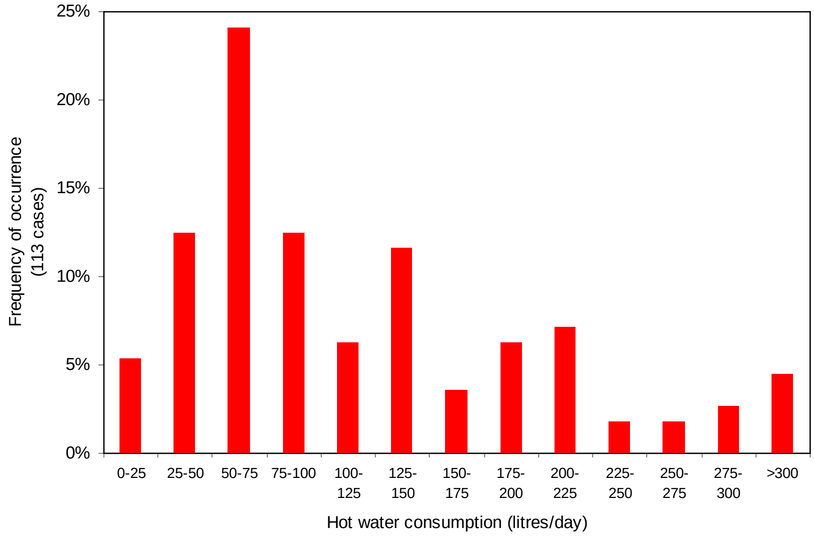

Figure 6.1: Distribution of daily hot water consumptions

The estimated mean consumption is 122 litres/day with a 95% confidence interval of ±18 litres/day. It is clearly of interest to determine which factors have a significant influence on consumption. As an example, Figure 6.2 shows a breakdown by boiler type.

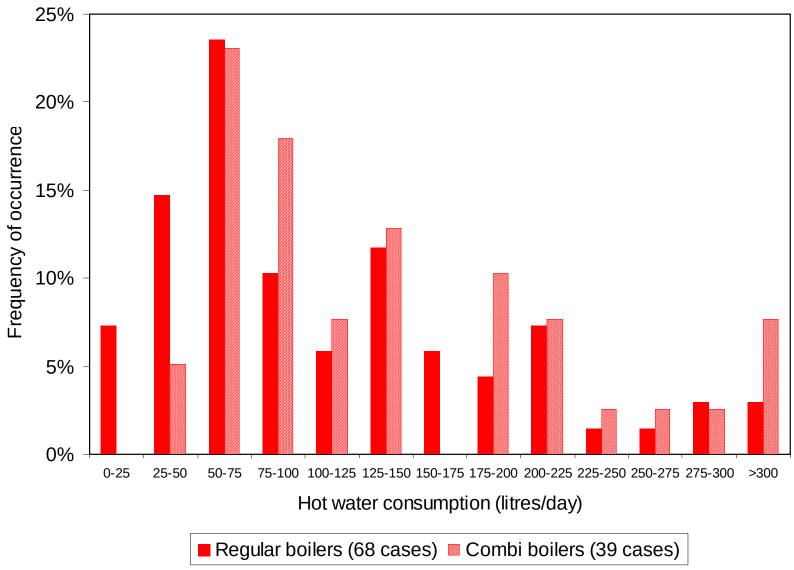

Figure 6.2: Hot water use by boiler type

The mean consumption in dwellings with a regular boiler is 116 ± 24 litres per day. In dwellings with a combi it is 142 ± 28 litres/day.

As well as boiler type (regular or combi), the geographical region, number of occupants and split between adults and children have been recorded, and may be of value when characterising water consumption. The resulting dataset therefore contains a mixture of continuous and categorical factors. To establish which of these factors have significant influence an Analysis of Covariance (ANCOVA) has been carried out. Initially the observed flow is modelled in terms of all the factors, and these are progressively eliminated until only those with significant impact remain. This process is described in detail in Appendix 2, and Table 6.1 below summarises the results. The probabilities in the table indicate how likely it is that removing a factor does not decrease the quality of the model of water consumption. Thus probabilities greater than 5% indicate that a factor can safely be dropped, and values below 5% indicate that the factor should be retained.

Factor	*Probability*
Number of children	67%
Region	20%
Boiler type	12%
Number of occupants	0%

Table 6.1: Relative importance of factors influencing volumetric consumption

The table demonstrates that number of children, region and boiler cannot be considered to be significant factors when trying to predict hot water consumption. In an ANCOVA analysis the order in which the factors are entered can make a significant difference to the conclusions if there are correlations between factors. Repeating the run above in reverse order shows that region and boiler type still have little effect, but that the number of children could be used in place of the total number of occupants to predict daily hot water use. However, the quality of the estimate is higher when only the number of occupants is used, and returning number of children to the analysis after this factor has been included provides no significant improvement.

The residuals which remain after the ANCOVA process are themselves skewed towards the left (like the original flow data shown on Figure 6.1). They are therefore not normally distributed, a requirement for the ANCOVA to be fully effective. The significances shown in the table are relatively robust to this, but to confirm which factors are important it is possible to use an alternative, weaker, analysis which does not make this assumption. The appropriate statistical test to determine whether the individual dependencies are significant is the non-parametric Kruskal-Wallis rank-sum test and the results, shown in Appendix 2, broadly confirm the conclusions from the parametric analysis. Interestingly the non-parametric tests imply that the influence of boiler type may be significant.

As described above, the most useful factor to use to predict hot water consumption is the number of occupants in the dwelling. Figure 6.3 shows the variation in consumption as a function of occupancy. The figure also shows two straight line fits to the data. The first is the fit to the whole dataset. The second was produced after eliminating points from the three dwellings featuring six or more occupants, which could potentially be exercising too much influence on the fitted line.

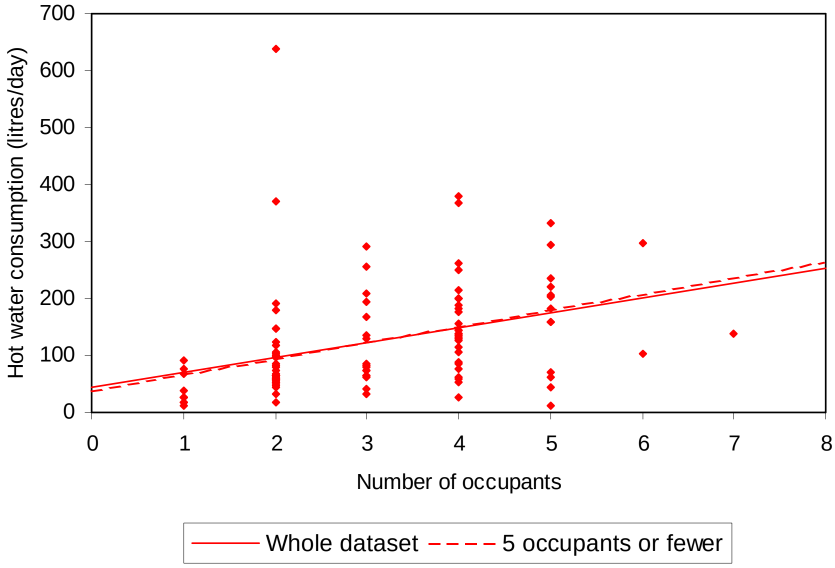

Figure 6.3: Consumption as a function of occupancy

Table 6.2 summarises the coefficients obtained from the two line fits, and their standard errors. It also shows the probability that the value of each coefficient might actually be zero. As discussed above these data are only approximately normal, so the confidence intervals should be considered approximate.

	Whole dataset		5 occupants or fewer	
	Value	p-value	Value	p-value
Intercept (litres/day)	46 ± 22	4.4%	40 ± 24	9.7%
Slope (litres/person.day)	26 ± 7	0.0%	28 ± 7	0.0%
Consumption model	46 + 26 N		40 + 28 N	

Table 6.2: Regression model of effect of number of occupants on consumption

The standard errors in the table indicate that neither the slopes or intercepts of both lines are significantly different, confirming the impression given by Figure 6.3. Further analysis can therefore continue with the data from dwellings with 6 or 7 occupants included.

Because data is collected at high speed when a run off is registered, it is possible to generate profiles showing the times of day at which water is most commonly used. Figure 6.4 shows a typical profile for a representative dwelling.

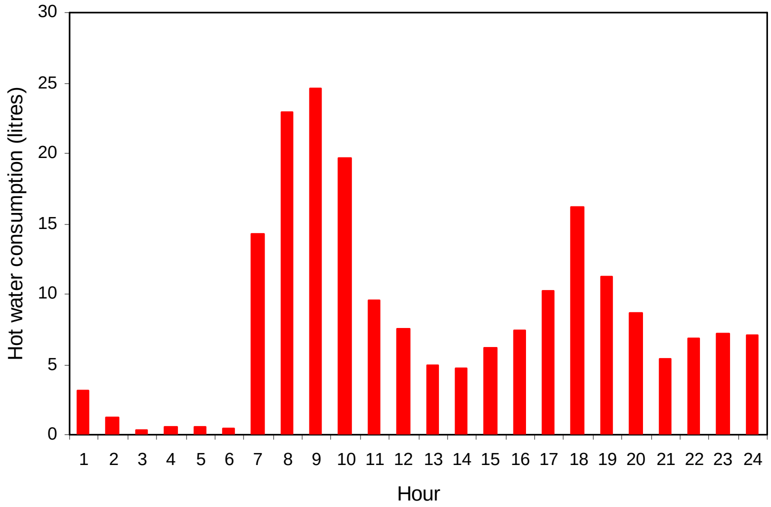

Figure 6.4: Daily run of profile of one dwelling

As with the total run off volume, there are wide variations in run off profile between dwellings, and the result of most interest is the average run off of the whole sample. Figure 6.5 shows the mean run off across the whole sample.

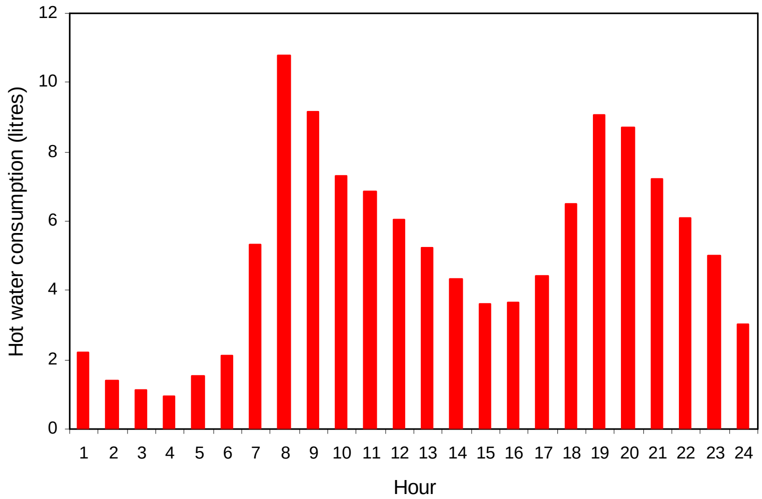

Figure 6.5: Daily run off profile of whole sample

It is also possible to divide the profiles between different boiler types. Figure 6.6 shows the result.

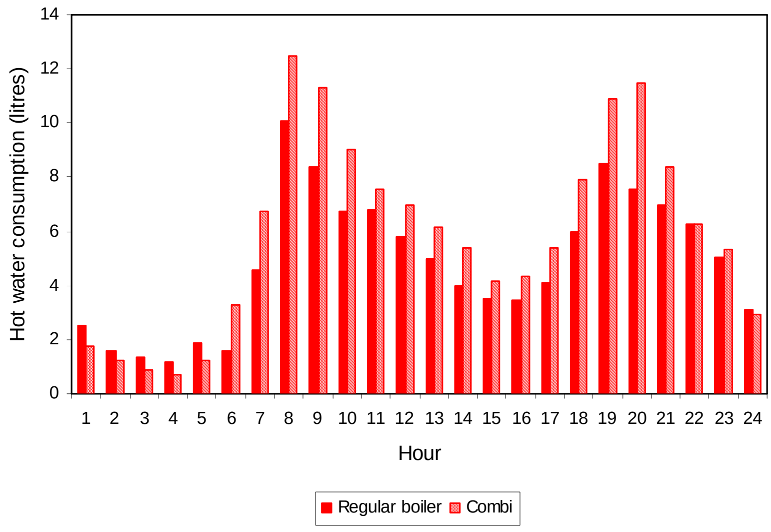

Figure 6.6: Daily run off profiles for regular boilers and combis

The figure reveals that the profiles from regular boilers and combis are very similar in shape, although as noted before flows are in general slightly higher in the case of combi boilers.

6.2 Energy consumption

The cold inlet and hot water delivery temperature measurements can be used in conjunction with the flow data to determine the energy content of the hot water used. Across the whole sample this has a mean value 16.8 $\pm$ 2.2 MJ/day. Figure 6.7 shows the distribution of energy consumption in the same sample dwelling used to generate Figure 6.4.

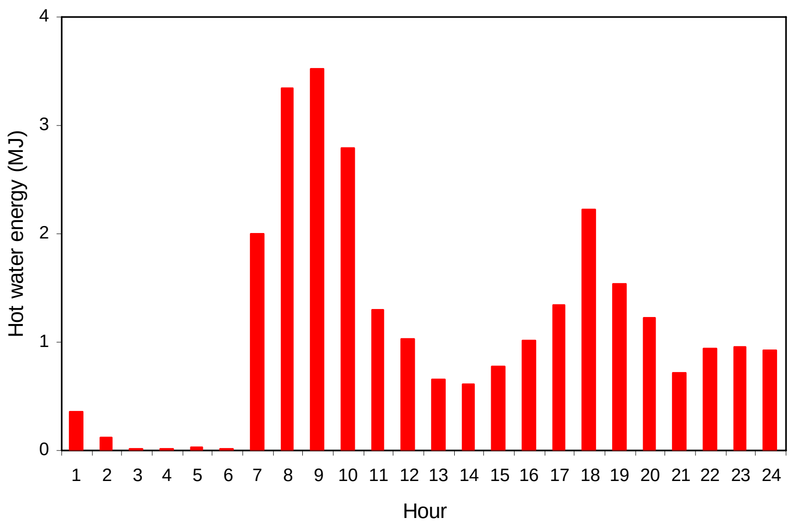

Figure 6.7: Distribution of energy delivered to hot water in one dwelling

As expected flow and energy consumption are closely related. If the difference between inlet and outlet temperatures was the same for each system they would be perfectly correlated. Figure 6.8 shows the relationship, and the scatter seen is due primarily to departures from this behaviour.

