UCL ENERGY INSTITUTE

INVESTIGATING VARIATIONS IN PERFORMANCE OF HEAT PUMPS INSTALLED VIA THE RENEWABLE HEAT PREMIUM PAYMENT (RHPP) SCHEME

Issued: February 2017

RAPID-HPC Authors: Jenny Love, Alex Summerfield, Phillip Biddulph, Jez Wingfield, Chris Martin, Colin Gleeson, Robert Lowe

Table of contents

List of figures

List of tables

Nomenclature

PERFORMANCE EFFICIENCY NOMENCLATURE

COP	Heat pump (HP) coefficient of performance
SPFHn	HP seasonal performance factor for heating at SEPEMO boundary Hn

MONITORED VARIABLES

Eb	Electricity for whole system boost only
Edhw	Electricity for domestic hot water (typically an immersion heater)
Ehp	Electricity for the heat pump unit (may include a booster heater and circulation pump)
Esp	Electricity for boost to space heating only
Fhp	Flow rate of water from heat pump (may be space heating only)
Fhw	Flow rate of water to DHW cylinder (if separately monitored)
Hhp	Heat from heat pump (may be space heating only)
Hhw	Heat to DHW cylinder (if separately monitored)
Tco	Temperature of water leaving the condenser
Tin	For ASHP: Temperature of refrigerant leaving the evaporator For GSHP: Temperature of ground loop water into the heat pump
Tsf	Flow temperature of water to space heating
Twf	Flow temperature of water to cylinder

(Note that external temperature, Tex, was not measured directly. Data from a publicly available database were used in the analysis.)

RHPP ENERGY AND POWER UNITS

Energy	J	Joule	SI unit of energy
Energy	kWh	3.6 MJ	Customary unit of energy for residential energy use
Energy	MWh, GWh	3.6 GJ, 3.6 TJ	
Power	W	Watt, J/s	SI unit of power and heat flow
Power	Wh/2 minutes	30 W	Base unit of energy for monitored data in RHPP trial, limit of resolution of power – note that power and heat have been recorded at 2 minute intervals
Power	kWh/year	3.6 MJ/year 0.11416 W	Customary unit for rate of residential energy use
Power	kW	1000 W	Typical unit for measurement of heating system ratings

KEY ACRONYMS AND ABBREVIATIONS

BEIS	Department of Energy and Climate Change (became part of the Department for Business, Energy & Industrial Strategy on 14th July 2016)
EST	Energy Saving Trust
Preliminary Assessment	Preliminary assessment of the RHPP data performed by BEIS (Wickins, 2014)

RAPID-HPC	Research and Analysis on Performance and Installation Data – Heat Pump Consortium
RHPP	Renewable Heat Premium Payment Scheme
MCS	Microgeneration Certification Scheme - a nationally recognised quality assurance scheme, supported by the BEIS. MCS certifies microgeneration technologies used to produce electricity and heat from renewable sources.
MIS	Microgeneration installation standards. MIS 3005 set out requirements for MCS contractors undertaking the supply, design, installation, set to work, commissioning and handover of microgeneration heat pump systems.
SEPEMO	SEasonal PErformance factor and Monitoring

Context

The RHPP policy provided subsidies for private householders, Registered Social Landlords and communities to install renewable heat measures in residential properties. Eligible measures included air and ground-source heat pumps, biomass boilers and solar thermal panels.

Around 14,000 heat pumps were installed via this scheme. DECC funded a detailed monitoring campaign, which covered 700 heat pumps (around 5% of the total). The aim of this monitoring campaign was to provide data to enable an assessment of the efficiencies of the heat pumps and to gain greater insight into their performance. The RHPP scheme was administered by the Energy Savings Trust (EST) who engaged the Buildings Research Establishment (BRE) to run the meter installation and data collection phases of the monitoring program. They collected data from 31 October 2013 to 31 March 2015.

RHPP funded heat pumps were installed between 2009 and 2014. Since the start of the RHPP Scheme, the installation requirements set by MCS standards and processes have been updated.

DECC contracted RAPID-HPC to analyse this data. The data provided to RAPID-HPC included physical monitoring data, and metadata describing the features of the heat pump installations and the dwellings in which they were installed.

The work of RAPID-HPC consisted of cleaning the data, selection of sites and data for analysis, analysis, and the development of conclusions and interpretations. The monitoring data and contextual information provided to RAPID-HPC are imperfect and the analyses presented in this report should be considered with this in mind. Discussion of the data limitations is provided in the reports and is essential to the conclusions and interpretations presented. This report does not assess the degree to which the heat pumps assessed are representative of a general sample of domestic heat pumps in the UK. Therefore these results should not be assumed to be representative of any sample of heat pumps other than that described.

Acknowledgements

The authors gladly acknowledge the inputs to this report of Roger Nordman of SP Technical Research Institute and Tom Garrigan of BSRIA. The work has been supported throughout by colleagues at BEIS, particularly by Penny Dunbabin, Amy Salisbury and Jon Saltmarsh.

Executive Summary

1.1 Introduction

The RHPP policy provided subsidies for private householders, Registered Social Landlords and communities to install renewable heat measures in residential properties. Eligible measures included air and ground source heat pumps, biomass boilers and solar thermal panels.

Around 14,000 heat pumps were installed via this scheme. BEIS funded a detailed monitoring campaign, which covered 700 heat pumps (around 5% of the total). The aim of this monitoring campaign was to provide data to enable an assessment of the efficiencies of the heat pumps and to gain greater insight into their performance. The RHPP scheme was administered by the Energy Savings Trust (EST) who engaged the Buildings Research Establishment (BRE) to run the meter installation and data collection phases of the monitoring program. Data were collected from 31 October 2013 to 31 March 2015.

RAPID-HPC were contracted by BEIS to analyse this data. The data provided to RAPID-HPC included physical monitoring data and metadata describing the features of the heat pump (HP) installations and the dwellings in which they were installed.

This report uses exploratory and statistical approaches to examine variations in HP performance (defined in terms of the set of seasonal performance factors (SPFs)). It should be read alongside other reports in the series, namely RAPID-HPC (2017b) "RHPP report on compliance with MCS installation standards", RAPID-HPC (2017c) "Case Studies Report from the RHPP Heat Pump Monitoring Campaign" and RAPID-HPC (2017d) "DECC RHPP - Note on Systematic Errors in Physical Monitoring Data".

1.2 Statistical analysis

Variation of performance was investigated as a function of heat pump type, heat emitter type and tenure. For this analysis, the largest possible samples were used. A comparison of two models of air source heat pump was also carried out.

1.3 Exploratory analysis

Sub-samples of the data were investigated to assess the impact on efficiency of:

- Heat pump cycling
- Supplementary heating (both domestic hot water immersion and supplementary space heating using the heat pump's internal boost heater, where present)
- Control of domestic hot water immersion and/or boost heating

- Load factor
- Flow temperatures.

1.4 Data quality and metering errors

In this report, the performance data have been filtered to remove the extreme performance values (SPFH4<1.5 and SPFH4>4.5) as well as a number of other filtering steps. Despite this, the remaining performance values are not error-free. RAPID-HPC has conducted a detailed analysis of the various kinds of metering errors that have been observed in the data. Some increase the apparent SPF, others decrease it. In addition, some of the metadata (e.g. on metering schematics) provided with the electricity and heat data are incomplete and/or faulty.

1.5 Characteristics of analysis dataset

After filtering, the resulting sample, referred to in this report as the "Sample B2 (Cropped)", has the following characteristics:

Heat Pump Type	Tenure	Number
ASHP	Private	78
	Registered Social Landlord (RSL)	215
GSHP	Private	39
	Registered Social Landlord (RSL)	53

Heat Pump Type	Emitter type	Number
ASHP	Radiators	257
	Underfloor heating	28
	Both	8
GSHP	Radiators	58
	Underfloor heating	25
	Both	9

1.6 Results

The principal results of this analysis are:

a) A wide distribution of seasonal performance factors (SPF) was observed.

This appears to be due to both metering errors (of various kinds) and real differences in efficiency, caused by, for example, variations in control and use of resistance heating (immersion heaters etc.).

b) The statistical analysis showed fewer clear results than might be expected.

Although GSHPs performed better than ASHPs, and ASHP sites with underfloor heating appeared to perform better than those without, the picture on tenure was more complex. It appears that there are many confounding factors.

c) Investigation of factors that would be expected to influence performance, such as flow temperature, cycle length and domestic hot water immersion produced the following results:

1. There was no single factor that accounted for good or poor performance.

2. A very large proportion of ASHPs have 10 minute on-to-on cycling patterns. This may be due to the use of boiler thermostats or other ways in which the heat pump controls interact with those in the rest of the heating system. The median on-to-on cycling time of GSHPs was longer, at 18 minutes. Previous lab tests by EA Technology indicated that ASHPs would be expected to show a reduction in efficiency as on-to-off times decreased below 6 minutes. RAPID-HPC's analysis did not show a correlation between median on-to-on cycling period and monthly COP but this lack of relationship may have been influenced by heat metering error or the definition of cycling period used.

3. Across the sample, average winter space heating flow temperatures were generally low (<45°C), with only a few sites showing average winter flow temperatures >50°C. Low flow temperatures indicate good design practice and would be expected to result in good efficiencies.

4. During winter, underfloor space heating flow temperatures were lower, on average, than those for systems using radiators. However, there were two underfloor sites with high maximum flow temperatures (>55°C).

5. Some sites showed excessive use of direct electric immersion for domestic hot water heating and this has an adverse effect on SPFH4. On average, where measured, immersion electricity was 12% of the total, but more than half of the sites with SPFH4<2 had immersion use > 20% of total electricity.

6. There was little evidence of the use of internal boost heating (using internal electric heating cassettes), which would be expected to reduce the heat pump efficiency. This is reassuring; the 2009-2010 EST heat pump field trials showed several examples of excessive use of internal electric cassettes. Note that many heat pump designs fo not contain these cassettes.

7. There were several clear examples of poor control; for example, domestic hot water immersion being used excessively.

1.7 Further work

Smaller scale, more focussed studies are recommended to understand phenomena not possible to fully investigate from the RHPP dataset, for example:

- The cause and performance effect of short timestep cycling in ASHPs, and possible means to mitigate this;
- The role of DHW cylinder temperature control and how to use immersion heating most efficiently;
- The large spread observed in the distribution of SPF for GSHP sites with underfloor heating, and whether optimum flow temperatures for underfloor heating systems are achieved in practice;
- Issues of longer term performance degradation (say after two or more years);
- Resilience of performance to changes in occupant behaviour.

These proposed investigations all require robust methods of performance measurement to minimise uncertainty introduced by metering error and ensure that estimates of the spread in heat pump performance from future studies are less affected by monitoring system issues.

Technical Summary

Context

The RHPP policy provided subsidies for private householders, Registered Social Landlords and communities to install renewable heat measures in residential properties. Eligible measures included air and ground source heat pumps, biomass boilers and solar thermal panels.

Around 14,000 heat pumps were installed via this scheme. BEIS funded a detailed monitoring campaign, which covered 700 heat pumps (around 5% of the total). The aim of this monitoring campaign was to provide data to enable an assessment of the efficiencies of the heat pumps and to gain greater insight into their performance. The RHPP scheme was administered by the Energy Savings Trust (EST) who contracted the Buildings Research Establishment (BRE) to run the meter installation and data collection phases of the monitoring program. Data were collected from 31 October 2013 to 31 March 2015.

RAPID-HPC were contracted by DECC to analyse this data. The data provided to RAPID-HPC included physical monitoring data and metadata describing the features of the heat pump installations and the dwellings in which they were installed.

This report uses exploratory and statistical approaches to examine variations in HP performance (defined in terms of the set of seasonal performance factors (SPFs)) in a number of ways, including basic characteristics of HP systems, HP operation, and issues of data quality. It should be read alongside other reports in the series.

Methodology

Three broad approaches were used to gain insight into the variations in performance observed:

1. Comparing SPF distributions of groups of heat pumps and using statistical tests for difference;

2. Further quantitative investigation into specific physical factors/operational features which have been found to lead to performance issues in previous field trials – for example the amount of immersion and boost electricity used.

3. Describing individual sites, illustrating physical factors that can influence performance – for example DHW heating and sterilisation cycles and heat pump control strategies

The Sample

The data sub-sample used in this analysis is labelled 'Sample B2 (cropped)'.

The RHPP trial provided high frequency (two minute) monitoring data from 699 sites with a variety of different air source heat pumps (ASHPs) and ground source heat pumps (GSHPs). Given the SPF requirements of relatively complete and stable data over a 12 month contiguous period, while balancing the need for as large a sample as possible across the different categories of HPs (such as those with radiators and those with underfloor heating), a number of *simple filters* for data quality and completeness were developed. This resulted in the selection of Sample B2 with 318 ASHPs and 99 GSHPs[1].

Based on an initial inspection of the degree of scatter in the data, the data from a small selection of sites was inspected in detail. This revealed some sites for which data may be erroneous. As a straightforward way to reduce the impact of sites with such data issues, a further restriction was applied to omit ('crop') sites outside the range for SPF_{H4} of 1.5 to 4.5. All but three of the 35 outliers removed were at the low performance end. This simple approach to removing outliers had been adopted in a previous heat pump field trial in Denmark (DTI, 2011), and is standard practice in statistical analysis.

The resultant *Sample B2 (cropped)* with 293 ASHPs and 92 GSHPs, which represents just over half (55%) of the sites in the original RHPP trial sample is used in the subsequent analysis in this report.

The presence of erroneous data within the data of a site may not just be due to metering error, as various operational or dwelling/occupant issues could be relevant. Some data patterns that appear erroneous may also be due to transient effects or spells of missing data, but these do not necessarily impact the SPF values of the site in a substantive way. So a manual inspection of data on a site by site basis would need to remove sites with signs of erroneous data from a year of measurement in a consistent way. This approach would face issues of selection bias, with a much smaller resultant sample size and likely omission of sites with genuine operational issues.

The mean heat output of the 32 sites omitted through the cropping process was less than half that for the included sites (3,787 kWh compared with 8,552 kWh), **even though installers' estimates of heat demand for the omitted sites were higher, as shown in Table 0-1.** This large discrepancy suggests there may be issues with heat metering in a sizable percentage of the omitted sites and that it is valid to exclude them from the analysis sample. However this does not mean that all sites with potentially erroneous data have been eliminated from Sample B2 (cropped); some of sites in the sample are likely to contain metering errors which may have an impact on SPF.

[1] This process is the same as was used for Sample B in the previous interim report (RAPID-HPC, 2016). The reason that Sample B2 is slightly larger than before is that an issue with one of the schematics was resolved and so a number of sites that had previously been excluded could be included in the selection process.

Table 0-1. Heat demand characteristics of sites included in and excluded from Sample B2 (cropped).

Category	N	Installer Estimated Heat Demand, kWh/yr	Measured Mean Heat Generation, kWh/yr
Included in cropped B2 sample	385	10,800	8,552
Omitted from cropped B2 sample	32	12,000	3,787

Table 0-2. Sample B2 (cropped): estimated SPFs for ASHPs and GSHPs

Sample	System boundary	HP type	N	Mean (95% CI)	Median (IQR)
B2 Cropped	**SPFH2**	ASHP	292	2.64 (2.60, 2.70)	2.65 (2.33 - 2.95)
		GSHP	92	2.93 (2.80, 3.06)	2.81 (2.63 - 3.14)
	SPFH4	ASHP	293	2.41 (2.37, 2.46)	2.44 (2.15 - 2.67)
		GSHP	92	2.77 (2.66, 2.89)	2.71 (2.48 - 3.02)

Key characteristics of the Sample B2 (cropped) are shown in Table 0-2. Note that when the term "SPF" is used in this report without qualification, it means the weighted average of both space and water heating according to the specified system boundary.

Types of Metering Error

Selection of the Sample B2 cropped sites may not have not eliminated all sites with measurement issues; a number of possible metering errors remain. The table below shows a non-exhaustive list of these.

Meter type	Potential Fault type	Description	How do we know that these faults exist?	Systematic error or an error that affects individual sites?	Effect on SPF or monthly COPs
Heat meters	Missing heat meter data, when electricity data is present.	Periods with zero or unusually low heat data were not filtered out in the data cleaning process.	Observed in data	Individual sites	Will have the effect of under-estimating SPF, by an estimated ~4% across the Sample B2 (cropped) but much higher for a few sites. Apparent slight effect on distribution of SPFs (statistical tests not carried out to confirm this).
Heat meters	Systematic under-reading due to meter installation.	Poor installation of strap-on sensors or pocket sensors RAPID-HPC removed 99 sites with known strap-on sensors at the start of the project, but suspect that others may exist.	Some suspiciously low COP readings observed in data (e.g. < 1)	Individual sites	Would reduce SPF and monthly COP but sites for which spfh4<1.5 have been filtered out of Sample B2 (cropped).
Heat meters	Systematic over-reading due to glycol correction not being applied.	Heat meters calibrated for water with no antifreeze.	Wickins (2014)	Likely to occur in many of the sites, in both ASHPs and GSHPs.	Likely to result in over-estimation of SPF by 4-7% - see separate report on systematic errors.

Heat meters	Limited to 18 kW	Up to 16 sites in Sample B2 (cropped) affected.	Observed in data	Individual sites	Expected to affect the SPFs of these sites slightly in cold weather.
Heat meters	Systematic over-reporting of heat output.	Probably due to heat meter temperature sensor offsets, exacerbated by circulation pump over-run.	Observed in data (heat output when no electricity input)	Individual sites	Will over-report SPF and COP.
Heat meters	Spikes in heat output when changing mode.	It is not known whether this is a metering problem or a real dynamic effect with no impact on estimates of heat.	Observed in data.	Individual sites. Not present in all sites, but for those in which this effect is present, it occurs every time there is a mode change.	Unknown. If real heat, no effect, if metering error, over-reports SPF and COP.
Heat meters	Transposition of Hhp and Hhw sensors.		Observed in data.	Individual sites	No effect on overall SPF or COP, but will affect space heating and DHW SPFs and COPs.
Heat meters	Flow decay over the dataset time period. Median 1.5% decay over year for Sample B2 (cropped).	Cause unknown.	Observed in data	Individual sites	Under-report SPF, and COP for later months.

Heat and electricity meters	Heat and electricity data missing at the same time.	Cause unknown.	Observed in data; of 34 sites investigated in detail, 16 had > 7 days of this.	Individual sites	Effect depends on the time of year at which the problem occurs.
Electricity meters	Suspected unmetered electricity – missing Eboost or Edhw	Temperature data shows unusual patterns which can't be explained by the existing heat and electricity data.	Observed in data	Individual sites	Over-report SPF.
Electricity meters	Transposition of electricity meters.	Transposition of Ehp and Edhw or Eboost, or Edhw and Eboost.	Observed in data. Automatic correction applied in code for cases where easily detectable but not all cases.	Individual sites	Effect depends on which sensors were involved. Overall SPFH4 unchanged but other boundaries affected. Space heating and DHW SPFs could also be affected.
Temperature	Sensors too close to other pipes.	This causes e.g. Tsf to be influenced by Twf and vice versa. This in turn affects which mode (space heating, DHW) gets attributed to each 2 minutes of data.	Observed in data and photos	Individual sites	Overall SPF and COPs are unaffected but space heating and DHW SPFs and COPs are affected.

Analysis carried out

This study investigated a range of factors relating to performance. Some investigations were based on statistical tests using the whole sample, while others were based on detailed analysis of individual sites. Distributions of SPF of the following groups are presented:

- Heat pump type (ASHP, GSHP)
- Seasonal variation in space and water heating for ASHP and GSHP
- Tenure (RSL versus Private)
- Heat Emitters (radiators versus underfloor heating)
- Different heat pump models

Quantitative investigations of physical factors that can in principle influence performance were undertaken:

- Prevalence of short-timestep cycling and length of cycling periods
- Supplementary heating: amount of domestic hot water immersion electricity used and strategy for controlling domestic hot water immersion use
- Supplementary heating: amount of internal direct electric heating used for space heating (referred to as "internal boost" electric heating)
- Heat output and load factor
- Winter flow temperatures for space heating (maximum and average)

Individual sites were explored to illustrate physical factors that can influence performance:

- DHW heating and control strategies
- Circulation pump operation
- Boost electric heating and control strategies

Analysis not carried out

Before the monitoring programme, DECC carried out a detailed cost-benefit analysis to decide which parameters should be monitored. This is discussed briefly in Wickins (2014).

It was not possible to monitor all components of heat pumps and dwellings at every installation. In particular, no data on the following parameters were recorded:

- Ground loop/borehole design
- Underfloor heating design
- Radiator sizing and design (the principles of heating system design are covered in the accompanying *RHPP MCS Compliance Report* [2])
- Type of buffer tank, sizing and control strategy
- Ground-loop circulation pump electricity use (although maximum ground loop pump power and settings are investigated in the accompanying *RHPP MCS Compliance Report*. Note that the electricity used by ground loop pumps should have been included in the measurement of electricity used by the heat pump as a whole.
- Evaporator temperatures
- Dwelling type
- Internal temperatures in each property
- Thermostat settings in each property (although this is available for the case studies)
- Supplementary heating used by householders (although this is available for the case studies).

Nonetheless, there is scope for additional analysis on the data, which will be made public on the UK Data Archive.

Results

Due to the prevalence of heat metering errors, it is pragmatic to group the results according to how much they depend on the heat meter data.

[2] Other aspects of the RHPP Field Trial are described in two parallel reports, the abbreviated titles of which are *RHPP Case Studies Report* (RAPID-HPC, 2017a) , and *RHPP MCS Compliance Report* (RAPID-HPC, 2017b).

Differences in SPF between groups of sites (using the heat meter data)

The following results, although using the heat meter data, mostly use large groups of sites. As such, the statistical analysis can still produce useful and clear results, particularly in terms of robust qualitative findings. The various metering issues appear to occur in all sub-groups, as evidenced by the similar degree of spread (e.g. from the interquartile range) seen in many of the SPF distributions. Evidence for *differences between* distributions tends to be more reliable than absolute summary statistics (magnitudes of mean, median etc.). Differences between means are given here, with their statistical confidence intervals. The credibility of results is also strengthened if the difference between groups is in the same direction as expected from a physics and/or thermodynamic perspective on heat pump performance.

Heat pump type

GSHPs in the sample tended to perform better than ASHPs at the H2 and H4 system boundaries. Including 95% confidence intervals, the difference between them was between 0.16 and 0.40 (centred on 0.28) at the H2 boundary and 0.22 to 0.45 (centred on 0.33) at the H4 boundary. Similar differences were observed in sub-groups of sites, e.g. between GSHPs and ASHPs with radiators (centred on 0.25, ranging from 0.11 to 0.38), and between GSHPs and ASHPs with underfloor heating systems (centred on 0.23, ranging from -0.15 to 0.61 as the confidence intervals here were wider). The similarity of these results for measuring the performance advantage of GSHPs over ASHPS, strengthens the quantitative findings since it suggests that metering issues tend to be spread across the dataset (at least in terms of emitter type), and are not restricted to any one group.

The GSHP and ASHP distributions have approximately the same spread (as evidenced by their interquartile ranges).

The performance advantage of GSHPs over ASHPs in the sample applies to both space heating and DHW. While the main efficiency advantage for space heating with GSHPs was during the winter months, this was no longer the case by April/May, as one would expect from thermodynamic principles. Conversely, GSHPs appeared to outperform ASHPs in their DHW heating COP all year round.

Emitter type

Underfloor heating and over-sized radiators are designed to operate with low flow temperatures. In the metadata, no information was provided on radiator sizing, so RAPID-HPC compared systems with radiators and systems with underfloor heating (hybrid sites that used both types were omitted due to small sample sizes).

On average, underfloor heating was shown to occur at lower flow temperatures than radiator heating. Underfloor systems with high flow temperatures (over 45°C) were present in 2 GSHP sites (representing

8% of the GSHP with undefloor heating sample, which was only 24 sites). It is more difficult to quantify the number of radiator systems with low flow temperatures, as the systems with the lowest flow temperatures off the condenser have some evidence indicating the presence of a boost heater between the condenser and the space heating circuit.

Figure 0-1 presents overall SPF (i.e. total heat out divided by total electricity in at the relevant boundary – which can also be thought of as a weighted average of space and water heating SPFs) as a function of heat emitter. We consider overall SPF to be a rough proxy for space heating SPF because, on average, DHW heating accounts for only 17% of the overall heat supplied.

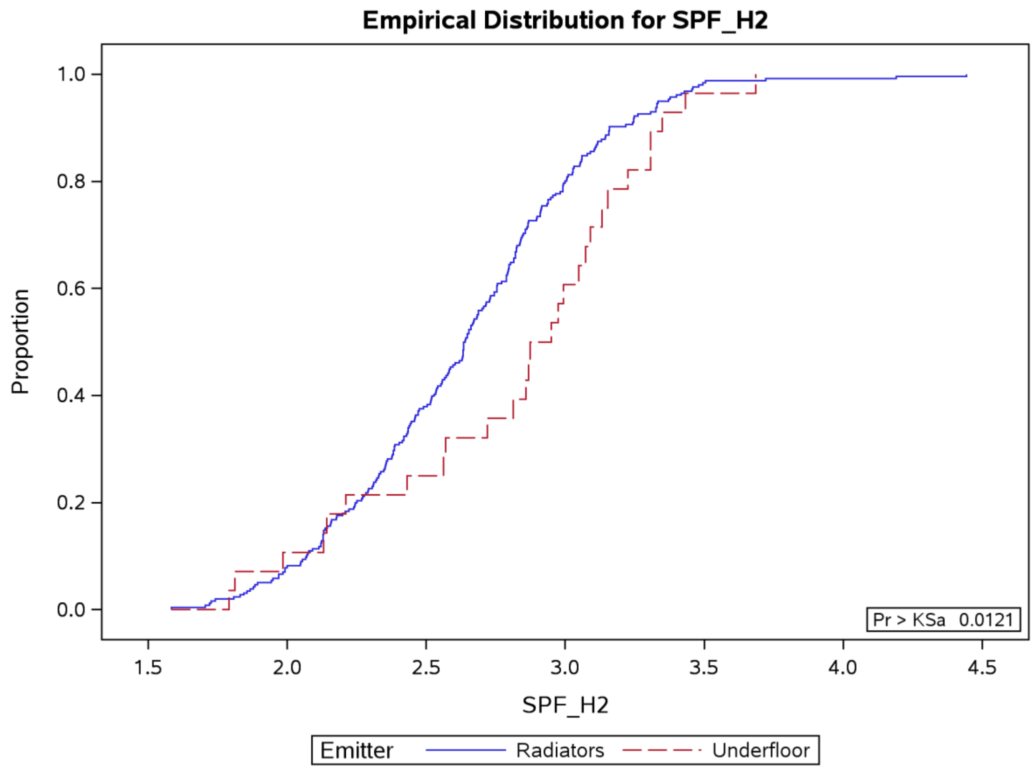


Figure 0-1. Comparison of SPFH2 by emitter type (radiators or underfloor) for ASHPs in Sample B2 (cropped).

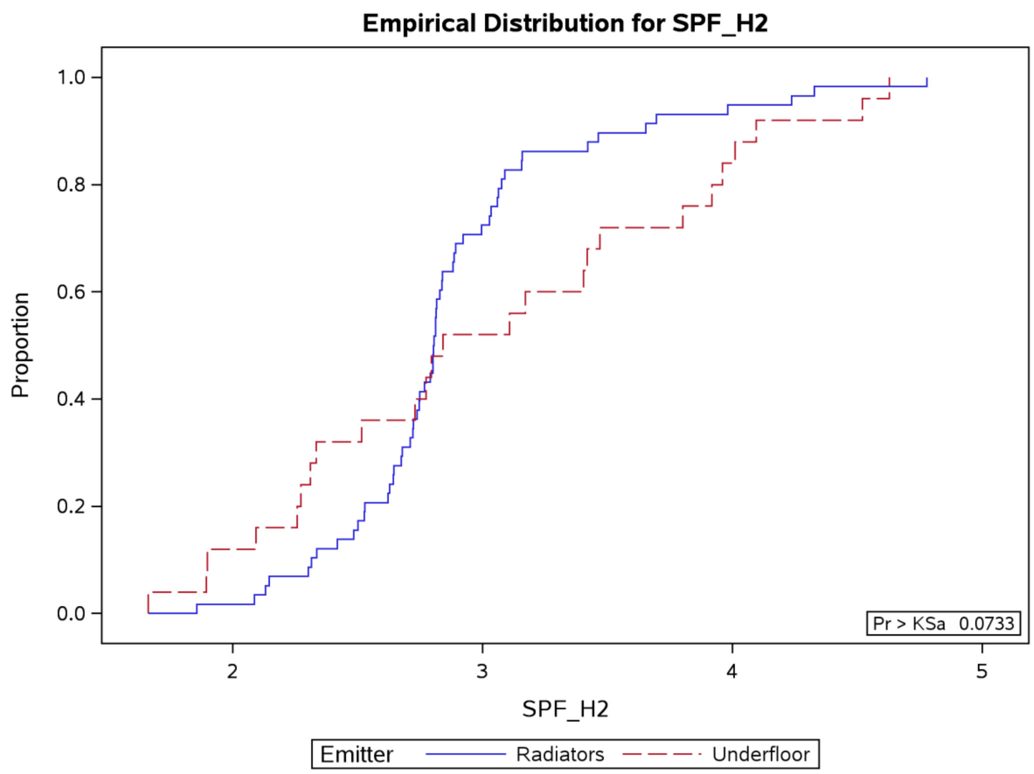


Figure 0-2. Comparison of SPFH2 by emitter type (radiators or underfloor) for GSHPs in Sample B2 (cropped).

The figure indicates that ASHPs in the sample with underfloor heating tended to have better overall performance than those with radiators at the H2 boundary. This difference was not observed at the H4 boundary which includes pumps, DHW immersion and boost heating (if present). There was also no difference in means observed at H2 or H4 for GSHPs, although the shape of the distribution is different (Figure 0-2). That is, SPFH4 performance for underfloor heating sites is very variable. The sample size was small as mentioned above, but even so, there is not even a peak in the middle of the GSHP+underfloor SPF distribution – further investigation is recommended as to why this is so much more variable than ASHP+underfloor sites.

Both figures suggest that the performance difference may be more evident for better performing sites (e.g. from around SPF > 2.5 for ASHPs), but further detailed on-site investigations would be needed to understand the underlying mechanisms at work.

Tenure

RSL and private domestic sites showed no difference in median performance. However, for GSHPs the distribution of SPFs for privately owned dwellings was wider than that of RSL sites. For ASHPs this was not the case and distributions were not significantly different.

One potential explanation for why the shape of the SPF distributions for GSHPs differs by tenure is an example of 'confounding'. From the previous section on emitter type, the GSHPs with underfloor heating were characterised by a wide variation in SPF. In this comparison by tenure for GSHPs, none of the 53 RSL sites had underfloor heating whereas 25 of the 39 Private sites had underfloor heating. So it is likely that the difference in SPF observed for RSL and Private sites actually reflects difference in emitter type. By contrast for ASHPs, only 17 out of 77 of the Private sites had underfloor heating, so their impact would be limited.

Heat pump model

Although the analysis was limited by the small sample sizes involved, some evidence for a small difference in performance was detected for one model compared with others of the same HP type. This was at the limits of what could be detected, with the median of SPFH2 for the model in question about 0.1 higher than for the median of the other models. However, another very similar model of ASHP showed no significant difference compared with other systems. The difference in performance between models was much less than the overall variation in performance – even for the better HP model – and suggests that currently other contextual factors play a more important role in affecting performance. Even if all the metering factors were addressed, few models had sufficient numbers for the analysis to have sufficient statistical power to detect a significant difference in SPF.

Characterisation of heat pump features which do not depend on heat meter data

Short timestep cycling

Two cycling metrics were investigated here: median on-to-on[3] cycle length per month per site, and proportion of cycles lasting less than 12 minutes. Cycling was investigating by applying an algorithm to the heat pump electricity data to find the time between compressor starts.

By far the most common median cycle length per month for ASHPs was 10 minutes. 49% of ASHP sites in Sample B2 (cropped) had at least one month where 10 minutes was the median cycle length. Using a slightly different metric, 25% of ASHP sites had more than 10% of their cycles less than 12 minutes long in all 12 months of the year. From visual inspection of the data, it appeared that most of the 10-minutely cycling was associated with space heating mode.

GSHPs did not show widespread occurrence of short timestep cycling. Most sites had no occurrences of months with more than 10% of cycles less than 12 minutes long, and only 1 site showed short timestep cycling all year round.

---

[3] Throughout this document, where the term "cycle length" is used with reference to the RHPP data, it means "on-to-on" cycle length.

Use of DHW immersion

Immersion heating for DHW is monitored in 68% of ASHPs and 13% of GSHPs in the sample. Immersion heaters appear to be the sole source of DHW heat in 1% of the ASHP sites and in none of the GSHP sites. Note however that these might not be the only systems which use immersion for all of the DHW provision; in some systems the immersion may not have been monitored.

Immersion heating was observed to be used for a wide range of purposes, including: legionella protection (weekly or less frequently), coming on after every HP DHW heating event possibly to boost the stored DHW temperature, being used instead of the HP according to a certain schedule, or (as noted above) being used to provide all the DHW. Of the 10 sites with the highest proportion of immersion to total electricity (these were all ASHPs), about half used immersion as well as the HP to provide DHW, and half used immersion instead of the HP to provide DHW. These latter sites all had relatively low DHW demand.

For the sites with monitored immersion, the mean contribution of immersion electricity to total electricity at the H4 bound is 12%.

Characterisation of heat pump features which depend on heat meter data

Internal boost

Monitoring electric resistance heaters internal to the heat pump unit is difficult. Because of this, internal boost heating was inferred using the relationship between heat and electricity data, and cross checked with information on whether the heat pump model actually contained an internal electric resistance heater (this was the case in 96 sites). This method is subject to error in heat meter data.

84% of these sites (87% of the ASHPs and 73% of the GSHPs) were calculated to have an estimated boost fraction of 10% or less. The median estimated boost fraction is 3.8% (3.7% for the ASHPs and 4.2% for the GSHPs). The highest estimated boost fraction for any individual system is 37%.

Heat output

Timeseries data from the sites with the lowest heat output were investigated, to determine whether there was evidence of potentially inefficient operation. The ASHPs with the lowest heat demand only came on for a few hours each day, without short cycling and with good modulation behaviour. The GSHPs with the lowest heat demand were on more continuously and did exhibit 10-minutely cycling, which may have affected their efficiency. Another mechanism which may associate low heat output with low SPF, dominance of parasitic electricity consumption, was not investigated at a sample level on this report due to differences between sites in exactly which pumps, fans and controls were monitored within the recorded electricity consumption.

Effect of above characteristics on SPF (depends on heat meter data)

Creating scatter plots of most of the above heat pump characteristics (cycle length, proportion of cycles shorter than 6 minutes, use of internal boost, heat output and load factor) did not reveal relationships. This is suspected to be at least partly because of noise in the heat meter data. Other contributing explanations could be: a) there was no expected trend, or b) to observe a trend, other variables would have had to be held constant.

There was an indication of a negative relationship between immersion use (as a fraction of total electricity) and SPFH4; this should indeed be expected.

Heat pump characteristics observed site-by-site (not using heat meter data)

As an alternative to exploring quantitative relationships between heat pump characteristics and SPF, a number of heat pump performance issues were explored qualitatively without using the heat meter data (the prevalence of phenomena in the sample was not quantified as some of them are difficult to detect algorithmically). Two of these concerned DHW heating. Evidence was found of: DHW cycles lasting unnecessarily long and causing efficiency to decrease substantially throughout the cycle, and of DHW events occurring more frequently than should ever be necessary given the presence of hot water stores, using both the heat pump and the immersion. There were also sites in which the circulation pump stayed on continuously. Further work is needed to ascertain why this was occurring. Finally, an example of space heating boost occurring with every heat pump space heating cycle was identified, indicating poor control.

Main Findings

The main findings of the investigation into variations in performance are summarised below along with the impact of suspected and observed metering errors.

Finding	Effect of metering error
Strong evidence that the mean SPF for GSHPs across the H2 and H4 boundaries are higher than those for ASHPs. For SPFH4 the difference in means was 0.28 (95% CI: 0.16 to 0.40) and at the H2 boundary 0.33 (0.22 to 0.45).	We are confident about the qualitative finding that there is a performance advantage for GSHPs. Metering errors may have impacted the estimated size of the difference in this sample, but this is also reflected in the uncertainty (confidence interval) of the difference provided.

On average GSHPs perform significantly better for both space heating (SH) and domestic hot water (DHW) at the H2 and H4 boundary	We are confident about the qualitative finding, but are more cautious as to estimates of the size of the difference, since additional sensors and data processing are needed to separate SH from DHW data.
Concerning space heating, the advantage of GSHPs over ASHPs disappears by springtime. Concerning DHW, the advantage continues throughout the year.	We are confident about the qualitative finding that these differences vary across the year.
For ASHPs, there is some evidence that sites with underfloor heating perform better than those with radiators. Although this is in the direction expected (median SPFH2 differs by 0.27) a statistically significant result was only found for SPFH2. The differences could not be detected for SPFH4. There is some limited evidence that better performing sites show more difference. For GSHPs, those with underfloor heating appear to show a wider distribution in performance than those with radiators.	For ASHPs, heat metering errors appear to be a factor in overlap at the bottom end of the distribution; for GSHPs, for GSHPs, heat metering errors have contributed to the spread of the whole distribution. Other factors, however also appear to be at work specifically for GSHP systems with underfloor emitters. The view that this reflects a real performance issue that is worthy of further investigation, is supported by the results that some underfloor systems had high flow temperatures.
Underfloor heating was shown to occur at lower flow temperatures than those in radiator systems, which is to be expected. However, high flow temperatures (over 45°C) were detected in 2 GSHP underfloor sites. Prevalence of low flow temperatures in sites with radiators is more difficult to quantify due to the indicated presence of boost heaters raising the flow temperature provided by the heat pump.	This analysis depends on the accuracy and correct positioning of the flow temperature sensors and the ability of the software to correctly detect space heating mode. The GSHP sites with underfloor heating > 45°C were verified in the timeseries data manually and found to be correct. The sites with the lowest flow temperatures observed in radiator systems were checked in the timeseries data and from there the presence of boost heaters was observed.

There is no evidence for a difference in performance between Privately owned and RSL sites with ASHPs. For GSHPs, the Private sites show a far more dispersed (spread out) distribution compared with the RSL sites, which for SPFH4 mainly lie in the 2.5-3.0 range. This may reflect aspects such as the diverse contexts of the private sites, compared with relative similarity of RSL sites.	While heat metering issues will have contributed to the spread in the distribution of both ASHPs groups, the distributions are sufficiently close that even without heat metering issues, it would be unlikely that a large difference would then be found. For GSHPs, it may be that heat metering errors caused the spread in private sites, but it may also be the case that the dwelling/occupant factors at work are also more variable.
It was only possible to undertake limited analysis of the performance of particular models against the rest of the sample. One example was shown where a model outperformed the rest by a small amount, but a very similar model showed no difference from the rest.	Heat metering issues have probably contributed to the spread in the distributions for the comparison of Model A to the rest, and possibly the size of the difference. The small sample sizes for most model types greatly limits the ability to make comparisons of performance.
Cycling time of 10 minutes was common in ASHPs, with 49% of ASHP sites having at least one month where 10 minutes was the median cycle length. This appeared to be predominantly associated with space heating mode (although this hypothesis has not been verified algorithmically), and could occur at any time of year. Further work is recommended to ascertain whether 10-minutely cycling comes about as a result of the HPs themselves or as an outcome of their installation into e.g. existing control systems of a dwelling. GSHPs did not show a modal monthly cycle length of 10 minutes but exhibited a flatter distribution of cycle lengths, with the most common at 18 minutes.	We are confident in this result with the proviso that if cycling of the order of 2 minutes were to occur it could not be detected by 2-minutely monitoring of electricity use of the compressor.