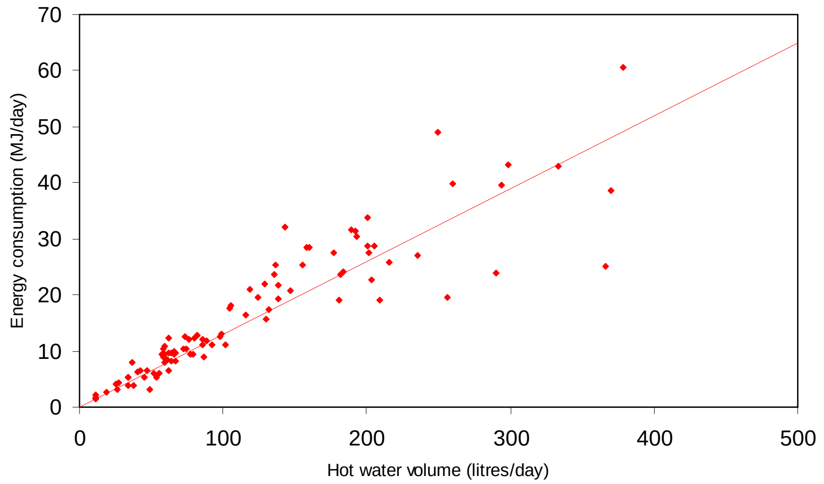

Figure 6.8: Relationship between volumetric and energy consumptions

This is confirmed by examining the energy consumption profile averaged over the whole sample, which is very similar to the whole-sample volume run off profile shown on Figure 6.5.

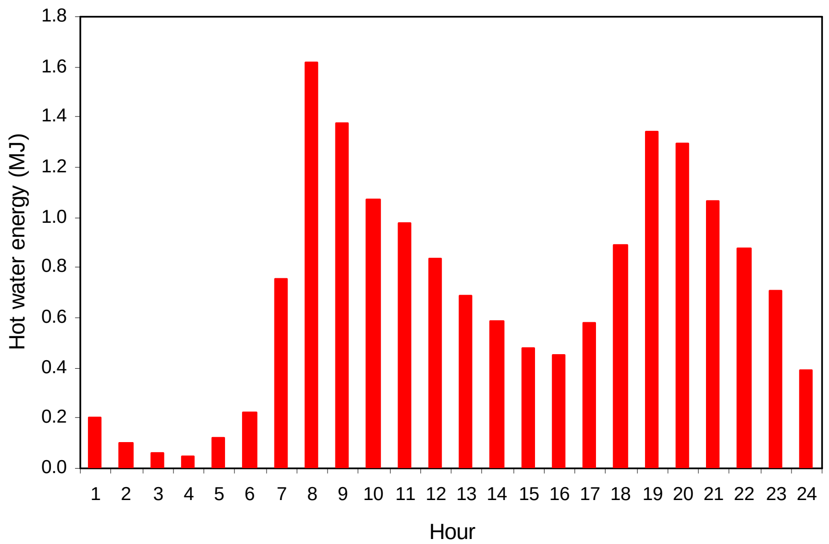

Figure 6.9: Energy delivery profile for whole sample

CEN Mandate 324 specifies a series of hot water run-off profiles, or tapping cycles, for testing and labelling hot water appliances. These are presented as a list of run-offs over a 24 hour measurement cycle. Appendix 3 shows one way in which such a tapping cycle can be derived from the data gathered in this project.

One of the key assumptions underlying the development of tapping cycles is the number of hot water draw-offs each day. Figure 6.10 shows the distribution of number of run-offs across the whole sample.

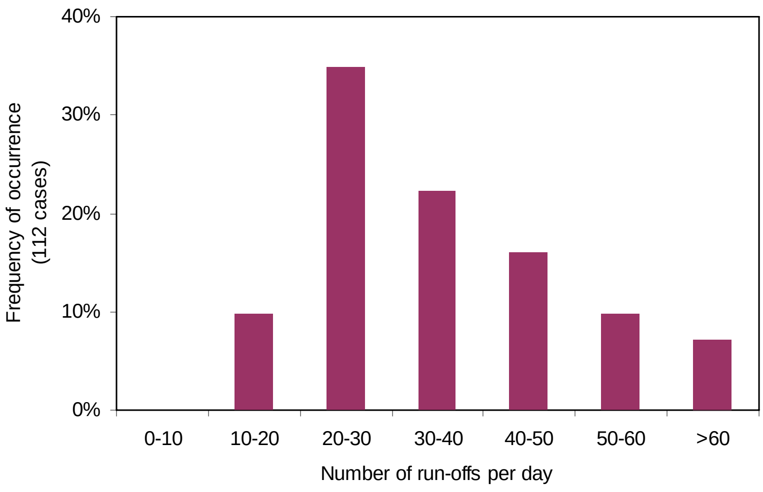

Figure 6.10: Distribution of number of run-offs per day

The data shown has an average of 28 run-offs/day, estimated with a 95% confidence interval of ±4. Once again, the data can be broken down by boiler type, and Figure 6.11 shows the result.

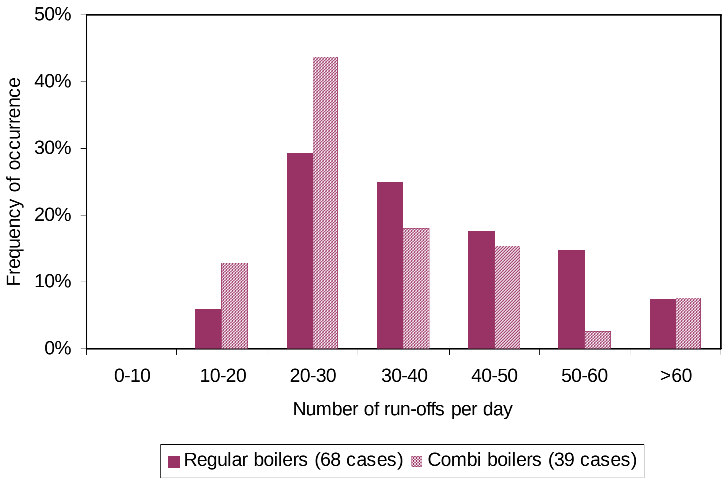

Figure 6.11: Number of run-offs per day broken down by boiler type

6.3 Identifying DHW heating patterns

The second goal of the project was to examine hot water heating patterns. This is done in terms of the delivery temperatures that occupants experience and, in the case of regular boilers, the times of day at which water is heated.

6.3.1 Hot water delivery temperatures

In general, the sensor measuring hot water delivery temperature is mounted some way from the hot water tank or combi boiler. When a run off occurs, it is necessary to allow a short period for the 'dead leg' of cooled water to be flushed from the connecting pipework before reading the temperature of the sensor. The sensitivity of the analysis to the magnitude of this delay has been analysed, and the results are summarised in Appendix 4. The most appropriate delay was found to be 35 seconds, and this is the value used throughout the analysis described here.

As with the volume of hot water delivered, there is a large variation in the temperature at which that water is used. The mean delivery temperature across the whole sample is 51.9°C, estimated with a 95% confidence interval of ±1.3°C. Figure 6.12 shows the distribution of delivery temperatures, each calculated as described above.

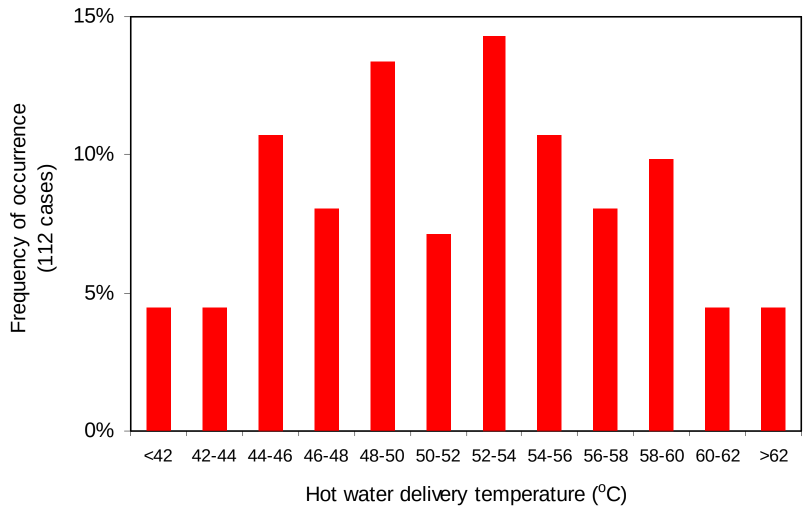

Figure 6.12: Distribution of hot water delivery temperatures

As expected from the figure, both the parametric t-test and the non-parametric Wilcoxon signed rank test indicate that the mean hot water temperature is significantly different from the conventionally assumed value of 60°C.

As with the volumetric data it is of interest to see whether delivery temperature varies with, for example, boiler type. Figure 6.13 shows the data for regular boilers and combis separately.

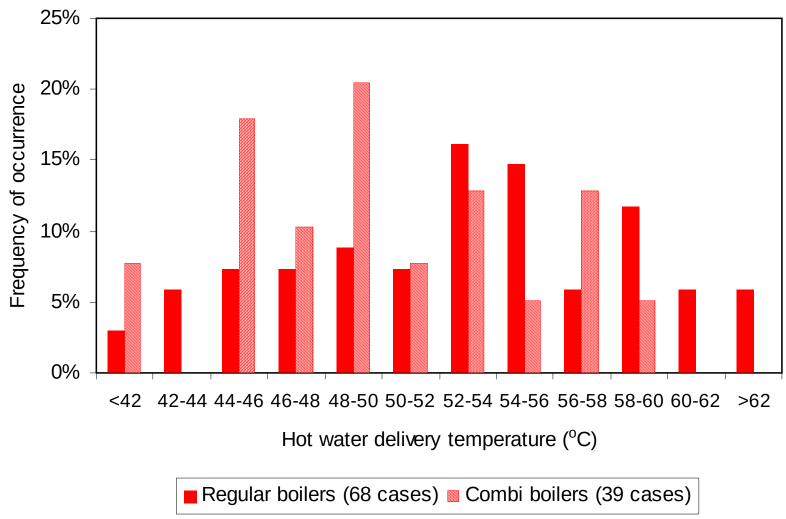

Figure 6.13: Distribution of delivery temperature by boiler type

The mean delivery temperature for regular boilers is 52.9°C with 95% confidence interval ±1.5°C. The corresponding value for combi boilers is 49.5 ± 2.0°C.

More detailed statistical analysis of delivery temperature, using the same approach that was applied to volumetric consumption, is described in Appendix 5. As before, the data does not appear to be normally distributed, and both parametric and non-parametric methods have been used. As expected from the mean values above, both tests indicate that the difference observed between households with regular and combi boilers is highly significant. It is concluded that combi boiler users routinely achieve lower hot water delivery temperatures than households with regular boilers.

6.3.2 Heating times

In the dwellings which were equipped with regular boilers a temperature probe was attached to the primary pipework which connected the boiler to the hot water cylinder. This allows the periods of boiler operation to be identified. Due to variations between systems and in the way that the sensor was applied considerable care must be taken when interpreting this data, and Appendix 4 describes how this was done. Figure 6.14 shows the estimated hot water heating profile in one dwelling.

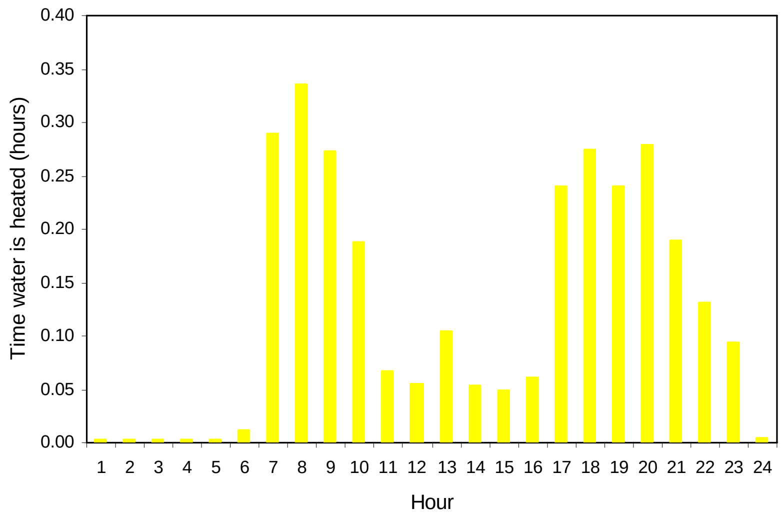

Figure 6.14: Water heating behaviour in a single dwelling

Once again, there is very large variation between dwellings. In particular, many seem to operate their water heating only when it is required. Figure 6.15 shows the distribution of heating time across all the regular boilers.

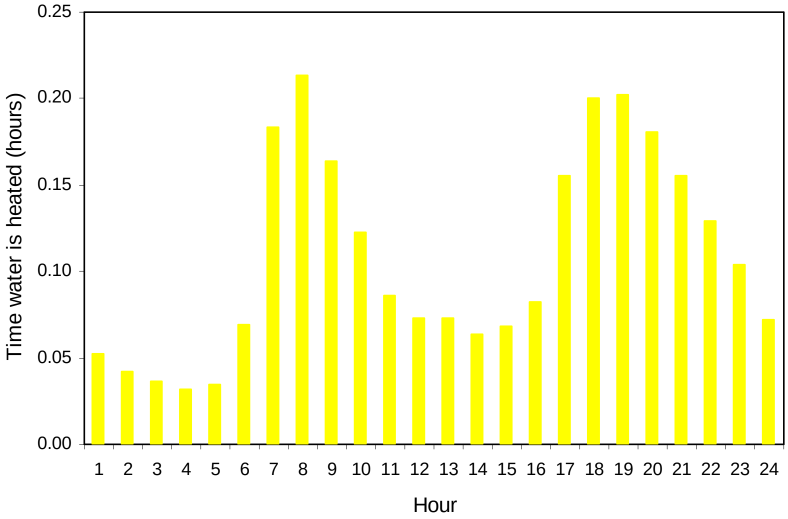

Figure 6.15: Water heating behaviour across whole sample

Figure 6.16 shows the average heating profile across all the regular boiler systems in the sample.

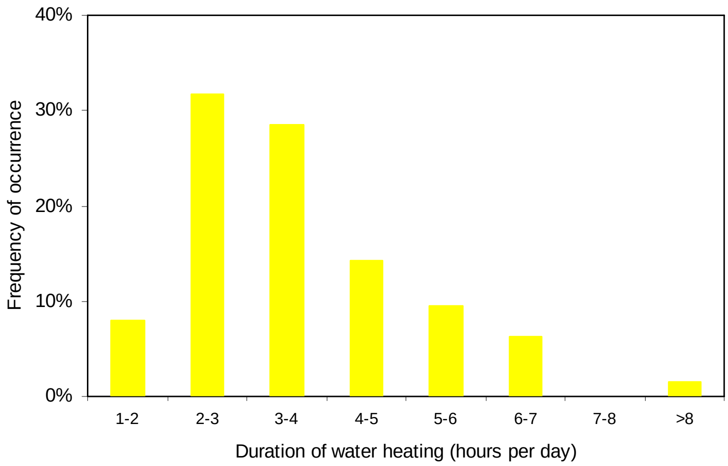

Figure 6.16: Distribution of time spent heating water across all regular boilers

The heating time has a mean of 2.60 hours/day, estimated with a 95% confidence interval of ± 0.35 hours/day.

6.3.3 Cold water inlet temperatures

Cold water inlet temperatures are not under the control of the householder, and therefore cannot strictly be considered to be to be part of the hot water use pattern. However, they play a vital role in determining the temperature rise required from the hot water system, and hence the energy content of delivered water. It is for this reason that their analysis is included in this section.

The cold water feed temperatures observed have mean value of 15.2°C with 95% confidence interval ± 0.5°C. Figure 6.17 shows the distribution of the measured data.

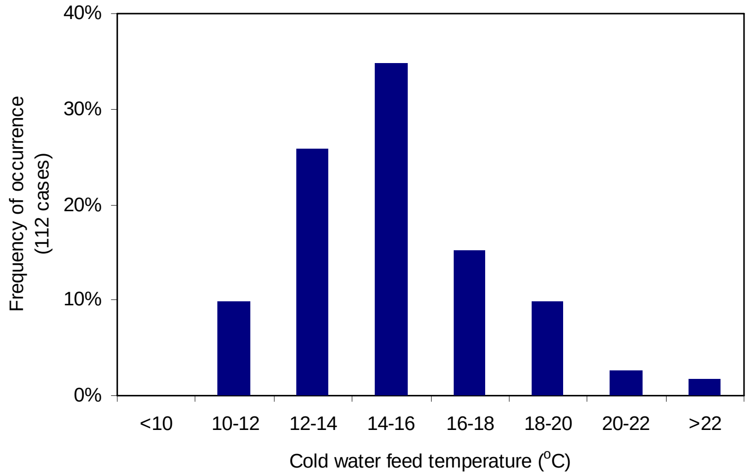

Figure 6.17: Cold water feed temperatures for whole sample

Breaking down the cold water temperatures by boiler type indicates that it is in general lower for combi boilers.

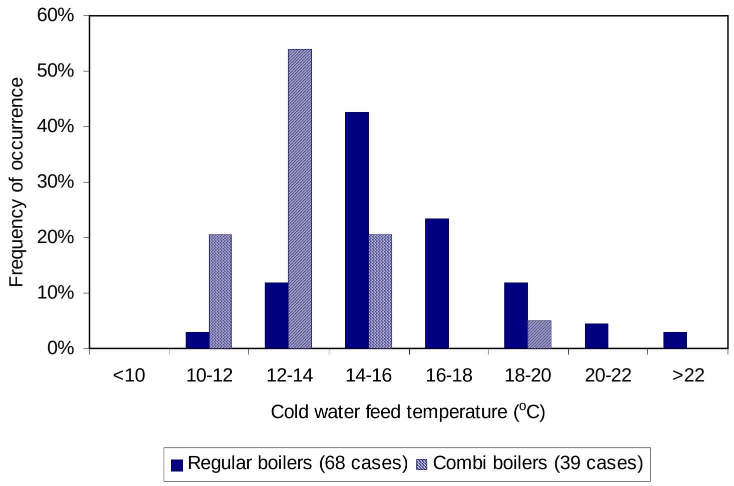

Figure 6.18: Cold water feed temperature by boiler type

For regular boilers the cold feed temperature is 16.2 ± 0.6°C, for combis it is 13.4 ± 0.6°C. Both parametric and non-parametric statistical tests indicate that this difference is highly significant. There is a simple explanation for this result. With a regular boiler water is often held in a cold water storage tank before use, where it may have a chance to warm up. With a combi, water is taken directly from the incoming cold main.

6.4 Seasonal variation of hot water consumption patterns

The data collected from each dwelling in most cases covers a period of approximately 12 months. Due to the time required to carry out each equipment installation the start dates are staggered, and the whole data collection period runs from March 2006 through to September 2007. With this data it is therefore possible to explore whether there are any variations in the key parameters over the course of a full year. Figure 6.19 shows the variation of volumetric consumption over a year.

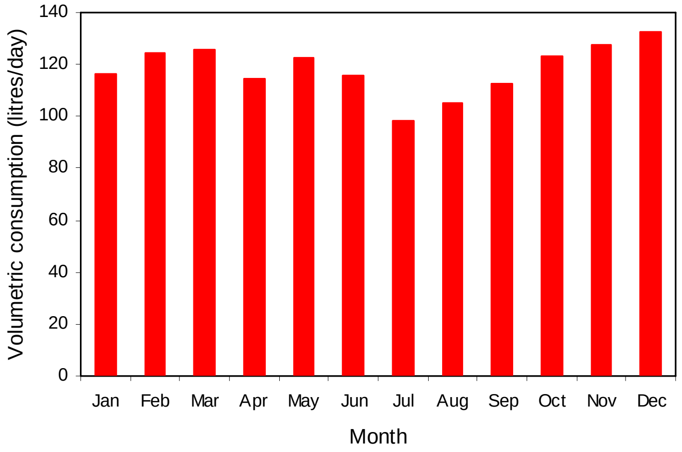

Figure 6.19: Annual pattern of whole sample volumetric consumption

The figure indicates that consumption is slightly reduced during July and August. This may be due to householders taking summer holidays. Next, Figure 6.20 shows how the mean incoming cold and hot water delivery temperatures vary.

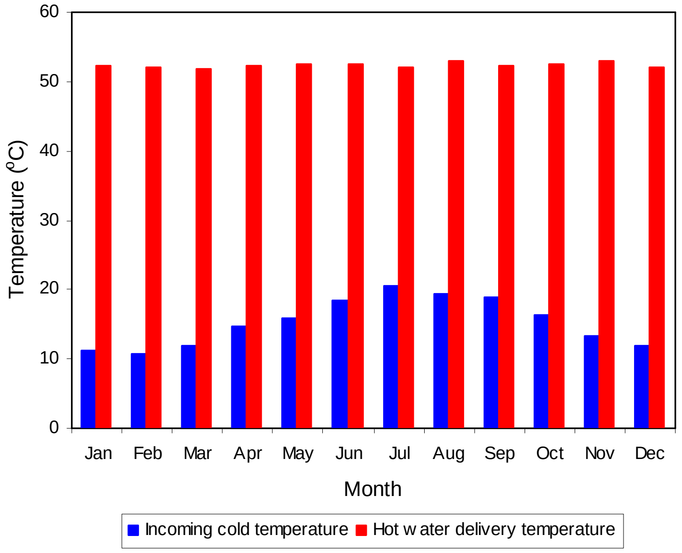

Figure 6.20: Annual pattern of cold water inlet and hot water delivery temperatures

As expected, there is marked variation in cold inlet temperature with season. Hot water delivery temperature is, however, extremely constant over the whole year. It was noted in Section 6.3.3 that cold inlet temperature also varied between regular and combi boilers, and Figure 6.21 shows the cold feed variations broken down by boiler type.

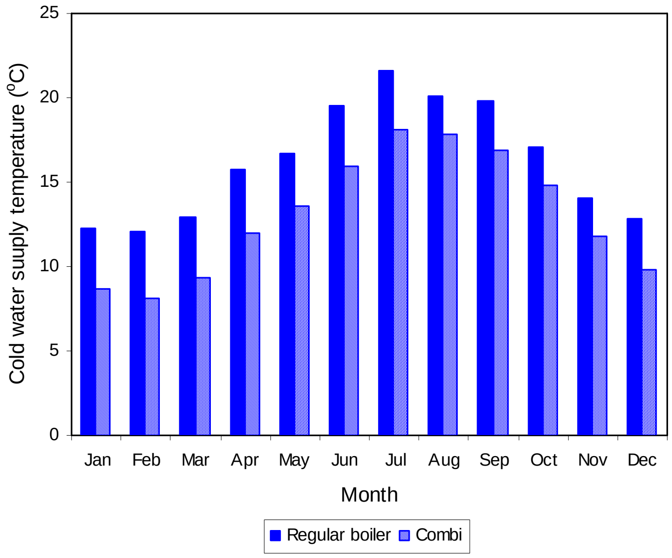

Figure 6.21: Annual variation of cold water inlet temperature by boiler type

The figure shows that although inlet temperature is consistently higher for regular boilers the difference is comparatively consistent over the course of the year.

Finally, Figure 6.22 shows how cold feed temperature varies over the year broken down by region.

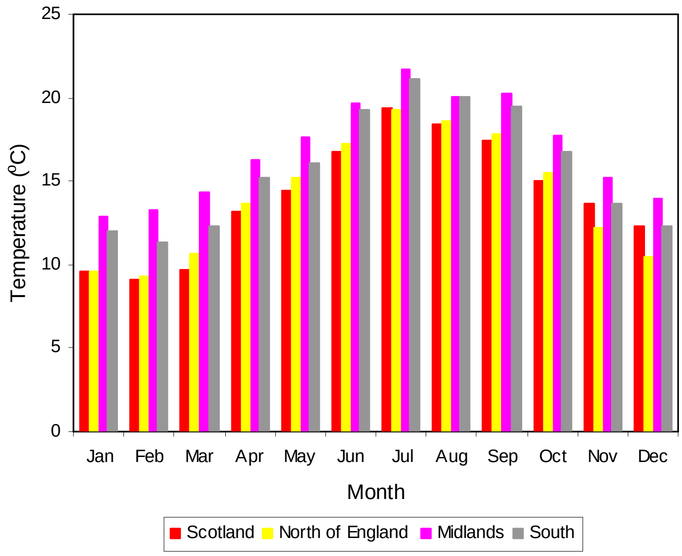

Figure 6.22: Regional variation of cold water inlet temperature

The values for Scotland and the North of England are similar. The Midlands have the highest temperatures, with the South of England falling somewhere in between.

6.5 Comparisons with current BREDEM assumptions

The calculation of hot water energy demand in BREDEM proceeds in a series of stages. The first step is to calculate the expected daily consumption. A fixed temperature difference of 50°C is then used to calculate energy content.

6.5.1 Comparison of measured flows with BREDEM

In BREDEM the daily volumetric hot water consumption is estimated using the formula:

Hot water demand (litres/day) = 38 + 25 N

where:

N is the number of occupants in the dwelling.

BREDEM applies this equation regardless of boiler type, and so the most appropriate comparison is with the straight lines derived over all dwellings. The analysis in Section 6.1 indicated that the best straight line fit to the consumptions measured was 46 + 26 N when all the data was considered, or 40 + 28 N for dwellings with five or fewer occupants. The figure below shows the data, together with the two fitted lines and the BREDEM equation. The figure demonstrates just how close to the BREDEM assumption the derived equations are.

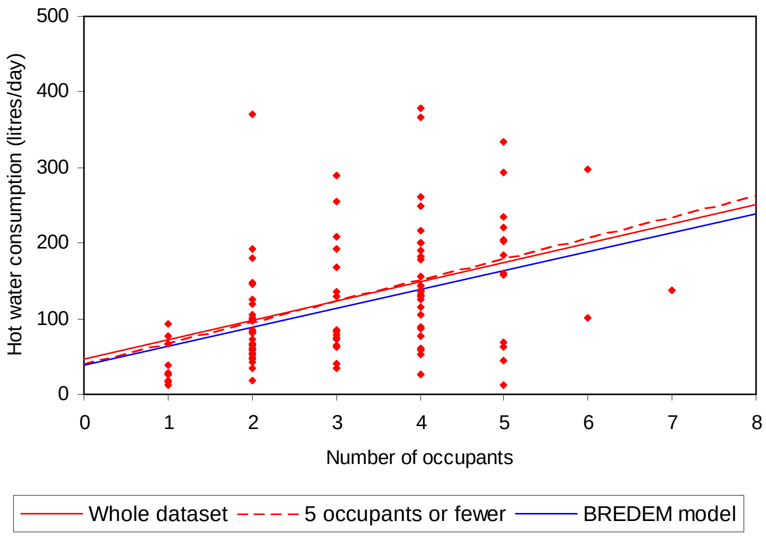

Figure 6.23: Comparison of measured flow data to BREDEM assumption

6.5.2 Comparison of temperature data with BREDEM

The next stage in the BREDEM calculation process estimates the energy delivered as hot water, based on a 50°C temperature rise. This figure was derived from an estimate of 10°C for the cold inlet temperature, and 60°C for the delivery temperature. We have already seen (in Section 6.2.1) that average hot water delivery temperature is significantly below 60°C. Calculating the mean temperature rise across the whole of the sample used in this project reveals a value of only 36.8°C, considerably lower that the BREDEM assumption. The distribution of the observed temperature rises is shown on Figure 6.24.

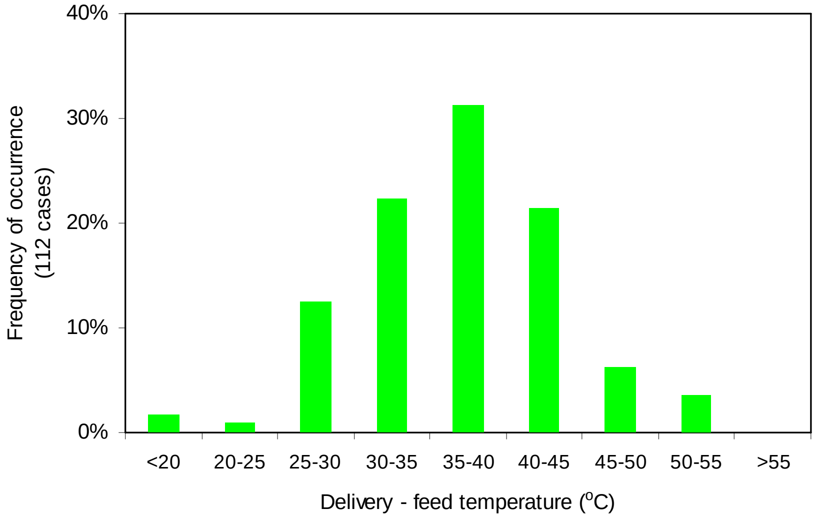

Figure 6.24: Distribution of hot water system temperature rises

This data does appear to be distributed in an approximately normal way, and so it is acceptable to use a t-test to compare it to the BREDEM assumption of a temperature rise of 50°C. In fact, both the t-test and the corresponding non-parametric Wilcoxon signed rank test indicate that the probability of this data coming from a population with mean 50°C is vanishingly small, implying that the difference observed is highly significant. This difference will result in BREDEM over-predicting the energy required by this sample of dwellings for hot water production by approximately 35%.