A related metric, proportion of cycles of length less than 12 minutes, showed most ASHPs had at least one month where at least 10% of cycles were shorter than 12 minutes, and in 25% of ASHPs this was the case all year round. GSHPs did not show this behaviour; only 1 site showed all year round short cycling.	
No effect on monthly COP was observed from median cycle length per month.	It is suspected that heat metering error could have masked a real trend in this case.
Immersion heating for DHW is monitored in 68% of ASHPs and 13% of GSHPs in the sample. 1% of ASHPs and 0% of GSHPs have monitored immersion and no DHW output from the heat pump. Note that these might not be the only systems which use immersion for all of their DHW; in some systems the immersion may not have been monitored. Immersion heating was observed to be used for a wide range of purposes, including: legionella protection (weekly or less frequently), coming on after every HP DHW heating event possibly to boost the stored DHW temperature, being used instead of the HP according to a certain schedule, or being used for all the DHW. Of the 10 sites with the highest proportion of immersion to total electricity (all ASHPs), about half used immersion as well as the HP to provide DHW, and half used immersion instead of the HP to provide DHW. The latter all had relatively low DHW demand. For the sites with monitored immersion, the mean contribution of immersion electricity to total electricity at the H4 bound was 12%.	These findings are considered robust to metering errors.

A negative relationship between immersion use and SPFH4 was observed. This is to be expected as immersion is a constituent of the SPFH4 metric.	Heat metering error is suspected to contribute to the noise in the relationship between SPFH4 and immersion use; however the relationship is strong enough to still be visible.
Around half of sites with SPFH4 < 2 were shown to have immersion contribution more than 20% of their total electricity.	We are confident in this result.
Internal boost heating was inferred using heat and electricity data and cross checked with information on whether the heat pump model actually contained an internal electric resistance heater (96 sites). 84% of these sites have an estimated boost fraction of 10% or less. (87% of the ASHPs, and 73% of the GSHPs); The median estimated boost fraction is 3.8% (3.7% for the ASHPs and 4.2% for the GSHPs). The highest estimated boost fractions were 37% (this occured in a GSHP); the highest boost fraction for ASHPs was 36%.	The method used to detect boost relied on detecting low values of the heat output which might equally have arisen from underreporting heat meters.
In sites with estimated boosts fraction below 10%, there is not a discernible trend of lower SPF for higher boost fractions. There is a minority of the subset of sites known to have boost heaters in which estimated boost fraction exceeds 10%; these sites all have SPFs below 2.5.	Noise in the heat meter data may have masked a real trend; it is unclear whether this is the case. This result may have been artificially created from under-reporting heat meters, which would both lower apparent SPF and increase boost fraction.
Monthly heat output and load factor did not yield a clear association with monthly COP.	Noise in the heat meter data may have masked a real trend; it is unclear whether this is the case.

The ASHPs with the lowest heat demand only came on for a few hours each day, without short cycling or any clear sign of oversizing. The GSHPs with the lowest heat demand displayed cycling behaviour at a frequency of around one cycle per 10 minutes which may lower SPF – however the investigation of SPF and cycling did not yield a trend in this dataset.	These findings were determined by detailed observation of timeseries data.
A number of real performance issues were identified by visual inspection of timeseries data and four were described here: DHW cycle going on too long, very frequent DHW heating events, circulation pump operating continuously and use of boost before compressor start-up.	These findings were determined by detailed observation of timeseries data.

Beyond the fundamental GSHP vs ASHP comparison, stating what is 'expected' in terms of differences between groups becomes more difficult. For example, the proposition that underfloor systems should be more efficient than radiator systems is plausible only if the flow temperatures in the former are lower than those in the latter. In this dataset, although this is generally the case, there are also examples of low temperature flow going to radiators (30-35°C) and relatively high temperature flow going to underfloor systems (45-55°C) – in other words, the groups overlap in terms of what one would assume would be the fundamental physical determinant of performance difference, making it unlikely that a difference would be detected. Another example is the proposition that systems with low heat output or low load factor should necessarily be less efficient; here it was shown that the ASHP sites with the lowest heat output displayed good modulation behaviour without short timestep cycling or other behaviours which would decrease their efficiency.

There was not evidence of any single factor – such as boost heating - being the key to explaining variations in performance, or of any specific HP model as showing outstanding SPF. Instead, the lack of clear relationships suggests that an array of factors, each capable of affecting performance, ranging from quality of installation to details of operation over the longer term and other dwelling/occupant related factors, are probably at work. These may be inherently difficult to unravel analytically from remotely monitored field data.

Nonetheless, the analysis in this report pointed towards the following areas for future research to help to improve systems and practice.

- Why was short timestep cycling so common in the ASHPs in the trial, does this decrease SPF, and if so how can this be mitigated?

- What are the most efficient DHW heating strategies in different dwelling types, and is there a case for DHW heating being carried out solely by electric immersion in sites with low DHW demand?

- How are DHW cylinder temperatures controlled in sites which do and do not exhibit unnecessarily frequent DHW heating events?

- How are high flow temperatures in sites with underfloor heating systems coming about?

- What is causing the especially large spread in the distribution of SPF for GSHP sites with underfloor heating?

In terms of further field testing and analytical work, considerable scope remains for small scale and far more detailed and focussed studies. These could include technical or other interventions to address any specific issues identified, including dwelling and occupant factors. Some key areas these studies could examine are methods of robust performance measurement, issues of longer term performance degradation (for example, after two or more years), the role of dwelling characteristics, and resilience of performance to changes in occupant behaviour. Recommendations for how to carry out this research are presented in the Final Report in this series.

1 Introduction

1.1 Context

The need to develop the UK supply for domestic heat pumps (HPs) and to evaluate the empirical performance of HP systems in the field has led to the establishment of two major UK field trials of HPs since 2000. The first took place in two phases: Phase I, conducted by the Energy Saving Trust between 2008 and 2010 (EST, 2010; Dunbabin & Wickins, 2012) and Phase II, conducted by EST and BEIS between 2011 and 2012 (Dunbabin et al. 2013). The second field trial, upon which this report is based, was established by BEIS in conjunction with the Renewable Heat Premium Payment (RHPP) grant scheme, which ran from 2011 to 2014 (DECC, 2014). This scheme was designed to support the replacement of fossil-fired and electric resistance heating systems with heat pumps in dwellings not supplied with natural gas. They included several makes and types of ground-source HP systems (GSHPs) and air-source HP systems (ASHPs), located in a range of domestic properties across Great Britain.

The RHPP heat pump trial monitored systems at just over 700 of these sites as the basis for an evaluation of their performance. Several reports on the trial have been published, on the following aspects:

- The data collection process (Wickins, 2014);
- The raw data quality, the cleaning process undertaken and the methods of construction of different samples for analysis, including a weather-adjusted sample (RAPD HPC, 2016);
- The overall performance of HP systems in the RHPP sample, in terms of SPF and other metrics such as renewable energy generation as defined under the EU Renewable Energy Directive (RAPID HPC, 2016).

The previous work identified a large range in measured HP performance (defined in terms of the set of SPF metrics) across the sample. This current report takes this observed performance variation as its starting point and uses statistical and exploratory techniques to identify evidence for differences in the distribution of performance for various groups. The report also discusses the important phenomena of metering and processing errors in the dataset.

1.2 Structure of this report

This report contains several types of analysis as follows:

- Section 1 continues with an update on the sample of sites and data selection adopted for the analysis, including a cropped version (where sites at the extremes of performance have been omitted) that will be used to detect statistically significant links between specific characteristics and variations in performance.

- Section 2 documents some of the issues that led to the decision to use a cropped version of the sample for the statistical analysis.

- Section 3 introduces the statistical methods used and reports findings on the type and magnitude of differences observed, according to a number of key HP system and site characteristics.

- Section 4 investigates a number of phenomena found in previous literature/experience to cause poor performance or which impact SPF by definition, quantifying their prevalence and, where possible, their effects on SPF.

- Section 5 highlights a number of heat pump performance issues observed within the data.

- Section 6 summarises the findings.

1.3 Sample B2: selection methods

Introduction to Sample B2: the updated 'broad dataset'

Detailed monitoring of HP performance is demanding from both a practical and methodological perspective, as the study requires numerous sensors in a range of configurations (categorised under a specific *schematic*) to suit the diverse systems and physical settings of each installation site. The sensors need to provide accurate high frequency data (in this case at two minute intervals) for at least 12 consecutive months.

Of the total of 699 sites in the RHPP sample supplied, 99 sites were excluded at the outset of the project due to technical issues relating to the installation of heat meter temperature sensors. A further 104 sites were omitted due to missing data streams needed for the calculation of SPFs, or where the correct schematic used could not be identified in a definitive way that permitted calculation of SPFs.

Sample B2, with **418 sites** (319 ASHPs and 99 GSHPs) was then formed from sites where the following were present:

- a period of 13 consecutive months where heat output and electricity input were recorded concurrently at some time on each of at least 5 days in each month and for which the difference in water flow rate through the flow sensor in the heat meter between the 1st and 13th month was minimised;

- where the reported schematic (system and monitoring equipment layout) matched the variables present.

Subsquent calculations (SPF etc.) are then based on the last 12 months of the 13 months of most stable data selected.

This process *is the same as used previously for the selection of Sample B*, which was the basis for some of the analysis in the previous report (RAPID-HPC, 2016). Sample B2 differs from the earlier version only in the sample size obtained, since further work on issues with schematics enabled additional sites to be selected.

The 12 month period used for SPF varies from site to site in Sample B2, which from a statistical perspective has the additional advantage that it tends to even out the impact of relatively cold or warm winters across the sample (as would not be the case if all sites used the same year start and endpoints for the data).

The selection process does not mean, however, that the data in Sample B2 are free from metering errors, given that the above filters do not filter out every possible type of problem. In particular, there remains a high degree of 'noise' (systematic errors of different directions and sizes) in the heat data, rendering some of the analysis difficult. Section 2 sets out the known and suspected issues with the dataset and their potential impacts on the results presented in this report.

However, it was decided that adding further filtering algorithms into the selection process would not only complicate matters but was likely to remove sites for which metering errors were not large enough to substantively impact on SPF, and may remove valid sites (for instance, where the occupants turned off the HP system whilst away from their dwelling). Moreover, carrying out blanket operations such as deleting all instances of electricity consumption with no heat output removes some genuine effects such as heat pump parasitics; it is difficult to design algorithms which can differentiate between genuine and erroneous reporting of a variable such as heat or electricity, in a complex system such as a heat pump, on the basis of remotely monitored data, in the absence of redundancy in monitoring systems and data.

At some point a balance has to be drawn between simplicity and transparency of the site selection and the need to maintain sample size, both to have sufficient numbers for sub-sample comparison and for the summary statistics to provide information about the RHPP sample as a whole. For example, we are still able to observe a statistically significant difference between the performance of GSHPs and ASHPs, which may not be possible with smaller samples of each. The current sample size of Sample B2 (N=417) is 60% of the original total RHPP dataset supplied.

1.4 Sample B2: summary statistics

The histograms of all sites in Sample B2 for SPFH2 and SPFH4 (Figures 1-1 and 1-2) both show that the distributions are skewed to the left compared with a normal distribution. This divergence from the normal distribution is also evident for the histograms of ASHPs and GSHPs separately (Figures 1-3 and 1-4), including a relatively narrow peak where more than 40% of the GSHPs lie in the 2.5 to 3.0 band for SPF_{H4}. One consequence, particularly when these groups are again split according to further categories, is that subsequent statistical analyses of SPFs need to apply non-parametric methods to identify statistically significant evidence of differences in performance (see Section 3). Specifically, the sample mean and standard deviations are difficult to interpret for non-normal distributions with an emphasis in this report instead placed on the **median**, **interquartile range** and other test statistics. However, for the purpose of comparison with previous results and those from other studies, estimates of population means (and 95% confidence intervals) are also presented. Note that throughout this report, where SPF is used without qualification it means the weighted average of both space and water heating according to the specified system boundary.

The updated summary statistics obtained for Sample B2 (Table 1-1) show no statistically significant change in the estimated population mean SPFs at the H2 and H4 boundaries compared with those for Sample B given in the interim report (RAPD-HPC, 2016). Specifically, the previous Sample B had mean SPFH2 of 2.59 (95% CI: 2.51-2.67) for ASHPs (N=297) and 2.91 (2.75-3.07) for GSHPs (N=94). For mean SPFH4, the previous findings had 2.36 (2.28-2.44) and 2.75 (2.61-2.89) for ASHPs and GSHPs respectively. Nor do these values differ significantly from the temperature corrected SPFs from Sample C (concurrent data), which were used in various analyses previously, such as estimates of mean annual CO_2 savings.

Table 1-1. Estimated SPFs for ASHPs and GSHPs in Sample B2

System boundary	HP type	N	Mean (95% CI)	Median (IQR)
SPFH2	ASHP	318	2.55 (2.49, 2.62)	2.63 (2.24 - 2.94)
	GSHP	99	2.89 (2.73, 3.06)	2.81 (2.52 - 3.16)
SPFH4	ASHP	319	2.33 (2.27, 2.38)	2.37 (2.07 - 2.65)
	GSHP	99	2.74 (2.59, 2.88)	2.7 (2.41 - 3.04)

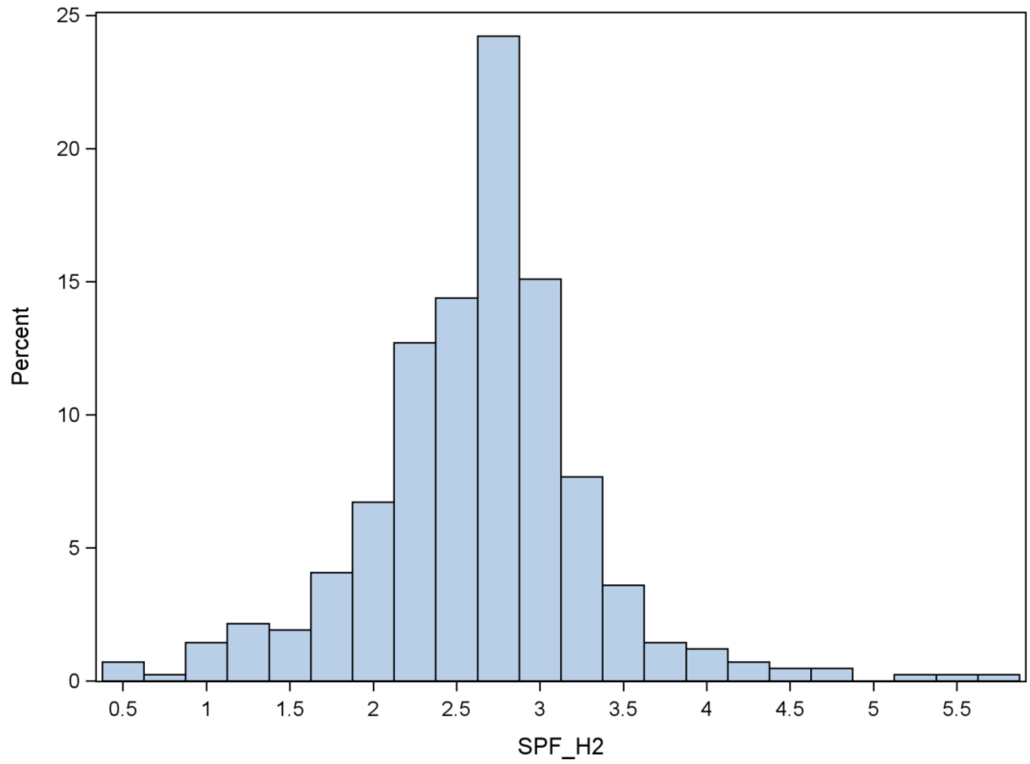


Figure 1-1. SPFH2 for all HP sites in Sample B2.

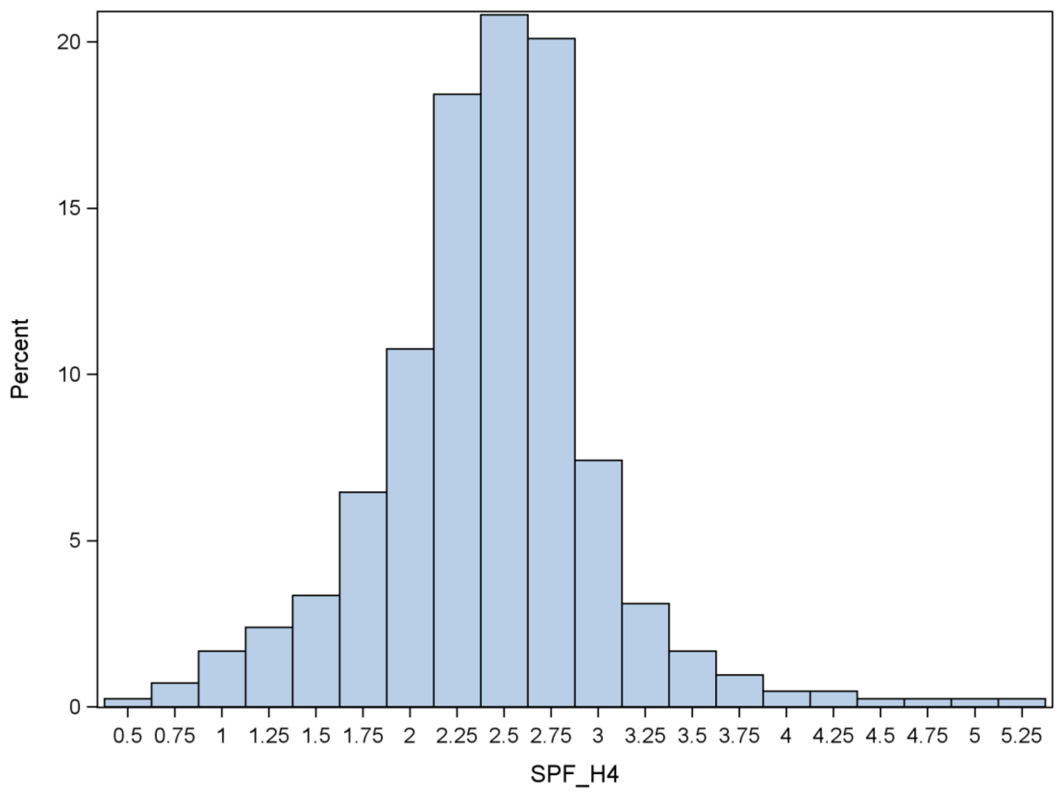


Figure 1-2. SPFH4 for all HP sites in Sample B2.

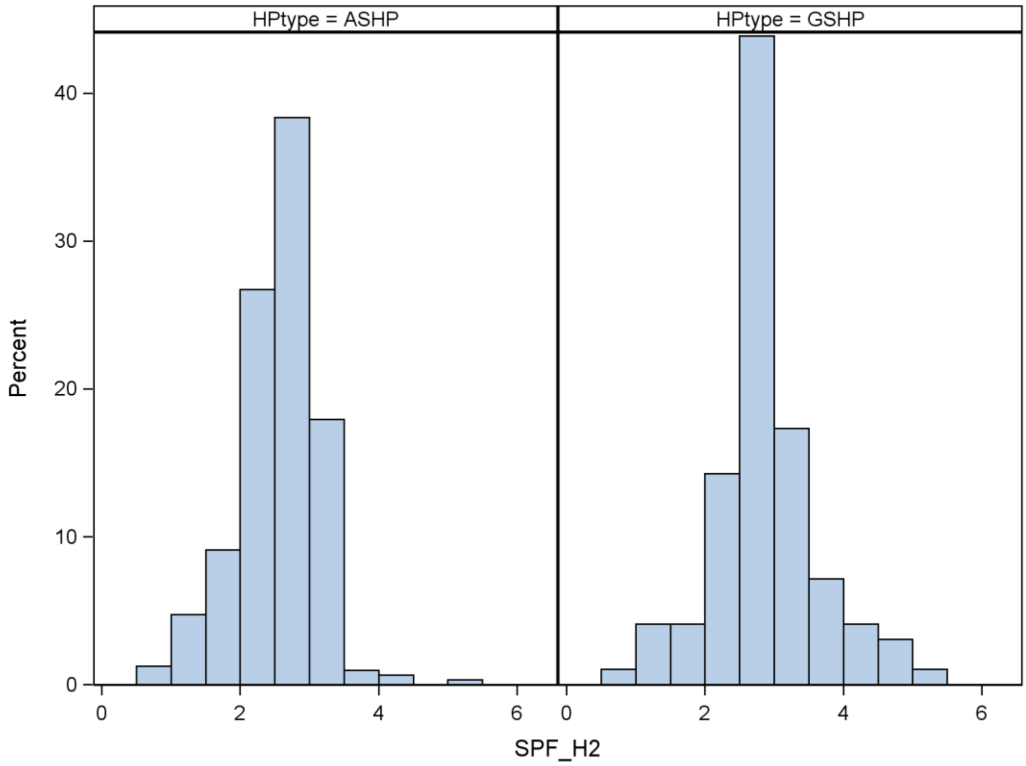


Figure 1-3. Histogram of SPFH2 for ASHPs and GSHPs in Sample B2.

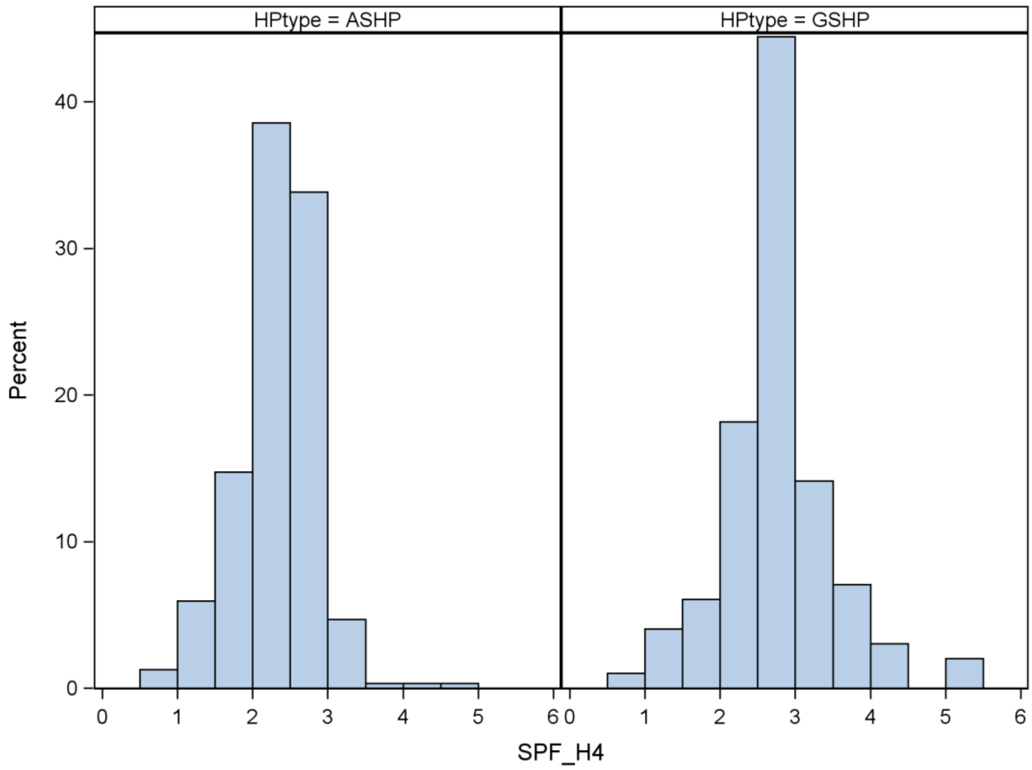


Figure 1-4. Histogram of SPFH4 for ASHPs and GSHPs in Sample B2.

1.5 Sample B2 (cropped)

Due to the level of noise in the data (discussed in detail in Section 2) the sensitivity needed to detect relationships for sub-samples of HP systems and small associated shifts in the distribution of SPFs poses considerable challenges. The key issue is that these relationships can be diluted and the statistical significance of any differences easily lost with noisy data. It is already observed that the sites show considerable variation in performance (far larger than the differences between ASHPs and GSHPs), with the distributions skewed to the lower performance. Further analysis on the data quality, presented in Section 2 to follow, leads to the conclusion that while data issues remain in some of the sites across the range of performance values, they are more likely to be present in sites at the edges of the distribution.

It was therefore decided to curtail the distribution at a lower bound **SPFH4 of 1.5 and a higher bound of 4.5**, in order to reduce the impact of sites with substantial metering issues and to enhance the ability to detect factors that affect performance for the remaining sites in the centre of the distribution. The resulting sample, henceforth referred to as 'Sample B2 (cropped)', has 25 fewer ASHPs and 7 fewer GSHPs (less than 10% reduction for both types). It should be noted that this approach is not without precedent, as the same technique of cropping the extreme results was used in a recent Danish field trial. This field trial and the cropping decision are further discussed in Section Error! Reference source not found.**3**. The summary statistics for Sample B2 (cropped) are provided in Table 1-3[4].

The excluded sites (i.e. those in Sample B2 but excluded from B2 cropped) are overwhelmingly those with SPFH4 < 1.5 (29 sites) as opposed to SPFH4 > 4.5 (3 sites). It is therefore not surprising that the median heat generation of the excluded sites is lower than that of the sites included in Sample B2 (cropped), while for these two groups there is not a large difference between the median heat demand estimated by installers, as shown in Table 1-2:

[4] Note that in Table 1-3 there are one fewer sites at the H2 bound than the H4 bound, this is because H4 is calculated first, then auxiliary heating and circulation pump energy are subtracted. In one site Ehp and one of the auxiliary heaters appear to have been transposed in the data so the H2 heat and electricity consumption do not make sense but the H4 heat and electricity consumption can be used.

Table 1-2. Estimated and measured mean heat output of heat pumps for sites included and omitted from Sample B2 (cropped)

Category	N	Installer Estimated Heat Demand, kWh/yr	Measured Mean Heat Generation, kWh.yr
Included	385	10,800	8,552
Omitted	32	12,000	3,787

This result indicates the possibility of problematic heat data in the excluded sites; this is further explored in Section 2.

Table 1-3. Sample B2 (cropped): mean SPFs for ASHPs and GSHPs

System boundary	HP type	N	Mean (95% CI)	Median (IQR)
SPFH2	ASHP	292	2.64 (2.60, 2.70)	2.65 (2.33 - 2.95)
	GSHP	92	2.93 (2.80, 3.06)	2.81 (2.63 - 3.14)
SPFH4	ASHP	293	2.41 (2.37, 2.46)	2.44 (2.15 - 2.67)
	GSHP	92	2.77 (2.66, 2.89)	2.71 (2.48 - 3.02)

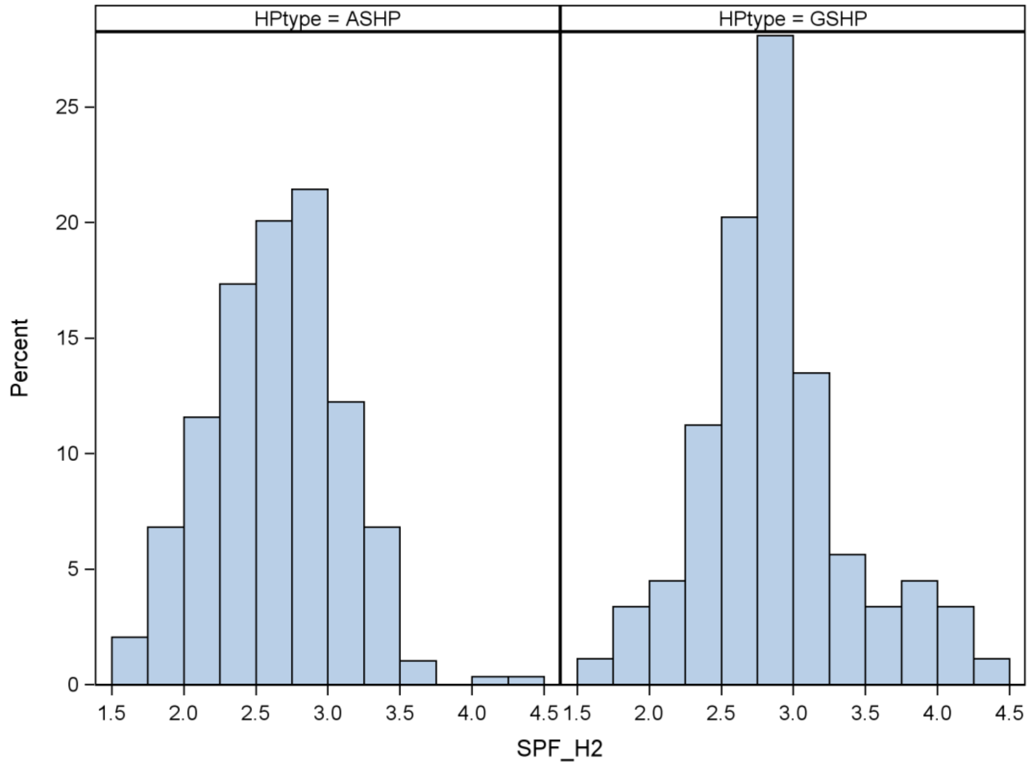


Figure 1-5. Histogram of SPF_{H2} for ASHPs and GSHPs in Sample B2 (cropped).

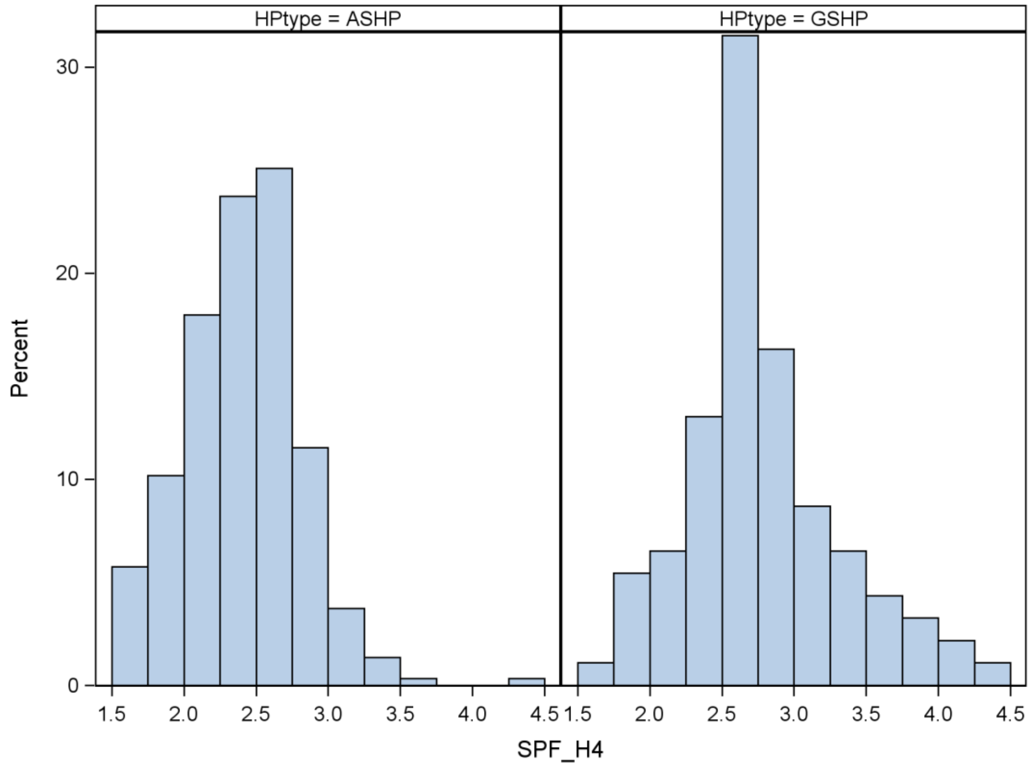


Figure 1-6. Histogram of SPF_{H4} for ASHPs and GSHPs in Sample B2 (cropped).

1.6 Overview of the sites in Sample B2 (cropped): heat output

Next, it is useful to set the context for the performance to be discussed throughout the rest of the report: what are the heat pumps providing, and when? Heat output is shown in this section, then seasonal performance in Section 1.7.

The previous discussion highlighted that the dataset to be used throughout this report is Sample B2 (cropped). The heat pump heat output of the 385 sites in Sample B (cropped) is shown in Figure 1-7, where the sites are ordered from left to right in terms of their total heat pump heat output, and their DHW heat output is stacked on top of their space heating heat output.

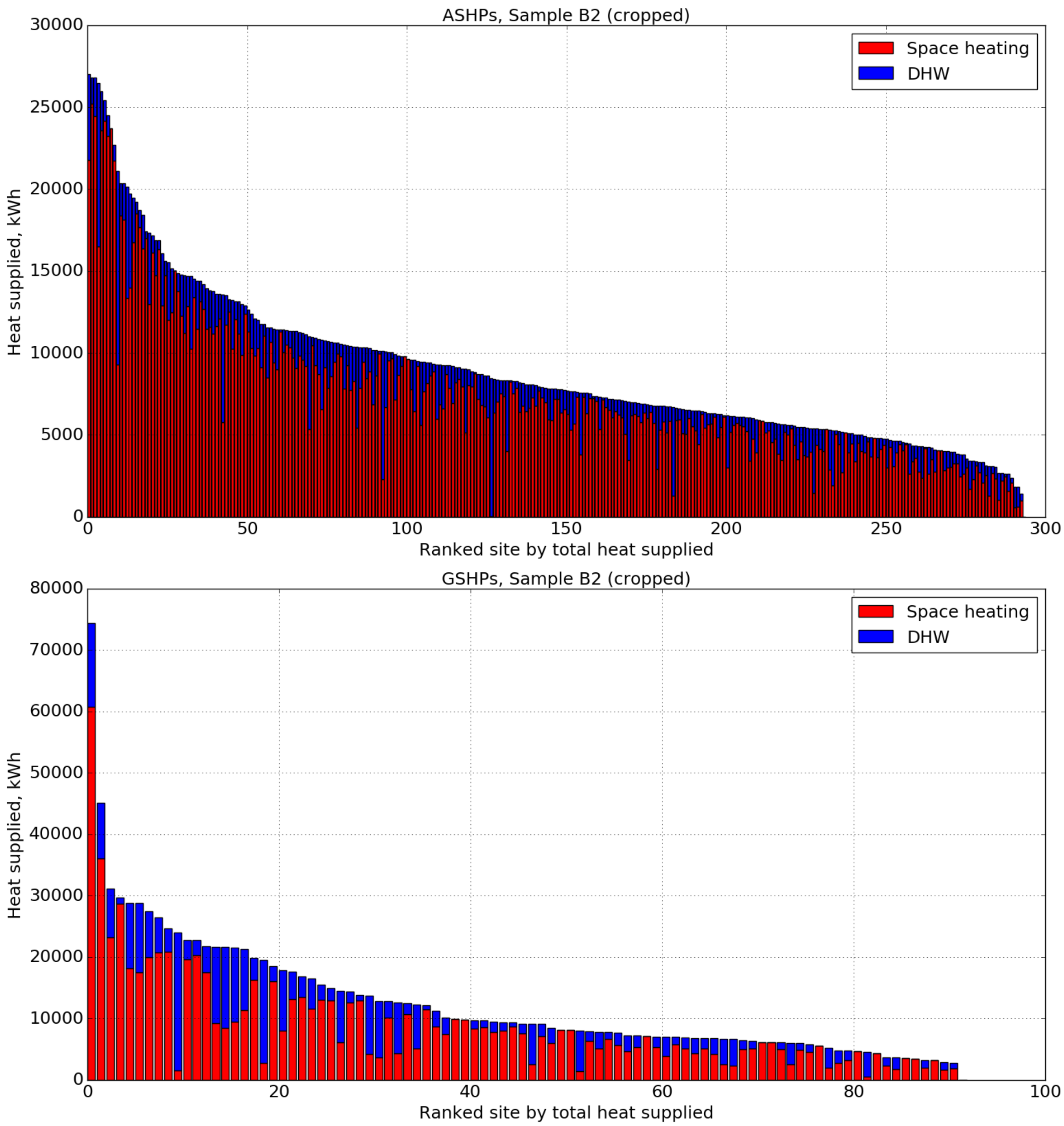


Figure 1-7. Annual heat from the HP (H2 bound: heat from HP only).

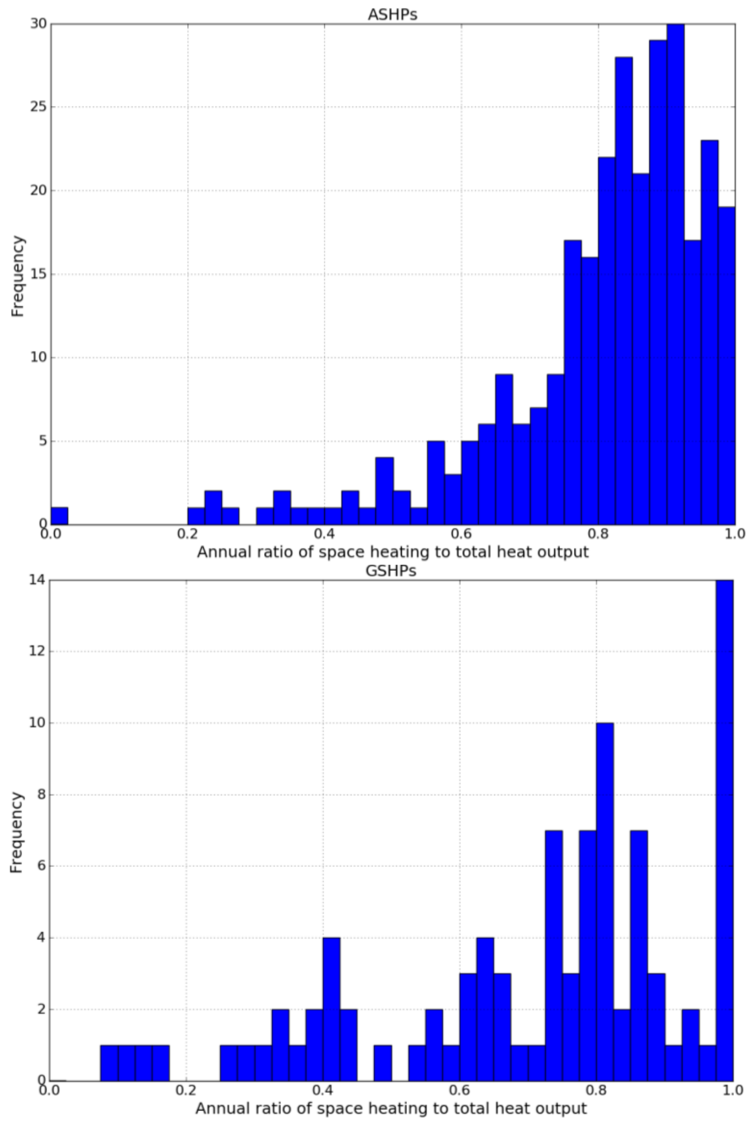


Figure 1-8. Annual ratio of space heating to total heat output from HP for GSHPs and ASHPs.

Figure 1-7 and Figure 1-8, which presents the same information on a frequency plot, show that in general around 80% of a site's heat pump heat output goes to the space heating and the rest to the DHW. There is however a lot of variation:

- 5% of ASHPs and 15% of GSHPs in Sample B2 (cropped) are used for space heating only (see Table 1-4) (one of these has been confirmed in a site visit – see *RHPP Case Studies Report*.
- There appear to be a small number of sites with higher DHW demand than space heating demand – this is likely to be an artefact of the way in which these systems were monitored and processed as opposed to a true representation of heat pump operation, and is covered in Section 2.2.

- As total heat output decreases (moving to the right along the x-axis of Figure 1-7), the ratio of space heating to total HP output decreases on average, but not in all cases. Two main phenomena resulting in total heat output decrease are likely to be: increasing dwelling thermal efficiency, leading to lower ratio of space heating to total heat, and decreasing dwelling size coupled with the tendency for smaller dwellings to have higher occupant densities, which may not affect the ratio of space heating to total heat.