In Section 6.3.1 it was shown that the mean hot water delivery temperature was lower than the value of 60°C used to derive the temperature difference used in BREDEM. This is obviously one contributor to the reduced temperature difference. It was noted that delivery temperature varied significantly between regular and combi boilers. In Section 6.3.3 it cold water inlet temperatures were found to be generally higher than the 10°C assumed in BREDEM. It was again noted that there are significant differences in cold water delivery temperature for the two types of boiler, but in opposite sense to the effect on hot water temperature. These effects thus tend to cancel out when calculating the temperature rise across the system. Figure 6.25 shows how the net temperature rise is distributed for the two boiler types, and confirms that to some extent this cancellation is taking place.

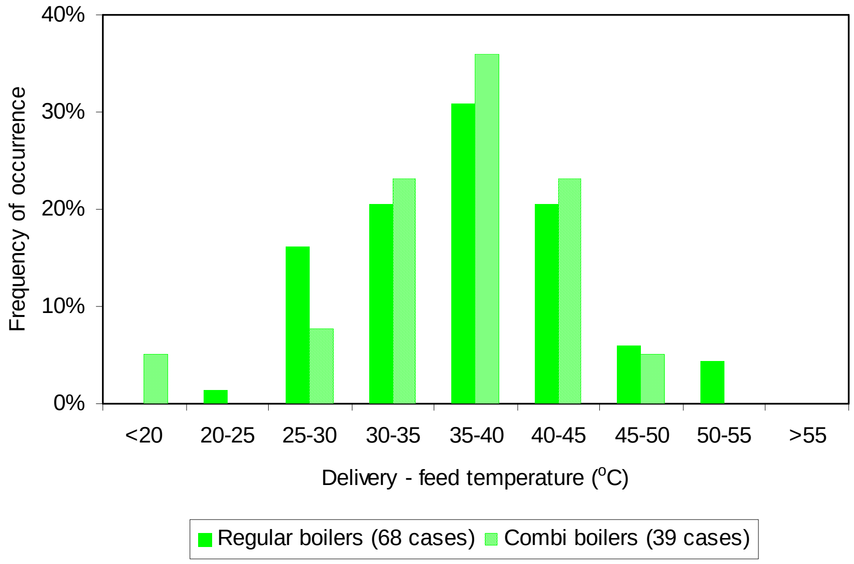

Figure 6.25: Temperature difference by boiler type

Both parametric and non-parametric statistical tests reveal that the effect of boiler type on temperature difference is not significant, and further analysis shows that none of the factors measured is useful when trying to characterise the temperature rise across water heating systems. The effect of different cold water temperatures has effectively cancelled out the structure previously seen in hot water delivery temperature.

In view of these results there are two routes which could be used to produce an improved model of domestic hot water consumption in BREDEM. The first would be to amalgamate the estimated temperature rise and volumetric flow and attempt to model energy consumption directly in terms of factors such as occupancy, boiler type, region and time of year. This would effectively be a ‘black box’ approach to estimating the energy required to produce hot water.

The second approach would be to model the fundamental variables determining hot water energy use separately. In particular the temperature rise would be separated into its two components: incoming cold temperature and hot water delivery temperature. In this model volumetric flow, cold inlet temperature and delivery temperature are each estimated in turn, and then combined to determine energy consumption. Table 6.3 summarises the dependencies which might be included, based on results reported here, and in Appendices 2, 5 and 6.

Quantity	Factors considered
Volumetric consumption	Number of occupants
Cold inlet temperature	Number of occupants Boiler type Region Time of year
Hot water delivery temperature	Boiler type

Table 6.3: Outline of factors in separated model of hot water energy requirement

This second approach is preferable for a number of reasons. Whilst considerable cancellation of the variations in cold and hot water temperatures has been seen in the present dataset, it is quite possible that this may not always occur. Furthermore, some of intermediate quantities may be useful in other parts of the BREDEM calculation. For example, even though it has been outside the scope of this project, hot water delivery temperature may be used to calculate cylinder and secondary pipework losses.

7 Analysis of flow destination data

A limited number of properties were equipped with additional instrumentation which allowed the destination of hot water to be determined. This section describes the analysis of that data.

7.1 Basic approach

The remote temperature sensors are generally clamped to the outside of the relevant pipe. This means that there is a marked time delay associated with their response, and also that they have some response to changes in ambient temperature. During each run off a block of data is accumulated. The analysis software then compares the temperature rise at each remote point and assigns the flow to the location with the largest. In the event that no temperature changes by more than 0.5°C over the course of the run off, the run off is deemed to be unallocated.

Table 7.1 summarises the locations monitored in the 21 dwellings from which useful data was obtained.

Location	Number monitored		
	Regular boiler	Combi boiler	Total
Kitchen sink	13	8	21
Bathroom basin	11	7	18
Bath	11	8	19
Washing machine	7	4	11
Shower	6	2	9
Downstairs basin	6	1	7
Upstairs basin	3	1	4

Table 7.1: Breakdown of hot water locations monitored

7.2 Dealing with unallocated flows

In practice there are inevitably flows which cannot be categorised using the procedure described above. There are two possible reasons for this. In a dwelling where all possible outlets have been instrumented there will be some run-offs which are so brief that the do not affect the relevant remote sensor. In this case the quantity of water unaccounted for (or 'unallocated') provides a useful check on the effectiveness of the method: in general the unallocated flow is in the region 5 to 15%.

A second possible reason for unallocated flows is that, for a variety of practical reasons, it was not possible to instrument every outlet. One such reason would be if it proved impossible to remove the side of a bath to gain access to the pipe serving the hot tap. In this case, all unallocated flows would be attributed to the bath. In such cases there is no check on the effectiveness of the method, but the results from the dwellings where no location was left unmetered serve to provide confidence that the monitoring strategy works.

7.3 Results

With these measures in place it is possible to generate profiles showing the time of day at which water is used at each location. Figure 7.1 shows the result for the same dwelling used as an example in Figure 6.4.

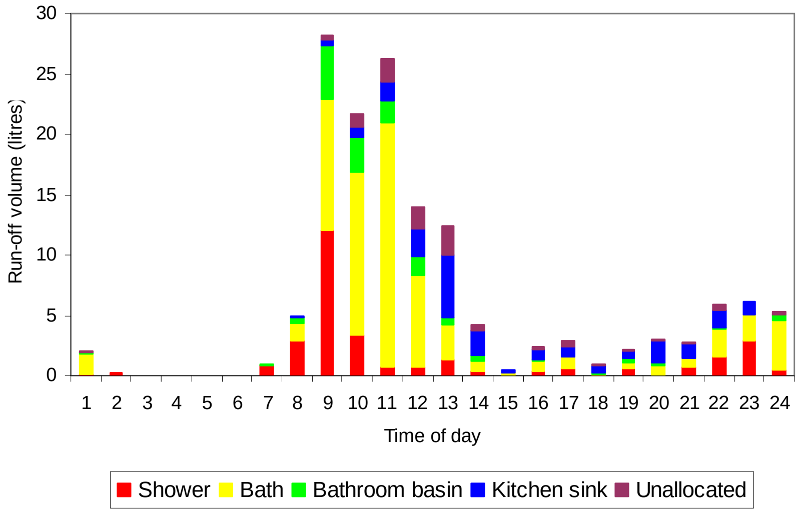

Figure 7.1: Run-off volume profiles at each location in a sample dwelling

Alternatively, the amount of water used at each point can be totalled over the day, and the relative sizes compared. Figure 7.2 shows the result.

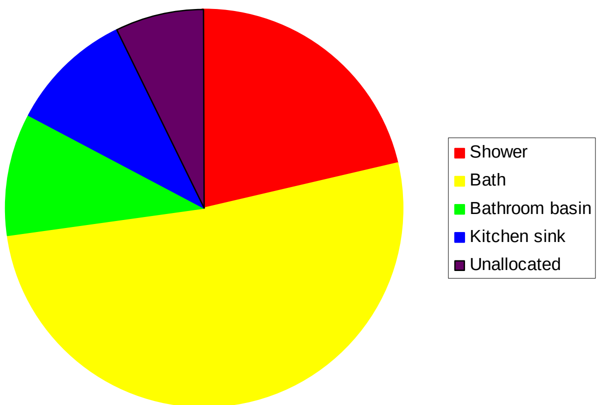

Figure 7.2: Relative run off volume at each location in a single dwelling

The figure shows that in this household by far the largest proportion of hot water is used in the bath in the morning. There are no unallocated draw off points in this dwelling, and the proportion of the flow which cannot be allocated is reassuringly small.

It is also possible to analyse the destination of hot water in terms of the energy delivered at each point. The corresponding profile for this sample dwelling is shown in Figure 7.3.

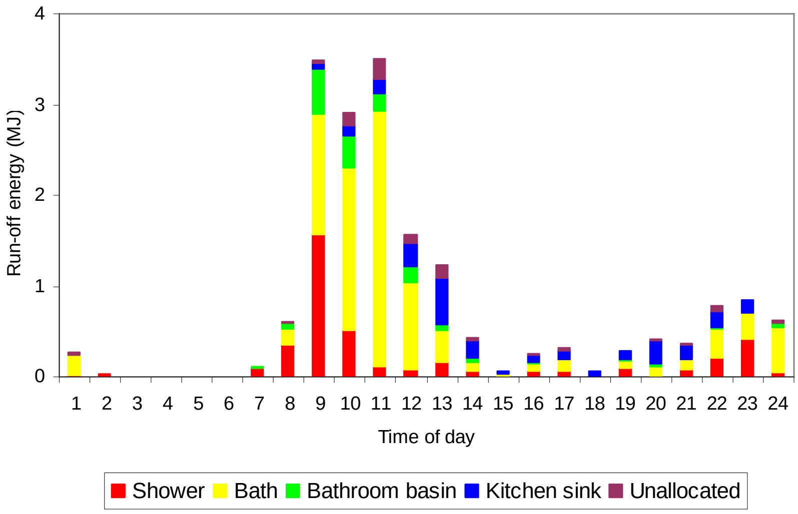

Figure 7.3: Run-off energy profiles at each location in a sample dwelling

Figure 7.4 shows the corresponding distribution of energy between the locations.

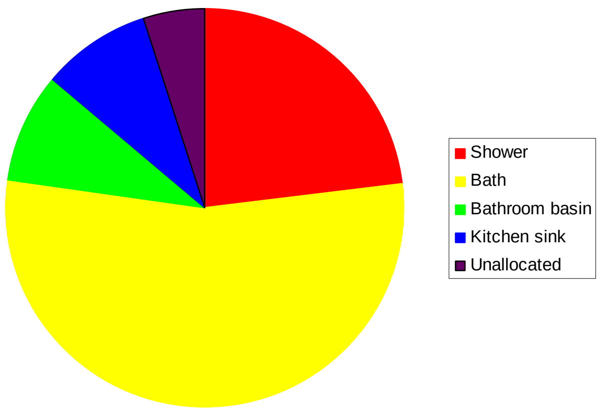

Figure 7.4: Relative run-off energy at each location in a single dwelling

The above figures showed the small amount of flow which was unallocated separately from the metered destination flows. From this point on, once the unallocated flow has been checked and found to be close to zero it is then attributed to each monitored location in the ratio of the flow at each location.

One of the reasons for location monitoring was to determine whether combi boilers change the characteristics of draw-offs at different points in the house. The analysis described above provides both the volume and energy delivered to each point. From these values it is possible to determine the energy cost of water delivery, expressed as the energy delivered divided by the volume. This quantity, which has units MJ/litre, can also be interpreted as an average temperature rise. An energy cost of 0.168 MJ/litre corresponds to a temperature rise of 40°C, and, as expected, most of the values calculated here are close to this.

Table 7.2 shows the flow, energy and cost of water at each of the four locations which provide enough points for further analysis.

Location	Flow (litres/day)		Energy (MJ/day)		Cost (MJ/litre)	
	Regular	Combi	Regular	Combi	Regular	Combi
Kitchen sink	15.6	38.0	2.44	5.02	0.145	0.130
Bathroom basin	12.5	18.3	1.94	2.49	0.150	0.132
Bath	43.9	36.5	7.33	4.95	0.160	0.138
Washing machine	2.6	4.1	0.42	0.53	0.144	0.127

Table 7.2: Differences in consumption characteristics as a function of boiler type

Figures 7.5 and 7.6 show the flow and energy results from the table graphically.

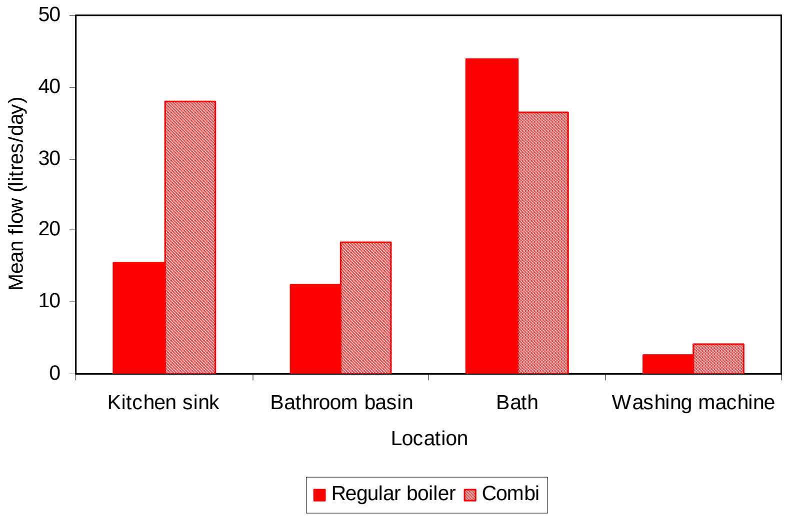

Figure 7.5: Volumetric consumptions at key locations

The figure shows that the discrepancy between boiler types is largest at the kitchen basin, with combi boilers resulting more than double the flow obtained with regular boilers. The reason for this may be that householders routinely require hotter water at the kitchen sink than at the other locations. As a result, water is wasted whilst a combi boiler responds to the request. At the other locations water can be accepted at a lower temperature.

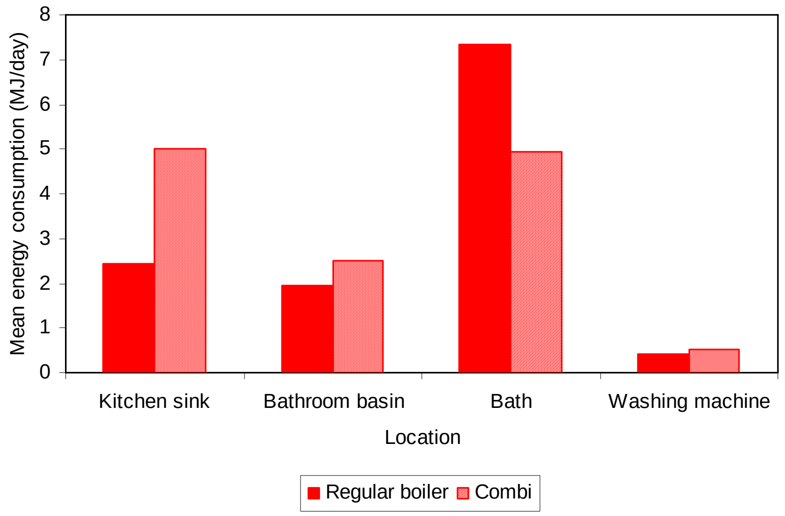

Figure 7.6: Energy consumptions at key locations

The energy use at each location follows the flow quite closely. As a result, the energy costs shown on Figure 7.7 show the much smaller variation with boiler type. To help with interpretation the right hand axis shows the temperature rise corresponding to the energy costs.

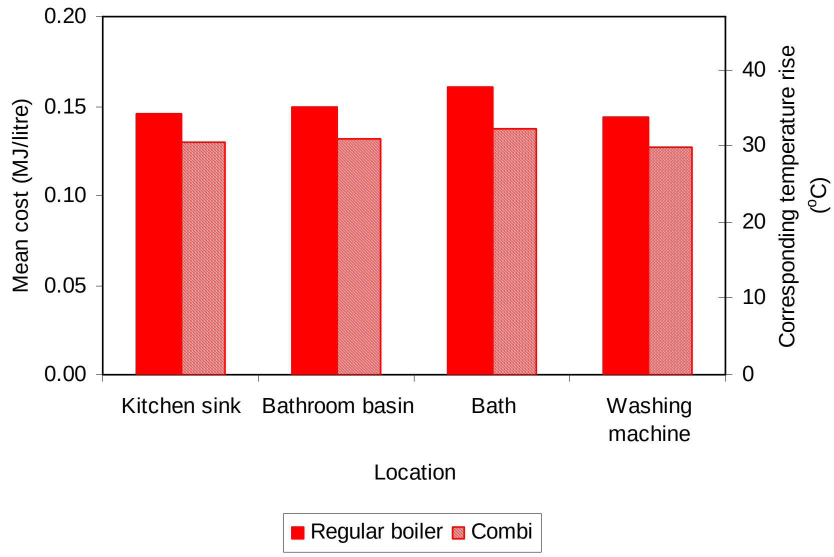

Figure 7.6: Energy consumptions at key locations

Table 7.3 shows the significance of the differences between regular and combi boilers for each of the four locations.

Location	Significance of regular/combi difference		
	Flow	Energy	Cost
Kitchen sink	0.4%	2.0%	26.1%
Bathroom basin	65.9%	79.1%	12.3%
Bath	60.0%	54.5%	8.2%
Washing machine	7.3%	16.4%	34.4%

Table 7.3: Statistical significance of consumption differences observed

The results in the table must be treated with considerable caution. Since they are based on such a small number of points it is not possible to determine whether the distribution from which they are drawn is normal. For this reason, only non-parametric statistics are shown in the above table.

A second danger in making inferences from such a small sample is that the addition (or removal) of a single point could make a very large difference to the conclusions: the analysis cannot be considered robust. However, the present information serves to demonstrate that the measurement method is successful, and could if desired be extended to a larger sample.

8 Conclusions

The first goal of this project was to measure DHW consumption and the associated energy delivered. Both overall consumption and consumption profile over the course of a day have been generated. Inspecting the results from single dwellings confirms that the monitored data is sensible, but it is the results taken across the whole sample which are principally of interest. The mean consumption has been found to be 122 litres/day, with a 95% confidence interval of ±18 litres/day. Statistical analysis of the flows from each dwelling has considered the impact of geographical region, boiler type, number of occupants, and the number of those occupants who are children. It has revealed that the key factor influencing consumption is the number of occupants. The energy content of the water delivered has also been analysed, and has also been found to depend only on number of occupants.

The next goal was to identify DHW heating patterns, both in terms of the delivery temperature and, in the case of regular boilers, the times at which water was heated. Across the whole sample delivery temperature has been found to be significantly below the widely assumed value of 60°C, with a mean value of 51.9°C estimated with a 95% confidence interval of ±1.3°C. The mean temperature among dwellings fitted with regular boilers is 52.9 ± 1.5°C. In houses with combi boilers it is 49.5 ± 2.0°C. This difference is highly statistically significant, and it is concluded that combi boiler owners routinely experience lower hot water delivery temperatures than regular boiler users.

In the case of regular boilers an additional temperature was measured on the primary circuit to the hot water cylinder. This has allowed an estimate to be made of the times of day at which householders heat their water. The heating time has a mean of 2.60 hours/day, estimated with a 95% confidence interval of ± 0.35 hours/day. The overall schedule reveals that some households heat water as and when it is required, and the remainder generally heat between 8:00 and 10:00am, and again between 6:00 and 11:00pm.

The results were next compared with the current BREDEM assumptions. This reveals that the current model of consumption (which is based on the number of occupants in a dwelling) is appropriate. Modifying this relationship to include boiler type cannot be recommended on the basis of the results of this work. However, analysis of the average temperature rise of water as it passes through the heating system (derived from the initial cold feed temperature and the hot water delivery temperature) shows a value of 36.7°C, significantly lower value than the 50°C currently assumed in BREDEM. On this sample of dwellings BREDEM would over-predict consumption by approximately 35%. The current temperature difference is based on an assumed cold water inlet at 10°C and hot water delivery at 60°C. The discrepancy with the measured result is due partly to hot water temperatures lower than assumed, and partly to cold feed temperatures higher than assumed. Statistical analysis has demonstrated that cold water inlet temperature could be better estimated using a model that takes into account occupancy, boiler type and region. Hot water delivery temperature prediction could also be improved by using a simple model based only on boiler type.

A subset of the dwellings monitored was equipped with remote sensors which allowed the location where hot water was being consumed to be identified. In dwellings where it was possible to monitor every hot water delivery point this scheme has proved capable of identifying the destination location of almost all the hot water used. In other dwellings, where practical considerations made it impossible to instrument one point the unallocated flow can therefore be attributed to that unmetered point with confidence. The relatively small number of dwellings which were equipped with this additional instrumentation means that it impossible to draw firm statistical conclusions. However, it can be seen that for locations such as the bathroom basin, bath and washing machine the difference in volume used is similar between regular and combi boilers. For the kitchen sink, however, the volume of water used by combi owners is significantly larger than in the dwellings with regular boilers. The most likely explanation for this is that users demand a higher temperature at the kitchen sink, this is harder to achieve with a combi boiler, and therefore more water is run off. At the other locations temperature is perhaps less critical, and that the combi systems can achieve this without the extended run-off.

Although the current sample of houses with the additional instrumentation required to determine the destination of hot water run-offs is small, it has served to prove that the measurement technique is effective. To obtain more robust information, however, it would have to carried out in a significantly larger number of dwellings.

APPENDIX 1: Details of the data cleaning process

This Appendix contains full details of the file format and the cleaning applied to the hot water consumption database.

A1.1 Data format

The file naming convention used for all the datafiles is:

\ssss\ssss _ yyyy mm dd n [t…t] [c] . x01

where:

ssss is the four digit dwelling serial number

yyyy is the year in which this particular file was downloaded

mm is the month in which this particular file was downloaded

dd is the day in which this particular file was downloaded

n is the number of the download on that particular day

[t…t] is optional additional text

[c] is an optional indication that the file has been cleaned, and

.x01 is the file extension.

A1.2 Data cleaning prior to generation of master files

All folders are renamed to comply with the convention described in Section A1.1. Data from dwelling 7978 appears as a sub-folder in 7984 and is moved back up to the root position

Two of the loggers used had duplicated serial numbers. The data was renamed using two new dwelling serial numbers, as described in Table A1.1.

Original dwelling id	New serial number
8026Mxxxxx	9026
8083Gxxxxxxxx	9083

Table A1.1: Dwelling numbers reallocated due to duplicate logger serial numbers

In one case data was correctly placed in a separate folder but was incorrectly named. All data in folder 8054 this has to be renamed in accordance with the file naming convention laid out above

A number of dwellings produced data which could not be used and were therefore abandoned. Table A1.2 summarises.

Dwelling
8009
8019
8026
8028
8090

Table A1.2: Dwellings removed from study at start of analysis

A1.2.1 Timing errors

Table A1.3 details the timing errors found in the data, and describes the corrective action.

Dwelling	Problem	Action
7980	File 7980_200607201.x01 runs from 18/07/06 to 20/07/06 File 7980_200610091.x01 runs from 18/07/06 to 03/10/06	Remove 7980_200607201.x01
7981	File 7981_200608011c.x01 and 7981_200608019c.x01 both run from 18/07/06 to 02/08/06 File 7981_200608091c.x01 starts on 01/08/06, duplicating some of the data in 7981_200608019c.x01	Remove 7981_200608011c.x01 and 7981_200608011.x01 Edit 7981_200608019c.x01 to remove excess data from end
7993	Files 7993_200607281c.x01 and 7993_200607311c.x01 both run from 16/07/06 to 25/07/06	Remove 7993_200607281.x01 and 7993_200607281c.x01
8000	File 8000_200605211test.x01 runs from 19/05/06 to 21/05/06 and file 800_200605281.x01 runs from 19/05/06 to 28/05/06	Remove 8000_200605211test.x01
8018	Files 8018_200607211.x01 and 8018_200607301.x01 both run from 16/07/06 to 21/07/06	Remove 8018_200607211.x01
8021	File 8021_200701121.x01 runs from 10/01/07 to 12/01/07 and file 8021_200701171.x01 runs from 10/01/07 to 17/01/07	Remove 8021_200701121.x01
8023	File 8023_200704201prob.x01 runs from 17/04/07 to 20/04/07 and file 8023_200704241.x01 runs from 17/04/07 to 24/04/07	Remove 8023_200704201prob.x01
8025	File 8025_200607201.x01 runs from 16/07/06 to 20/07/06 and file 8025_200607251c.x01 runs from 16/07/06 to 25/07/06	Remove 8025_200607201.x01
8033	File 8033_200608131.x01 runs from 16/07/06 to 12/08/06 and file 8033_200608251.x01 runs from 16/07/06 to 13/08/06	Remove 8033_200608131.x01
8035	File 8035_200607201.x01 runs from 16/07/06 to 20/07/06 and file 8035_200608141.x01 runs from 16/07/06 to 14/08/06	Remove 8035_200607201.x01
8036	File 8036_200607281.x01 runs from 26/07/06 to 28/07/06 and file 8036_200607311.x01 runs from 26/07/06 to 31/07/06	Remove 8036_200607281.x01
8037	File 8037_20060811.x01 runs from 09/08/06 to 11/08/06 and file 8037_200608111 runs from 09/08/06 to 11/08/06	Remove 8037_20060811.x01
8052	File 8052_200606221.x01 runs from 20/06/06 to 22/06/06 and 8052_200606261.x01 runs from 20/06/06 to 26/06/06 File 8052_2006062811.x01 runs from 26/07/06 to 28/07/06 and file 8052_200607031.x01 runs from 26/07/06 to 31/07/06	Remove 8052_200606221.x01 Remove 8052_200607031.x01
8058	File 8058_200607201.x01 runs from 16/07/06 to 20/07/06 and file 8058_200607251.x01 runs from 16/07/06 to 25/07/06	Remove 8058_200607201.x01

8061	File 8061_200607201.x01 runs from 19/07/06 to 20/07/06 and file 8061_200608131.x01 runs from 19/07/06 to 13/08/06	Remove 8061_200607201.x01
8066	File 8066_200609011.x01 runs from 29/08/06 to 01/09/06 and file 8066_200609051.x01 runs from 29/08/06 to 05/09/06	Remove 8066_200609011.x01
8068	File 8068_200608131.x01 runs from 16/07/06 to 03/08/06 and file 8068_200608141.x01 runs from 16/07/06 to 13/08/06	Remove 8068_200608131.x01
8084	File 8084_200608121.x01 runs from 16/07/06 to 07/08/06 and file 8084_200608141.x01 runs from 16/07/06 to 12/08/06	Remove 8084_200608121.x01
8085	File 8085_200607201.x01 runs from 16/07/06 to 20/07/06 and file 8085_200608141.x01 runs from 16/07/06 to 14/08/06	Remove 8085_200607201.x01
8089	File 8089_20060708.x01 runs from 24/06/06 to 04/07/06 and file 8089_200607151 runs from 24/06/06 to xxxxxx	Remove 8089_20060708.x01
9026	File 9026_200608141.x01(M*****) runs from 31/07/06 to 14/08/06 and file 902608141.x01 also runs from 13/07/06 to 14/08/06	Remove 9026_200608141.x01(M*****)

Table A1.3: Modifications to rectify data overlaps

As described in the body of this report, a number of loggers had additional channels enabled at the start of logging. The relevant files are edited as described in Table A1.4.

Dwelling	Problem	Action
7998	Channel 3 logged in error in early files	Channel 3 data removed from files: 7998_200605071.x01 7998_200605141.x01 7998_200605141c.x01 7998_200605281.x01 7998_200606041.x01 7998_200606181.x01 7998_200606261.x01 7998_200606261c.x01 7998_200607021.x01 7998_200607091.x01 7998_200607161.x01 7998_200607311.x01 7998_200608081.x01
8017	Time and date information in file 8017_200607191c.x01 in separate channels in incorrect format	Time and date channels combined and result reformatted
8039	Channel 3 logged in error in early files	Channel 3 data removed from file: 8039_200607201.x01
8070	Channels 3 and 12 logged in error in early files	Channels 3 and 12 removed from files: 8070_200605071.x01 8070_200605141.x01 8070_200605281.x01 8070_200606041.x01 8070_200606181.x01 8070_200607021.x01 8070_200607091.x01 8070_200607161.x01 8070_200608111.x01 8070_200608131.x01 8070_200608181.x01

Table A1.4: Modifications to database to rectify channel errors

A1.3 Data cleaning at the analysis stage

The remainder of the data cleaning was carried out as the data was analysed.

A1.3.1 Reversed cold water feed and hot water delivery temperatures

The dwellings for which cold water feed and hot water delivery temperatures are reversed are listed in Table A1.5. Any data analysis software should therefore swap channels 1 and 2 before using data from these dwellings.

Dwelling
7972
7984
7997
8001
8002
8003
8030
8031
8084

Table A1.5: Dwellings with reversed cold water feed and hot water delivery temperatures

A1.3.2 Data with inconsistent behaviour over time

A number of datasets displayed behaviour which changed suddenly over the course of a year. Most commonly this was due to the measured daily flow suddenly dropping to zero, probably indicating that the flowmeter had failed part way through the monitoring period. Checking the relevant data confirms that this was the case. The problem was rectified by truncating the data from that particular dwelling. Table A1.6 summarises the actions taken to remove the faulty data from the analysis.