Table 1-4: Overview of functionality of heat pumps in Sample B2 (cropped)

HP Type	Number of sites in Sample B2	Number of sites in which HP provides both space heat & DHW	Number of sites in which HP provides space heat only[5]
ASHP	292	277	15
GSHP	92	78	14

1.7 Overview of the sites: seasonal performance

Figure 1-9 shows monthly COPH2, broken down into months, modes (space heating and DHW) and HP type.

[5] Note that this category includes sites in which there is no DHW monitored, and those in which an immersion heater (and in one case, an immersion heater/active solar heating system combination) is used for DHW.

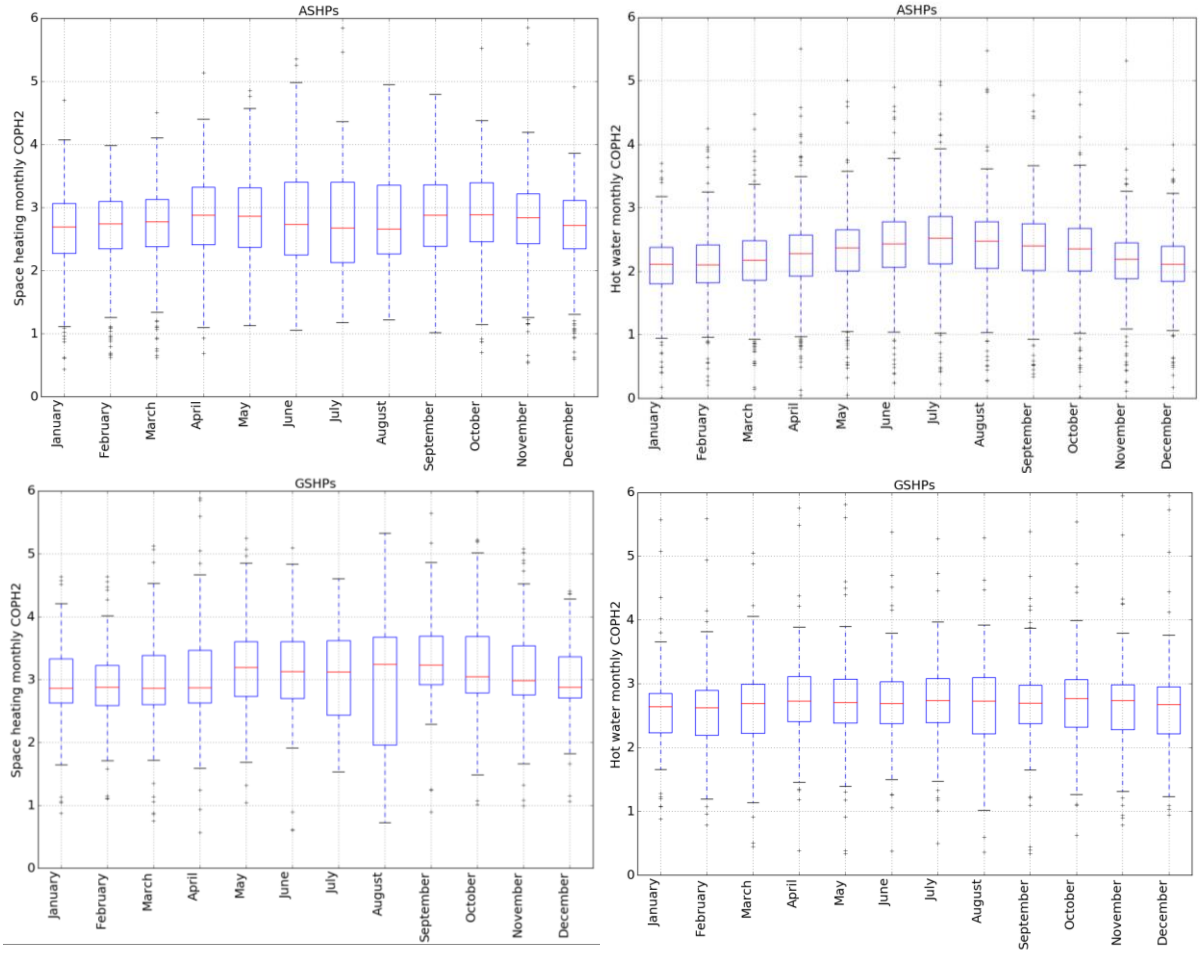


Figure 1-9. COP breakdowns by month, mode and HP type.

Note that not all of the sites in Sample B2 (cropped) are included in Figure 1-9:

- On the DHW plots (right hand side), only those sites with DHW provided at least in part by the HP are included;
- On the space heating plots (left hand side), although every site has space heating provided by the heat pump, it was found that only a minority of sites have space heating in summer. However, the algorithms determining whether a heat pump is in space heating or DHW mode can in some cases imply that there is a small amount of space heating when in reality there is not. This is a very small effect and so is only apparent in summer in sites where there is no genuine space heating, otherwise it is negligible compared to the real demand. To remove this effect, a lower limit on space heating output was determined (100 kWh/month) and only sites with this or greater space heating output were included in Figure 1-9.
- This reduces the sample size in summer months down to a minimum of 12 (GSHPs, July) and 49 (ASHPs, July).

Taking the above into account, Figure 1-9 shows that:

- DHW provision behaves as would be expected from a heat pump, in both ASHPs and GSHPs. That is, the monthly COP rises gradually towards the summer and decreases again towards the winter, as the source temperature changes. This effect is larger for ASHPs than GSHPs, which again is what would be expected given that the air temperature changes throughout the year more than the ground temperature.
- DHW COP tends to be lower than space heating COP in winter, which would be expected as generally DHW is expected to be provided at a higher temperature. This is also likely to be the case in summer although the sample sizes for space heating are smaller and there is a lot of variation.
- A small dip is observed in summer space heating COPs for ASHPs, which is not present for GSHPs – although again, note the small sample sizes in summer, as explained above.

2 Data quality and implications of metering and processing issues on study of heat pump performance

In this section, outstanding sources of error within the sample are discussed and their implications for the conclusions which can be drawn regarding reasons for good and bad performance are set out.

Firstly, it is necessary to discuss the different issues affecting the dataset which affect estimates of heat pump performance. Two categories of error are discussed here: metering error and processing error.

2.1 Metering error

As set out in Section 1.3, Sample B2 and Sample B2 (cropped) were created by applying a number of filters to the original datasets, requiring for example that there be at least 5 days per month in which heat and electricity data are recorded simultaneously. However, filters (applied automatically by software) do not pick up all potential metering errors within the dataset. Whilst carrying out the analysis for this report, a number of new issues were discovered, which affected a large number of sites.

Apparent metering issues present in the data are categorised here by their effect on heat pump performance metrics. Some, for example missing data from a site, can be quantified in terms of their effect on SPF if the rest of the data from a site is present, whilst others (such as apparently underreporting heat meters) are more difficult to quantify as the extent of deviation from the 'true value' of a heat pump's performance is not known.

Due to the large sample size, there was neither time nor resource to examine each site in detail to diagnose metering error. Even with more time, it is not always clear from inspection of the time series data whether an anomaly is a heat pump performance issue or a metering error, without going to the site and conducting further work.

The discussion in this chapter is therefore on the basis of:

- A detailed and systematic investigation of the timeseries data from a random sample of 34 sites (7 GSHPs and 27 ASHPs), representing 9% of Sample B2 (cropped).
- Other issues detected by the data analysis team arising and being noted not in a systematic way like above but as and when they were observed.

The sources of metering error observed are also summarised in Table 2-1.

Error leading to underreporting of SPF

Five sites in Sample B2 report SPFs below 1. Examination of the raw data led to the observation that these extremely low SPFs are not caused by periods of missing heat data or high parasitic loads of pumps staying on between heating events. Instead, the *instantaneous COP* (ratio of heat output to electrical input at a given snapshot in time) is below 1 for extended durations. This is physically unlikely[6] for periods longer than a few minutes, and thus suggests erroneous recording of either the heat output or electricity consumption. Furthermore, there is no reason for the erroneous reading to solely take place at these times of COP < 1; for example if a heat meter is systematically under-recording, this could also affect the rest of the data from a given site (the same is true for over-recording which is discussed shortly).

It should be noted that electricity consumption is easier to measure than heat and therefore that heat data problems are more likely that electricity data problems. The sites with SPFH4 < 1 sites are omitted from Sample B2 (cropped) which requires that SPFH4 >= 1.5.

Twenty two sites in Sample B2 report SPFs between 1 and 1.5. Again, visual inspection of raw data enables the explanation of missing data to be ruled out and instantaneous COP to be observed as very low. In some cases, visual inspection leads to a hypothesis of metering error: for example a hypothesis that there is further electric heating after the heat meter, whose electricity is captured but not its resulting heat, or that a heat meter is under-reporting due to poor installation (note that these issues can affect sites with SPF >= 1.5 too). However, instantaneous COP in this range is not necessarily a result of metering error; it could be that internal electric boost heating (not submetered) is the cause, as is further discussed in Section 4.3. Where instantaneous COP is very low all year round, this is unlikely to be the case but there is usually not enough information to determine whether metering error or poor performance is causing the low COP. These sites are, however, not in Sample B2 (cropped) which requires that SPFH4 >= 1.5.

Missing heat data, **electricity data or both** still occur in Sample B2 cropped. When both are missing, then whether this leads to under- or over-reporting of SPF depends on which is missing and what time of year. For example, if both heat and electricity readings are missing for 30 days during the coldest time of year, this can be estimated to result in *over-reporting* of SPF values by approximately 0.1.

Error leading to overreporting of SPF

A number of sites show **heat output where there is no electricity input**: heat is apparently being produced and assigned to space heating many minutes - and sometimes hours - after the compressor has stopped. The magnitude of the heat output (normally 1-3kW) is too great to be caused by gains to the circulation fluid from the circulation pump.

[6] But not impossible. A situation in which heat was provided entirely by resistance heating, but in which a significant parasitic load (a fan motor in the external unit of an ASHP, or a ground loop circulation pump for an GSHP) continued to operate, could lead to an SPF of less than unity.

Figure 2-1 shows around 12 hours of data from a GSHP system with underfloor heating in which this phenomenon is visible. The third subplot shows times of heat output from the HP (Hhp>0) with no electrical input (Ehp). The first subplot shows the monitored temperatures on the hot side of the heat pump and the bottom subplot shows those on the cold side (note that Tin is the ground loop temperature, not the building internal temperature), and are useful for cross-checking the electricity and heat data in the third subplot.

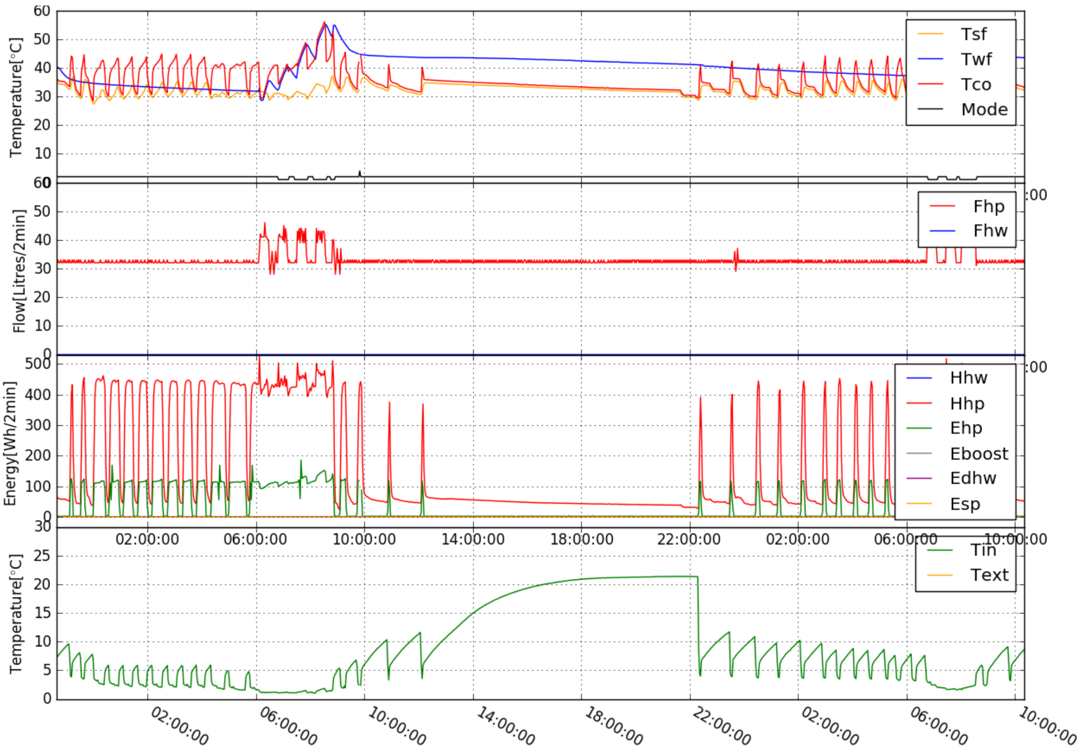


Figure 2-1: Example of apparent heat output without the compressor running.

A number of hypotheses can be formed as to why there is heat output at times of no electrical input:

- Heat being extracted from hot water cylinder or other heat store;
- Heat being extracted from radiators;
- Heat extracted from a buffer vessel between the heat pump and the heat meter;
- A calibration offset between the temperature sensors on the heat meter.

The problem here, as with some other instances of suspected metering error, is that since the cause of the apparent heat output is unknown, it is not clear whether this is a true effect or a metering error. If it is a metering error, it is also not known whether it just applies to the times when there is heat out and no electrity input or whether the error is present at other times too (for example when the compressor is running). This is an example of where deleting anomalous-looking sections of data might not remove the whole of the metering error.

Error affecting performance metrics other than SPF

In a number of sites there is evidence that the location of two sensors was accidentally swapped around when they were installed. The most common examples are: Eboost and Edhw, Tsf and Twf, and Fhp and Fhw. This error, observed in around 18% of the 34 sites examined in detail, does not show up in the filters applied to create Sample B2. Although overall SPFH4 is normally not affected, SPFH2 can be, and any further breakdown of these overall metrics such as space heating SPF / DHW SPF are then based on the wrong inputs.

In some cases, sensors are on the correct pipes, but not properly isolated from other pipes such that that they pick up signals from other phenomena more strongly than those which they are intended to measure. For example, sites were found in which temperature sensors placed on the DHW flow show a peak whenever the system is in space heating mode. The reasons are unknown: perhaps to do with diverter valves or conduction along pipes, or proximity of sensors to pipes from which they should not receive a signal. This information flows through the processing algorithms applied with the result that the system is labelled as in DHW mode when it is actually in space heating mode, which in turn leads to heat being attributed to DHW instead of space heating.

2.2 Processing error

As the analysis proceeded it became apparent that the heterogeneity within the HP systems and their installation configurations was greater than was anticipated by the software designed to process the data from all the different systems.

An example of this is the algorithm which determines whether the heat pump is in space heating or DHW mode. In certain HP systems, for example those using desuperheating (8 sites in Sample B2), the heat pump is capable of providing both space heating and DHW at once. The software has not been written for this case and as such the heat output from these HPs is all, or almost all, attributed to the DHW. Similarly, in one particular configuration of space heating-only heat pump, the hot water flow temperature sensor was not needed and was placed near the evaporator of the heat pump instead (9 sites in Sample B2) which leads to apparent DHW use when it is clear that there could have been none. Beyond these examples of drastic misattribution, a large number of sites display more modest levels of misattribution of DHW heating to space heating and vice versa. These errors do not affect overall SPF, but do affect the computed estimates of heat, electricity and SPF for space and water heating.

These particular errors were uncovered at a late stage in RAPID-HPC's work, and highlight the difficulty of designing software suitable to process a very heterogeneous dataset, describing a complex physical system, and performing post-processing using indirect variables such as pipe temperatures to determine heat pump mode.

Table 2-1. Main sources of suspected metering error in the datasets.

Meter type	Potential Fault type	Description	How do we know that these faults exist?	Systematic error or an error that affects individual sites?	Effect on SPF or monthly COPs
Heat meters	Missing heat meter data, when electricity data is present.	Periods with zero or unusually low heat data were not filtered out in the data cleaning process.	Observed in data	Individual sites	Will have the effect of under-estimating SPF, by an estimated ~4% across the Sample B2 (cropped) but much higher for a few sites. Apparent slight effect on distribution of SPFs (statistical tests not carried out to confirm this).
Heat meters	Systematic under-reading due to meter installation.	Poor installation of strap-on sensors or pocket sensors RAPID-HPC removed 99 sites with known strap-on sensors at the start of the project, but suspect that others may exist.	Some suspiciously low COP readings observed in data (e.g. < 1)	Individual sites	Would reduce SPF and monthly COP but sites for which spfh4<1.5 have been filtered out of Sample B2 (cropped).
Heat meters	Systematic over-reading due to glycol correction not being applied.	Heat meters calibrated for water with no antifreeze.	Wickins (2014)	Likely to occur in many of the sites, in both ASHPs and GSHPs.	Likely to result in over-estimation of SPF by 4-7% - see separate report on systematic errors.

Heat meters	Limited to 18 kW	Up to 16 sites in Sample B2 (cropped) affected.	Observed in data	Individual sites	Expected to affect the SPFs of these sites slightly in cold weather.
Heat meters	Systematic over-reporting of heat output.	Probably due to heat meter temperature sensor offsets, exacerbated by circulation pump over-run.	Observed in data (heat output when no electricity input)	Individual sites	Will over-report SPF and COP.
Heat meters	Spikes in heat output when changing mode.	It is not known whether this is a metering problem or a real dynamic effect with no impact on estimates of heat.	Observed in data.	Individual sites. Not present in all sites, but for those in which this effect is present, it occurs every time there is a mode change.	Unknown. If real heat, no effect, if metering error, over-reports SPF and COP.
Heat meters	Transposition of Hhp and Hhw sensors.		Observed in data.	Individual sites	No effect on overall SPF or COP, but will affect space heating and DHW SPFs and COPs.
Heat meters	Flow decay over the dataset time period. Median 1.5% decay over year for Sample B2 (cropped).	Cause unknown.	Observed in data	Individual sites	Under-report SPF, and COP for later months.

Heat and electricity meters	Heat and electricity data missing at the same time.	Cause unknown.	Observed in data; of 34 sites investigated in detail, 16 had > 7 days of this.	Individual sites	Effect depends on the time of year at which the problem occurs.
Electricity meters	Suspected unmetered electricity – missing Eboost or Edhw	Temperature data shows unusual patterns which can't be explained by the existing heat and electricity data.	Observed in data	Individual sites	Over-report SPF.
Electricity meters	Transposition of electricity meters.	Transposition of Ehp and Edhw or Eboost, or Edhw and Eboost.	Observed in data. Automatic correction applied in code for cases where easily detectable but not all cases.	Individual sites	Effect depends on which sensors were involved. Overall SPFH4 unchanged but other boundaries affected. Space heating and DHW SPFs could also be affected.
Temperature	Sensors too close to other pipes.	This causes e.g. Tsf to be influenced by Twf and vice versa. This in turn affects which mode (space heating, DHW) gets attributed to each 2 minutes of data.	Observed in data and photos	Individual sites	Overall SPF and COPs are unaffected but space heating and DHW SPFs and COPs are affected.

2.3 Partial solution: Sample B2 (cropped)

To reiterate from Section 1, a simple and transparent final filter has been applied to the data to create Sample B2 (cropped):

Sample B2 (cropped) contains sites with overall SPFH4 >=1.5 and <=4.5

The reasoning for this additional filter was as follows:

- The sites with SPF outside of this range are almost certainly incorrectly metered, as opposed to those within this range whose reported performance is usually more difficult to validate/invalidate;

- A precedent can be found in a previous study published in 2011 by the Danish Technological Institute (Pederson and Jacobsen, 2011), in which the boundaries were 1.5 and 5.5, with the assumption that any site outside of these boundaries was incorrectly metered. The upper boundary of 5.5 was not used for Sample B2 since inspection of the 3 sites between 4.5 and 5.5 yielded anomalies in the data.

- The alternative was to carry out a detailed inspection of data from every site, which was not possible in the allotted time and still would not have led to definitive conclusions about whether a site is reporting correctly.

- This filter does not rely on Tco, Twf or Tsf or the algorithms which, as discussed above, sometimes incorrectly attribute space heating to hot water and vice versa.

- It was found that removing the extremes of the distributions in some cases made effects clearer to see in the middle of the distributions.

However, the use of Sample B2 (cropped) did not reduce the noise in the dataset as much as necessary to observe all expected trends; this is discussed in the next section.

2.4 Implications of error on conclusions made attempting to explain variations in performance

Distributions

A distribution is a frequency plot of all the values of one result, such as SPF. In this report we use distributions as a way to show the overall performance of either the whole sample or different groups within it.

A number of different types of metering issues have been detected in the heat (and to a lesser extent electricity) data, which can affect the value of SPFs in either direction, so the expected overall impact across the sample would be a **widening of the distribution**. There is an assumption, consistent with observation to date, that sites with some degree of metering error are scattered throughout the sample and do not just occur in one particular group. Since errors have also been found which affect SPF in both directions, it is not possible to state whether the net effects on the key results (such as sample median SPF for ASHPs and GSHPs) lead to underreporting or overreporting of what they would have been in the absence of metering error.

Widening of SPF distributions due to metering error, and any other factors at work, can mean that underlying differences between the SPF distributions of each group have to be larger before it can be concluded that a statistically significant difference exists; that is, small differences between groups cannot be detected. However, the larger the group size, the smaller the difference can be whilst still being detected[7].

For example, due to the large sample size for ASHPs, assessment of the overall results suggest that the **statistical tests can detect a systematic difference in SPF** between two groups of ASHPs of ~8% (i.e. ~0.2) or more, though the exact figure depends on the size of each group. In fact, it was just possible to detect a difference between the shape of the SPFH2 distributions from the 28 ASHPs with underfloor emitters (median 2.80) and those for radiators (median 2.64). Similarly, there was some evidence that ASHPs with Model A installed tended to be higher (by ~0.12) across the distribution than other ASHPs.

While both of these findings would not be characterised as strong evidence of a difference, the size of the systematic difference detected is just 5% of typical SPF values of 2.5 to 3.0. Similarly, despite of the effects of metering errors in widening the distribution, if a reasonably large group is characterised by some feature that produced a consistent 10-15% change in SPF (or ~0.25 or more), then it is highly likely that this

[7] The minimum detectable effect size varies roughly as the inverse square root of the number of cases in each group. To halve the minimum detectable difference between two groups requires roughly four times as many cases.

would have been clearly detected as strong statistical evidence of a difference. An example is the differences in SPFs found between GSHPs and ASHPs.

The sample size of GSHPs (n=92) is less than a third that of ASHPs, so the statistical power to detect differences of groups within the GSHPs sample is considerably reduced. Overall the distribution of GSHPs showed a very similar degree of spread as the ASHPs, which suggests that metering errors are also present as with the ASHPs. It is arguable, however, that due to the small sample size for GSHPs and various factors likely at work in affecting SPF, then **even without metering errors it would be challenging to detect differences in SPF** that would be reasonably be expected for most characteristics. Across the various characteristics that were examined, the only evidence identified was for the difference in the shape of the distributions between RSL (n=53) and private sites (n=39) with GSHPs.

One standard way to at least provide a sense check of the results is to compare key summary statistics with those from previous field trials with a similar study design. In a meta-analysis of European heat pump field trials (Gleeson and Lowe, 2013), the results here for ASHPs compare closely with SPFH4 of 2.4, but for GSHPs are considerably lower than the 3.2 reported in the meta-analysis. One potential explanation is that those results include a study from Germany on new dwellings with GSHPs that had an exceptionally high mean value for SPFH4 of 3.7. In another study published in 2011 by the Danish Technological Institute (DTI, 2011), a mean SPFH4 of 2.33 was reported for ASHPs (N=12) and of 3.03 for GSHPs (N=138). Little information is provided about the results or the sample, including the uncertainties, except that they involved existing dwellings. The Danish study was conducted under external conditions that overlap with those in the north of the UK. However the DTI sample of GSHPs contained a high proportion of sites with underfloor emitters, which would be expected to perform better, so this may provide a partial explanation for the higher mean SPFH4 figure than was found in in Sample B2 cropped (mean SPFH4 of 2.77).

Scatter plots

Scatter plots are used to investigate the relationship between two variables, with the outcome variable on the y-axis (such as SPF) plotted against an explanatory variable on the x-axis (such as flow temperature). Typically the statistical tests applied are aimed at detecting a linear correlation, such as with linear regression, where the findings will indicate if there is evidence that for every change in the explanatory variable there is a corresponding change in the outcome variable. The explanatory variables (x-axis) are usually continuous, such as flow temperature. It would not be valid to do a scatter plot with a variable such as heat pump type (just two categories: ASHP or GSHP) on the x-axis.

Scatter plots address a different type of research question, and require more information content in the data, than would be needed simply to identify a difference in the SPF distributions of two groups, such as between ASHPs and GSHPs. In the distributions section, non-parametric tests were chosen due to the small sample sizes for some groups and due to the fact that the distributions were not 'normal' distributions (ie. Gaussian). These tests only showed if there was statistical evidence that one group tended to have higher or lower SPF than the other, and if the distribution differed in shape.

In the scatter plots shown in this study, the effects of range of factors other than the explanatory variable, including metering errors, can add noise to the SPF values such that this tends to mask both visually and statistically any sign of a relationship between the explanatory variable and SPF. It is worth noting that SPF is a measurement of performance designed to cover the full range of external conditions (and hence variations in operational performance) over the year, so this provides considerable opportunity for an array of factors to affect performance, including for metering issues to occur at some point during the measurement period.

In the RHPP dataset there are clearly some outstanding factors which mask trends which would otherwise be expected to be visible. For example, Figure 2-2 shows a scatter plot of monthly COPH2 against monthly proportion of cycles shorter than 12 minutes (where cycle time defined as time from one compressor start to the next) for all ASHPs in Sample B2 (cropped). It would be expected that sites with a high proportion of short cycles exhibit low COPs but this cannot be seen in the Figure.

Scatter plots shown in this report are all caveated with the potential masking effect of metering error and other factors and the reader is advised to bear this in mind where expected trends are not visible.

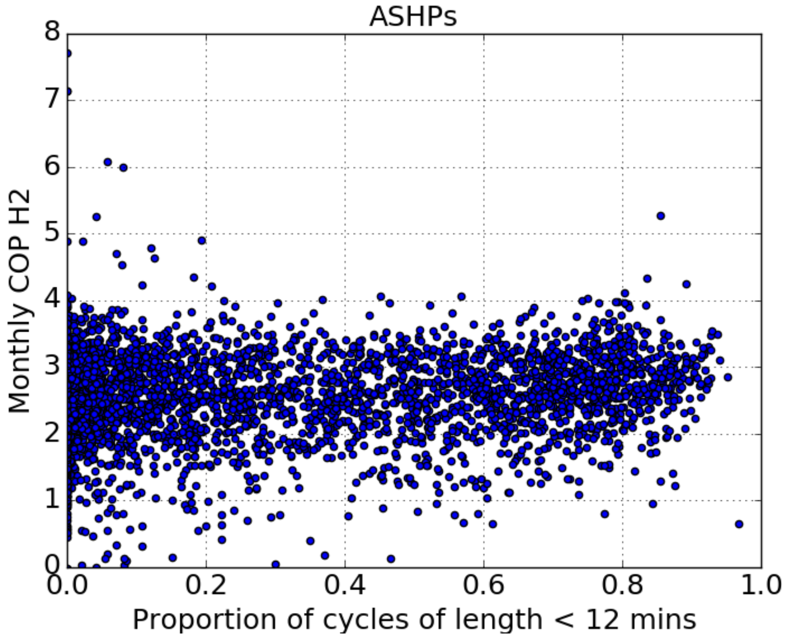


Figure 2-2. ASHPs' monthly COPH2 versus proportion of on-to-on cycles shorter than 12 minutes.

Summary and next steps

In order to proceed and investigate variations between sites taking into account the issues set out in this chapter, we use the following techniques:

- A statistical section (Section 3) examining performance using cumulative distribution functions to attempt to detect whether there are differences between groups. If the various sources of error affect subgroups similarly, and if subgroup sample sizes are large enough, then evidence for differences between groups can yield insights on performance.

- An exploratory section (Section 4) which does not focus on performance as defined by SPF per se; instead, we simply report on occurrence in the data of phenomena shown in previous literature to influence performance (e.g. prevalence of short-timestep cycling), even if their effect on SPF cannot be observed in this dataset.

- Each result is caveated with reference to the type and severity of error which may affect it. Where all results are drawn together in the Conclusion chapter, a table is presented of each finding and its sensitivity to heat meter or other error.

3 Results

For many of the factors or characteristics investigated, the sub-sample sizes in the categories are small and the SPFs do not form normal distributions, hence non-parametric analytical methods often need to be applied (rather than using statistics based on means, standard deviations and confidence intervals). Furthermore, in seeking to understand of differences in performance, one is not only concerned about whether one group has a higher mean or median SPF than another group, but also whether the distributions of SPF differ. In the following sections we introduce the two analytical methods used to detect statistically significant differences in the distribution of SPF, and hence identify potential factors that are associated with variations in performance. Specifically we use:

- Wilcoxon T-tests to determine if the distribution of SPFs is shifted significantly in any particular direction, i.e. does one group of HP systems tend to have lower or higher SPFs than another.
- Plots of the empirical distribution (also known as cumulative distribution) of each group of HP systems under investigation to illustrate graphically the difference between their SPF distributions in a way that is far more informative than comparing overlapping histograms.
- The Kolmogorov-Smirnov (KS) test which indicates if the empirical distributions have significantly different shapes (rather than just a lateral shift). For instance, a distribution with the same median as another may be more peaked and less spread out, or may even be bi-modal.

For both the Wilcoxon T-tests and the KS statistics we use the p-value from each test to indicate if the distributions are significantly different at 95% confidence. Formally, this means that the null hypothesis (that the empirical distributions are from the same population distribution) is *rejected* for p-values ≤0.05, and hence provides some evidence that the two distribution are indeed different (rather than just due to the natural variation expected for that sample size). Both tests have strengths and weaknesses: the Wilcoxon T-tests and the KS statistics tend to provide both greater sensitivity to detecting differences that would be missed using 95% confidence intervals or even student T-tests, and have general applicability including to variables that are not normally distributed.

For instance, if the Wilcoxon T test or KS test on the SPFs for Group A produced a p-value <0.001, this would be regarded as strong evidence that the characteristics that define group A are associated or linked with a distinct distribution of SPFs compared with other HP systems. In this way, the finding suggests that the characteristic under investigation may be one of the explanatory factors at work in the observed variation in SPFs for the sample as a whole. The case is strengthened somewhat if it agrees with well understood physical or other mechanisms and if the difference has been observed in previous studies.

Note that we still cannot state that the characteristic or factor causes the difference in SPF distribution, as other correlated or confounding factors may also be involved.

3.1 ASHPs vs GSHPs

The categorisation of HP systems as either ASHPs or GSHPs represents a fundamental defining characteristic of HP systems, based on the external source from which they draw heat and the consequent differences in technical and installation specifications. From a thermodynamic perspective, the higher source temperature in Winter for GSHPs compared with ASHPs should correspond to generally higher values for SPFs (all other things being equal). But note that source temperatures are likely to be higher for ASHPs in Summer.

The tables of summary statistics given in the introduction to Sample B2 and Sample B2 (cropped) have already shown that the mean SPF for GSHPs across the H2 and H4 boundaries are indeed higher than those for ASHPs. Table 3-1 expands on this evidence for SPFH2, with the results showing that on average, GSHPs perform significantly better for both space heating (SH) and domestic hot water (DHW) at the H2 boundary.

In this first comparison, the large sample sizes for each group and the relatively large difference in SPFs involved has permitted the use of estimates of the *population means* (and their 95% confidence intervals) to identify evidence of differences between the groups. Including 95% confidence intervals, the SPFH2 difference between GSHPs and ASHPs was between 0.16 and 0.40 (centred on 0.28) and 0.22 to 0.45 (centred on 0.33) at the H4 boundary. To check this further in sub-groups, it was also found that a difference was observed in between GSHPs and ASHPs with radiators (centred on 0.25, from 0.11 to 0.38), and between GSHPs and ASHPs with underfloor heating systems (centred on 0.23, -0.15 to 0.61 as the confidence intervals here were wider). Furthermore, Wilcoxon T tests and the KS tests have p-values that are all below 0.05 (KS p-values are shown in the bottom left corner of the Figures 3-1 to 3-3) and therefore also indicate that ASHPs and GSHPs have different distribution of SPFs.

Table 3-1. Sample B2 (cropped): SPFs for ASHPs and GSHPs

SPFH2	HP type	N	Mean (95% CI)	Median (IQR)
Overall	ASHP	292	2.65 (2.6, 2.7)	2.65 (2.33, 2.95)
	GSHP	92	2.93 (2.8, 3.06)	2.81 (2.63, 3.14)
SH	ASHP	292	2.72 (2.66, 2.78)	2.74 (2.36, 3.09)
	GSHP	92	3.03 (2.86, 3.2)	2.89 (2.59, 3.34)
DHW	ASHP	284	2.3 (2.24, 2.36)	2.31 (2.05, 2.56)
	GSHP	78	2.7 (2.56, 2.85)	2.71 (2.29, 2.99)

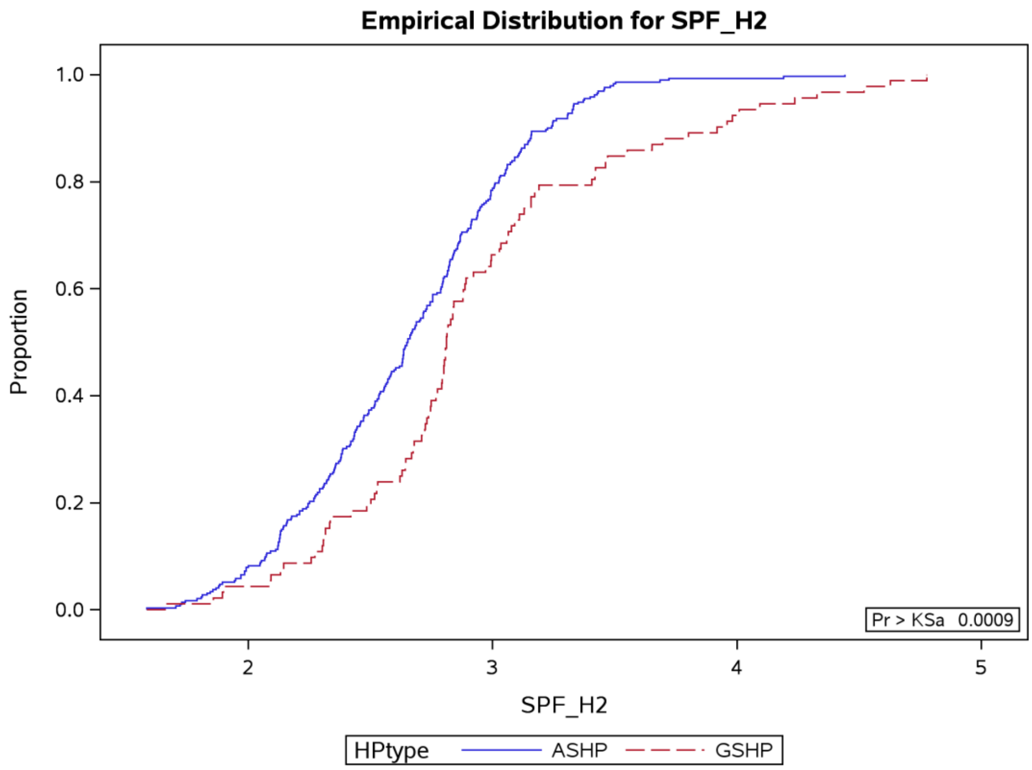


Figure 3-1. Comparison of SPFH2 for ASHPs and GSHPs in Sample B2 (cropped).

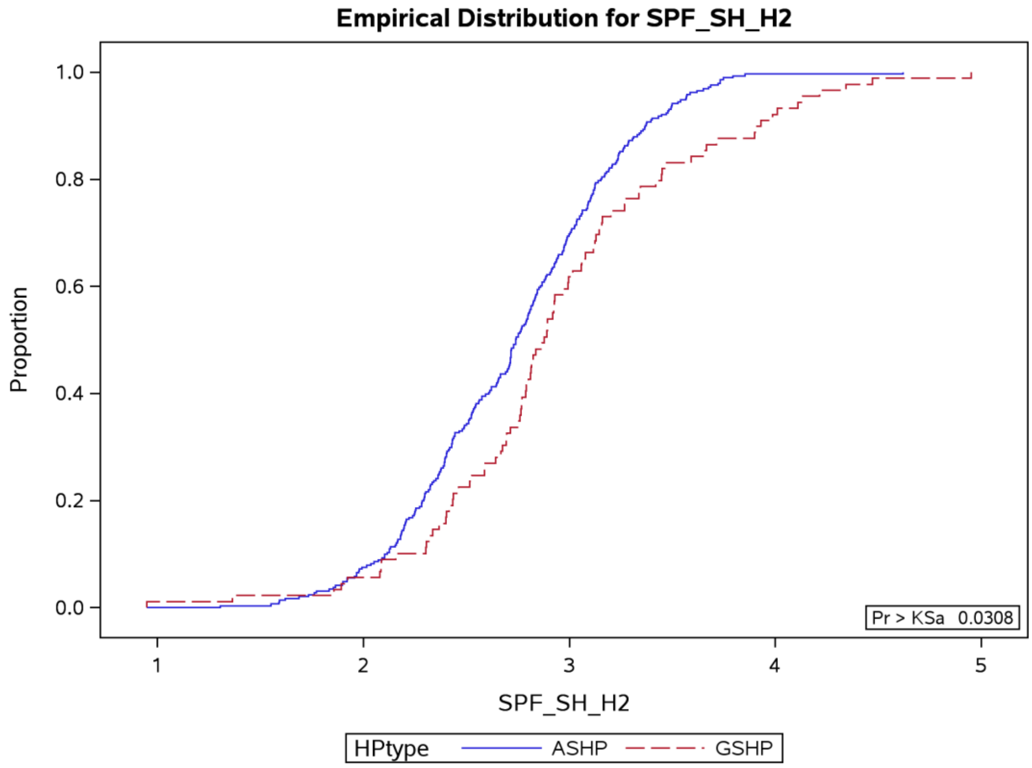


Figure 3-2. Comparison of SPFH2 for space heating for ASHPs and GSHPs in Sample B2 (cropped).