Dwelling	Data removed
7971	Data collected in March 2006 removed (this is believed to have been collected during instrumentation pilot) All data from start of August 2007 removed
7980	All data from start of December 2006 removed
7981	All data from start of July 2007 removed
7989	All data from start of July 2007 removed
7994	All data from start of July 2007 removed
7996	All data from start August 2007 removed
7997	All data from start of April 2007 removed
7998	All data from start of March 2007 removed
8000	All data from start of July 2007 removed
8002	All data from end of April 2007 removed
8007	All data from start of January 2007 removed
8014	All data from start of July 2007 removed
8015	All data before start of December 2006 removed All data after start of September 2007 removed
8016	All data from start of July 2007 removed
8023	All data before start September 2006 removed
8024	All data from start of April 2007 removed
8027	All data before start of August 2006 removed
8035	All data from start of Feb 2007 removed
8037	All data to start of November 2006 removed
8038	All data from start of July 2007 removed
8039	All data from start of April 2007 removed
8040	All data from start of March 2007 removed
8044	All data from start of April 2007 removed
8046	All data after end of June 2007 removed
8052	All data before start of August 2006 removed
8054	All data from start of May 2007 removed
8056	All data before start of December 2006 removed
8059	All data from start of February 2007 removed
8067	All data from start of August 2007 removed
8068	All data from start of July 2007 removed
8069	All data from start of March 2007 removed
8074	All data from start of June 2007 removed
8077	All data from start of Jan 2007 removed
8079	All data from start of September 2006 removed
8081	All data up to start of July 2006 removed
8082	All data from start of October 2006 removed
8083	All data from start of September 2006 removed
8087	All data from start of September 2007 removed
8089	All data before start of August 2006 removed
9026	All data from start of January 2007 removed

Table A1.6: Datasets truncated due to faults with flow meters

Dwelling 8083 was removed from the dataset completely as a result of inspecting the monthly consumptions.

APPENDIX 2: Further statistical analysis of hot water consumption

This Appendix describes the preliminary statistical analysis of the hot water volume and energy use of the sample. The model seeks to explain the variation in two continuous quantities (water volume and energy content) using a mixture of continuous variables (occupancy level and number of children) and categorical quantities (boiler type and geographical region). The appropriate modelling tool is therefore an analysis of covariance (ANCOVA). The dataset comes mostly from regular and condensing boilers. However in one dwelling the boiler type is not known, three dwellings had no boiler and one had a multipoint. Since the analysis shown here includes the effect of boiler type these cases have been excluded.

A2.1: Analysis of volumetric consumption

The analysis begins by fitting a model using all of the available input variables:

```
        *** Linear Model ***

Call: lm(formula = Flow ~ Occupants + Children + Boiler + Region, data = dhw,
        na.action = na.exclude)
Residuals:
    Min     1Q Median    3Q   Max
 -150.1 -49.36 -14.48 32.83 511.3

Coefficients:
               Value Std. Error  t value Pr(>|t|)
(Intercept)  43.8226  25.6306     1.7098   0.0904
  Occupants  30.6836   9.9779     3.0752   0.0027
   Children  -5.4010  12.5452    -0.4305   0.6677
     Boiler -11.7032  10.2132    -1.1459   0.2546
    Region1  -4.5081  14.3132    -0.3150   0.7534
    Region2  13.5319   8.4584     1.5998   0.1128
    Region3  -6.9391   4.8740    -1.4237   0.1576

Residual standard error: 90.47 on 100 degrees of freedom
Multiple R-Squared: 0.1848
F-statistic: 3.778 on 6 and 100 degrees of freedom, the p-value is 0.00197
```

From these results it is clear that the inclusion of number of children in the model probably cannot be justified: the underlined probability term indicates that in this sample the observed dependency could very easily have happened by chance. The model is simplified by removing this variable:

```
        *** Linear Model ***

Call: lm(formula = Flow ~ Occupants + Boiler + Region, data = dhw, na.action =
        na.exclude)
Residuals:
    Min     1Q Median    3Q   Max
 -153.5 -50.67 -12.31 33.47 511.9

Coefficients:
               Value Std. Error  t value Pr(>|t|)
(Intercept)  49.2777  22.1896     2.2208   0.0286
  Occupants  27.4828   6.6276     4.1467   0.0001
     Boiler -11.3014  10.1293    -1.1157   0.2672
    Region1  -4.1081  14.2253    -0.2888   0.7733
    Region2  13.4953   8.4238     1.6020   0.1123
    Region3  -7.0366   4.8491    -1.4511   0.1498

Residual standard error: 90.1 on 101 degrees of freedom
Multiple R-Squared: 0.1833
F-statistic: 4.533 on 5 and 101 degrees of freedom, the p-value is 0.0009057
```

In order to check that this step is justified, the two models are compared using an analysis of the variances of the original and the simplified models:

```
Analysis of Variance Table

Response: Flow

                                   Terms Resid. Df      RSS      Test Df
1 Occupants + Children + Boiler + Region       100 818466.8
2            Occupants + Boiler + Region       101 819983.8 -Children -1
  Sum of Sq   F Value     Pr(F)
1
2 -1517.027 0.1853498 0.6677419
```

The result confirms that the very slight increase in error variance resulting from removing number of children could be expected to occur 67% of the time if there was no dependence: again the relevant term is underlined. The simplified model is accepted.

In the new model it appears that knowing the region a dwelling is in does not improve the estimation of water consumption. Once again the model is simplified, and the result is shown below:

```
       *** Linear Model ***

Call: lm(formula = Flow ~ Occupants + Boiler, data = dhw, na.action = na.exclude)
Resduals:
    Min     1Q Median    3Q   Max
 -153.5 -51.09 -22.19 32.32 552.5

Coefficients:
               Value Std. Error  t value Pr(>|t|)
(Intercept)  48.2082  22.3050     2.1613   0.0330
  Occupants  26.1853   6.6107     3.9611   0.0001
     Boiler -14.3085   9.1291    -1.5673   0.1201

Residual standard error: 90.83 on 104 degrees of freedom
Multiple R-Squared: 0.1454
F-statistic: 8.847 on 2 and 104 degrees of freedom, the p-value is 0.0002831
```

Again, the simpler of the two models is compared with its predecessor using analysis of variance:

```
Analysis of Variance Table

Response: Flow

                        Terms Resid. Df      RSS    Test Df Sum of Sq
1 Occupants + Boiler + Region       101 819983.8
2          Occupants + Boiler       104 858006.8 -Region -3 -38022.98
   F Value     Pr(F)
1
2 1.561137 0.2034946
```

The simplification is found to be justified. The simpler model reveals that boiler type may not be a significant driving force. Removing boiler type model results in a model of water flow as a function of the number of occupants alone (as currently used in BREDEM):

```
       *** Linear Model ***

Call: lm(formula = Flow ~ Occupants, data = dhw, na.action = na.exclude)
Residuals:
    Min     1Q Median    3Q   Max
 -163.2 -48.29 -17.35 29.47 541.7

Coefficients:
              Value Std. Error t value Pr(>|t|)
(Intercept) 45.5686 22.3951     2.0348  0.0444
  Occupants 25.7850  6.6514     3.8766  0.0002

Residual standard error: 91.46 on 105 degrees of freedom
Multiple R-Squared: 0.1252
F-statistic: 15.03 on 1 and 105 degrees of freedom, the p-value is 0.0001848
```

This is exactly the model fitted in the main text of this report, and the coefficients are the same as those reported in Table 6.2.

As expected, comparison between the two models confirms that this final simplification can be justified: there is a 12% chance that the observed dependency could have occurred by chance:

```
Analysis of Variance Table

Response: Flow

               Terms Resid. Df      RSS    Test Df Sum of Sq  F Value
1 Occupants + Boiler       104 858006.8
2          Occupants       105 878273.7 -Boiler -1 -20266.92 2.456577
      Pr(F)
1
2 0.1200713
```

It may be that a more sophisticated model, which takes into account the interactions between number of occupants and boiler type can provide a better fit to the data. The final model re-introduces boiler type, but also includes those interactions:

```
       *** Linear Model ***

Call: lm(formula = Flow ~ Occupants * Boiler, data = dhw, na.action = na.exclude)
Residuals:
    Min     1Q Median    3Q   Max
 -140.2 -50.33 -24.92 20.14 544.5

Coefficients:
                    Value Std. Error  t value Pr(>|t|)
     (Intercept)  31.6837  24.0258     1.3187   0.1902
       Occupants  31.7728   7.2851     4.3614   0.0000
          Boiler  24.6018  24.0258      .0240   0.3082
Occupants:Boiler -12.7342   7.2851    -1.7480   0.0834

Residual standard error: 89.95 on 103 degrees of freedom
Multiple R-Squared: 0.17
F-statistic: 7.033 on 3 and 103 degrees of freedom, the p-value is 0.0002387
```

This is now tested against both the model which includes boiler type only as a main variable, and also against the previous model which does not include boiler type at all:

```
Analysis of Variance Table

Response: Flow

               Terms Resid. Df      RSS                     Test Df Sum of Sq
1          Occupants       105 878273.7
2 Occupants * Boiler       103 833287.8 +Boiler+Occupants:Boiler  2  44985.93
   F Value      Pr(F)
1
2 2.780283 0.06668131
```

```
Analysis of Variance Table

Response: Flow

               Terms Resid. Df      RSS              Test Df Sum of Sq
1 Occupants + Boiler       104 858006.8
2 Occupants * Boiler       103 833287.8 +Occupants:Boiler  1  24719.01
   F Value      Pr(F)
1
2 3.055437 0.08344719
```

These tests confirm that in neither case can the inclusion of boiler type in the model be justified.

The residuals which remain after the ANCOVA process are skewed towards the right (like the original flow data shown on Figure 6.1). They are therefore not normally distributed, a requirement for the ANCOVA to be fully effective. There are two possible approaches to this problem: either use a test designed for non-normal data (in this case the Kruskal-Wallis test replaces the ANOVA part of the ANCOVA) or transform the data until it is normal. Carrying this test out for each driving variable yields the following results:

```
        Kruskal-Wallis rank sum test

data:  Flow and Occupants from data set dhw
Kruskal-Wallis chi-square = 28.4156, df = 6, p-value = 0.0001
alternative hypothesis: two.sided


        Kruskal-Wallis rank sum test

data:  Flow and Children from data set dhw
Kruskal-Wallis chi-square = 11.1957, df = 3, p-value = 0.0107
alternative hypothesis: two.sided


        Kruskal-Wallis rank sum test

data:  Flow and Boiler from data set dhw
Kruskal-Wallis chi-square = 4.2635, df = 1, p-value = 0.0389
alternative hypothesis: two.sided


        Kruskal-Wallis rank sum test

data:  Flow and Region from data set dhw
Kruskal-Wallis chi-square = 3.2156, df = 3, p-value = 0.3596
alternative hypothesis: two.sided
```

This analysis supports the results above demonstrating that, when taken individually, number of occupants is the most useful factor when modelling daily volumetric use. The previous analysis has already shown that once this variable has been used the number of children is no longer of any significance. Unlike the parametric analysis above, these tests suggest that boiler type could possibly be a useful factor, but its impact cannot be further analysed without more data.

A2.2: Analysis of energy use

The second modelling exercise uses the same approach to discover how the variation in hot water energy use depends on the possible driving factors. As described in the main text, volumetric and energy consumptions are closely correlated, and as expected the second modelling exercise follows quite closely that in Appendix 2. As before, the first model fitted uses all the available factors:

```
        *** Linear Model ***

Call: lm(formula = Energy ~ Occupants + Children + Boiler + Region, data = dhw,
        na.action = na.exclude)
Residuals:
    Min     1Q Median    3Q   Max
 -22.51 -6.693 -1.378 5.406 48.84

Coefficients:
              Value Std. Error t value Pr(>|t|)
(Intercept)  5.6176  3.2244     1.7422  0.0845
  Occupants  3.9173  1.2552     3.1207  0.0024
   Children  0.2078  1.5782     0.1317  0.8955
     Boiler -0.9469  1.2848    -0.7370  0.4629
    Region1  0.5425  1.8006     0.3013  0.7638
    Region2  1.3688  1.0641     1.2863  0.2013
    Region3 -0.5924  0.6132    -0.9661  0.3363

Residual standard error: 11.38 on 100 degrees of freedom
Multiple R-Squared: 0.2089
F-statistic: 4.401 on 6 and 100 degrees of freedom, the p-value is 0.0005468
```

As with flow rate, the model indicates that Region can be dropped from the list of factors.

```
        *** Linear Model ***

Call: lm(formula = Energy ~ Occupants + Children + Boiler, data = dhw, na.action =
        na.exclude)
Residuals:
   Min     1Q Median    3Q   Max
 -21.8 -6.273 -1.874 5.097 52.79

Coefficients:
              Value Std. Error t value Pr(>|t|)
(Intercept)  5.5007  3.2149     1.7110  0.0901
  Occupants  3.8755  1.2449     3.1130  0.0024
   Children  0.1419  1.5689     0.0904  0.9281
     Boiler -1.3650  1.1434    -1.1938  0.2353

Residual standard error: 11.35 on 103 degrees of freedom
Multiple R-Squared: 0.189
F-statistic: 8 on 3 and 103 degrees of freedom, the p-value is 0.00007612
```

Comparison of the two models confirms that this simplification is appropriate:

```
Analysis of Variance Table

Response: Energy

                                   Terms Resid. Df      RSS    Test Df
1 Occupants + Children + Boiler + Region       100 12953.22
2          Occupants + Children + Boiler       103 13279.54 -Region -3
  Sum of Sq   F Value     Pr(F)
1
2 -326.3241 0.8397502 0.4752182
```

As before, the next factor to be dispensed with is the number of children in the dwelling:

```
        *** Linear Model ***

Call: lm(formula = Energy ~ Occupants + Boiler, data = dhw, na.action = na.exclude)
Residuals:
    Min     1Q Median    3Q   Max
 -21.65 -6.275 -1.891 5.033 52.77

Coefficients:
              Value Std. Eror t value Pr(>|t|)
(Intercept)  5.3560  2.7750     1.9301  0.0563
  Occupants  3.9597  0.8225     4.8145  0.0000
     Boiler -1.3713  1.1358    -1.2074  0.2300

Residual standard error: 11.3 on 104 degrees of freedom
Multiple R-Squared: 0.1889
F-statistic: 12.11 on 2 and 104 degrees of freedom, the p-value is 0.00001868
```

The analysis of variance again supports the simpler model:

```
Analysis of Variance Table

Response: Energy

                          Terms Resid. Df      RSS      Test Df Sum of Sq
1 Occupants + Children + Boiler       103 13279.54
2           Occupants + Boiler       104 13280.60 -Children -1 -1.054627
      F Value     Pr(F)
1
2 0.008179989 0.9281106
```

In this case, the probability associated with boiler type indicates that this too might be removed from the model, leaving straightforward regression on occupancy:

```
        *** Linear Model ***

Call: lm(formula = Energy ~ Occupants, data = dhw, na.action = na.exclude)
Residuals:
    Min     1Q Median    3Q   Max
 -22.57 -6.723 -1.185 4.774 51.73

Coefficients:
             Value Std. Error t value Pr(>|t|)
(Intercept) 5.1030 2.7731     1.8402  0.0686
  Occupants 3.9213 0.8236     4.7611  0.0000

Residual standard error: 11.32 on 105 degrees of freedom
Multiple R-Squared: 0.1776
F-statistic: 22.67 on 1 and 105 degrees of freedom, the p-value is 6.19e-006
```

Comparison with the previous model indicates that the type of boiler is indeed not significant when predicting the energy content of the hot water consumed:

```
Analysis of Variance Table

Response: Energy

               Terms Resid. Df      RSS    Test Df Sum of Sq  F Value
1 Occupants + Boiler       104 13280.60
2          Occupants       105 13466.76 -Boiler -1 -186.1566 1.457787
      Pr(F)
1
2 0.2300214
```

Once again, inspection of the residuals remaining after the ANCOVA analysis indicates that they are not normally distributed. The non-parametric Kruskal-Wallis tests confirm the conclusions outlined above:

```
        Kruskal-Wallis rank sum test

data:  Energy and Occupants from data set dhw
Kruskal-Wallis chi-square = 28.509, df = 6, p-value = 0.0001
alternative hypothesis: two.sided


        Kruskal-Wallis rank sum test

data:  Energy and Children from data set dhw
Kruskal-Wallis chi-square = 11.2176, df = 3, p-value = 0.0106
alternative hypothesis: two.sided


        Kruskal-Wallis rank sum test

data:  Energy and Boiler from data set dhw
Kruskal-Wallis chi-square = 2.3732, df = 1, p-value = 0.1234
alternative hypothesis: two.sided


        Kruskal-Wallis rank sum test

data:  Energy and Region from data set dhw
Kruskal-Wallis chi-square = 2.9107, df = 3, p-value = 0.4056
alternative hypothesis: two.sided
```

BREDEM currently approaches the task of calculating daily hot water energy requirement by first estimating flow and then applying a temperature rise. More detailed modelling of the temperature data gathered in this project is presented in Appendices 5 and 6.