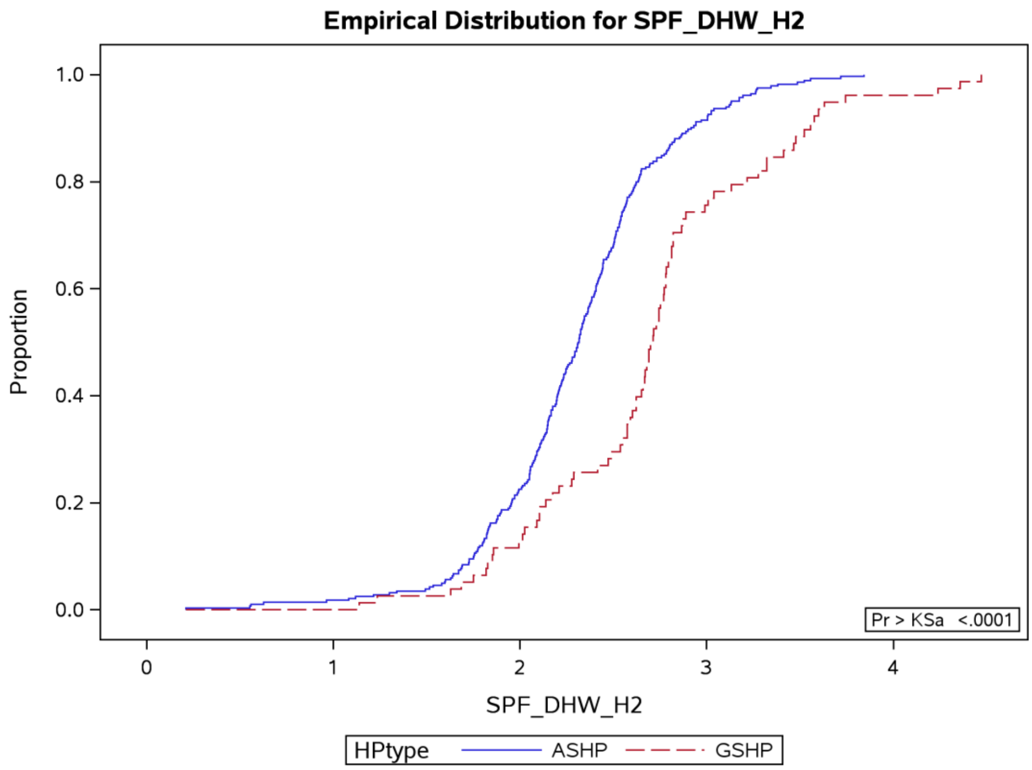


Figure 3-3. Comparison of SPFH2 for hot water for ASHPs and GSHPs in Sample B2 (cropped).

One aspect notable in the Figures is that the difference between ASHPs and GSHPs appears more clear-cut regarding the DHW component, whereas GSHPs have a significant but smaller advantage for the space heating component.

For the vast majority of sites, space heating (which occurs primarily in winter) will tend to make the largest contribution to the calculation of overall SPF values. Thus a further question arises regarding if the difference in space heating performance for GSHPs applies across the year. Figure 3-4 compares the average monthly COP over January and February, as an indicator of space heating performance during the winter heating season, between ASHPs and GSHPs. As expected it shows distributions that are very similar to the annual SPF for space heating, with GSHPs having a distinct shape (KS p-value = 0.022) and tending to have higher performance (Wilcoxon p-value = 0.005). In contrast, Figure 3-5 finds no significant difference in the distributions for the average monthly COP over April and May, and suggests that by this time of year there is no substantive space heating advantage for GSHPs.

Unlike for space heating, the performance advantage of GSHPs over ASHPs for DHW provision is maintained across the year.

A good deal of caution needs to be applied when examining the details of the COP components, as the method of identifying when the HP system is in SH and DHW mode has varying reliability across sites (see Section 2.2).

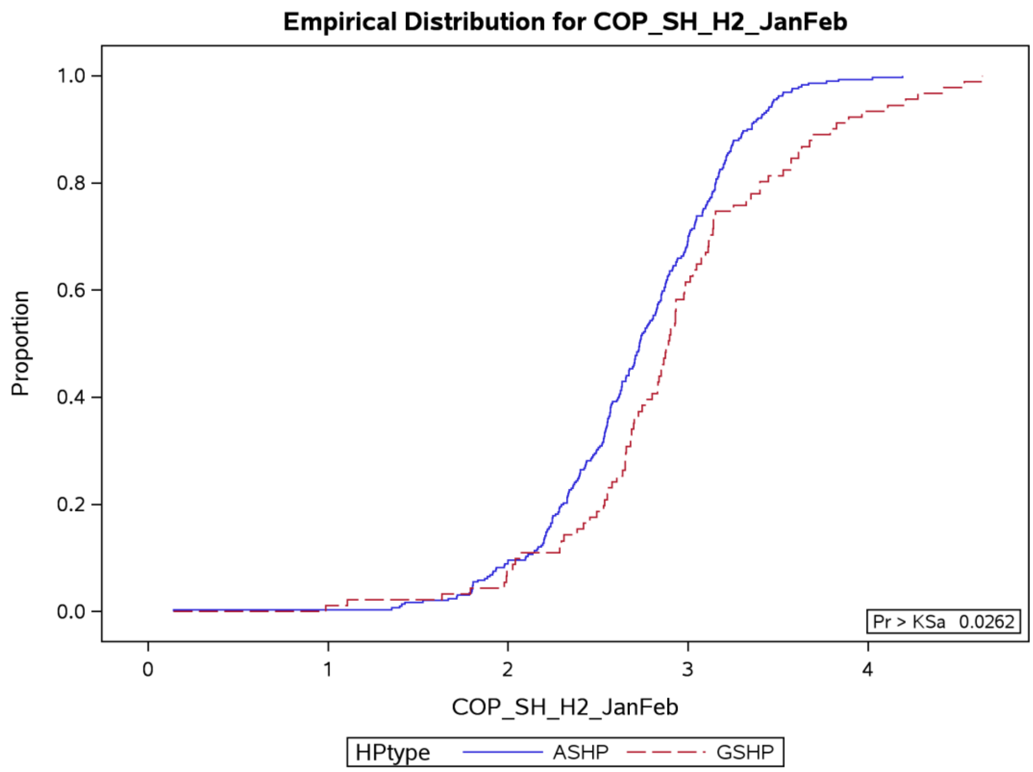


Figure 3-4. Comparison of COPH2 for space heating in January and February for ASHPs and GSHPs in Sample B2 (cropped).

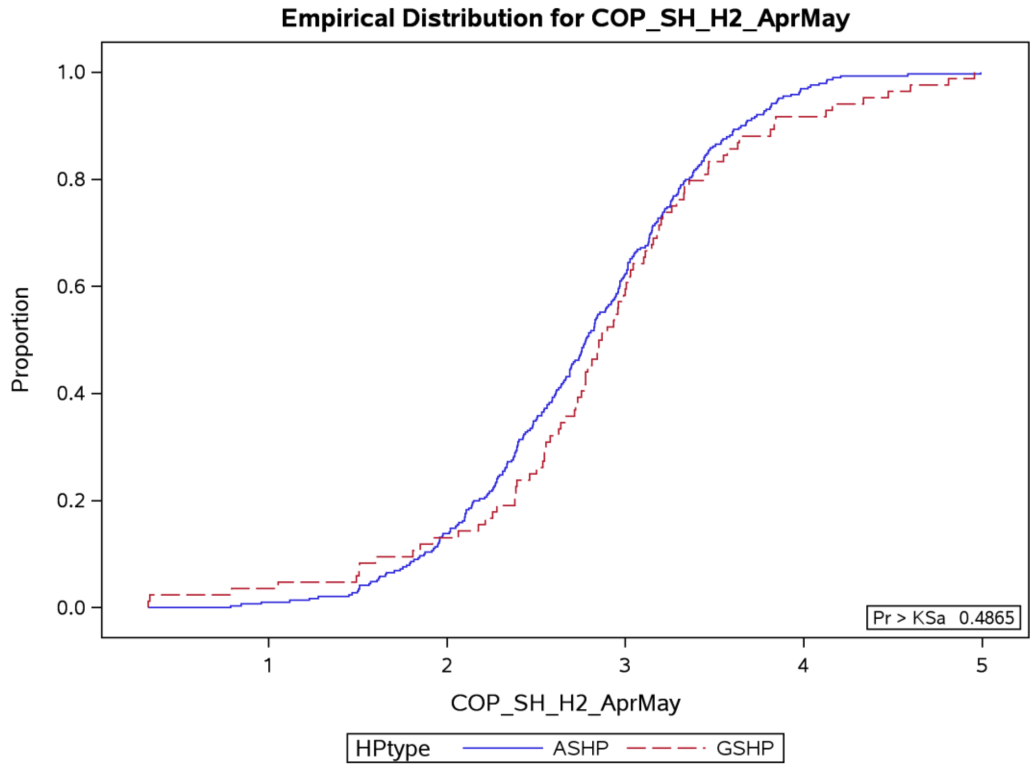


Figure 3-5. Comparison of COP_{H2} for space heating in April and May for ASHPs and GSHPs in Sample B2 (cropped).

Table 3-2. Sample B2 (cropped): SPFH4 for ASHPs and GSHPs

SPFH4	HP type	N	Mean (95% CI)	Median (IQR)
Overall	ASHP	293	2.41 (2.37, 2.46)	2.44 (2.15 - 2.67)
	GSHP	92	2.77 (2.66, 2.89)	2.71 (2.48 - 3.02)
SH	ASHP	293	2.60 (2.55, 2.66)	2.62 (2.3 - 2.95)
	GSHP	92	2.81 (2.67, 2.94)	2.73 (2.46 - 3.10)
DHW	ASHP	289	1.92 (1.86, 1.98)	1.92 (1.55 - 2.26)
	GSHP	78	2.61 (2.47, 2.75)	2.61 (2.19 - 2.87)

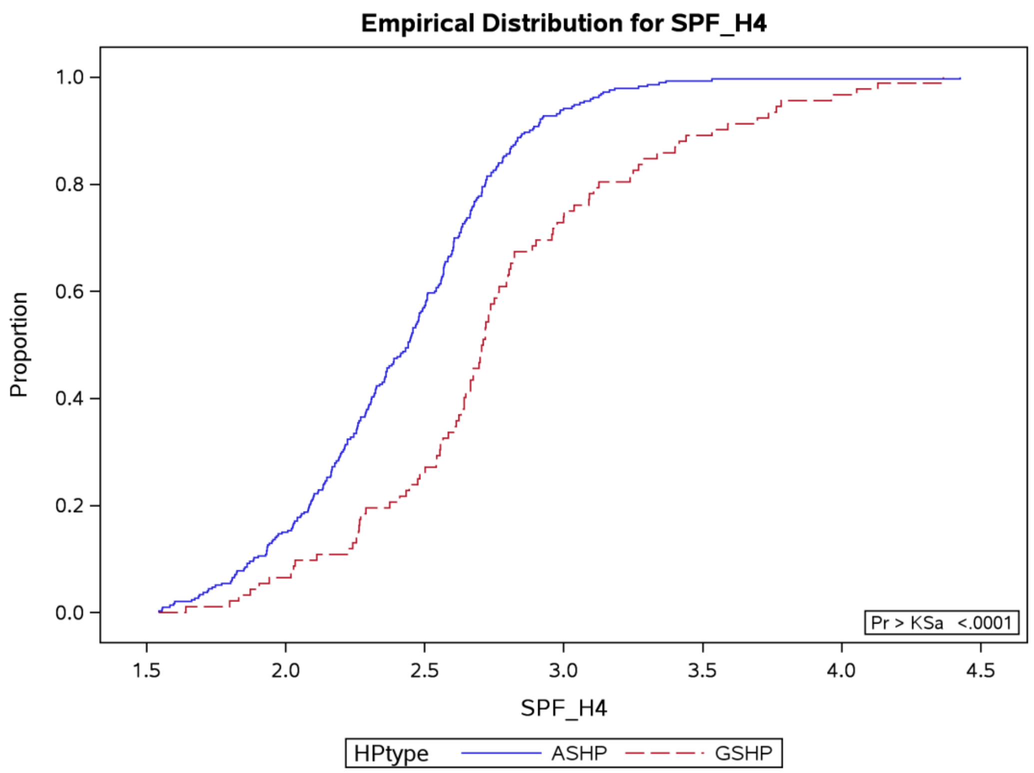


Figure 3-6. Comparison of SPFH4 for ASHPs and GSHPs in Sample B2 (cropped)

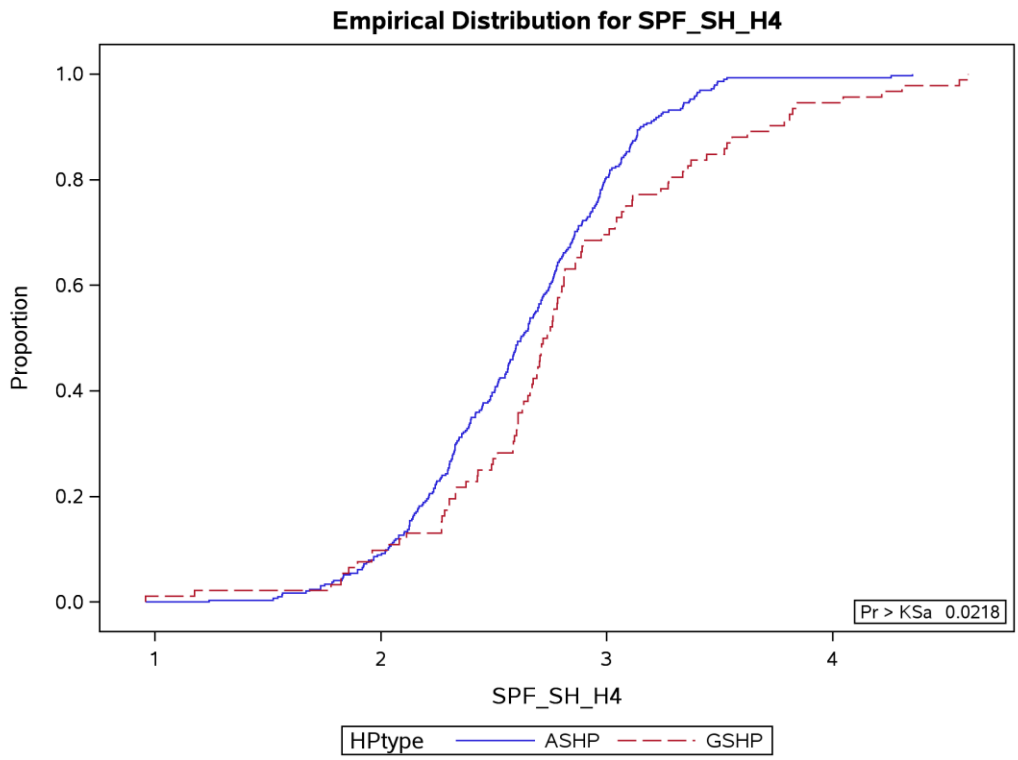


Figure 3-7. Comparison of SPFH4 for space heating for ASHPs and GSHPs in Sample B2 (cropped)

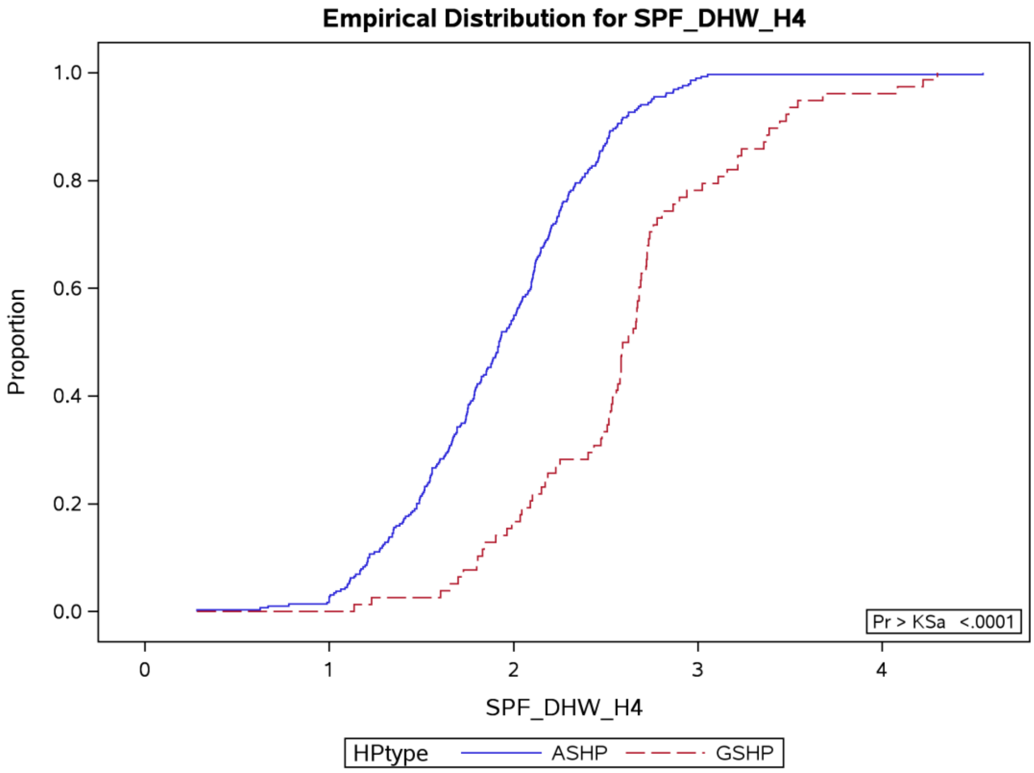


Figure 3-8. Comparison of SPF_{H4} for hot water for ASHPs and GSHPs in Sample B2 (cropped)

The results for SPFH4 (Table 3-2, Figures 3-7 and 3-8) are similar to those for SPFH2, with evidence that distributions are different overall, for space heating, and for DHW (KS p-value < 0.001 overall, 0.02 for space heating and < 0.001 for DHW).

In summary, the higher annual performance for GSHPs strongly suggests that the difference in the mean performance, including their advantage for space heating during the winter and hot water heating through the year, reflects the intrinsic thermodynamic attributes of these systems of HP but probably also reflects differences in the types of property into which they were installed; for example, GSHPs were on average installed in larger properties than ASHPs, and in a higher proportion of private dwellings.

It should also be noted that the difference in the means of around 0.3-0.4, compares with the interquartile variation (the difference between the first and third quartile) of around 1.5-1.6 for ASHPs and 0.5-0.6 for GSHPs (even after cropping), so the variations identified with one of the most basic characteristics describing HP systems are less than those due to other factors (including measurement issues).

3.2 Comparison of emitter types: Underfloor heating vs radiators

Within ASHPs and GSHPs, the type of emitter represents another key technical characteristic that would be expected to result in differences in SPF. The simplified physical explanation is that:

- the area of floor in most dwelling types exceeds the likely area of radiators by a factor between 5 and 10;
- the heat transfer coefficient (Watts per Kelvin difference between room and heat emitter temperatures) is therefore likely to be higher for underfloor heating than for radiators;
- the higher heat transfer coefficient can be traded off against a lower operating temperature for underfloor heating.

However, this mechanism is not expected to apply in all cases. Some underfloor systems are expected to require higher temperature heat to be delivered to them. One variable influencing flow temperature in underfloor systems is floor construction: for example, suspended wooden floors are likely to require higher flow temperatures. Conversely, very-well insulated dwellings with underfloor heating may be able to operate at lower flow temperatures, other things being equal.

The emitter types for this analysis were categorised as being either Radiators or Underfloor heating, with too few sites in the 'Both' (radiators and underfloor heating) group for inclusion in the analysis (n=8 for ASHPs and n=9 for GSHPs). The relatively small sample size for the underfloor group shifts the analytical focus to the use of the Wilcoxon T-test, and the KS test to detect differences in the SPF distributions.

Table 3-3. SPFH2 by emitter type for ASHPs in Sample B2 (cropped)

HP type	Emitter	N	Mean	Median (IQR)
ASHP	Radiators	256	2.64	2.64 (2.33, 2.91)
	Underfloor	28	2.80	2.91 (2.50, 3.14)
GSHP	Radiators	58	2.88	2.80 (2.64, 3.03)
	Underfloor	25	3.04	2.84 (2.31, 3.8)

Table 3-4. SPFH4 by emitter type for ASHPs in Sample B2 (cropped)

HP type	Emitter	N	Mean	Median (IQR)
ASHP	Radiators	257	2.40	2.43 (2.15, 2.65)
	Underfloor	28	2.52	2.53 (2.2, 2.82)
GSHP	Radiators	58	2.73	2.70 (2.54, 2.89)
	Underfloor	25	2.85	2.67 (2.26, 3.34)

Considering ASHPs first of all, initial inspection of figures in Tables 3-3 and 3-4 suggest some improvement in performance for sites with underfloor heating, such that the median SPFH2 for ASHPs with underfloor heating appears higher than than for sites with radiators. However, these differences were only significant for SPFH2 and not SPFH4. The empirical distributions (Figure 3-7 and 3-8) also support this, with the KS test only identifying a significant difference in the shape of the distribution for SPFH2 (KS p-value=0.01).

For GSHPs, with smaller sample sizes in the two groups, only weak evidence of differences in the distributions was found, and only for SPFH4. One aspect that is suggested from the visual inspection of all the empirical distributions (Figure 3-9 to 3-13) is that the divergence between groups tends to occur for better performing sites. Specifically, when the analysis for GSHPs was repeated with the reduced sample of only those sites with SPFH4 > 2.5, those with underfloor heating were found to have significantly higher SPFs at the both H2 and H4 boundary. Using the same restriction for ASHPs, however, produced no equivalent change in results.

In other words, only small differences in performance between underfloor heating systems and radiator systems were observed. Larger differences *might* have been expected on thermodynamic grounds, and have been observed in other studies – for example, the DTI study (see Section 1.4) showed that GSHPs performed 22% better with UFH than with radiators (SPF_{H4} of 2.7 (radiators: N=18) compared to 3.3 (underfloor: N=23)[8,9]. For Sample B2 (cropped), a similar magnitude for the difference in performance

[8] Unfortunately the low number of ASHPs in the DTI study does not permit a reliable comparison for ASHPs.
[9] Flow temperatures reported by DTI are around 10°C higher with radiators than with UFH. However flow temperatures with radiators have short term peaks of 60°C, about 17°C higher than the 43°C peak temperatures for UFH.

for GSHPs seen in the DTI study only occurs when the comparison is carried out on sites with SPFH4 > 2.5.

The absence of this signal in the RHPP dataset is most pronounced for the HP sites with below average performance, where the distributions tend to converge. This may be due to metering issues and resultant noisy data, but one must also consider the possibiity that the performance of some of the underfloor systems in the RHPP study had been compromised. The heat transfer coefficient of an underfloor heating system would be reduced if:

- the system occupied a smaller proportion of floor area of the dwelling than intended;
- pipes in areas heated with UFH were spaced further apart than recommended;
- floor coverings such as carpets that were not considered during the design of the system had been added to areas of floor with UFH;
- the system had been fitted with a three-port mixing or blending valve (this would be contrary to MCS guidance, but such systems are commonly installed in dwellings heated by gas or oil boilers).

Anything that reduced the heat transfer coefficient of an underfloor heating system, would require that system to be operated at a higher than intended temperature to deliver the same quantity of heat. Note that three of the above four performance degradation mechanisms would be difficult to confirm without detailed investigation and access to design calculations.

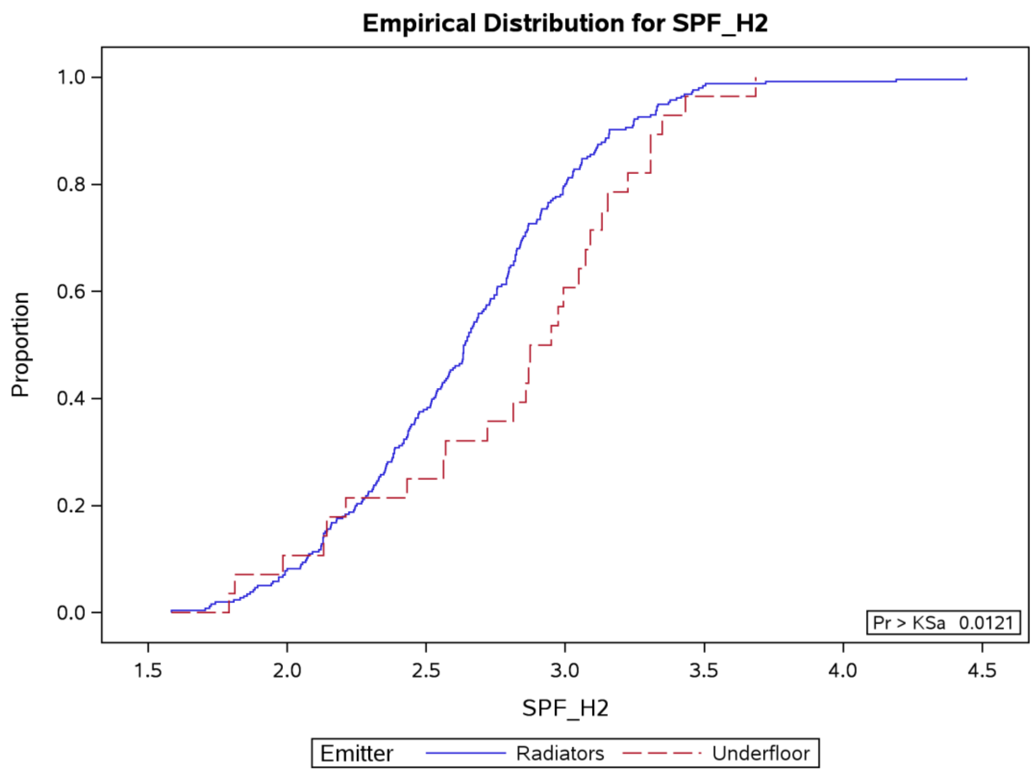


Figure 3-9. Comparison of SPFH2 by emitter type (radiators or underfloor) for ASHPs in Sample B2 (cropped).

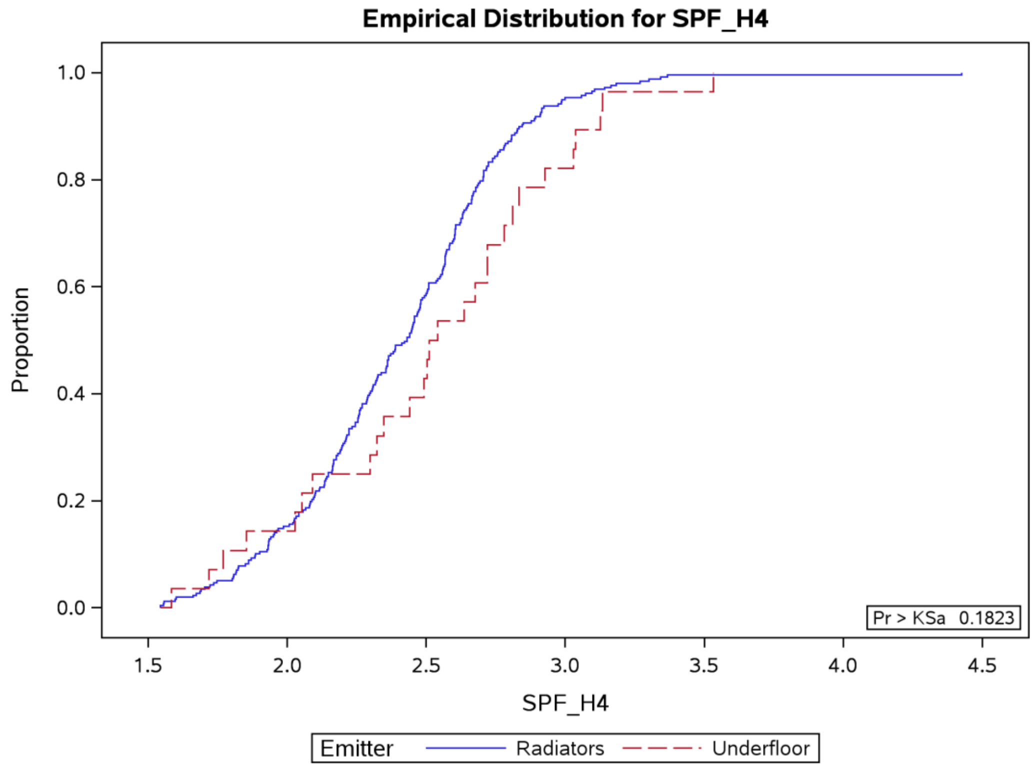


Figure 3-10. Comparison of SPF_{H4} by emitter type (radiators or underfloor) for ASHPs in Sample B2 (cropped).

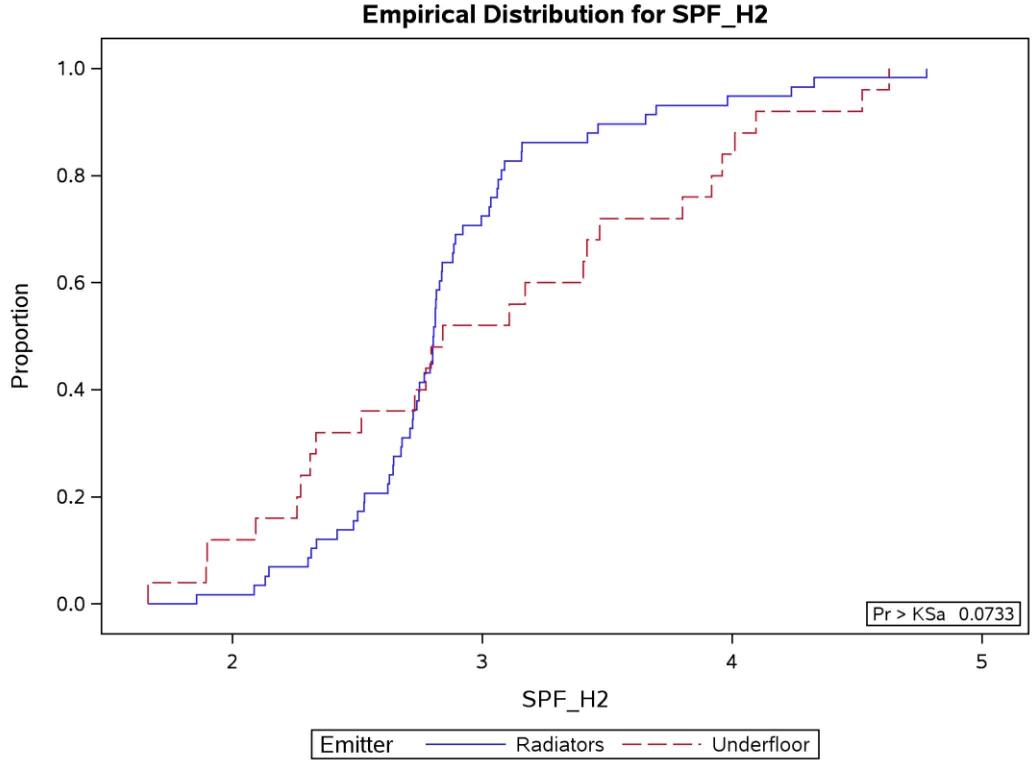


Figure 3-11. Comparison of SPFH2 by emitter type (radiators or underfloor) for GSHPs in Sample B2 (cropped).

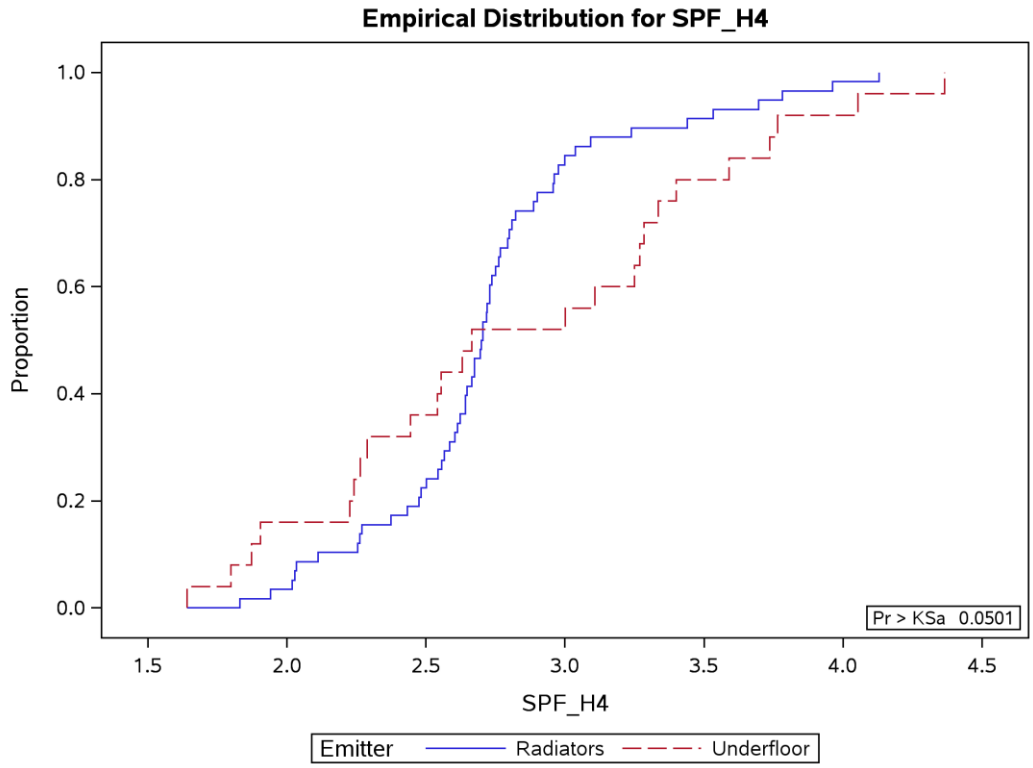


Figure 3-12. Comparison of SPFH4 by emitter type (radiators or underfloor) for GSHPs in Sample B2 (cropped).

In summary, while ASHPs with underfloor heating tended to perform slightly better than those with radiators at the H2 boundary, with the inclusion of other inputs and metering at the H4 boundary the distinction was no longer significant. For GSHPs, the slightly better performance of underfloor heating was only evident for those sites with higher values for SPFH4. It appears that the effect of emitter type on performance is largely overshadowed by other factors, especially at the low end of performance.

Noisy data monitoring data may wholly or partially explain why the expected effect is not clearer. But it may not be the whole story. Factors associated with design and installation of UFH and subsequent occupant actions could compromise the performance of such systems and lead to a convergence of their performance with radiator systems.

In Section 4.5, flow temperatures are used to further investigate the performance of underfloor and radiator systems, and there is a discussion of possible reasons for the performance differences not being larger.

3.3 Comparison of Site or Housing type

This section compares HP performance according to the site type or housing type, categorised as Registered Social Landlords (RSL or social housing) and Private (essentially privately owned dwellings), which is one of key variables related to the type of site into which the HP systems were installed. Housing type is likely to be correlated with a wide-ranging social, sociotechnical, and technical factors, including: socioeconomic status and educational attainment, dwelling size and heating patterns. One point to note is that occupants of RSL housing receive no direct financial incentives (apart from potentially saving on energy bills) related to the HP system, whereas the occupants of private dwellings receive RHPP payment.

Regardless of the details, it is reasonable to expect that the occupants of two groups of sites are may have quite different motivations and levels of engagement with the operation of their HP system. It is unclear how these varied factors might combine to influence overall performance levels - it could go in either direction. For instance, if private homes are more varied they may pose difficulties in ensuring the quality of installation, whereas installation in similar RSL sites may benefit from 'learning' resulting in better or at least more consistent installation quality. Nevertheless, it is worth investigating if performance by housing type – one of the defining site characteristics available – is associated with distinct variations in SPF in the Sample B2 (cropped) dataset.

From Table 3-5 and Figures 3-15 and 3-16, it is clear that for ASHPs in RSL sites the distribution of SPFs for both H2 and H4 boundaries appear to closely match those of the Private group.

Table 3-5. SPFS by site type for ASHPs in Sample B2 (cropped)

SPF	HP type	Site Type	N	Mean	Median (IQR)
H2	ASHP	Private	77	2.60	2.57 (2.27, 2.92)
		RSL	215	2.67	2.66 (2.38, 2.97)
H4	ASHP	Private	78	2.50	2.45 (2.1, 2.78)
		RSL	215	2.38	2.43 (2.16, 2.63)

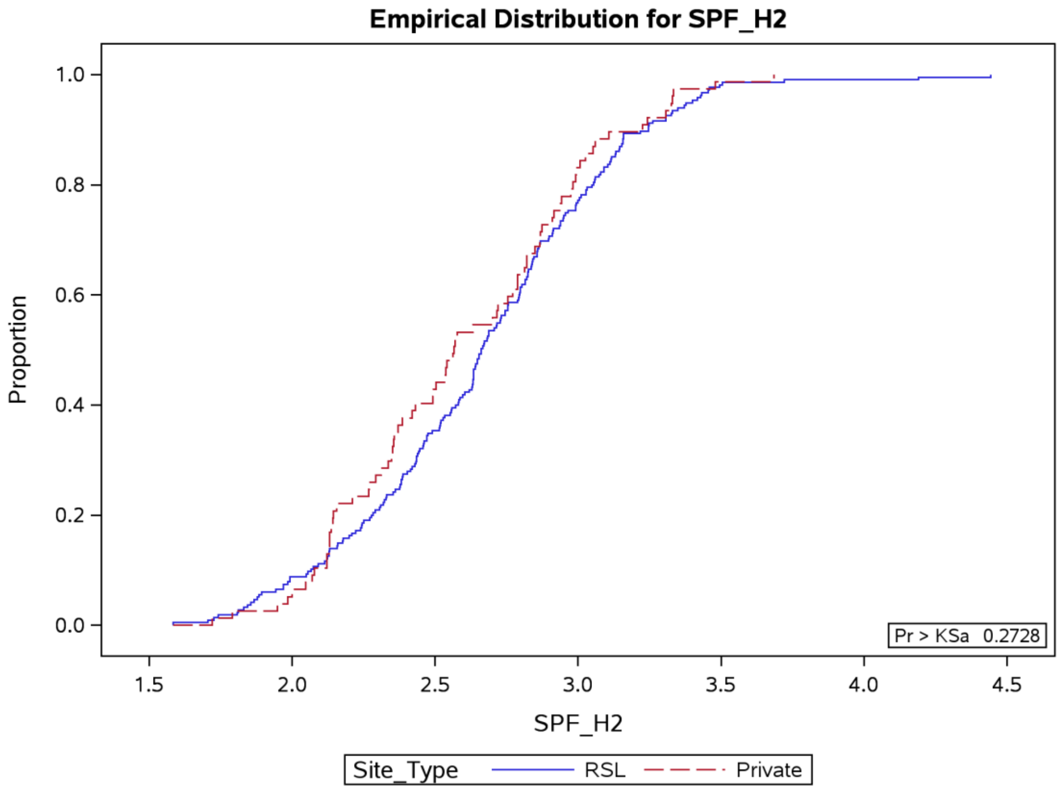


Figure 3-13. Comparison of SPFH2 by site type (RSL or Private) for ASHPs in Sample B2 (cropped).

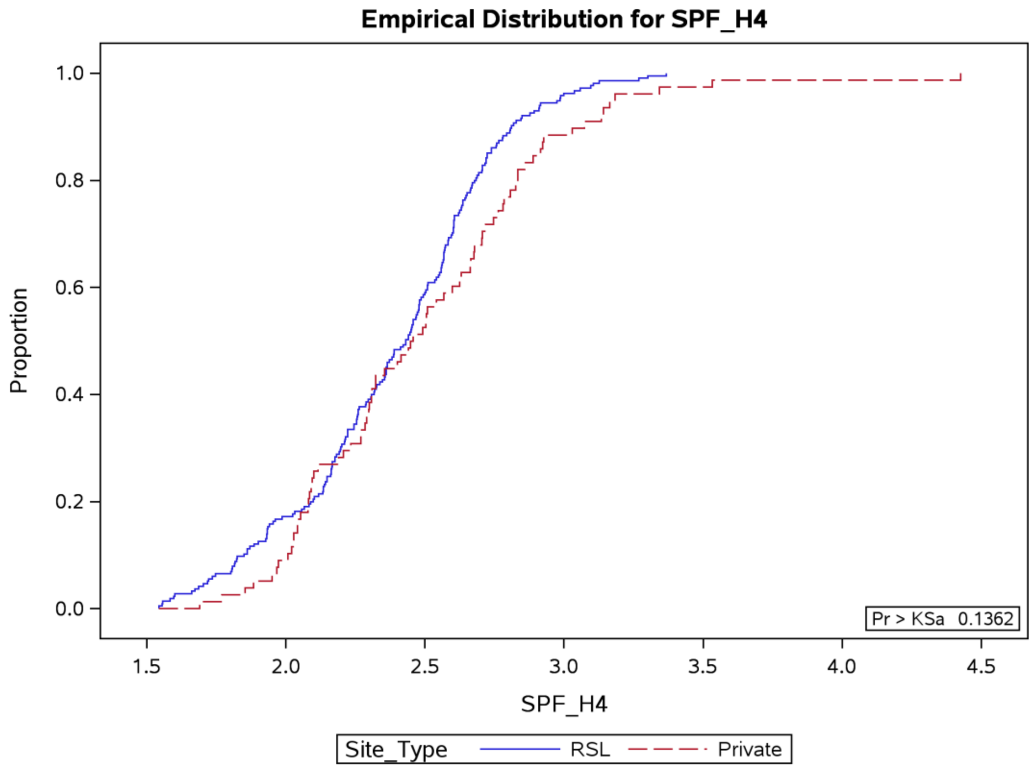


Figure 3-14. Comparison of SPFH4 by site type (RSL or Private) for ASHPs in Sample B2 (cropped).