APPENDIX 3: Presenting run-off data in CEN Mandate 324 format

CEN Mandate 324 specifies a series of hot water run of profiles, or tapping cycles, to be used for testing and labelling hot water appliances. The profiles are presented as a list of run-offs over a 24 hour measurement cycle. Within that cycle the starting times of each run off and the total energy content of hot water tapped are specified.

Three different profiles are specified, containing different numbers of run-offs. Table A3.1 summarises the properties of the three sequences.

Cycle number	Number of run-offs	Total energy (kWh)	Total volume (litres)
1	11	2.100	36.0
2	23	5.845	100.2
3	24	11.655	199.8

Table A3.1: EU reference tapping cycles

Preliminary examination indicates that the average energy consumption observed in this study, 15.5 MJ or 4.3 kWh/day) and the average volumetric consumption of 116 litres/day, place the data in the middle of the range covered by the CEN cycles.

To convert the average energy profile developed in Section 6.1.2 to this format it is simply necessary to convert the energy content of the water used at each stage in the day from the SI units of MJ to kWh. Deriving the required flow is slightly more taxing. For each run-off the CEN mandate specifies a flow rate equal to either the specified flow rate of the heating regular under test, or a reduced rate which is 2/3 that value. Finally, it presents an equivalent run off volume, based on the assumption that all the hot water required has been heated for 10 to 60°C.

In the data generated for this project there is no specified flow rate. Furthermore, because the mean energy use profile is averaged across all dwellings it inevitably becomes spread across the full 24 hour period (whatever the time of day, there will always be someone using hot water). It would therefore not make sense to test or model a water heating system using this data in its raw state.

The following procedure has been used to generate a profile which attempts to resolve these problems, whilst still representing the data obtained in this trial. The volumetric flow in each hour long bin is classified according to whether it is greater than 2/3 of the maximum flow, in the range 1/3 to 2/3 maximum flow or below 1/3. The run off is then classified as type S, type R or a null. For periods classified as null run-offs the energy in the measured profile for that hour is carried forward to the next non-null period. In this way the total energy content of the tapping cycle is the same as that in the measured profile. The net effect of this rather arbitrary treatment of the data is that all hot water use before 5:00 am or after 10:00 pm is re-allocated to the period between 5:00 and 6:00 am in the morning.

It has already been observed that the temperature rise discovered in this study is significantly lower than 50°C, and the effect of this is that when the ‘Equivalent volume’ is calculated the value obtained is significantly lower than that actually observed.

Table A3.2 summarises the tapping cycle derived in this way from the average profiles across the whole sample used in this study.

EST reference tapping cycle			
	Start time	Energy (kWh)	Flow rate (Specific or reduced)
1	06:00	0.529	R
2	07:00	0.450	S
3	08:00	0.381	S
4	09:00	0.297	S
5	10:00	0.272	R
6	11:00	0.233	R
7	12:00	0.191	R
8	13:00	0.164	R
9	14:00	0.133	R
10	15:00	0.125	R
11	16:00	0.160	R
12	17:00	0.248	R
13	18:00	0.373	S
14	19:00	0.360	S
15	20:00	0.296	S
16	21:00	0.244	R
17	22:00	0.197	R
TOTAL		4.653	
Equivalent hot water volume at 60°C: 79.8 litres			

Table A3.2: Tapping cycle based on average profiles measured in EST study

APPENDIX 4: Parameter sensitivity analyses

Two of the analyses presented in this report rely on user defined parameters:

- the calculation of hot water delivery temperature requires that a value is chosen for the length of the delay before temperature averaging starts; and

- deriving the amount of time for which water is heated each day initially used a simple temperature threshold. As a result of the studies described in this Appendix this was subsequently refined to use a filter to track the mean temperature, and an offset with which to compare the measured primary temperature. This requires the choice of two parameters: the filter forgettal factor (or equivalently, time constant) and the offset.

When an analysis relies on a user defined parameter is used it is clearly important to determine that the sensitivity to that choice is small, or that the analysis is robust. If not, relatively small adjustments to the parameter may be used to obtain whatever result may be required.

In this appendix a simple set of sensitivity studies is described. The data analysis described in the body of the report is repeated for a series of different parameter choices, and the sensitivity is assessed. Fortunately, the two calculations described above are independent of each other, and it suffices to determine their most appropriate values individually.

A4.1: Sensitivity to hot water response time

As described in the main text, the measured hot water delivery temperature is slightly delayed from the moment a run off starts. The first set of sensitivity runs explored the importance of the chosen delay.

The analysis of the data from regular boilers was repeated using delay parameters from one to twelve records. Figure A4.1 shows the resulting mean hot water temperature across the whole sample.

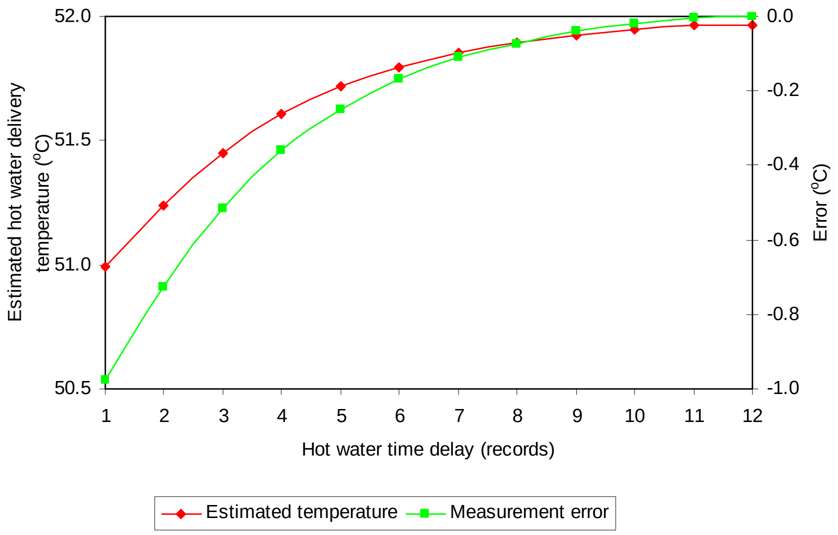

Figure A4.1: Sensitivity of estimated hot water temperature to starting delay

The figure shows that if the delay was only one record (5 seconds) then the estimated temperature would be 1°C below the value derived when a delay of 12 records (one minute) is used.

In this trial, temperatures were recorded to a resolution of 0.1°C. The figure reveals that to limit the 'error' in the estimated temperature to a similar value a wait of 7 records (35 seconds) should be used. This is the value used to produce the analysis presented in Sections 6.3.1 and 6.4.1.

A4.2: Sensitivity to primary circuit threshold temperature

The first method used to determine the number of hours for which regular boilers were used each day relied on a simple temperature threshold: if the measured temperature was above this the heating was deemed to be on, below it was assumed to be off.

Figure A4.2 shows the sensitivity of the estimated number of heated hours per day to this parameter. A total of thirteen sets of analyses was carried out using primary temperature thresholds over the full range of possible choices: from 20 to 80°C. Figure A4.2 shows the resultant estimates of number of hours heated across the regular boilers in the sample.

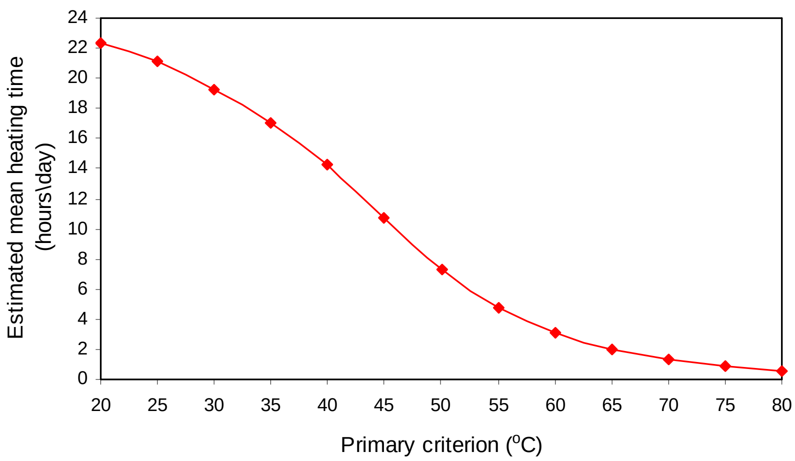

Figure A4.2: Sensitivity of estimate of hours heated to primary temperature parameter

The figure shows a very pronounced dependence on the choice of parameter, indicating that this analysis is not robust. It is possible to obtain any answer between one and thirteen hours per day by judicious choice of the threshold temperature. As described in the body of this report, this observation prompted the development of an alternative analysis.

A4.3: Sensitivity to choice of filter parameter

The next analysis method tried for the estimation of heating period used a threshold which was allowed slowly to adapt to the incoming data. This filter can be characterised by either a forgettal factor, or by the corresponding time constant. Figure A4.3 shows how the estimated heating time varies. To facilitate interpretation the time constant has been used to represent the filter. The figure shows the result of 24 analysis runs, with time constants varying from one to twenty four hours. Once again, the 'error' has been defined by comparison with the result obtained at the longest time constant.

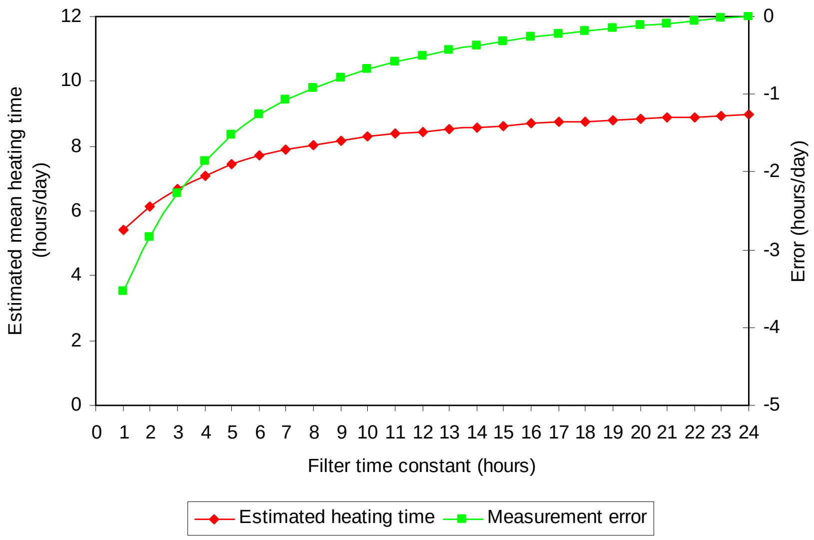

Figure A4.3: Sensitivity of estimate of hours heated to filter time constant

The data used for this analysis are the background scans, which occur every ten minutes. To reduce the error to a value less than this requires a filter with time constant 20 hours or more. The results presented all use a 20 hour filter. This corresponds to a forgettal factor of 0.0083.

When this algorithm was used, and the resulting estimates of whether heating was on or off compared to the original temperature data, it was found that the analysis was biased towards indicating that the heating was operational. The reason for this was that after any relatively long period without heating (during which the primary temperature is relatively constant) the filter output would approach that constant value. At this stage the smallest upward fluctuation in temperature would cause the algorithm to deduce that water was being heated. The effect can be seen on Figure A4.3 which indicates approximately 9 hours heating per day, a very high value. Furthermore, the estimation of whether the heating was on at this stage showed very noisy behaviour, as it flicked between on and off due to very small temperature fluctuations.

To counteract this problem an offset was introduced into the analysis. If the heating was previously on, then a temperature below the filtered average plus the offset was taken to mean that the heating had gone off. If, on the other hand, the heating was believed to be previously off, then the temperature would have to rise to a value of the filtered average plus twice the offset. In this way a dead band was introduced, and the resulting hysteresis served to eliminate the noisy behaviour. Figure A4.4 shows the sensitivity to the magnitude of this offset.

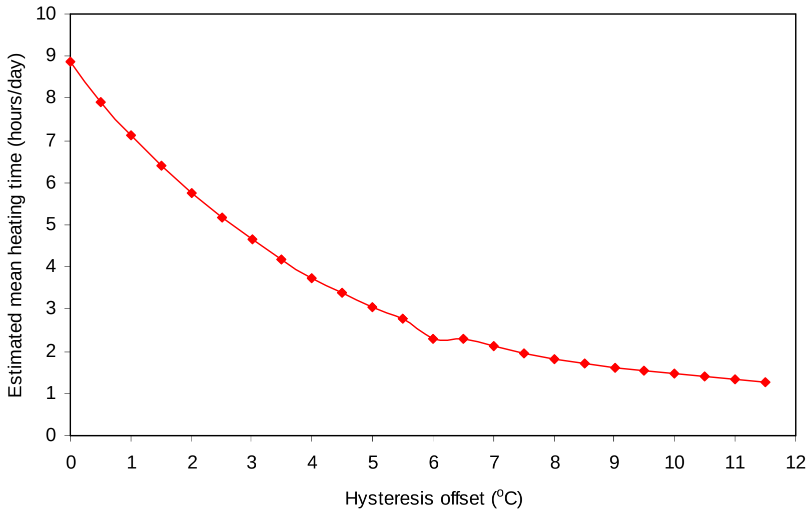

Figure A4.4: Sensitivity of number of hours heated to temperature offset

The figure shows that at an offset of 6°C there is a small plateau, indicating that in this region the estimate is not sensitive to small changes in this parameter. It would obviously be very dangerous to base the whole analysis on this single graph. The physical interpretation of the chosen value of 6°C is when the primary temperature falls below the average plus 6°C the heating has gone off, and when the temperature rises above average plus 12°C it has come back on. This is intuitively reasonable. Furthermore, detailed examination of the raw data indicates that the algorithm responds correctly.