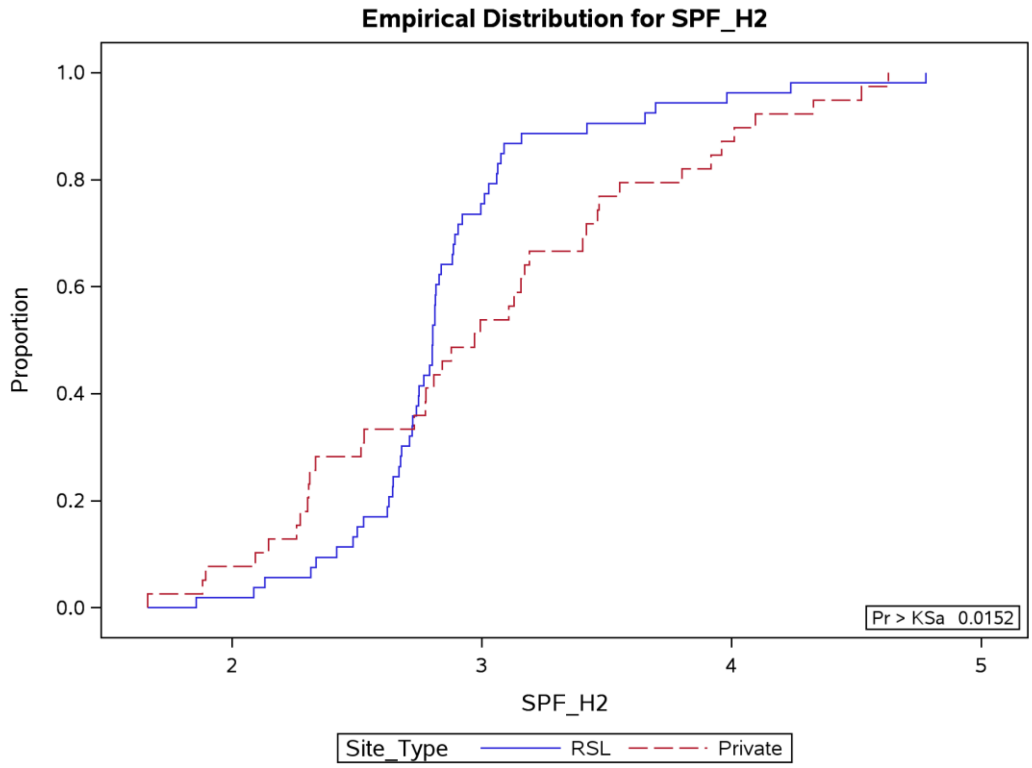


Figure 3-15. Comparison of SPFH2 by site type (RSL or Private) for GSHPs in Sample B2 (cropped).

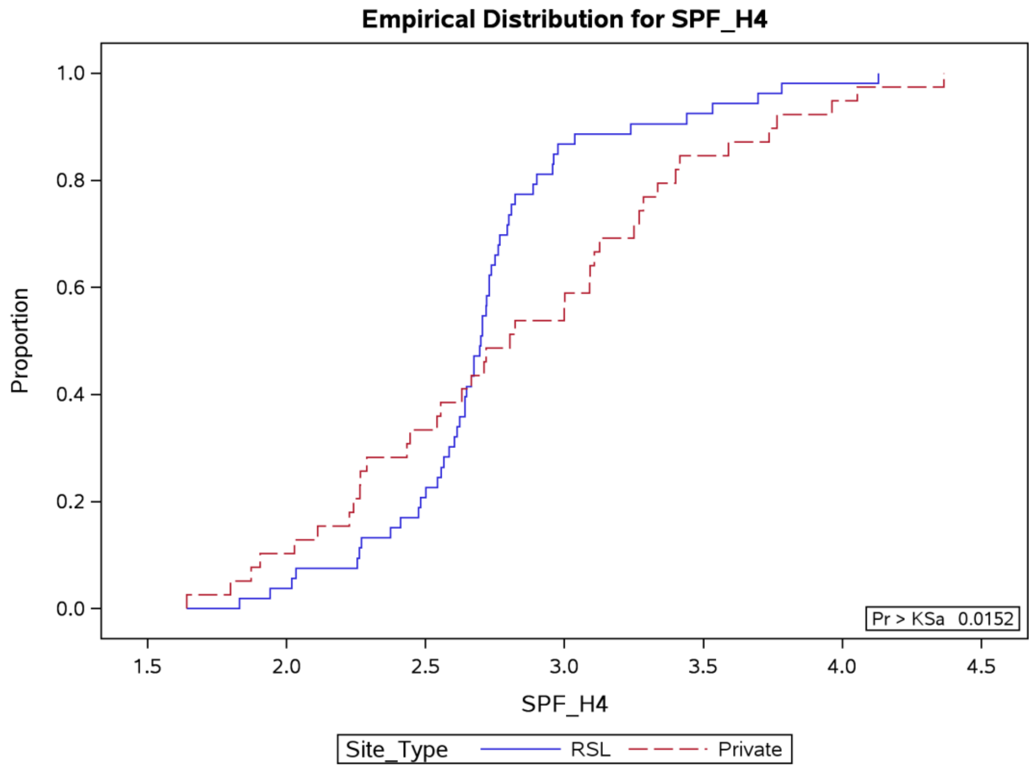


Figure 3-16. Comparison of SPFH4 by site type (RSL or Private) for GSHPs in Sample B2 (cropped).

Table 3-6. SPFS by site type for GSHPs in Sample B2 (cropped)

SPF	HP type	Site Type	N	Mean	Median (IQR)
H2	GSHP	Private	39	3.02	2.97 (2.31 - 3.47)
		RSL	53	2.87	2.80 (2.67 – 3.00)
H4	GSHP	Private	39	2.84	2.8 (2.27 - 3.28)
		RSL	53	2.72	2.7 (2.56 - 2.82)

In contrast with the ASHPs, the shape of the distribution for GSHPs is significantly different (Figures 3-17 and 3-18, K-S p-value = 0.015) between RSL and Private sites. The SPFs at H2 or H4 *do not* tend to be higher or lower, that is the distribution is not shifted in any particular direction, instead the distribution for RSL sites with GSHPs is narrower and more peaked than those in Private sites. This is also evident in the width of the interquartile range (IQR) of 0.33 (2.67 to 3.00) for RSL sites compared with 1.16 (2.31 to 3.47) for Private sites.

One potential explanation for why the shape of the SPF distributions for GSHPs differs by tenure is an example of 'confounding'. From the previous section on emitter type, the GSHPs with underfloor heating was characterised by a wide variation in SPF. In this comparison by tenure for GSHPs, none of the 53 RSL sites had underfloor heating whereas 25 of the 39 Private sites had underfloor heating. So it is likely that the difference in SPF observed for RSL and Private sites reflects this disparity in the emitter type. By contrast for ASHPs, only 17 out of 77 of the Private sites had underfloor heating, so their impact would be limited.

These results are consistent with idea of a wider variety of circumstances or contexts at work with the Private dwellings, and by contrast a more 'methodical' installation and/or operation and/or building types for RSL sites even if this has not resulted in better performance. The question then arises, why would this pattern only be evident for GSHPs but not for the ASHPs? It may be because the population of GSHP installations in RSL dwellings is fairly homogeneous in terms of building type, compared to the population of ASHP installations in RSL dwellings.

In summary, the results for ASHPs suggest that for all the potential sociotechnical factors that may come into play between the groups, they did not amount to overall differences in performance between the two sub-samples. For GSHPs the findings suggest that *data issues, operational, and/or installation factors,* particularly those with underground heating, may have come into play in the Private site context (including sociotechnical factors such as heating patterns in large dwellings) that resulted in far more variation in performance for those sites but not an overall performance improvement or penalty over RSL sites.

3.4 Comparison of HP models

A distinguishing characteristic worth checking for any signal in the performance data is the specific HP model installed. Only a few models had sufficient sub-sample size in Sample B2 (cropped) to permit a statistical analysis. For instance, if a specific HP model had consistently good intrinsic performance (e.g. at the point where each unit leaves the factory floor) then the analysis might reveal this group as having a distinct and shifted distribution for SPF. Further, we might expect this difference as more likely to occur at the H2 boundary, which includes less peripheral components in the performance calculation, than at the H4 boundary. RAPID-HPC selected two models of heat pumps, by the same manufacturer and compared performance; Model A has a capacity of 5kW while model B has a capacity of 8.5 kW.

Table 3-7. SPFs for Model A and Model B ASHPs in Sample B2 (cropped)

SPF	HP type	Model	N	Mean	Median (IQR)
H2	ASHP	Model A	>40	2.74	2.72 (2.46, 3.11)
		Model B	>40	2.61	2.63 (2.38, 2.92)
H4	ASHP	Model A	>40	2.47	2.51 (2.26, 2.67)
		Model B	>40	2.39	2.43 (2.2, 2.68)

In the first case examined, for ASHPs labelled as Model A the SPFH2 tends to be slightly higher (by around 0.1) than other ASHPs (Wilcoxon p-value=0.03). The distribution of SPF_{H2} for Model A is shifted slightly to the right across the range (Table 3-9 and Figure 4-23) compared with other ASHPs, but the KS test finds no change of *shape* in the distribution (KS p-value=0.1).

In contrast, there was no evidence of higher SPFH4 for Model A, and again no clear evidence of a difference in the distribution shape (KS p-value=0.06). Specifically, the part of the distribution for Model A with high performing sites tends to overlap with other ASHPs, but as SPFH4 declines the two distributions diverged until SPF_{H4} is ~2 or below. Given that there is evidence for a difference at the H2 boundary, the most that can be said about Model A at the H4 boundary is that it appears to have relatively fewer poorer performing sites.

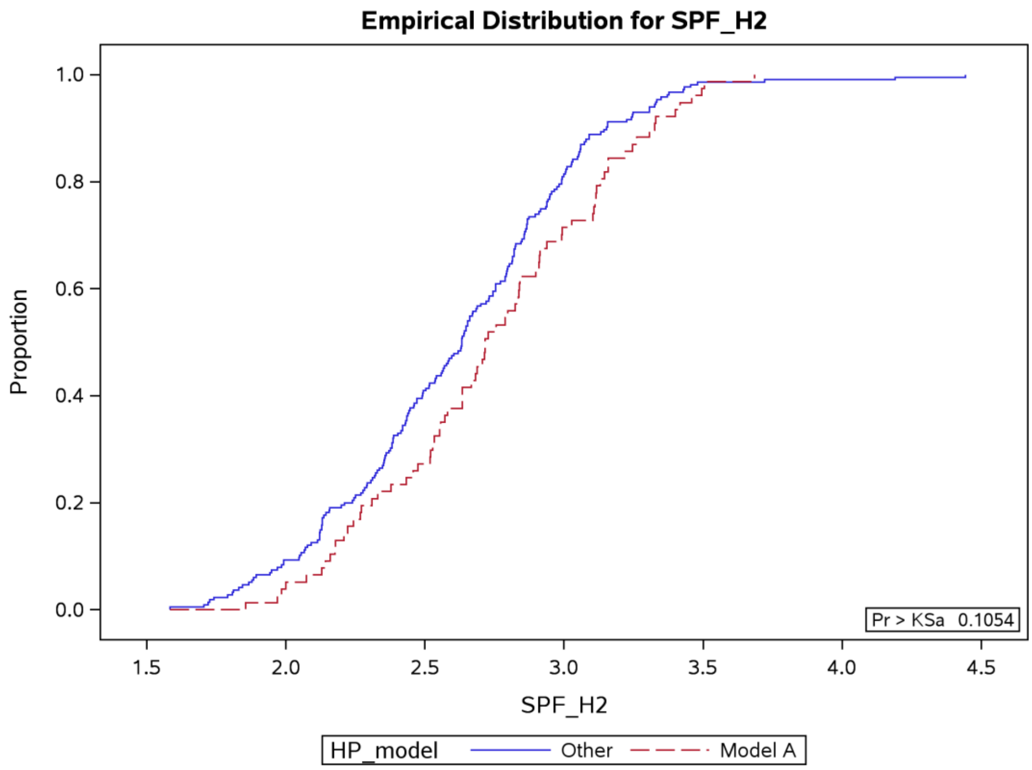


Figure 3-17. Comparison of SPFH2 for Model A with other ASHPs sites in Sample B2 (cropped).

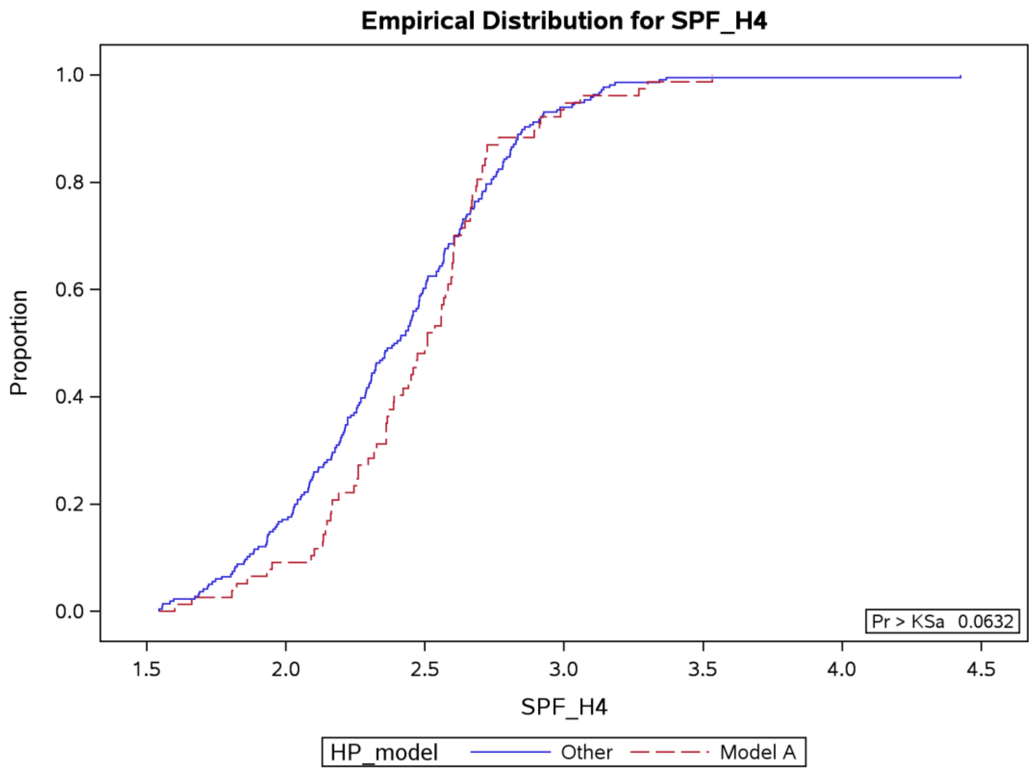


Figure 3-18. Comparison of SPFH4 for ASHP Model A with other ASHPs in Sample B2 (cropped).

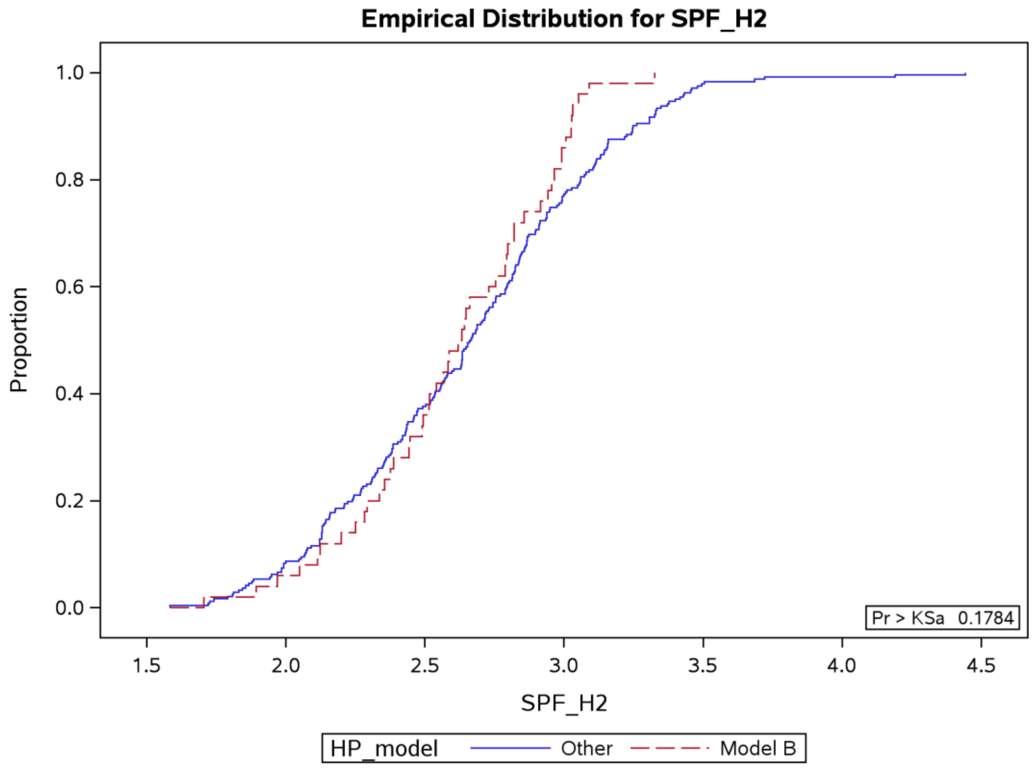


Figure 3-19. Comparison of SPFH2 for Model B with other ASHPs sites in Sample B2 (cropped).

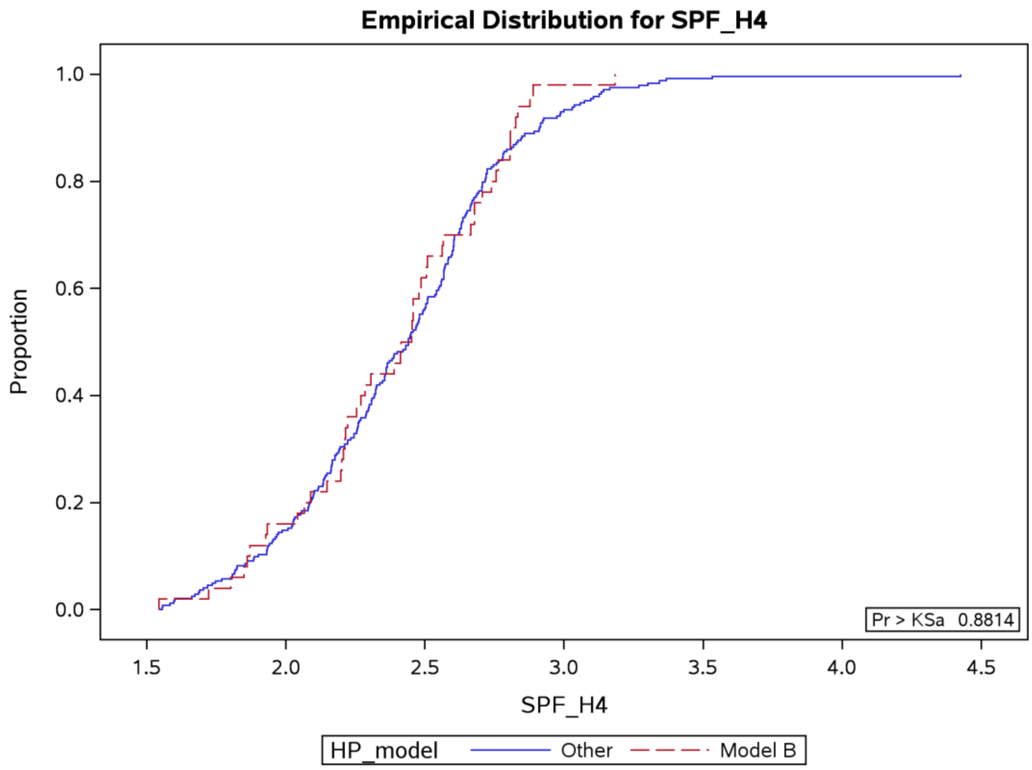


Figure 3-20. Comparison of SPFH4 for Model B with other ASHPs sites in Sample B2 (cropped).

When the analysis is carried out for Model B (Table 3-9), which as far as we are aware has similar specification to Model A except for a higher output capacity (8.5kW compared with 5kW for model A), no evidence is found of a significant difference in the distribution for SPFH2 and SPFH4 with other ASHPs (Figures 3-23 and 3-24).

In summary, there is little evidence from this analysis of two similar models that points to the performance advantage of one HP model over another as being a key factor: rather the overall evidence suggests that other factors may be more important in their impact on performance. The scope of this analysis is greatly constrained by the small sample sizes for most models meaning that only a few could be tested here. However it does suggest that at least one model had more robust performance at the H2 boundary, where the combination of factors from the model to the installation on site, to metering, and operation have tended to result in fewer underperforming sites compared with other models. To determine if there is a specific underlying factor that has enabled that outcome to occur (such as relative simplicity of installation) requires further detailed on site investigation. But it is important to highlight that the advantage in performance (of around 0.1 between medians) of Model A over the other ASHPs was much less than the overall variation in performance. This suggests the need to ensure the set of broader factors influencing measured performance should not be overlooked in preference to focussing on technical advantages within any specific HP 'unit'.

Issues in metering quality certainly contribute to the variation in measured SPF. Even if all the metering factors were addressed, however, only few models of heat pump had sufficient sample size for any analysis to have sufficient statistical power to detect a significant difference in SPF.

4 Further analysis

Section 2 explained some of the metering and processing issues apparent in the RHPP dataset, especially in the heat meter data. Section 3 demonstrated that using the cleanest dataset (Sample B2 cropped) to find and explain statistically significant differences in SPF between groups is difficult due to the amount of noise in the dataset. In this section, exploratory investigation is undertaken using metrics other than SPF, since SPF is highly dependent on the heat meter data. This allows other markers of good and bad performance to be discussed.

4.1 Cycling

Method

An algorithm was developed to determine cycle length from the electricity consumption of each heat pump. This allowed calculation of median cycle length for each site, and investigation into the prevalence of cycle lengths short enough to cause detrimental impact on SPF.

The definition of 'cycle length' here is the time from the compressor coming on until the next time it comes on. This is illustrated in Figure 4-1, which demonstrates a cycle length of 38 minutes. This number only applies to the highlighted cycle; cycle length is variable over any one site.

In this report, the term 'cycling' is simply taken to mean the compressor switching on and off but not necessarily on a timescale leading to inefficient performance; the term 'short-timestep cycling' is taken to mean cycling over a timescale of around 12 minutes or less. Previous research identified negative impacts on SPF when compressor-on time is under 6 minutes (Green, 2012). The amount of time the compressor is on is not visible to a high precision in the RHPP dataset due to its 2-minutely resolution, therefore a 12-minute cycle length according to the definition used here is taken to be the closest possible equivalent to 6-minute compressor-on time in the aforementioned previous research. It should be noted that some heat pumps have control strategies that prevent more than 6 restarts per hour.

Note that since the definition of cycling is based on the compressor's operation, it does not distinguish between space heating and DHW modes.

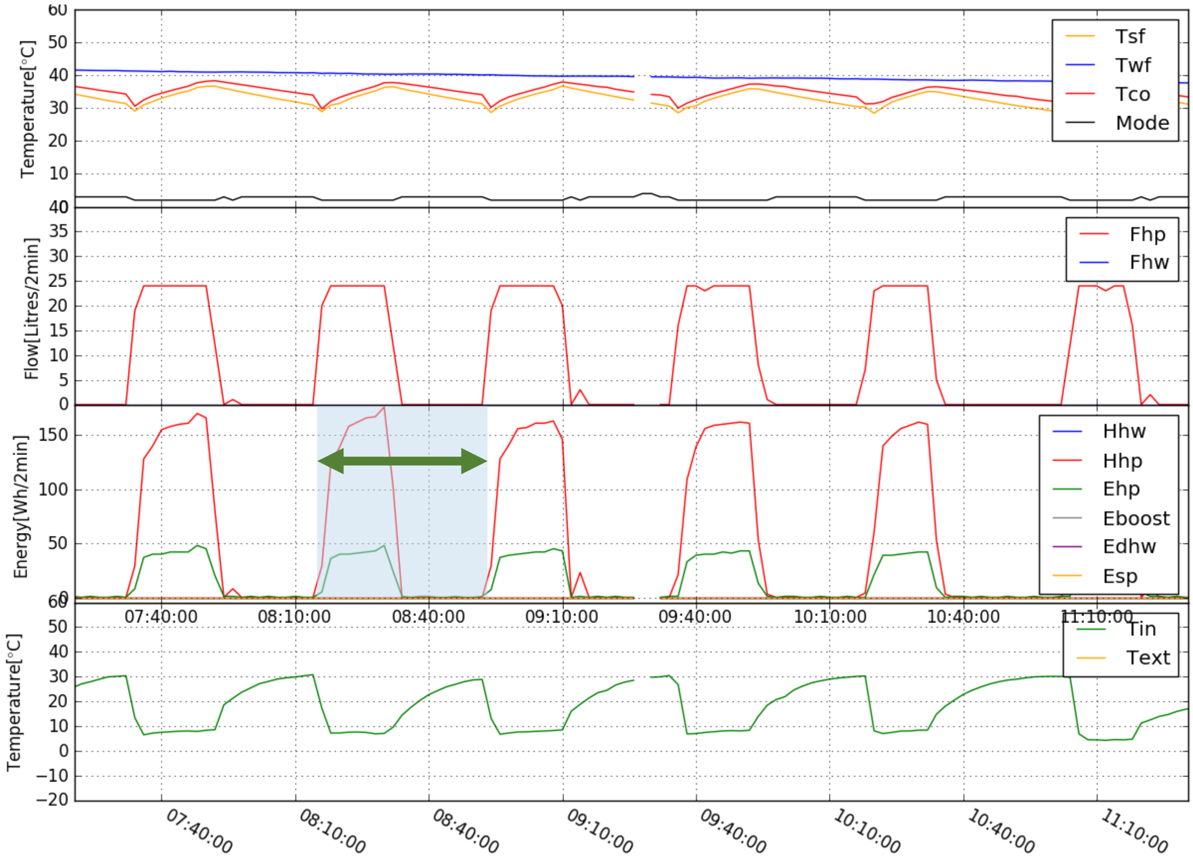


Figure 4-1. Illustration of cycling, with definition of cycle length shown in green.

In many sites, calculating each cycle length is not as simple as finding the time between two consecutive occurrences of Ehp rising from zero to a positive number, as parasitic loads recorded within Ehp cause a baseload electricity consumption even when the compressor is not functioning as such. To automate the calculation of cycle length, the algorithm used here estimates a threshold which signals the difference between ‘on’ and ‘off’ using a frequency distribution of Ehp. The threshold is taken to be the midpoint between Ehp=15 Wh/2 mins (450 W) and the largest peak to the right, as demonstrated in Figure 4-2. The use of the Ehp = 15 point (as opposed to Ehp = 0) is to make sure the largest peak found is not the one near zero which is likely to be composed of parasitic electricity consumption as opposed to the compressor coming on.

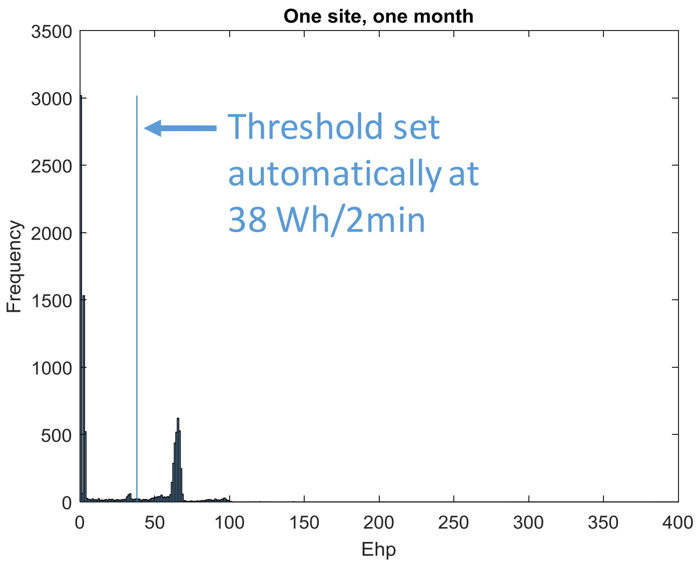


Figure 4-2. Threshold Ehp in Wh set by the algorithhm; one example month and site.

This threshold is calculated separately for each month of data for each site. The differentiation between months was deemed necessary as cycle lengths were expected to change between months for one site. The example in Figure 4-2 is shown in timeseries form in Figure 4-3, in which the threshold can be used to determine when the heat pump is on and when it is off.

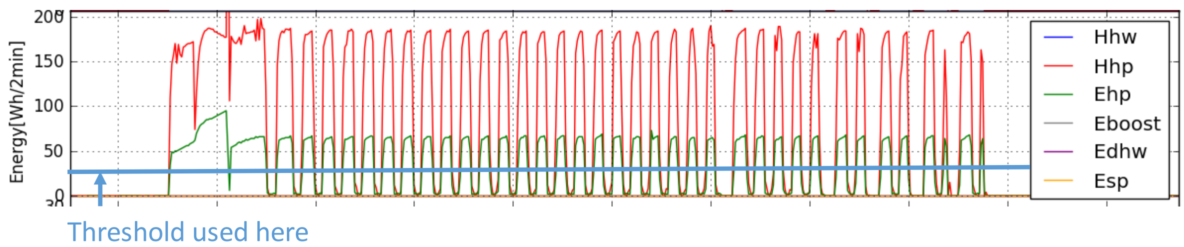


Figure 4-3. Use of the threshold to determine on/off states of the HP.

Cycle lengths for every cycle are then found by calculating the distance from each occurrence of Ehp crossing the threshold in a positive direction to the next occurrence of this. Note that this definition does not provide information about the proportion of time within the cycle that the compressor was running – an on-time of 4 minutes followed by an off-time of 16 minutes is treated the same as an on-time of 16 minutes and an off-time of 4 minutes.

Two metrics are used below to summarise the information thus gained: median cycle length for each site for each month (from herein one month of data from one site is termed one '*site-month*'), and proportion of cycles of length less than 12 minutes, as found by Green (2012) to be detrimental to SPF.

Results across the sample of site-months

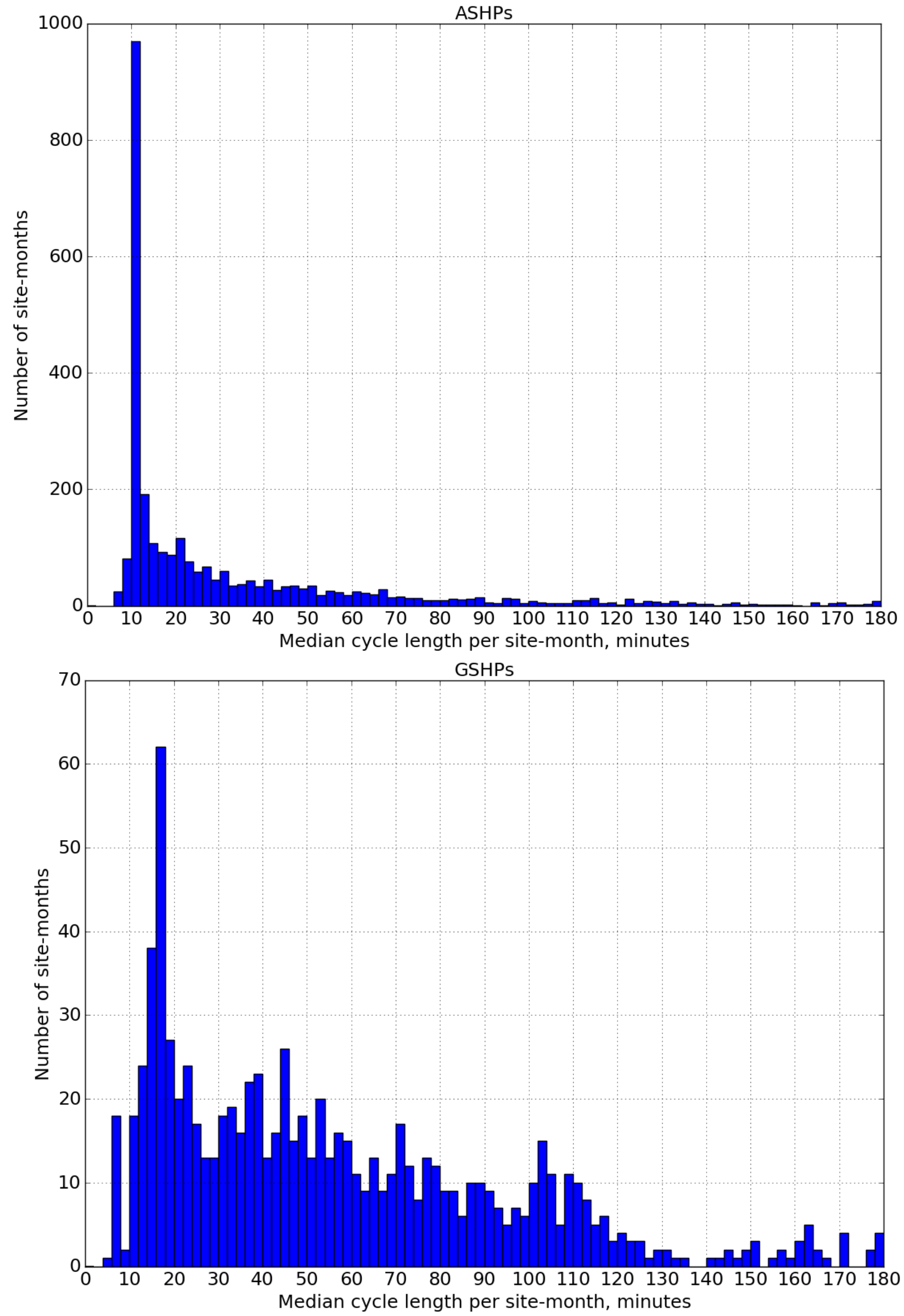


Figure 4-4. Median on-to-on cycle lengths for each site-month.

Figure 4-4 shows the distribution of median cycle lengths, for ASHPs and GSHPs respectively. It can be seen that for ASHPs there is a very distinctive peak at 10 minutes; for GSHPs the most common median cycle length is 18 minutes and there are more cycles of greater lengths.

Concerning the 10-minutely cycling of ASHPs, 49% of ASHP sites in Sample B2 (cropped) have at least one month where this is the median cycle length – although if it occurs at all it tends to occur in several months and can occur at any time of year. From visual inspection of the data, 10-minutely cycling of ASHPs appears to occur predominantly in space heating mode. For example, Figure 4-5 illustrates 10-minutely cycling behaviour in an ASHP in space heating mode in August.

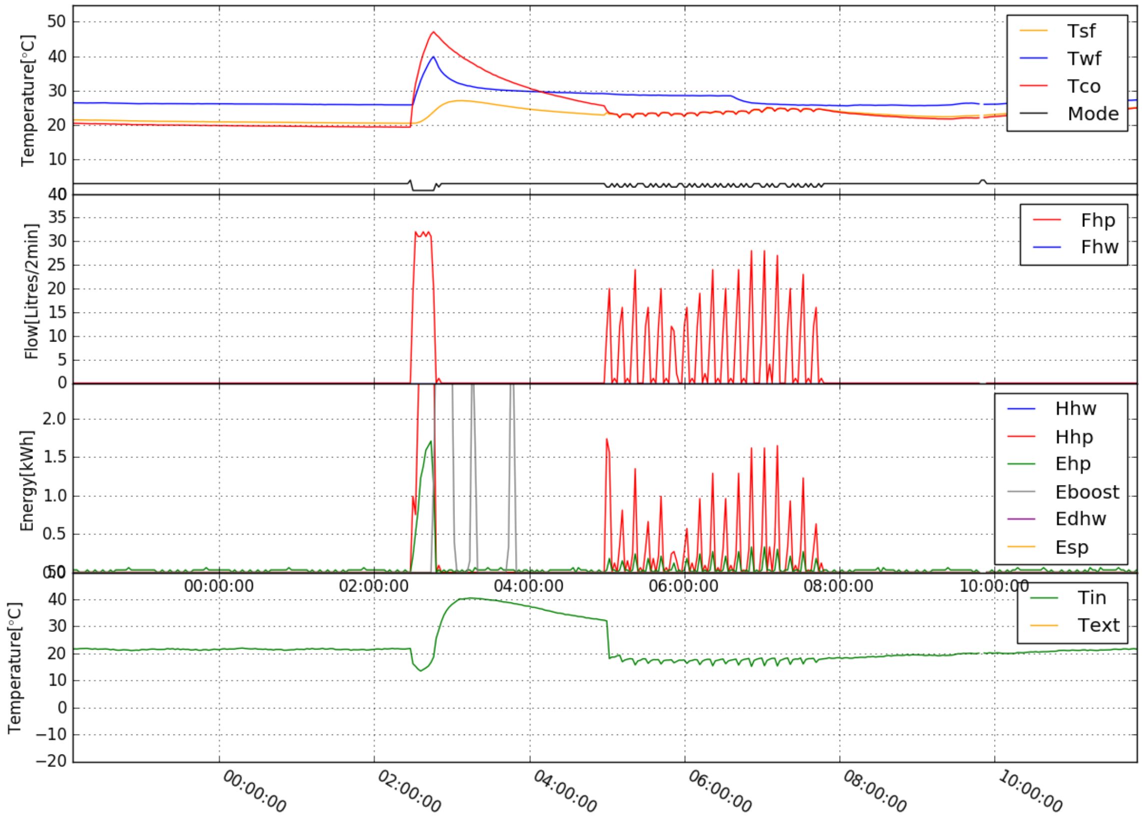


Figure 4-5. Example of 10-minutely cycling in an ASHP in space heating mode in summer.

A 10-minutely frequency is considered as high, and further work is recommended to understand its cause. Given that most installations were in existing dwellings with previous heating systems, the widespread short timestep cycling could have arisen from the interaction of HPs with the rest of the heating system.

In terms of whether monthly median cycle length affects monthly COP, Figure 4-6 shows no clear relationship – this does not mean that none exists but instead indicates that other factors (including heat metering error) contribute more to COP than median cycle length. It can be seen that those ASHPs with cycle length of 10 minutes and GSHPs with cycle length of 18 minutes (the peaks in Figure 4-4) appear to yield a range of COPs.

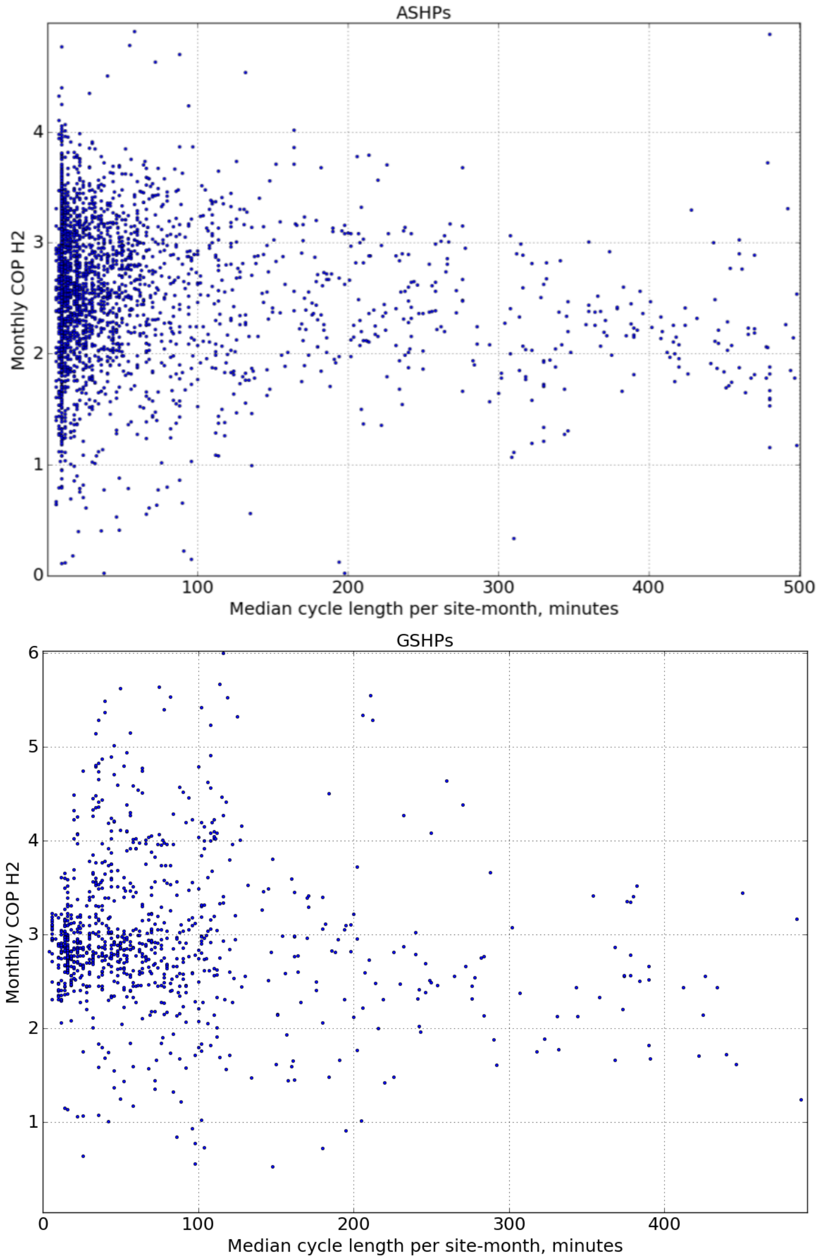


Figure 4-6. Monthly COP (H2 bound) versus median on-to-on cycle length per site-month.

A related metric to examine is proportion of cycles of length less than 12 minutes. Note that since the sampling frequency was 2 minutes, shorter cycles will not reliably show up in the timeseries data (one would expect aliasing and beating effects that would be hard to disentangle). Therefore the following results may be an underestimation of the actual extent of cycling occurring.

The annual extent of sub 12 minute cycling in the sample is shown in Figure 4-7. Most ASHP sites have at least one month per year in which more than 10% of all cycles are shorter than 12 minutes, and in 25% of sites, all months have more than 10% of cycles shorter than 12 minutes. This is not true for GSHPs where only 1 site showed the latter behaviour.

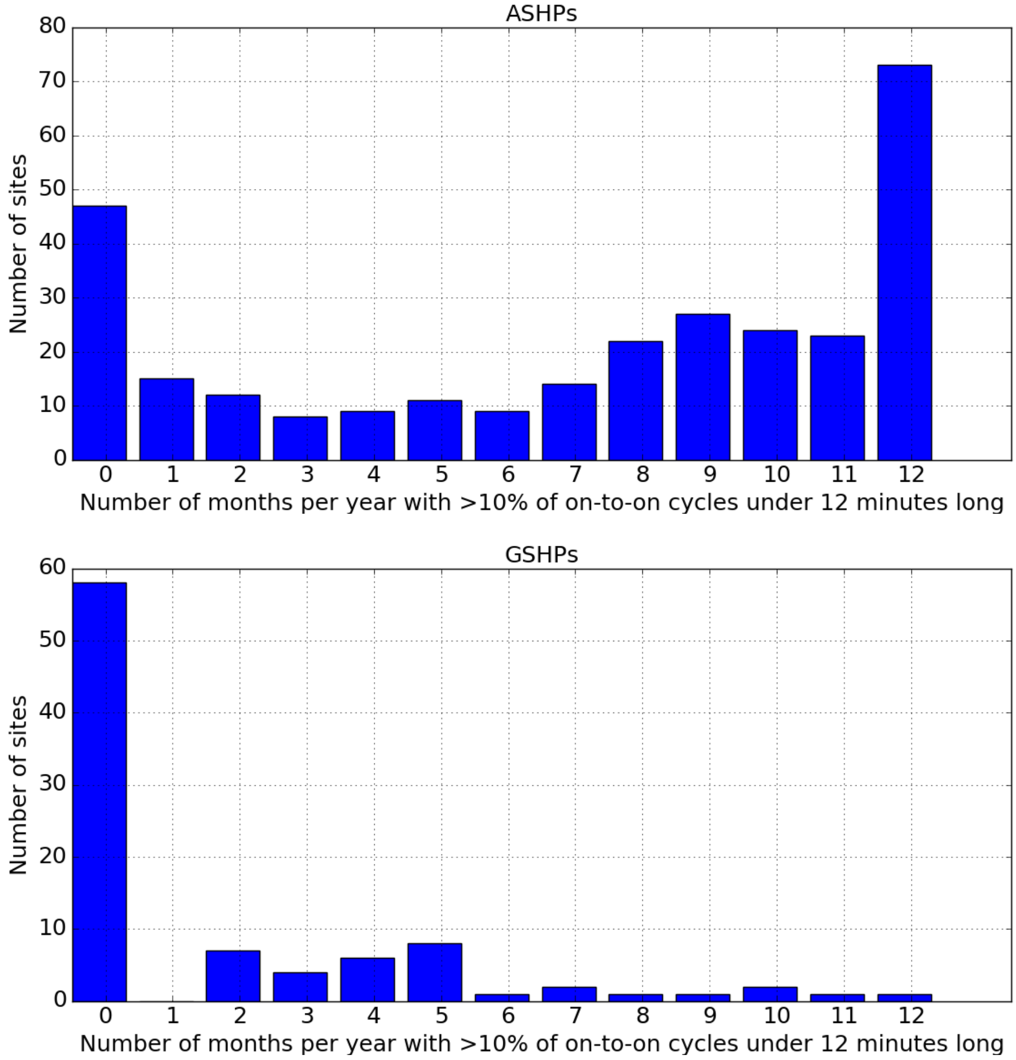


Figure 4-7. Showing the extent of sub 12 minute on-to-on cycling in the sample.

As shown by the top subplot in Figure 4-7, ASHP sites can have multiple and even all months with a high occurrence of short timestep cycling. This gives evidence that short timestep cycling can occur at any time of year. For GSHP sites, short timestep cycling tended to occur more in the winter months but again could occur in any month.

From identifying the worst sites in terms of extent of short timestep cycling, further investigation can be carried out on a site-by-site basis to illustrate particular types of cycling. These are described, but not quantified in terms of their prevalence in the sample as a whole.

Cycling in DHW mode

The ASHP with the highest occurrence of cycle lengths less than 12 minutes is RHPP5802, in July. Examining the timeseries data for this site in this month, as in Figure 4-8, shows that rapid cycling is occurring whilst the HP is in DHW mode. Instead of the compressor turning off and staying off the first time that the cylinder thermostat is satisfied, the thermostat appears to call for heat again very quickly, and is then satisfied very quickly. This occurs for around 2 hours at a time. Possible reasons include the cylinder stat being set higher than the maximum flow temperature, or the deadband on the cylinder stat being too narrow.

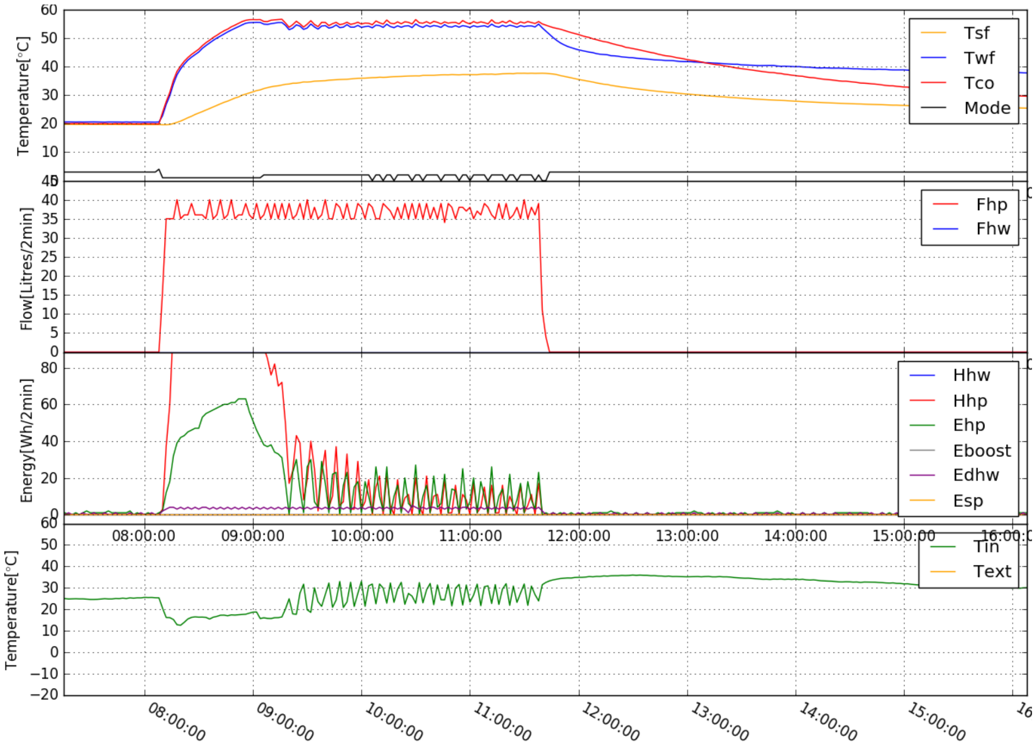


Figure 4-8. Example of short timestep cycling in DHW mode in summer in an ASHP.

Cycling during space heating mode in winter

The GSHP with the highest occurrence of cycle lengths less than 12 minutes is RHPP5763, in February. An hour of data from this site-month is shown in Figure 4-9.

Figure 4-9. example of short timestep cycling in a GSHP in winter.

Figure 4-9, from a GSHP system with radiators, shows very low condenser and space heating flow temperatures (Tco, Tsf, around 20°C) for space heating from Feb-Aug 2014 but not for the periods before and after this. These low temperatures indicate an error with the heat pump system and/or controls during this period. A different example shows more typical winter cycling: Figure 4-10 shows an ASHP system with underfloor heating whose condenser and space heating flow temperatures rise to around 40/~35°C respectively, then dropping, then rising again on a timescale of around 6 minutes.

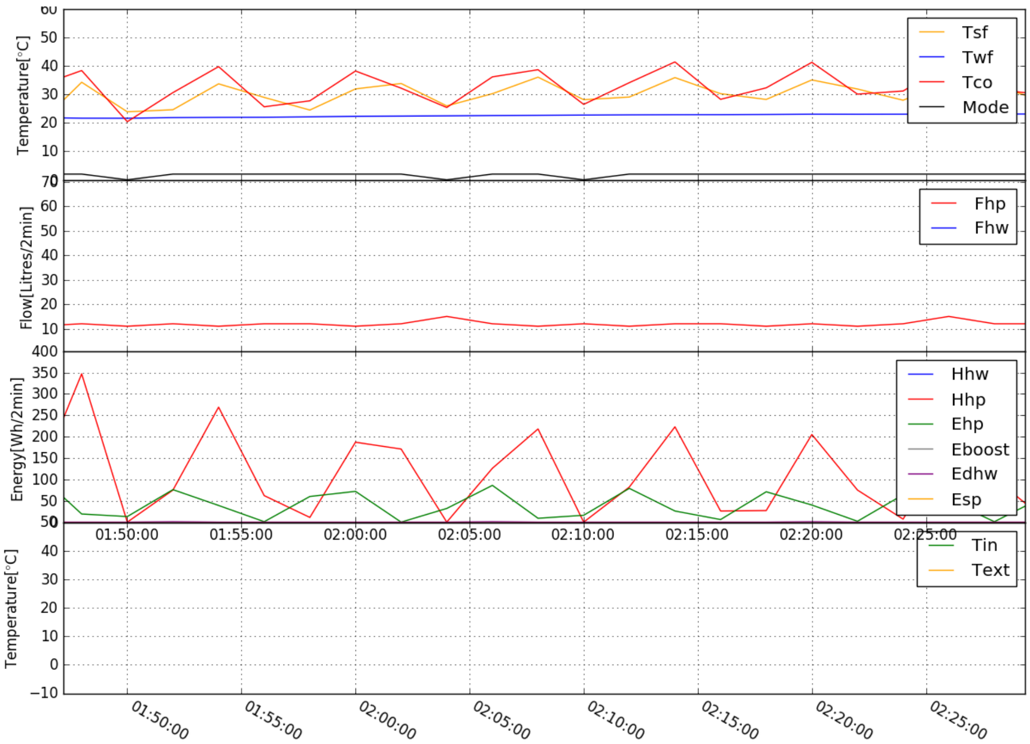


Figure 4-10. Example of short timestep cycling of an ASHP in winter with rapid flow temperature change.

Cycling during space heating mode in summer

It is the normal assumption that people would tend to turn off their gas boiler heating systems in summer. However, this may not necessarily be the case with heat pump systems; this depends on how the controls and room thermostat are set. In many of the sites the space heating functionality is turned off in summer[10]. For those sites in which it is enabled, ASHPs and GSHPs may respond differently:

- Most of the ASHPs in the sample have variable speed compressors, so can operate at part load up to an extent;
- Most of the GSHPs in the sample have fixed speed compressors, so at times of very low space heating demand are unlikely to come on; if they did, cycling would occur. GSHPs should however have controls which stop the time period of the cycling being short enough to decrease efficiency.

[10] It is difficult to state the exact proportion of sites with space heating switched off in summer due to some false apparent space heating demand arising as a result of the processing difficulties described in Section 2.2.

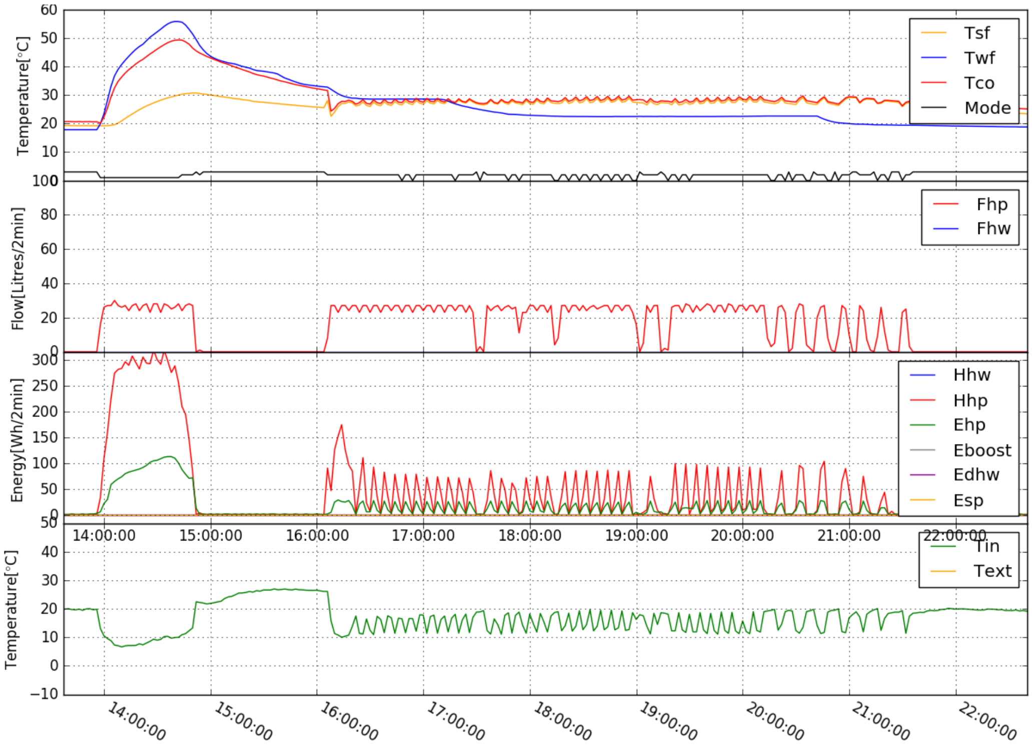


Figure 4-11. ASHP carrying out summer space heating (July).

Figure 4-11 shows an ASHP commencing space heating mode around 16:00 and exhibiting some degree of modulation downwards as the load is low – however this modulation is not sufficient to prevent short-timestep cycling (7-8 cycles per hour).

As for a GSHP example, Figure 4-12 shows a GSHP performing space heating in July; the top plot shows the entire month, with space heating coming on on certain days, and the bottom plot is a magnified version of one heating event. This system does not have the capability to modulate.

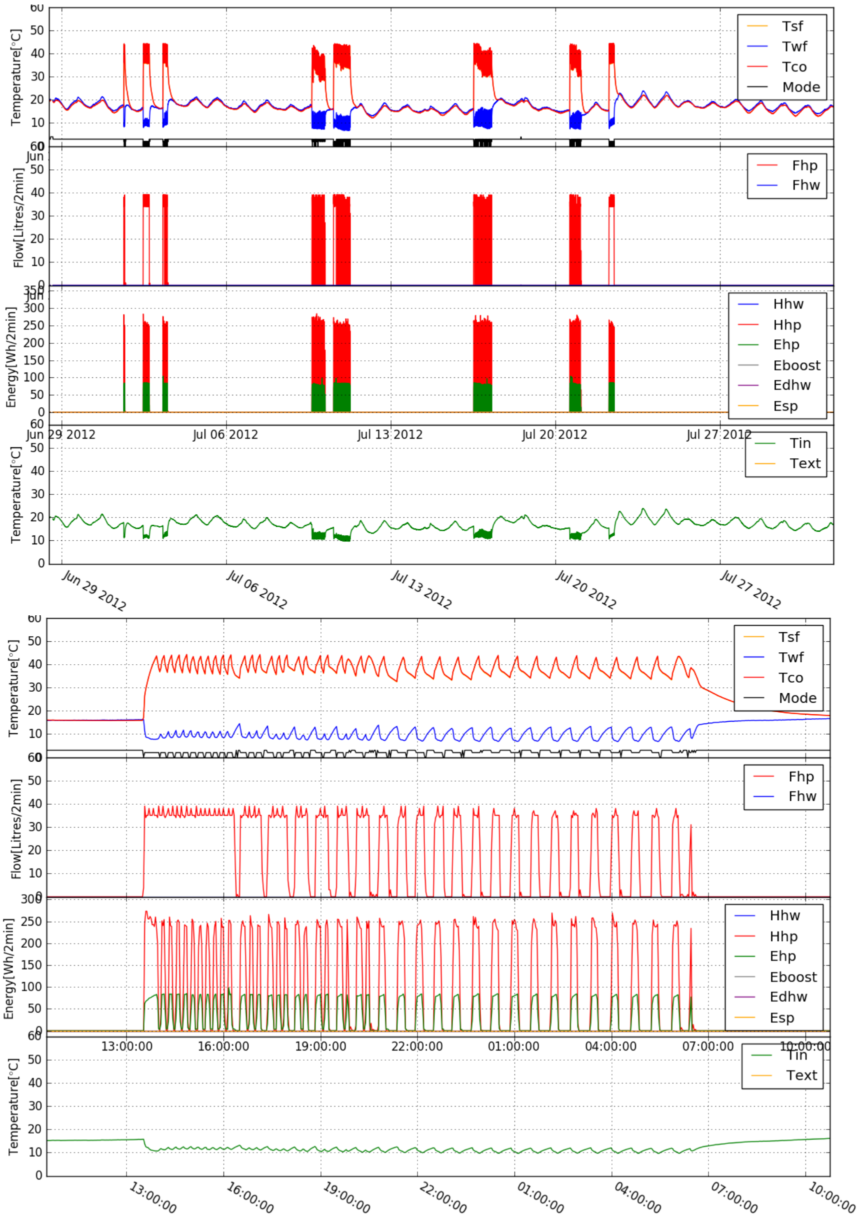


Figure 4-12. GSHP carrying out summer space heating (July). Zoomed-out plot (top) shows one month of data; zoomed-in plot (bottom) shows 1 day.

Good examples of avoiding short timestep cycling

As discussed above, most of the ASHPs in the sample have variable speed compressors and as such if they are correctly working, then their output can modulate down depending on the instantaneous demand for heat. An example of good operation is shown in Figure 4-13. This heat pump turns on just after 06.00 and, by gradually decreasing its heat output, stays on almost continuously until around 15.00.

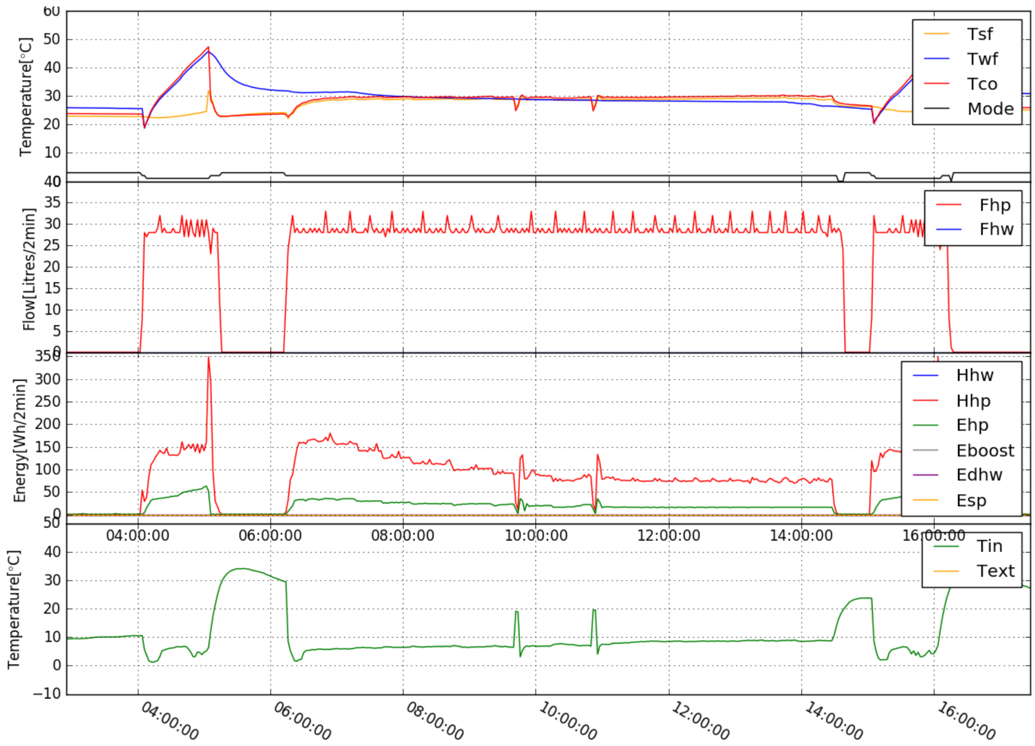


Figure 4-13. Example of an ASHP modulating its output over several hours.

Conversely, most of the GSHPs in the sample have fixed speed compressors so cannot carry out the modulation behaviour illustrated above. However, it is possible to identify well performing GSHP sites whose HPs stay on for around half an hour, then turn off for about the same time, as shown in Figure 4-14 (GSHP with radiators).

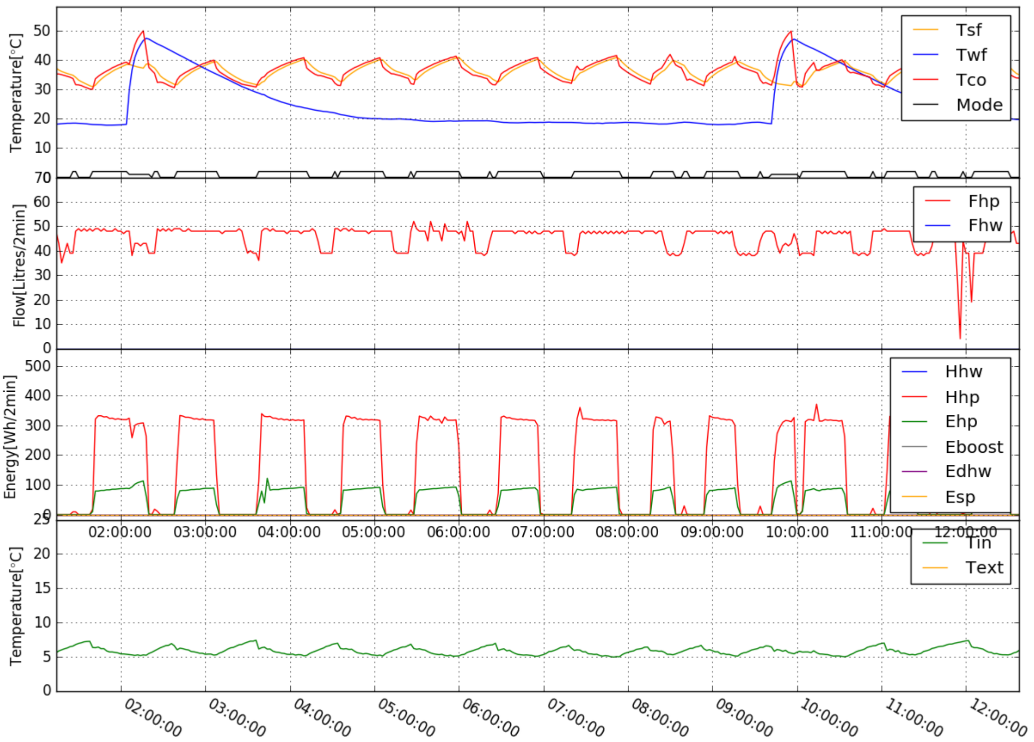


Figure 4-14. Example of an hourly cycle time and COP around 3.

Conclusion to cycling analysis

In conclusion:

- On-to-on cycles with length less than 12 minutes are common in the sample of ASHPs but not GSHPs.
- The most common median on-to-on cycle length per site-month is 10 minutes for ASHPs and 18 minutes for GSHPs.
- A relationship was not observed between median cycle length per site-month and monthly COP. Either this is genuinely the case or the high degree of noise in the COP data is masking a signal. More theoretical and empirical work is needed to determine whether median cycle length is the most relevant metric to use to investigate effect of cycling on COP.
- Short-timestep cycling can occur in DHW mode or space heating mode, and at any time of year. Visual inspection enables attribution of most ASHP short timestep cycling as occurring in space heating mode, but this has not been verified algorithmically.
- Well performing GSHP sites can 'cycle' in space heating mode, with cycle lengths of around an hour, achieving good instantaneous COPs.

4.2 Supplementary heating: Use of DHW immersion

The analysis now moves on to two types of supplementary electric resistance heating; the first being DHW cylinder heating.

Types of DHW immersion use

Four types of site were identified within the broad dataset. The characteristics and numbers of sites of each type within the broad dataset are shown in Table 4-1:

Table 4-1. Breakdown of DHW immersion use types in Sample B2 (cropped).

	ASHPs	GSHPs	DHW characteristics
Type 1	84	66	HP provides space heating and DHW. No evidence of DHW immersion heating.
Type 2	195	12	HP provides space heating and DHW, with extra immersion heating for DHW. In some cases the DHW immersion is rarely used, but in some cases immersion comes on instead of the heat pump when DHW is required[11].
Type 3	5	0	HP provides space heating, immersion provides DHW.
Type 4	11	14	HP provides space heating, origin of DHW is unknown because it is not metered.

Three examples of Type 2 heat pumps using DHW immersion in different ways are shown in Figure 4-15-Figure 4-17 inclusive, in which immersion use is shown as the **purple series** on the bottom subplot. These Figures illustrate the range of immersion uses/frequencies from:

- Weekly use of immersion, for legionella protection (Figure 4-15). Please refer to the accompanying, *RHPP Case Studies Report*, for further analysis on this.
- Immersion used just after HP to heat DHW (Figure 4-16).

[11] Note that most immersion heaters have their own integral thermostats. If the set point of the immersion heater thermostat is below that of the cylinder thermostat, then depending on the status of other parts of the control system, the immersion heater may provide all of the hot water heating.

- Immersion used instead of HP to heat for two-day periods at a time, but HP used the other 5 days per week (Figure 4-17).

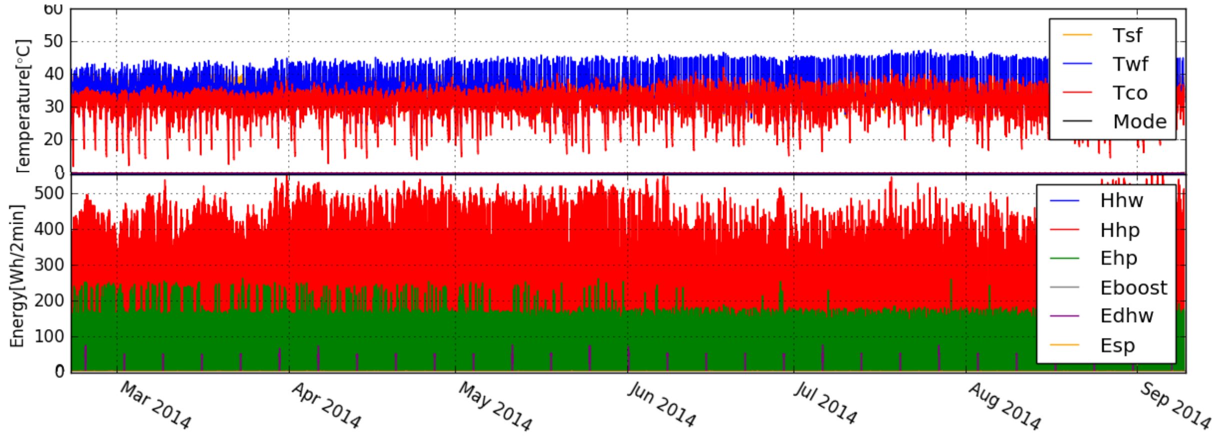


Figure 4-15. Immersion used for pasteurisation cycle, weekly.

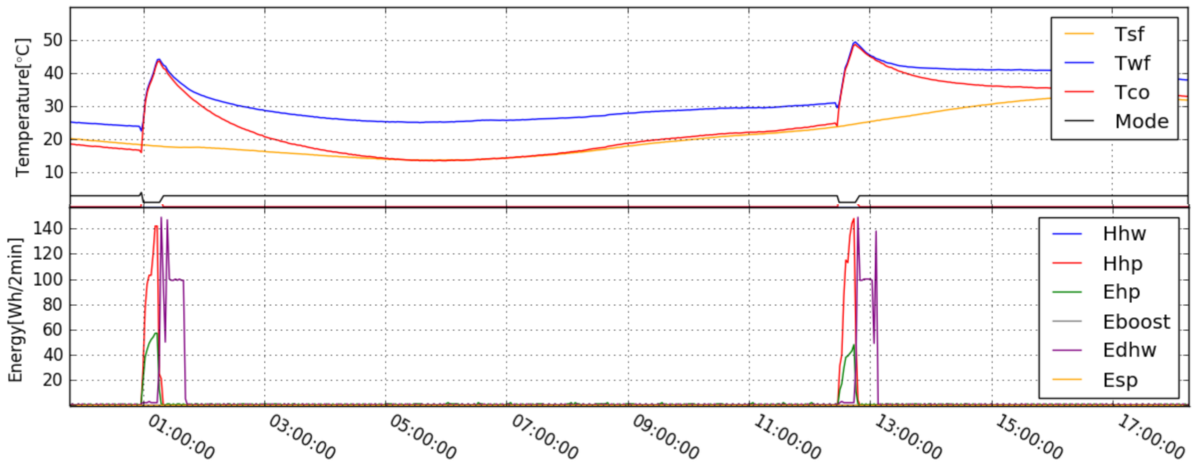


Figure 4-16. Immersion used just after heat pump carries out DHW operation each time.

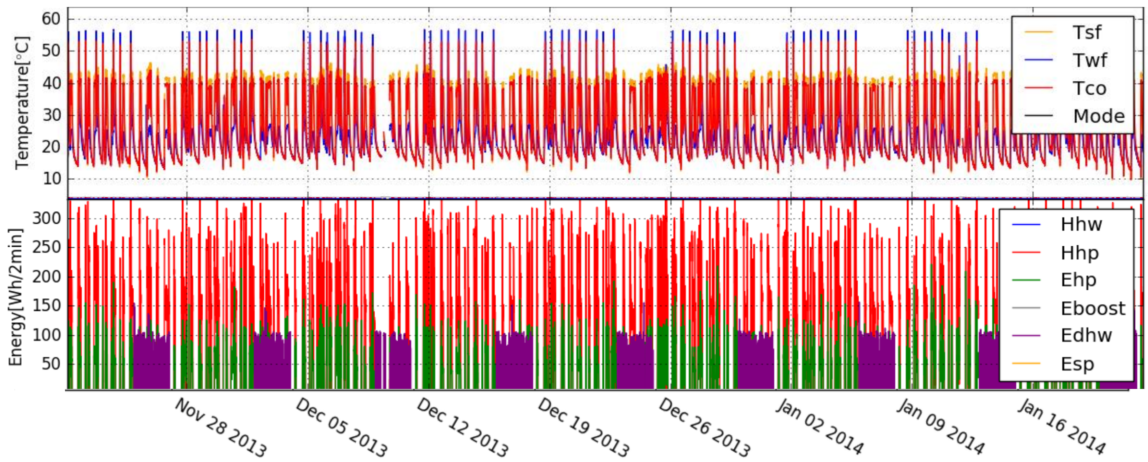


Figure 4-17. ASHP provides DHW on certain days, immersion used on other days.

Type 2 systems (i.e. those with DHW provided by both the heat pump and immersion) are now examined in more detail. Figure 4-18 ranks the Type 2 sites in order of the contribution of DHW immersion electricity to total electricity use, and contribution of DHW immersion heat to total DHW heat (ie not total heat).

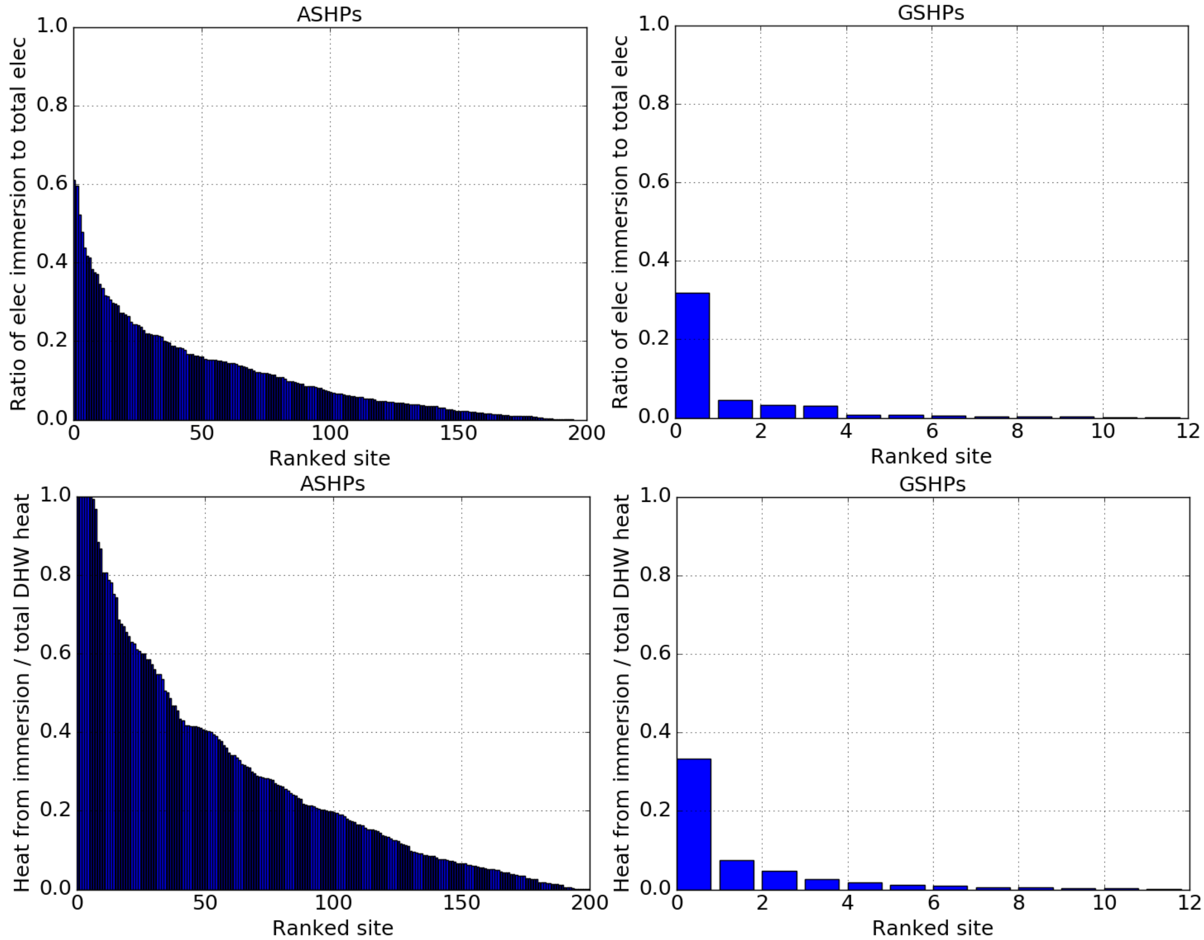


Figure 4-18. Contribution of DHW immersion electricity to total electricity use, and of immersion heat to total DHW heat.

Figure 4-18 shows that of the sites with immersion heating (of which there are proportionally more in the ASHP group than the GSHP group), the contribution of immersion heating is higher in the ASHP group.

Main reasons for high DHW immersion use

Immersion use is only a concern if it is high. High immersion use can occur because the immersion heater is used in addition to the heat pump to provide DHW; or because the immersion heater is used instead of the heat pump to supply DHW.

ASHP sites with the highest immersion use are primarily those using the immersion instead of the heat pump to provide either all of the DHW, or (in a minority of cases) all of the DHW in certain seasons (e.g. summer)/on certain days. In these sites, the schematics show that the heat pumps are capable of providing space heating and DHW, so it is unknown why or how they have been set up to provide space heating only and use immersion for DHW. The above also applies to the one GSHP with markedly higher use of DHW immersion than the others. Examining the timeseries data from this site showed a switch from using the HP for DHW heat to using the immersion, after a number of months. It is not known whether this was due to a fault with the HP, or was the result of an accident intervention in the control system, or was a deliberate strategy.

To examine why the DHW is being provided by immersion heating instead of the heat pump, it is useful to consider whether sites with proportionally high immersion use have proportionally low DHW demand. Sites with DHW demand anticipated to not constitute a large fraction of overall heat demand could be set up to use immersion for DHW without increasing energy bills by a large fraction, as the efficiency penalty resulting from using immersion instead of the heat pump for the DHW would not result a large fractional increase in electricity use. Figure 4-19 shows that indeed, sites with a high proportion of the DHW provided by immersion are sites with space heating dominating total heat demand and therefore proportionally low DHW demand.

Please note that as described in Setion 2, the split between space heating and DHW demand is not always possible to determine correctly.