APPENDIX 5: Further statistical analysis of hot water temperature

This Appendix describes detailed statistical analysis of hot water delivery temperature. The modelling approach is the same as that used to model volumetric and energy consumption in Appendix 2. It begins by building a model which uses all of the available factors:

```
        *** Linear Model ***

Call: lm(formula = HotT ~ Occupants + Children + Boiler + Region, data = dhw,
        na.action = na.exclude)
Residuals:
    Min     1Q  Median    3Q   Max
 -17.23 -4.047 -0.1136 3.708 17.95

Coefficients:
               Value Std. Error  t value Pr(>|t|)
(Intercept)  52.9309   1.7650    29.9892   0.0000
  Occupants  -0.8275   0.6871    -1.2043   0.2313
   Children   0.5644   0.8639     0.6533   0.5150
     Boiler   2.1244   0.7033     3.0206   0.0032
    Region1   0.9753   0.9856     0.9895   0.3248
    Region2  -1.3628   0.5825    -2.3397   0.0213
    Region3   0.0125   0.3356     0.0373   0.9703

Residual standard error: 6.23 on 100 degrees of freedom
Multiple R-Squared: 0.1403
F-statistic: 2.721 on 6 and 100 degrees of freedom, the p-value is 0.01724
```

The model indicates that the least useful factor when predicting hot water temperature is the number of children in the household. Removing this factor produces the following model:

```
        *** Linear Model ***

Call: lm(formula = HotT ~ Occupants + Boiler + Region, data = dhw, na.action =
        na.exclude)
Residuals:
    Min     1Q  Median    3Q   Max
 -17.37 -4.008 -0.2647 3.969 17.19

Coefficients:
               Value Std. Error  t value Pr(>|t|)
(Intercept)  52.3609   1.5299    34.2256   0.0000
  Occupants  -0.4930   0.4569    -1.0788   0.2832
     Boiler   2.0824   0.6984     2.9818   0.0036
    Region1   0.9335   0.9808     0.9518   0.3435
    Region2  -1.3590   0.5808    -2.3399   0.0213
    Region3   0.0227   0.3343     0.0680   0.9459

Residual standard error: 6.212 on 101 degrees of freedom
Multiple R-Squared: 0.1367
F-statistic: 3.198 on 5 and 101 degrees of freedom, the p-value is 0.01011
```

As before, the two models are compared using an Analysis of Variance:

```
Analysis of Variance Table

Response: HotT

                                   Terms Resid. Df      RSS      Test Df
1 Occupants + Children + Boiler + Region       100 3881.235
2            Occupants + Boiler + Region       101 3897.802 -Children -1
  Sum of Sq   F Value     Pr(F)
1
2 -16.56671 0.4268413 0.5150423
```

The comparison confirms that the inclusion of number of children in the model cannot be justified. The next least significant factor is the number of occupants. Removing this from the model yields:

```
        *** Linear Model ***

Call: lm(formula = HotT ~ Boiler + Region, data = dhw, na.action = na.exclude)
Resiuals:
    Min    1Q  Median    3Q   Max
 -16.91 -3.81 -0.1783 3.837 16.85

Coefficients:
               Value Std. Error  t value Pr(>|t|)
(Intercept)  50.8730   0.6628    76.7537   0.0000
     Boiler   2.0753   0.6989     2.9694   0.0037
    Region1   0.8564   0.9790     0.8748   0.3837
    Region2  -1.3636   0.5812    -2.3459   0.0209
    Region3  -0.0235   0.3318    -0.0707   0.9437

Residual standard error: 6.217 on 102 degrees of freedom
Multiple R-Squared: 0.1267
F-statistic: 3.7 on 4 and 102 degrees of freedom, the p-value is 0.00744
```

Comparing this with the previous model again indicates that the information discarded by this simplification is not significant, and that the number of occupants may be dropped from the model:

```
Analysis of Variance Table

Response: HotT

                        Terms Resid. Df      RSS       Test Df Sum of Sq
1 Occupants + Boiler + Region       101 3897.802
2             Boiler + Region       102 3942.719 -Occupants -1 -44.91716
   F Value     Pr(F)
1
2 1.163895 0.2832289
```

The final model removes region and considers a model which uses only boiler type:

```
        *** Linear Model ***

Call: lm(formula = HotT ~ Boiler, data = dhw, na.action = na.exclude)
Residuals:
    Min     1Q  Median    3Q   Max
 -20.27 -3.824 -0.1471 4.141 16.79

Coefficients:
              Value Std. Error t value Pr(>|t|)
(Intercept) 51.1939  0.6362    80.4655  0.0000
     Boiler  1.7431  0.6362     2.7398  0.0072

Residual standard error: 6.335 on 105 degrees of freedom
Multiple R-Squared: 0.06672
F-statistic: 7.507 on 1 and 105 degrees of freedom, the p-value is 0.007225
```

Once again, an Analysis of Variance indicates that this final simplification is justified:

```
Analysis of Variance Table

Response: HotT

            Terms Resid. Df      RSS    Test Df Sum of Sq  F Value      Pr(F)
1 Boiler + Region       102 3942.719
2          Boiler       105 4213.631 -Region -3 -270.9126 2.336213 0.07816358
```

We conclude that the appropriate model for hot water temperature is one which considers only the type of boiler, and that this model is highly significant (the probability of observing this result by chance is less than 0.01%).

As with flow data, the residuals remaining after this process show some non-normality. Using a series of Kruskall-Wallis tests to explore the individual effect of each of the factors produces the following results:

```
        Kruskal-Wallis rank sum test

data:  HotT and Occupants from data set dhw
Kruskal-Wallis chi-square = 4.9339, df = 6, p-value = 0.5523
alternative hypothesis: two.sided


        Kruskal-Wallis rank sum test

data:  HotT and Children from data set dhw
Kruskal-Wallis chi-square = 0.94, df = 3, p-value = 0.8158
alternative hypothesis: two.sided


        Kruskal-Wallis rank sum test

data:  HotT and Boiler from data set dhw
Kruskal-Wallis chi-square = 5.5818, df = 1, p-value = 0.0181
alternative hypothesis: two.sided


        Kruskal-Wallis rank sum test

data:  HotT and Region from data set dhw
Kruskal-Wallis chi-square = 2.4521, df = 3, p-value = 0.484
alternative hypothesis: two.sided
```

This non-parametric analysis confirms the results from the ANCOVA: only boiler type is significant when explaining hot water delivery temperature, and that it is highly significant.

APPENDIX 6: Further statistical analysis of cold water temperature

The model BREDEM derives the energy required for hot water production by first predicting volumetric consumption, and then assuming that the water is heated through a fixed temperature rise. In order to calculate this rise it is necessary to estimate both the temperature of hot water delivery and that of the incoming cold feed. This Appendix presents a statistical analysis of observed cold feed temperatures.

As before, the cold water supply temperatures measured in this project are first modelled as a function of all the available factors:

```
        *** Linear Model ***

Call: lm(formula = ColdT ~ Occupants + Children + Boiler + Region, data = dhw,
        na.action = na.exclude)
Residuals:
    Min     1Q  Median    3Q   Max
 -4.012 -1.258 -0.3256 0.935 7.416

Coefficients:
               Value Std. Error  t value Pr(>|t|)
(Intercept)  16.6599   0.6157    27.0602   0.0000
  Occupants  -0.6289   0.2397    -2.6240   0.0101
   Children   0.2209   0.3013     0.7329   0.4653
     Boiler   1.2645   0.2453     5.1545   0.0000
    Region1  -0.8893   0.3438    -2.5865   0.0111
    Region2  -0.2875   0.2032    -1.4150   0.1602
    Region3  -0.0091   0.1171    -0.0777   0.9382

Residual standard error: 2.173 on 100 degrees of freedom
Multiple R-Squared: 0.3732
F-statistic: 9.922 on 6 and 100 degrees of freedom, the p-value is 1.422e-008
```

From this preliminary fit it is clear that it should be acceptable to drop number of children from the model:

```
        *** Linear Model ***

Call: lm(formula = ColdT ~ Occupants + Boiler + Region, data = dhw, na.action =
        na.exclude)
Residuals:
    Min     1Q  Median    3Q   Max
 -4.046 -1.236 -0.2763 1.014 7.418

Coefficients:
               Value Std. Error  t value Pr(>|t|)
(Intercept)  16.4368   0.5339    30.7841   0.0000
  Occupants  -0.4980   0.1595    -3.1228   0.0023
     Boiler   1.2481   0.2437     5.1206   0.0000
    Region1  -0.9056   0.3423    -2.6457   0.0095
    Region2  -0.2860   0.2027    -1.4109   0.1613
    Region3  -0.0051   0.1167    -0.0438   0.9651

Residual standard error: 2.168 on 101 degrees of freedom
Multiple R-Squared: 0.3698
F-statistic: 11.85 on 5 and 101 degrees of freedom, the p-value is 4.949e-009
```

Comparison of the two models using Analysis of Variance confirms that this factor can indeed be dispensed with:

```
Analysis of Variance Table

Response: ColdT

                                   Terms Resid. Df      RSS      Test Df
1 Occupants + Children + Boiler + Region       100 472.2405
2            Occupants + Boiler + Region       101 474.7773 -Children -1
  Sum of Sq   F Value     Pr(F)
1
2  -2.53677 0.5371775 0.4653194
```

The next candidate for removal from the model is region:

```
        *** Linear Model ***

Call: lm(formula = ColdT ~ Occupants + Boiler, data = dhw, na.action = na.exclude)
Residuals:
    Min     1Q  Median     3Q  Max
 -4.416 -1.269 -0.3298 0.9604 7.31

Coefficients:
               Value Std. Error  t value Pr(>|t|)
(Intercept)  16.4032   0.5454    30.0767   0.0000
  Occupants  -0.5279   0.1616    -3.2657   0.0015
     Boiler   1.4045   0.2232     6.2923   0.0000

Residual standard error: 2.221 on 104 degrees of freedom
Multiple R-Squared: 0.3191
F-statistic: 24.37 on 2 and 104 degrees of freedom, the p-value is 2.089e-009
```

However, the comparison with the previous model reveals that the contribution made by region is significant, and it should therefore remain in the model. This is as expected, given the differences which were noted in Section 6.4:

```
Analysis of Variance Table

Response: ColdT

                        Terms Resid. Df      RSS    Test Df Sum of Sq
1 Occupants + Boiler + Region       101 474.7773
2          Occupants + Boiler       104 512.9556 -Region -3 -38.17829
   F Value      Pr(F)
1
2 2.707239 0.04921675
```

The next most possible candidate for exclusion is the number of occupants:

```
        *** Linear Model ***

Call: lm(formula = ColdT ~ Boiler + Region, data = dhw, na.action = na.excude)
Residuals:
    Min     1Q  Median     3Q   Max
 -4.365 -1.377 -0.4753 0.9607 8.571

Coefficients:
               Value Std. Error  t value Pr(>|t|)
(Intercept)  14.9337   0.2409    62.0037   0.0000
     Boiler   1.2409   0.2540     4.8861   0.0000
    Region1  -0.9835   0.3557    -2.7647   0.0068
    Region2  -0.2906   0.2112    -1.3759   0.1719
    Region3  -0.0518   0.1206    -0.4295   0.6685

Residual standard error: 2.259 on 102 degrees of freedom
Multiple R-Squared: 0.3089
F-statistic: 11.4 on 4 and 102 degrees of freedom, the p-value is 1.095e-007
```

Once again, the comparison with the previous model indicates that details of occupancy should be retained:

```
Analysis of Variance Table

Response: ColdT

                        Terms Resid. Df      RSS       Test Df Sum of Sq
1 Occupants + Boiler + Region       101 474.7773
2             Boiler + Region       102 520.6195 -Occupants -1 -45.84219
   F Value       Pr(F)
1
2 9.752069 0.002337132
```

The final possibility for model simplification is to drop boiler type from the model:

```
       *** Linear Model ***

Call: lm(formula = ColdT ~ Occupants + Region, data = dhw, na.action = na.exclude)
Residuals:
    Min     1Q  Median    3Q   Max
 -4.885 -1.402 -0.3065 1.14 9.088

Coefficients:
               Value Std. Error  t value Pr(>|t|)
(Intercept)  16.6956   0.5936    28.1246   0.0000
  Occupants  -0.4903   0.1781    -2.7531   0.0070
    Region1  -1.3718   0.3685    -3.7224   0.0003
    Region2  -0.2562   0.2263    -1.1320   0.2603
    Region3   0.1973   0.1226     1.6093   0.1106

Residual standard error: 2.421 on 102 degrees of freedom
Multiple R-Squared: 0.2062
F-statistic: 6.623 on 4 and 102 degrees of freedom, the p-value is 0.00008852
```

Again, the dependencies observed in Section 6.3.3 suggest that boiler type should be an important factor, and this is confirmed by the Analysis of Variance.

```
Analysis of Variance Table

Response: ColdT

                        Terms Resid. Df      RSS    Test Df Sum of Sq
1 Occupants + Boiler + Region       101 474.7773
2          Occupants + Region       102 598.0352 -Boiler -1 -123.2579
   F Value         Pr(F)
1
2 26.22083 1.461271e-006
```

It is concluded that none of these variables should be dropped, and that a model based on occupancy, boiler type and region is most appropriate. In this case the non-parametric tests of association reveal that the links between boiler type and region are significant, but warn that without the assumption of Normality the dependence on occupants may not be significant.

```
        Kruskal-Wallis rank sum test

data:  ColdT and Occupants from data set dhw
Kruskal-Wallis chi-square = 8.745, df = 6, p-value = 0.1884
alternative hypothesis: two.sided


        Kruskal-Wallis rank sum test

data:  ColdT and Children from data set dhw
Kruskal-Wallis chi-square = 4.4914, df = 3, p-value = 0.2131
alternative hypothesis: two.sided


        Kruskal-Wallis rank sum test

data:  ColdT and Boiler from data set dhw
Kruskal-Wallis chi-square = 36.8258, df = 1, p-value = 0
alternative hypothesis: two.sided


        Kruskal-Wallis rank sum test

data:  ColdT and Region from data set dhw
Kruskal-Wallis chi-square = 20.2523, df = 3, p-value = 0.0002
alternative hypothesis: two.sided
```