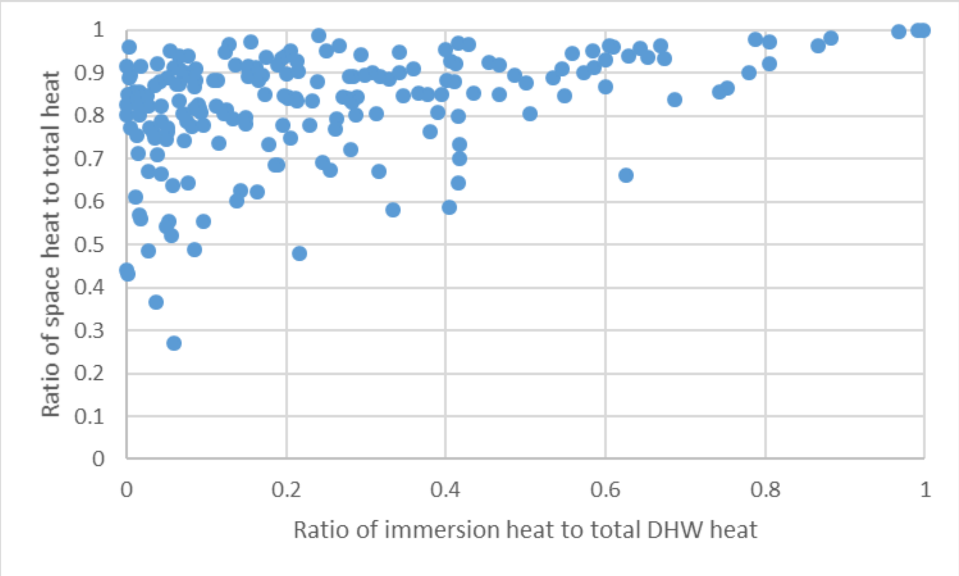


Figure 4-19. Exploring whether immersion is used more in sites with relatively low DHW demand.

The group of sites with the second highest immersion use are those in which the immersion comes on immediately after each episode of DHW heating by the heat pump. This can be daily or more than once per day. Examination of a few of these sites showed that the heat pump provides DHW up to around 53°C, after which the immersion comes on. This could be a simple, if inefficient way of sterilising the DHW cylinder, depending on the length of the period for which the immersion operates. A small number of sites were found in which the immersion heater remains on for several hours per day, apparently cycling on its own internal thermostat. This would almost certainly lead to unnecesarily high use of the immersion heater.

Effect of DHW immersion use on SPFH4

The relationship between the proportional electricity consumption from immersion heater and SPFH4 is shown in Figure 4-20 for ASHPs (the GSHP sample of 12 HPs is too small to show a meaningful scatter plot):

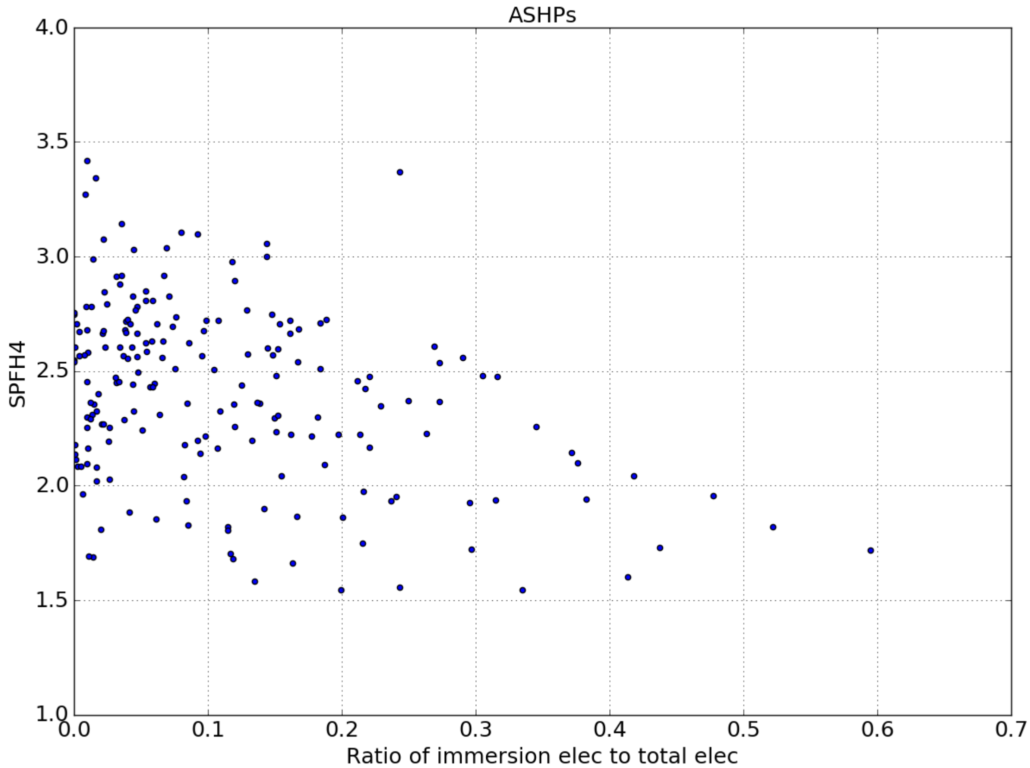


Figure 4-20. Proportional electricity consumed by immersion heater, and SPFH4.

The negative relationship implied in Figure 4-20 is to be expected, since immersion electricity is a constituent of the SPFH4 metric. Out of the 17% of ASHPs in Figure 4-20 with SPFH4 < 2, 48% have immersion contributing over 20% of their total electricity. The mean ratio of immersion electricity use to total electricity here is 0.12.

There is a lot of variability in the SPFH4 data not associated with immersion heating use; however, a relationship is still visible, implying firstly that metering error is not dominating this plot and secondly that it appears difficult for sites with high immersion use to achieve high SPFH4.

Of the 10 ASHP sites with the highest ratio of immersion to total electricity as shown in Figure 4-20, 6 are of Type 2 (immersion heater supplements the DHW provision by the HP) and 4 are of Type 1 (immersion heater carries out all of the DHW heating). These latter 4 Type 1 immersion heating users all had relatively low DHW heat use as a fraction of total heat (< 20%) while the 6 Type 2 immersion heating users did not.

4.3 Supplementary heating: Use of internal boost heating

Many heat pumps have electric boost heating built into them, referred to here as *internal boost heating*. The operation of such boost heaters internal to the heat pump enclosure was not monitored as part of the RHPP scheme, as this would have required modifying wiring inside the heat pump enclosure, with consequences for heat pump and system warranties. Instead, the electricity used by the internal boost is included in Ehp (and therefore SPFH2, 3, 4 and 5) but there is no metered record of how much it contributes to Ehp.

The Appendix to this report describes two post-processing methods to attempt to estimate the fraction of Ehp which is internal boost (*estimated boost fraction*). Both methods operate over relatively short timescales, typically using 20 minute average values. The first method attempts to classify points on a plot of heat output against total electricity input into regions corresponding to different modes of boost operation: either internal boost heating on top of the heat pump's output, or resistive heating instead of the heat pump's compressor running. The second method uses temperature data to separate out periods of operation with internal boost.

The first method has proved the more robust, and the results are summarised below. Since this is a post processing exercise using the same heat and electricity data as used in the rest of the results in this report, the same caveats around data quality, and especially noisy heat data, apply. In particular, note that since one of the modes of internal boost heating is resistive heating without the compressor running, and that this is identified from periods where the twenty-minutely COP = 1 (see the Appendix for a full description of this), then a heat metering error leading to apparently low COP would lead to both high estimated boost fraction and low SPF.

In Sample B2 cropped, there are 77 ASHPs and 19 GSHPs which satisfy the following conditions:

- Are heat pump models which are known to have an internal resistance heater;
- The boost detection algorithm functions satisfactorily.

In Figure 4-21 the sites satisfying the above conditions are ranked according to their estimated boost electricity as a percentage of the total heat pump unit electricity consumption, Ehp.

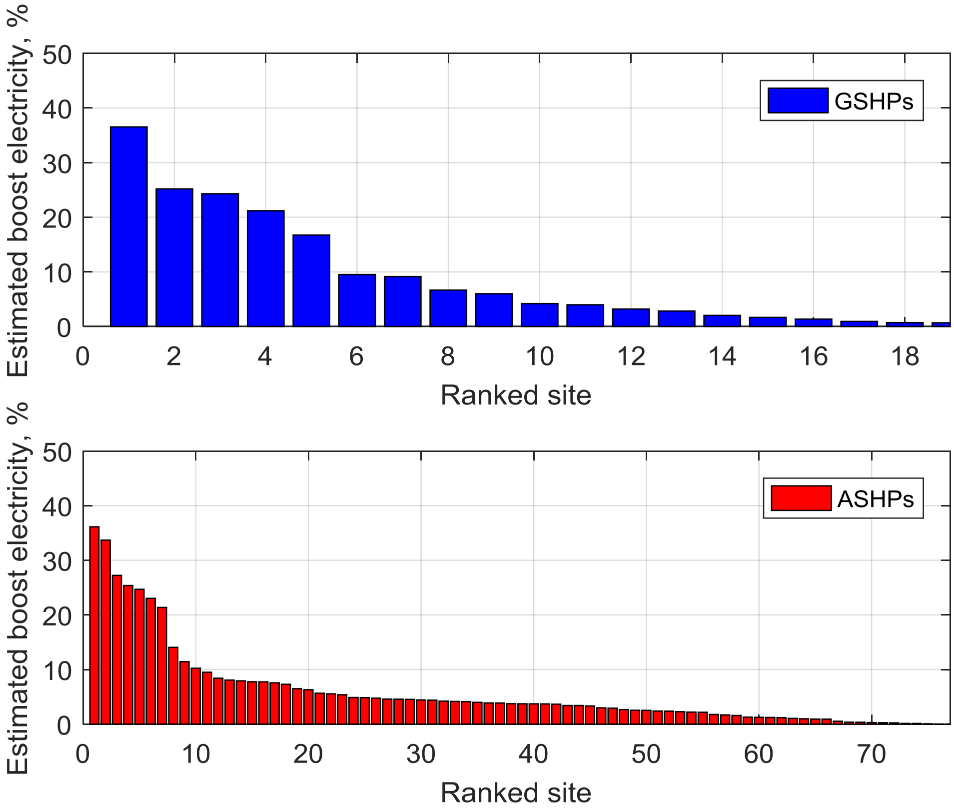


Figure 4-21. Ranked distribution of estimated boost electricity, Sample B2 (cropped), for those sites which are known to have internal resistance heaters.

Figure 4-21 shows that, of the sites which are known to have an internal boost heater and for which the boost detection algorithm functions satisfactorily:

- 84% of these sites have an estimated boost fraction of 10% or less. (87% of the ASHPs, and 73% of the GSHPs);
- The median estimated boost fraction is 3.8% (3.7% for the ASHPs and 4.2% for the GSHPs).
- The highest estimated boost fraction is 37% (this occurs in a GSHP; the highest for ASHPs is 36%).

Figure 4-22 plots the estimated boost fractions of the sites with their associated SPFH4.

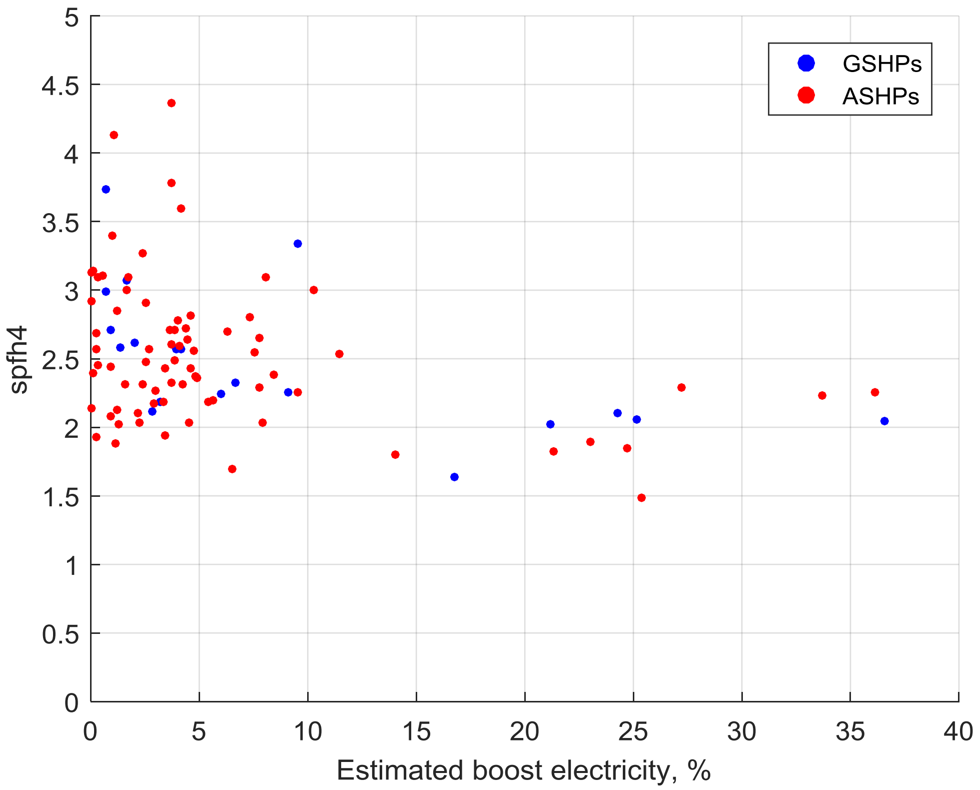


Figure 4-22: SPFH4 vs estimated boost electricity, for sites in Sample B2 (cropped) known to have an internal boost heater.

A minority of the subset of sites known to have boost heaters have estimated boost fractions exceeding 10%; these sites all have SPFs below 2.5. As for the sites with estimated boosts fraction below 10%, there is not a discernible trend of lower SPF for higher boost fractions.

Please see the Appendix for a full discussion of the methods and results used to estimate the boost fraction for the sites in Sample B2 (cropped).

4.4 Heat output and load factor

In this section, heat output and a related metric, load factor, of the HPs are examined in terms of their association with performance. Modes of operation associated with very low and very high heat output are then discussed.

The annual heat output of all the sites in Sample B2 (cropped) was shown in the Introduction, and is repeated in Figure 4-23, for context. The GSHPs in the sample have a higher median heat output, of 9087 kWh, than the ASHPs, at 7793 kWh.

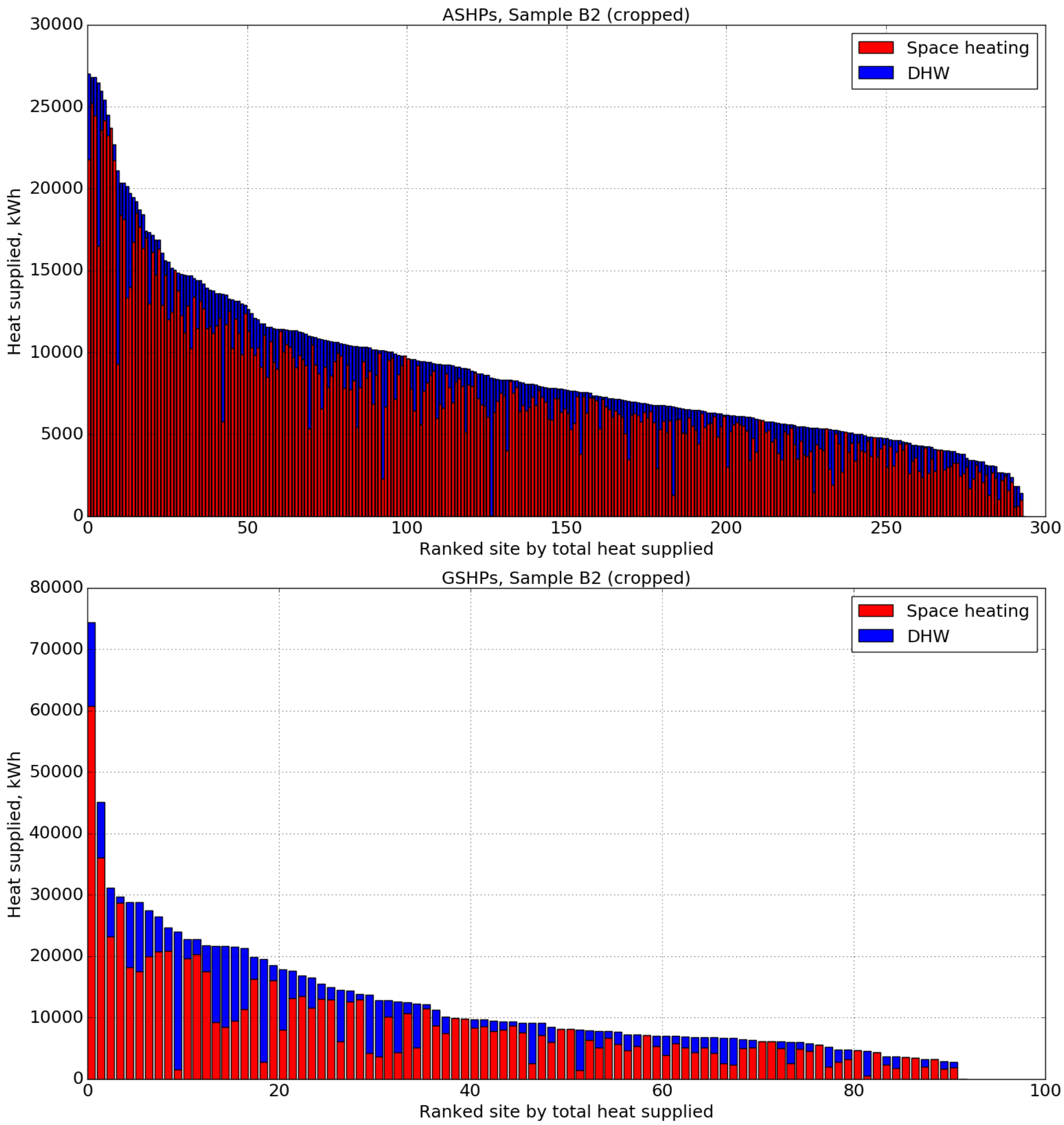


Figure 4-23: Sites ranked by total heat demand, Sample B2 (cropped).

Load factor is the ratio of actual heat generation to maximum heat generation if a heat pump operated at full power all the time. Low load factor can indicate oversizing of a heat pump, and can be a more useful metric than heat output as a large heat pump working at low load factor can be less efficient than a small heat pump working at high load factor but producing the same amount of heat.

The relationship between SPF and monthly load factor is shown in Figure 4-24 and Figure 4-25 below, for ASHPs and GSHPs.

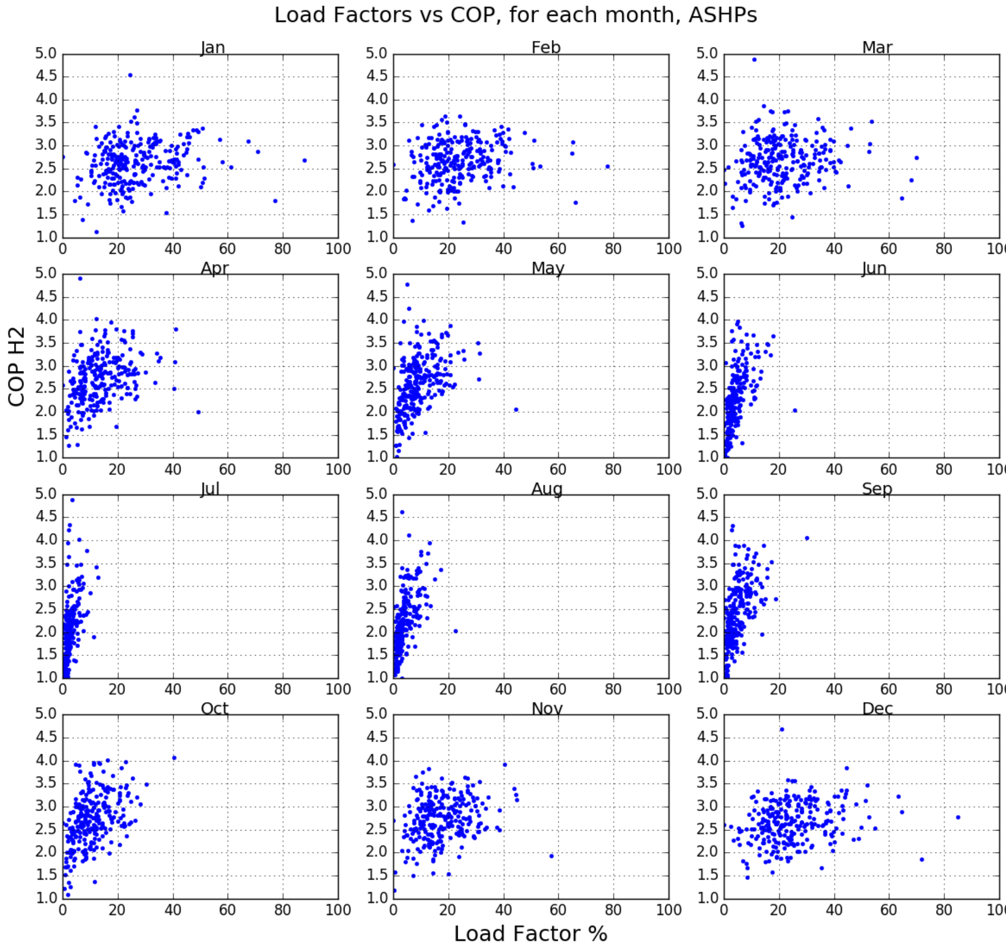


Figure 4-24. Monthly load factors and association with COP at the H2 bound, ASHPs.

The x-axis of the two Figures reveals the months with high space heating demand as January, February, March, November and December, followed by April, then October. The degree of scatter during these months is such that although GSHPs generally show higher monthly COPs, they do not show an obviously higher load factor. There may be a (weak) association between load factor and COP at the H2 bound, but only at load factors below about 20%.

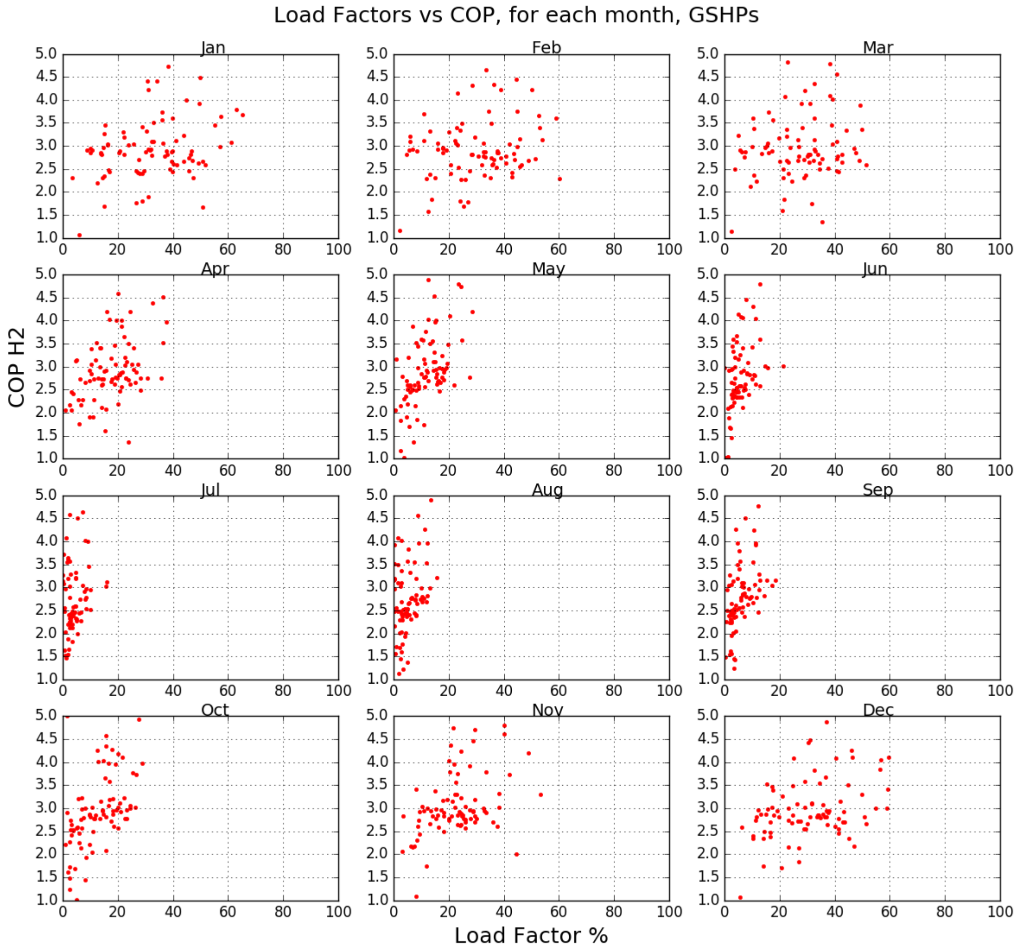


Figure 4-25. Monthly load factors and association with COP at the H2 bound, GSHPs.

It might have been expected that there would be a stronger association between sites with low load factor and sites with low COP. However, this would depend on the presence of efficiency-decreasing effects such as short timestep cycling. These mechanisms do not necessarily occur. For example, consider the operation of the ASHP with the 3rd lowest heat demand (at the right hand side of Figure 4-23). This site has a winter-averaged load factor of 5%, which is low. A typical day in winter for this site is shown in Figure 4-26. It can be seen that the low space heating demand is caused by few hours of operation per day, but during these hours, the instantaneous COP (instantaneous ratio of heat to electricity in the 3rd subplot) shows a reasonable efficiency.

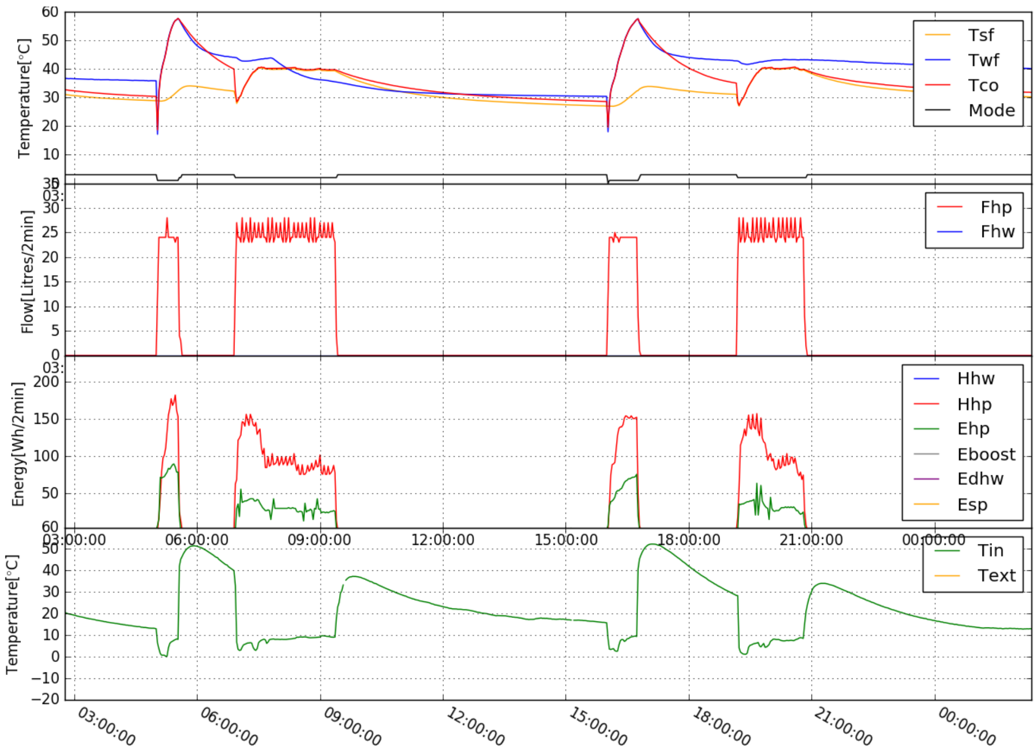


Figure 4-26. RHPP5229: Example of an ASHP with low annual space heating demand and low load factor, coming on for a few hours in a day.

More broadly, it is useful to examine a number of the top and bottom sites on to examine their modes of operation. Table 4-2 gives an overview of the operation modes of the sites with the lowest and highest heat demand.

Table 4-2. Sites with lowest and highest heat output from the HP.

Space heating output	ASHPs	GSHPs
5 lowest	Either space heating is hardly ever used (~once per week) or is only used for a few hours each day (e.g. Figure 4-26).	In three out of the five sites, space heating is on continuously with cycling on timescales of the order of 10 minutes.
5 highest	On either continuously or continuously except a few hours at night.	On continuously with long cycles (30 mins to 1 hr).

For ASHPs, Table 4-2 shows a large difference in operation mode between the systems with highest heat output, operating continuously or almost-continuously, and those with the lowest heat output, operating occasionally. For GSHPs, it is not the operation mode which differs between high and low heat output sites but the cycle length.

Finally, Figure 4-27 shows that there is no clear association between heat output and COP in the winter months. Effects such as that shown in Figure 4-26 may explain this in part – in that particular example the low heat output still yielded a reasonable instantaneous COP.

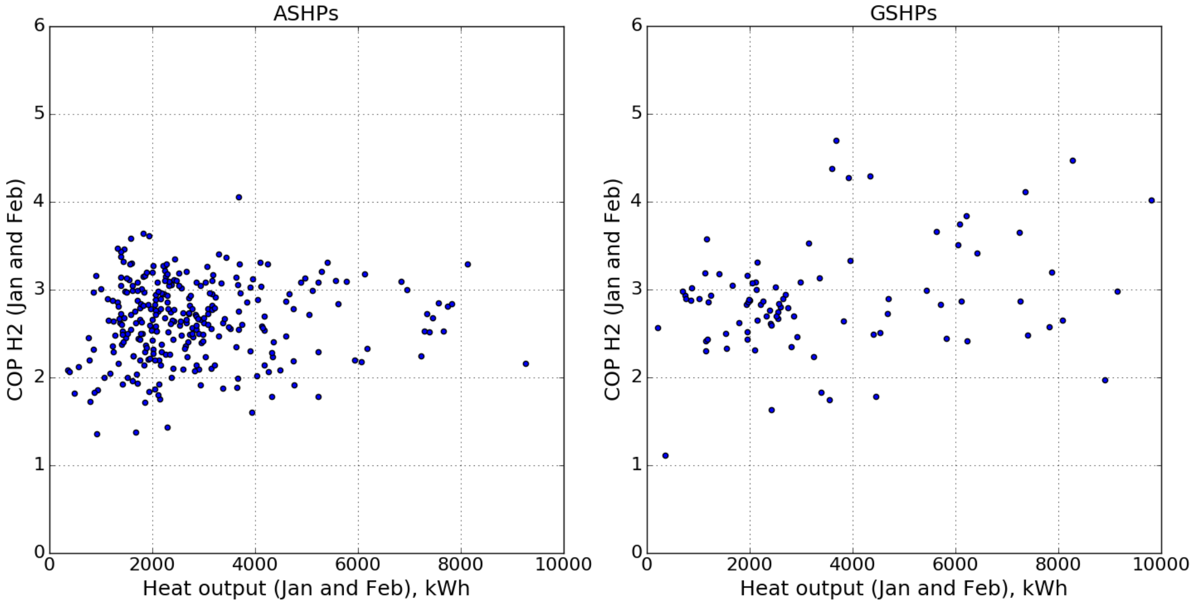


Figure 4-27. Winter COP vs winter heat output.

It should be noted that another efficiency-decreasing mechanism which should be associated with low heat output is the relative importance of parasitic electricity consumption (circulation pumps, fans, controls and other electricity consuming processes). The magnitude of parasitic electricity consumption is difficult to quantify at a sample level using the RHPP data because it is not always clear which electronic devices apart from the heat pump compressor are included within the monitored electricity consumption. However a site with obviously high parasitic electricity consumption is discussed in Section 5.3.

4.5 Flow temperatures for underfloor heating and radiator systems

In Section 3.2, it was demonstrated that the relative performance of underfloor systems compared to those with radiators was not as different as may have been expected. Since the main benefit of underfloor systems is *in principle*[12] that lower flow temperatures for space heating can be achieved, flow temperatures in space heating mode are the focus of this section.

Calculation of average monthly flow temperature per site

Average flow temperature for one heat pump is defined here as the average over all of the flow temperatures when the heat pump is in space heating mode (according to the algorithm which determines a heat pump's mode).

The top subplot of Figure 4-28 shows a day of data from an example site. The periods in which the HP is in space heating mode during the day are highlighted in blue. Flow temperatures are shown in the top subplot. The periods of space heating mode end when either the compressor turns off (third subplot), or DHW mode begins, or defrost mode begins (not shown in this example). The Figure also illustrates that this particular metric includes both the warm-up time of Tco rising, and the time afterwards when it falls back, roughly exponentially, towards the ambient temperature local to the sensor.

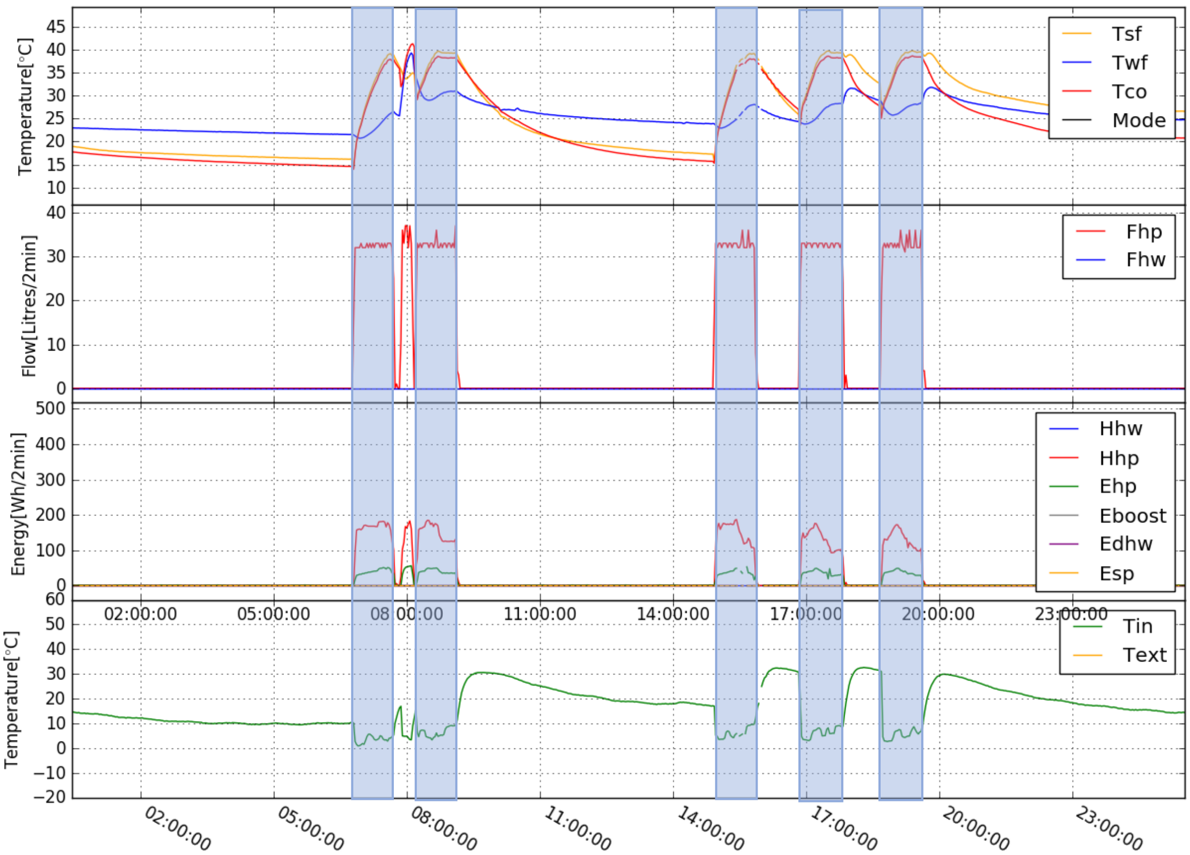


Figure 4-28. An example timeseries of space heating mode from which average flow temperature is calculated.

[12] This caveat is important. As noted in section 3.4, underfloor systems that are poorly designed or installed may not perform significantly better than radiators. It is hard to determine the quality of design and installation of UFH from observations on site, and impossible to do so from the remotely monitored data.

Each month, the average of the highlighted flow temperatures is calculated to give a monthly average space heating flow temperature for each site. However, problems with the mode determination algorithm as introduced in Section 2.2 lead to the summer data being unreliable: DHW heating is sometimes erroneously labelled as space heating. At times of low actual space heating demand, this error dominates the flow temperature data, whereas at times of high space heating demand (i.e. winter), the error is negligible. Since it is winter space heating which dominates SPF (see Section 3.1), below we limit the flow temperature results to wintertime only.

Figure 4-29 shows average monthly space heating flow temperatures for underfloor and radiator systems in winter. Not all sites had temperature data suitable to use in the analysis so the sample sizes for this analysis are stated in below. Note the considerably larger size of the ASHP + radiator group than the other groups.

Table 4-3. Numbers of ASHPs and GSHPs with radiators and Under-Floor Heating (UFH).

HP type	Heat emitters	Sample size in flow temperature analysis
ASHPs	Radiators	248
	UFH	28
GSHPs	Radiators	58
	UFH	24

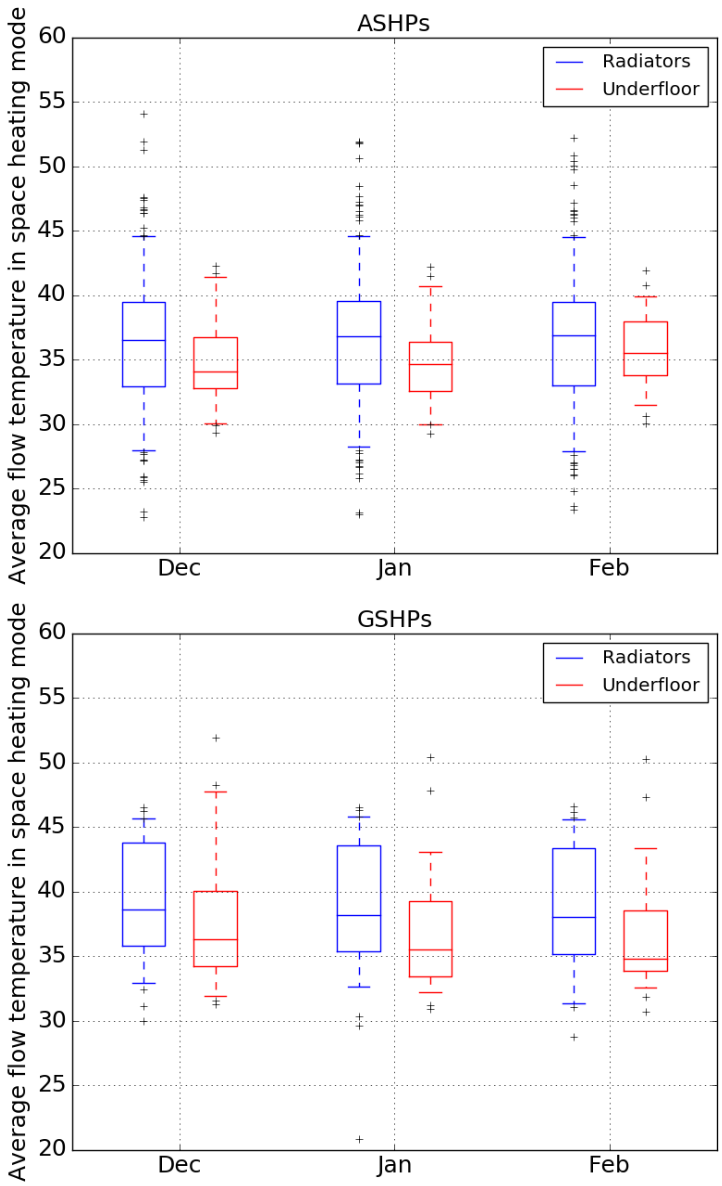


Figure 4-29. Winter space heating average flow temperatures.

Underfloor systems show lower average flow temperatures in winter. However, there is a high degree of variation between sites, visible in the spread of Figure 4-29, and in 2 GSHP sites the flow temperatures are high (over 45°C). Moreover, in the case of GSHPs, the highest flow temperatures occur not in dwellings with radiators, but in dwellings with underfloor heating – two possible explanations for this come to mind: mislabelling of underfloor heating systems, and poor design and/or installation of UFH.

Section 3.2 demonstrated a very weak signal showing the SPF of underfloor systems as being higher than that of radiator systems, in GSHPs and ASHPs. Below in Figure 4-30 it is demonstrated that there is some weak evidence to support an inverse relationship between average flow temperature and monthly space heating COP. Metering error in the heat data or misclassification of sites by heat emitter, could account for some of the noise in the vertical axis.

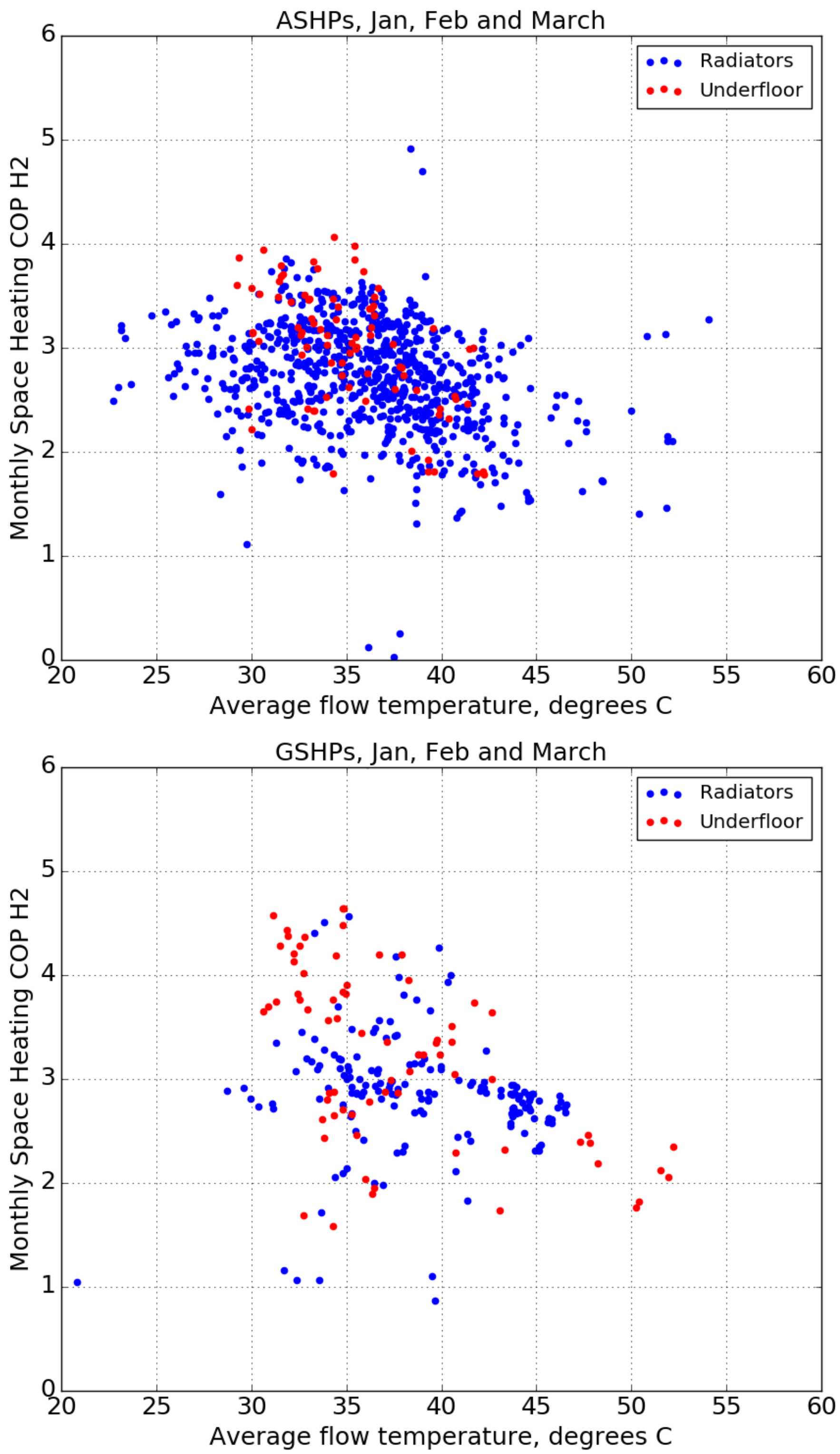


Figure 4-30. Monthly space heating COP against average flow temperature (each dot represents one site-month).

Some sites appear to have high space heating flow temperatures. To analyse this further, the next metric concerns maximum space heating flow temperatures. This is calculated by finding the 99th percentile of flow temperatures in space heating mode, discounting the first ten minutes of each heating event (because if the previous state of the system was DHW mode, the flow temperature starts off high when the mode switches over to space heating).

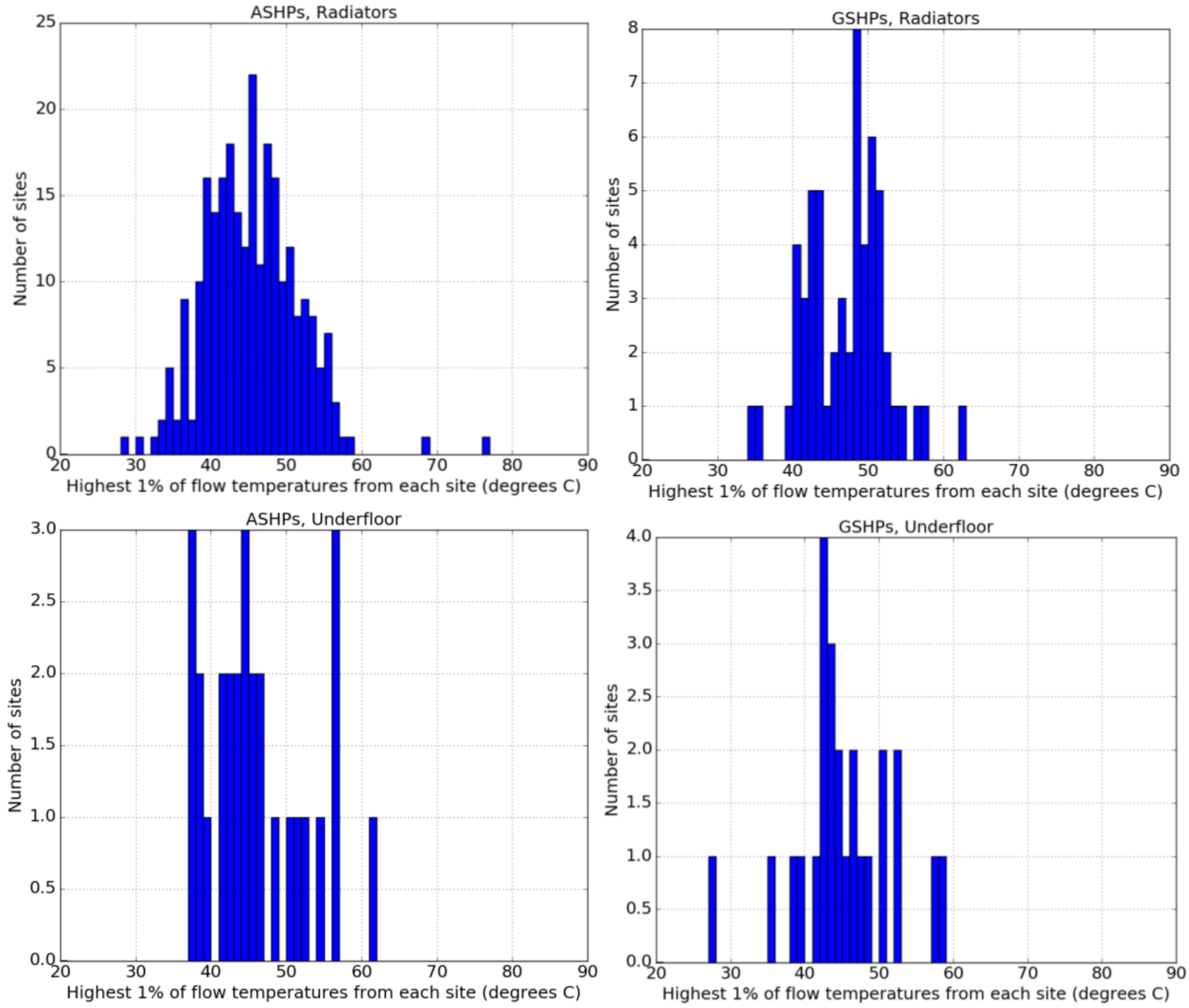


Figure 4-31. 99th percentile (highest 1%) of flow temperatures in space heating mode.

HPs supplying underfloor heating should deliver heat from the condenser at no more than 45-50°C. If T_{co} is above this during space heating mode, then (as noted in section 3.4) one of a number of possibilities may be indicated:

- the UFH system occupies a smaller proportion of floor area of the dwelling than intended;
- pipes in areas heated with UFH are spaced further apart than recommended;
- floor coverings such as carpets that were not considered during the design of the system have been added to areas of floor with UFH;

- the system had been fitted with a three-port mixing or blending valve (this would be contrary to MCS guidance, but such systems are commonly installed in dwellings heated by gas or oil boilers).

Further work is recommended to investigate how high flow temperatures to underfloor heating systems arose in the RHPP sample.

5 Other selected performance issues

In this section a limited number (four) of performance issues identified as the data was examined are shown. These are chosen based on being the most confident these are real performance issues as opposed to metering artefacts. Their prevalence in the sample is not quantified here.

5.1 DHW cycle going on too long

Figure 5-1 gives an example DHW cycle from an ASHP site in which the cycle goes on for long enough that the COP is affected. The first subplot shows the temperature of the flow to the water cylinder (Twf, blue), and the third subplot shows the heat output (red) and electricity input (green) during this time. It can be seen that the ratio of heat to electricity (the COP) starts to decrease after the DHW flow has reached a constant temperature, eventually approaching 1. The probable mechanism is the return temperature rising as the HP delivers heat to the temperature it is set to deliver at, but the cylinder thermostat continuing to call for heat. This phenomenon has been referred to elsewhere as "stalling".

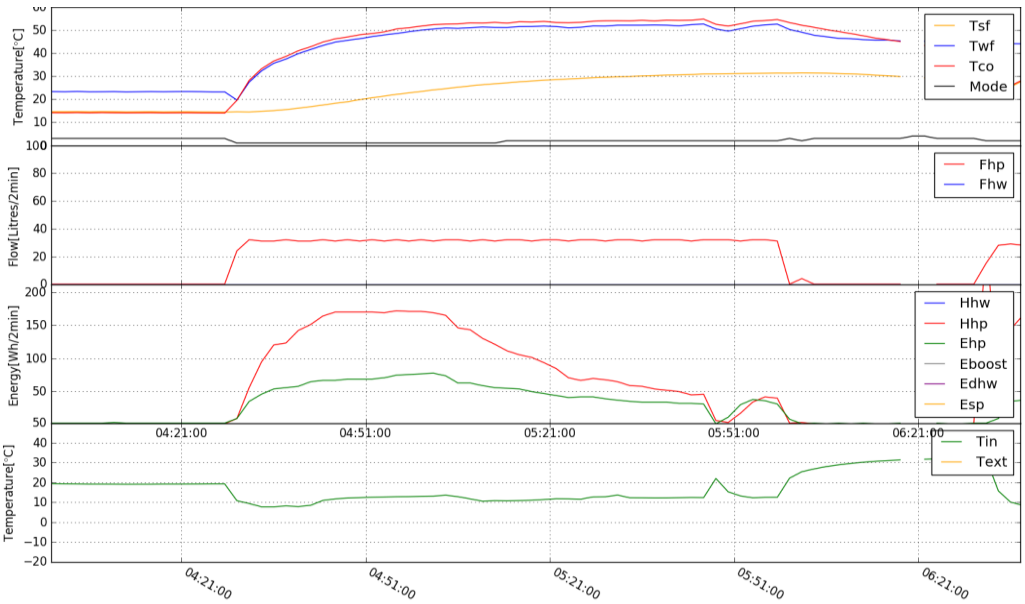


Figure 5-1. ASHP in winter. DHW cycle carries on too long and COP decreases.

The same phenomenon is shown in summer in Figure 5-2. This time, after the DHW cycle continues long enough that the COP decreases to 1, cycling takes place.

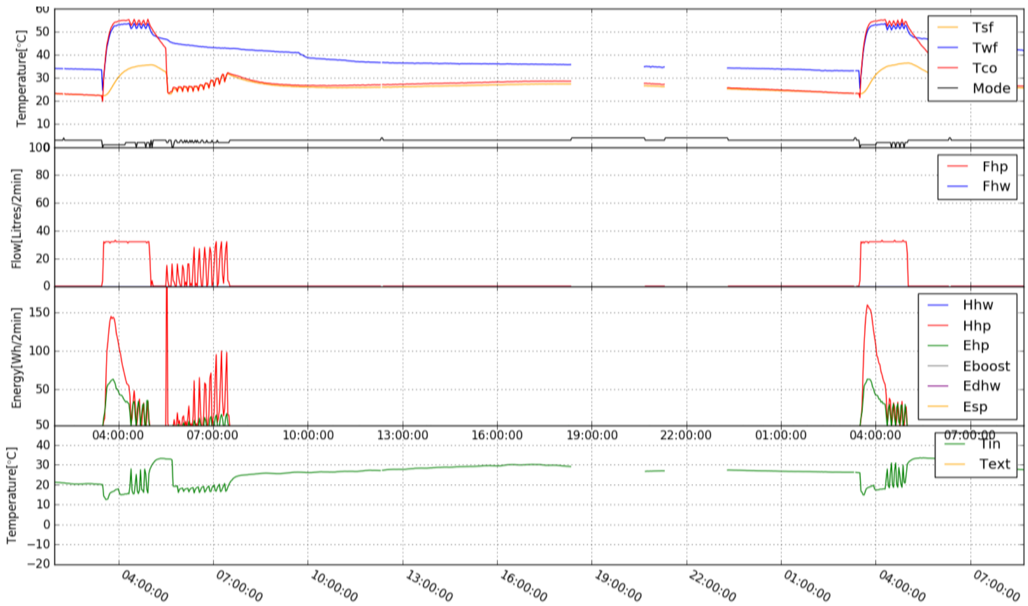


Figure 5-2. ASHP in summer. DHW cycle carries on too long and cycling occurs.

5.2 Very frequent DHW heating from both heat pump and immersion

Figure 5-3 shows an example of an ASHP installation which has multiple (4-5) timed DHW periods per day, with each cycle lasting around 30 minutes. During each DHW heating period the HP switches off and on several times. There is also frequent use of immersion, coming on in short bursts several times after the heat pump DHW heating period. It is not clear why both the heat pump and immersion provide heat to the DHW so frequently.

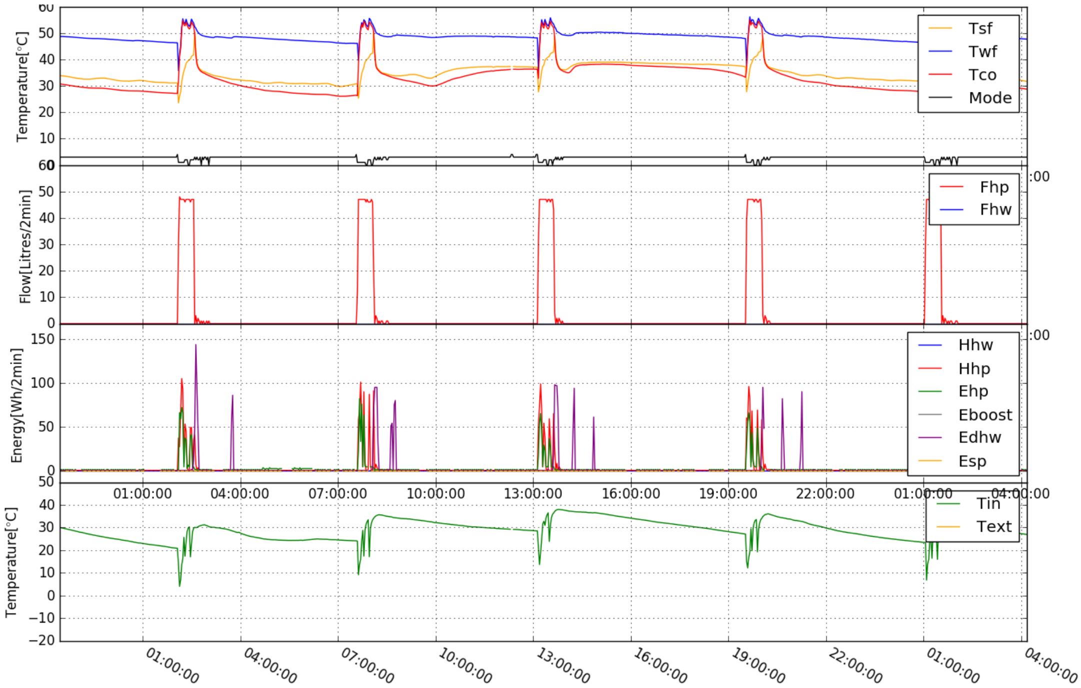


Figure 5-3. ASHP in summer. Frequent DHW heating and immersion use.

5.3 Circulation pump on continuously

The second subplot in Figure 5-4 shows that the heat flow is almost continuous for this GSHP; thus the circulation pump must only occasionally turn off. This is the case even when the heat pump compressor is off. This would cause high parasitic electricity consumption. It is not clear why the pump is on all year round: it could be that this is a requirement of the control strategy (e.g. a return flow temperature is needed), or it could be that the system has not been correctly commissioned.

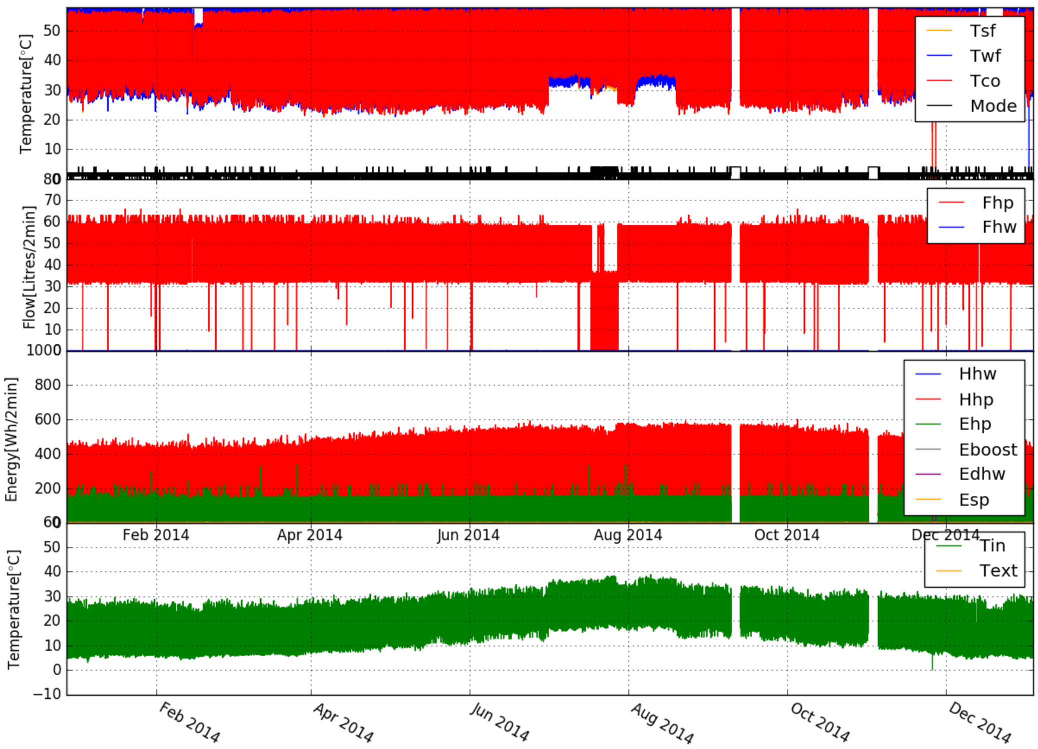


Figure 5-4. GSHP with continuous flow: circulation pump on all the time.

5.4 Use of boost before compressor start-up in space heating mode

The ASHP in Figure 5-5 has a separately metered space heating electric boost heater, (E_{sp}, shown in orange in the third chart), which comes on just before the compressor each time the heat pump starts up in space heating mode (first and third periods of operation in Figure 5-5). This does not occur in DHW mode (middle period of operation of Figure 5-5).

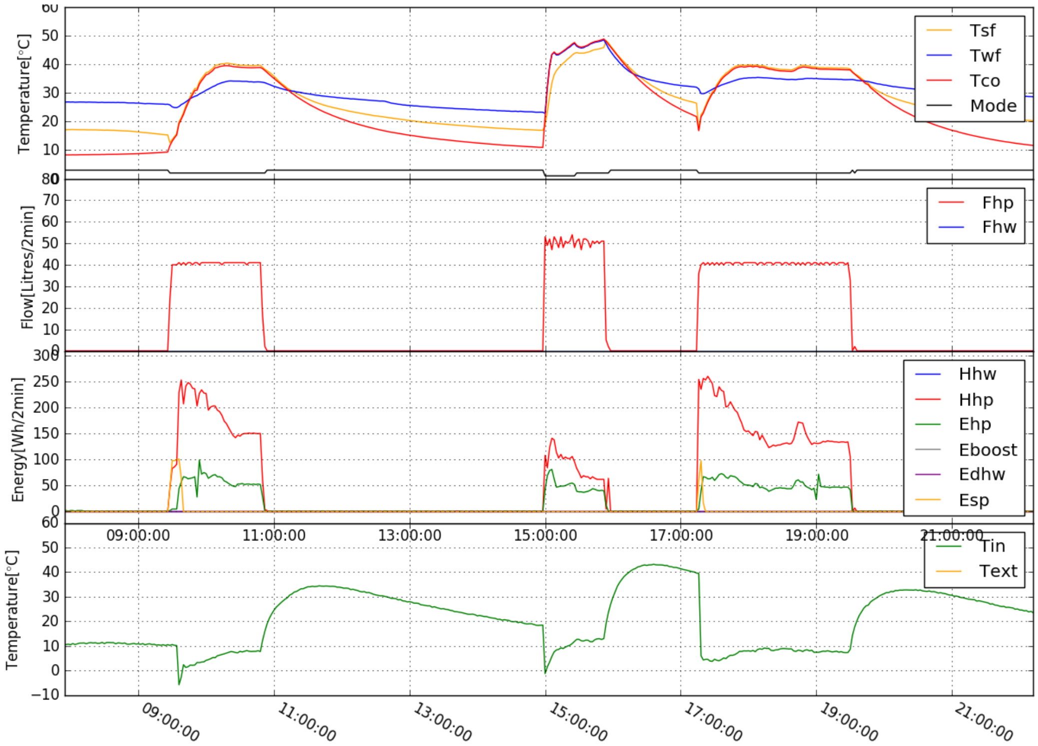


Figure 5-5. Space heating boost used every time space heating mode begins.

6 Summary of findings

The quantitative findings are stated below, with the potential effect of metering error made explicit in the right hand column. Where 'heat meter error' is mentioned, its presence has been suspected due to noise in the SPFs or heat data, and so this includes errors which artificially increase or decrease heat output.

Finding	Effect of metering error
Strong evidence that the mean SPF for GSHPs across the H2 and H4 boundaries are higher than those for ASHPs. For SPFH4 the difference in means was 0.28 (95% CI: 0.16 to 0.40) and at the H2 boundary 0.33 (0.22 to 0.45).	We are confident about the qualitative finding that there is a performance advantage for GSHPs. Metering errors may have impacted the estimated size of the difference in this sample, but this is also reflected in the uncertainty (confidence interval) of the difference provided.
On average GSHPs perform significantly better for both space heating (SH) and domestic hot water (DHW) at the H2 and h4 boundary	We are confident about the qualitative finding, but are more cautious as to estimates of the size of the difference, since additional sensors and data processing are needed to separate SH from DHW data.
Concerning space heating, the advantage of GSHPs over ASHPs disappears by springtime. Concerning DHW, the advantage continues throughout the year.	We are confident about the qualitative finding that these differences vary across the year.
For ASHPs, there is some evidence that sites with underfloor heating perform better than those radiators. Although this is in the direction expected (median SPFH2 differs by 0.27) a statistically significant result was only found for SPFH2. The differences could not be detected for	For ASHPs, heat metering errors appear to be a factor in overlap at the bottom end of the distribution; for GSHPs, for GSHPs, heat metering errors have contributed to the spread of the whole distribution.

SPFH4. There is some limited evidence that better performing sites show more difference. For GSHPs, those with underfloor heating appear to show a wider distribution in performance than those with radiators.	Other factors, however also appear to be at work specifically for GSHP systems with underfloor emitters. The view that this reflects a real performance issue that is worthy of further investigation, is supported by the results that some underfloor systems had high flow temperatures.
Underfloor heating was shown to occur at lower flow temperatures than those in radiator systems, which is to be expected. However, high flow temperatures (over 45°C) were detected in 2 GSHP underfloor sites. Prevalence of low flow temperatures in sites with radiators is more difficult to quantify due to the indicated presence of boost heaters raising the flow temperature provided by the heat pump.	This analysis depends on the accuracy and correct positioning of the flow temperature sensors and the ability of the software to correctly detect space heating mode. The GSHP sites with underfloor heating > 45°C were verified in the timeseries data by eye and found to be correct. The sites with the lowest flow temperatures observed in radiator systems were checked in the timeseries data and from there the presence of boost heaters was observed.
There is no evidence for a difference in performance between Privately owned and RSL sites with ASHPs. For GSHPs, the Private sites show a far more dispersed (spread out) distribution compared with the RSL sites, which for SPFH4 mainly lie in the 2.5-3.0 range. This may reflect aspects such as the diverse contexts of the private sites, compared with relative similarity of RSL sites.	While heat metering issues will have contributed to the spread in the distribution of both ASHPs groups, the distributions are sufficiently close that even without heat metering issues, it would be unlikely that a large difference would then be found. For GSHPs, it may be that heat metering errors caused the spread in private sites, but it may also be the case that the dwelling/occupant factors at work are also far more variable.
It was only possible to undertake limited analysis of the performance of particular models against the rest of the sample. One example was shown	Heat metering issues have probably contributed to the spread in the distributions for the comparison of Model A to the rest, and possibly the size of the

where a model outperformed the rest by a small amount, but a very similar model showed no difference from the rest.	difference. There was not strong evidence for a difference, at this result was at the limits of what we could detect. The small sample sizes for most model types greatly limits the ability to make comparisons of performance.
Cycling time of 10 minutes was common in ASHPs, with 49% of ASHP sites having at least one month where 10 minutes was the median cycle length. This appeared to be predominantly associated with space heating mode (although this hypothesis has not been verified algorithmically), and could occur at any time of year. Further work is recommended to ascertain whether 10-minutely cycling comes about as a result of the HPs themselves or as an outcome of their installation into e.g. existing control systems of a dwelling. GSHPs did not show a modal monthly cycle length of 10 minutes but exhibited a flatter distribution of cycle lengths, with the most common at 18 minutes. A related metric, proportion of cycles of length less than 12 minutes, showed most ASHPs had at least one month where at least 10% of cycles were shorter than 12 minutes, and in 25% of ASHPs this was the case all year round. GSHPs did not show this behaviour; only 1 site showed all year round short cycling.	We are confident in this result with the proviso that if cycling of the order of 2 minutes were to occur it could not be detected by 2-minutely monitoring of electricity use of the compressor.
No effect on monthly COP was observed from median cycle length per month.	It is suspected that heat metering errors could have masked a real trend in this case.
Immersion heating for DHW is monitored in 68% of ASHPs and 13% of GSHPs in the sample. 1% of ASHPs and 0% of GSHPs have monitored	These findings are considered robust to metering error

immersion and no DHW output from the heat pump. Note that these might not be the only systems which use immersion for all of their DHW; in some systems the immersion may not have been monitored. Immersion heating was observed to be used for a wide range of purposes, including: legionella protection (weekly or less frequently), coming on after every HP DHW heating event possibly to boost the stored DHW temperature, being used instead of the HP according to a certain schedule, or being used for all the DHW. Of the 10 sites with the highest proportion of immersion to total electricity (all ASHPs), about half used immersion as well as the HP to provide DHW, and half used immersion instead of the HP to provide DHW. The latter all had relatively low DHW demand. For the sites with monitored immersion, the mean contribution of immersion electricity to total electricity at the H4 bound was 12%.	
A negative relationship between immersion use and SPFH4 was observed. This is to be expected as immersion is a constituent of the SPFH4 metric.	Heat metering error is suspected to contribute to the noise in the relationship between SPFH4 and immersion use; however the relationship is strong enough to still be visible.
Around half of sites with SPFH4 < 2 were shown to have immersion contribution more than 20% of their total electricity.	We are confident in this result.
Internal boost heating was inferred using heat and electricity data and cross checked with information on whether the heat pump model actually contained an internal electric resistance heater (96 sites). 84% of these sites have an estimated boost	The method used to detect boost relied on detecting low values of the heat ouput which might equally have arisen from underreporting heat meters.

fraction of 10% or less. (87% of the ASHPs, and 73% of the GSHPs); The median estimated boost fraction is 3.8% (3.7% for the ASHPs and 4.2% for the GSHPs). The highest estimated boost fractions were 37% (this occured in a GSHP); the highest boost fraction for ASHPs was 36%.	
In sites with estimated boosts fraction below 10%, there is not a discernible trend of lower SPF for higher boost fractions. There is a minority of the subset of sites known to have boost heaters in which estimated boost fraction exceeds 10%; these sites all have SPFs below 2.5.	Noise in the heat meter data may have masked a real trend; it is unclear whether this is the case. This result may have been artificially created from underreporting heat meters, which would both lower apparent SPF and increase boost fraction.
Monthly heat output and load factor did not yield a clear association with monthly COP.	Noise in the heat meter data may have masked a real trend; it is unclear whether this is the case.
The ASHPs with the lowest heat demand only came on for a few hours each day, without short cycling or any clear sign of oversizing. The GSHPs with the lowest heat demand displayed cycling behaviour at a frequency of around one cycle per 10 minutes which may lower SPF – however the investigation of SPF and cycling did not yield a trend in this dataset.	These findings were determined by detailed observation of timeseries data.
A number of real performance issues were identified by visual inspection of timeseries data and four were described here: DHW cycle going on too long, very frequent DHW heating events, circulation pump operating continuously and use of boost before compressor start-up.	These findings were determined by detailed observation of timeseries data.

7 Discussion and areas for further research

The RHPP project has presented the RAPID-HPC consortium with the challenge of extracting as much value as possible from a major investment of public funds aimed at understanding the field performance of a strategic, low carbon technology, despite a problematic dataset. This has required painstaking analysis, and close attention to the framing and caveating of results. The resulting reports, of which this is one, are complex and require careful reading. Despite this, the authors believe that they contain many useful insights with respect to the status of the technology and opportunities for future development and improvement, for the heat pump industry, Government and other stakeholders. These include 'lessons learned' in terms of design of future field trials, which are to be published in the Final Report of this series.

This report has focused on variations in performance across the RHPP dataset, and the possible factors that give rise to them. It needs to be stated at the outset that the ability to comment on variation in performance between sites or groups of sites has been limited by the quality of the data available to the consortium, and especially of the data on heat output of heat pumps. Issues of data quality have turned out to be hard to distinguish from underlying uncertainty regarding the technical characteristics and attributes of individual sites. These two factors introduces significant uncertainty into estimates of SPF and other performance indicators for each site in the field trial.

Nevertheless, in large enough groups of sites and with large enough performance differences between groups, it has been possible to detect statistically significant differences. For example, the expected performance advantage of GSHPs over ASHPs has been identified, with clear non-overlapping confidence intervals for the GSHP and ASHP distributions.

Beyond the fundamental GSHP vs ASHP comparison, stating what is 'expected' in terms of differences between groups becomes more difficult. For example, the proposition that underfloor systems should be more efficient than radiator systems is only likely to be true if the flow temperatures in the former are lower than those in the latter. In this dataset, although this is generally the case, there are also examples of low temperature flow going to radiators (30-35°C) and relatively high temperature flow going to underfloor systems (45-55°C) – in other words, the two groups appear to overlap in this key respect. Another example is the proposition that systems with low heat output or low load factor are necessarily less efficient; here it was shown that the ASHP sites with the lowest heat output displayed good modulation behaviour without short timestep cycling or other behaviours which would decrease their efficiency.

There was no evidence of any single factor – such as boost heating - being the key to explaining variations in performance, or of any specific HP model as showing outstanding SPF. Instead, the lack of clear relationships suggests that an array of factors, ranging across correct installation to operation over the longer term and other dwelling/occupant related factors, are probably at work. These may be inherently difficult to unravel analytically, just from limited remote monitoring data.

To obtain optimal performance in practice it is probably going require a greater focus on all aspects of the supply chain, especially post manufacture due to the bespoke nature of each project. Attention needs to be paid to design, installation, the final commissioning and handover instructions, ensuring all are in harmony. Picking up on the principle of soft landings[13], post occupancy checks for the first few months and through subsequent years would advisable and would represent a business opportunity for heat pump suppliers and installers.

Nonetheless, the analysis in this report pointed towards the following questions that may help to define directions for future research aimed at improving systems and practice:

- Why was 10-minutely cycling so common in the ASHPs in the trial, and does this decrease SPF?
- What are the most efficient DHW heating strategies in different dwelling types, and is there a case for DHW heating being carried out solely by electric immersion in sites with low DHW demand?
- How are DHW cylinder temperatures controlled in sites which do and do not exhibit unnecessarily frequent or extended DHW heating events?
- How are high flow temperatures in sites with underfloor heating systems coming about?
- What is causing the especially large spread in the distribution of SPF for GSHP sites with underfloor heating?

In terms of further field testing and analytical work, considerable scope remains for small scale and far more detailed and focussed studies. These could include technical or other interventions to address any specific issues identified, including dwelling and occupant factors. Some key areas these studies could examine concern methods of robust performance measurement, issues of longer term performance degradation (say after two or more years), the role of dwelling characteristics, and resilience of performance to changes in occupant behaviour.

[13] Soft landings is a concept that emerged from the PROBE project (Bordass et al. 2010). It acknowledges that modern buildings often require an extended commisioning period to achieve full integration and performance of complex building services systems. The principle has not been widely applied in dwellings, but in the context of heat pump installations evidence presented in this report suggests that it would be valuable to do so. Guidance on soft landings has been published by BSRIA - https://www.bsria.co.uk/services/design/soft-landings/free-guidance/.

8 References

Bordass, B., Leaman, A. & Ruyssevelt, P., 2001. Assessing building performance in use 5: conclusions and implications, *Building Research & Information*, 29 (2) (Special Issiue on Post-occupancy Evaluation) 144-157.

EST 2010 *"Getting Warmer, A field Trial of Heat Pumps*", London. EST.

Dunbabin, P. & Wickins, C. 2012 *"Detailed Analysis from the First Phase of the Energy Saving Trust's Heat Pump Field Trial"*, London. DECC.

Dunbabin, P. Charlick, H. & Green, R. 2013. *"Detailed Analysis from the Second Phase of the Energy Saving Trust's Heat Pump Field Trial"*, London, DECC.

Gleeson and Lowe, 2013. Meta-analysis of European heat pump field trial efficiencies. *Energy and Buildings*, 66(637-647).

Green, R., 2012. *The Effect of Cycling on Heat Pump Performance*. London:BEIS.

Hughes, D., 2016. *Monitoring of Non-Domestic Renewable Heat Incentive Ground-Source and Water-Source Heat Pumps*. London: BEIS.

Pedersen SV & Jacobsen E, 2011. *Approval Of Systems Entitled To Subsidies. Measurements, Data Collection And Dissemination*. Taastrup (Denmark): Danish Technological Institute.

Pickup, G.A. & Miles, A.J., 1977. The performance of domestic wet heat heating systems in contemporary and future housing, I.G.E. Communication 1041. *J.I.G.E.* 18 (6) 188-204, and reproduced in British Gas (1979) *Studies in Energy Efficiency in Buildings*.

RAPID-HPC, 2016. *BEIS RHPP Detailed Analysis Report*. London: BEIS.

RAPID-HPC (2017a) *Investigating variations in performance of heat pumps installed via the Renewable Heat Premium Payment Scheme*. London: BEIS.

RAPID-HPC (2017b) *RHPP report on compliance with MCS installation standards*. London: BEIS.

RAPID-HPC (2017c) *Case Studies Report from the RHPP Heat Pump Monitoring Campaign*. London. BEIS

RAPID-HPC (2017d*) RHPP Heat Pump Study- Note on Systematic Errors in Physical Monitoring Data*. London. BEIS

Wickins, C., 2014. *Preliminary report on the Renewable Heat Premium Payment metering programme*. London: BEIS.

Appendix 1: Internal boost analysis

BEIS RHPP heat pump monitoring:

Identifying operation of internal boost heating

Prepared by: Chris Martin

Introduction

As part of a recent review of the software produced to analyse data from the recent RHPP Heat Pump Field Trial it was established that the operation of boost heaters internal to the heat pump enclosure was not being treated [1]. This has the effect that SPF_{H2} is under-reported.

Many heat pumps have electric boost heating built into them, and we refer to this as internal boost heating. Often it will be controlled automatically by the heat pump electronics. It is therefore likely that it will operate under at least some conditions. For the RHPP monitoring it was not considered feasible to meter this energy separately, as this would have required modifying wiring inside the heat pump enclosure, with consequences for heat pump and system warranties.

The original RHPP monitoring specification [2] required a temperature sensor placed after the heat pump compressor and before any boost heating, denoted T_{co}. The intention was that, in combination with the space and water heating flow temperatures (T_{sf} and T_{wf}), this could be used to detect operation of an internal boost heater. This would be done by detecting the temperature rise across the heater, which for a typical design of heat pump would be expected to be in the region of about 4°. The accuracy of the sensors used (each ±0.25°C) should have made this easily achievable, and analysis software produced within BEIS used this approach [3]. Unfortunately installation constraints often meant that the sensors were mounted close to heat sources (such as the compressor, the boost heater or hot water storage tanks) and this approach was subsequently decided to be unreliable and was eventually discontinued.

This report presents an alternative, algorithmic, approach to detecting internal boost heater operation. Two algorithms are explored – a very simple one which determines thresholds for total heat pump electricity input, and a more detailed one which concentrates on the distribution of the observed SPF of the heat pump.

2 Underlying model of internal boost heating

In the schematics developed to guide monitoring equipment installers [4], E_{boost} refers to a boost heater mounted before the heat meter used to measure heat pump output, H_{hp}. Thus its heat output is included in H_{hp}, but its electricity consumption is measured separately and is not included in E_{hp}. In cases where the boost heater is internal its electricity consumption is not metered separately and is instead included in E_{hp}. We refer to such a boost heater as 'internal boost', and refer to the part of E_{hp} which relates to its electricity consumption as E_{iboost}.

In contrast, the monitoring specification uses E_{sp} and E_{dhw} to denote the electricity consumption of heaters which are provided specifically for space heating and hot water respectively. Often these will be in the form of immersion heaters in storage tanks, although in the case of E_{sp} a flow boiler may also be used. Unlike E_{boost} and E_{iboost} these heaters are mounted after the heat meter recording heat pump output. Their electricity consumption is not included in E_{hp}, and their output is not included in H_{hp}. To preserve this distinction we refer to these as auxiliary heaters.

A given system may have any combination of E_{boost}, E_{iboost}, E_{sp} and E_{dhw}. Figure 2.1 shows the model assumed for a heat pump with both internal and external boost heaters.

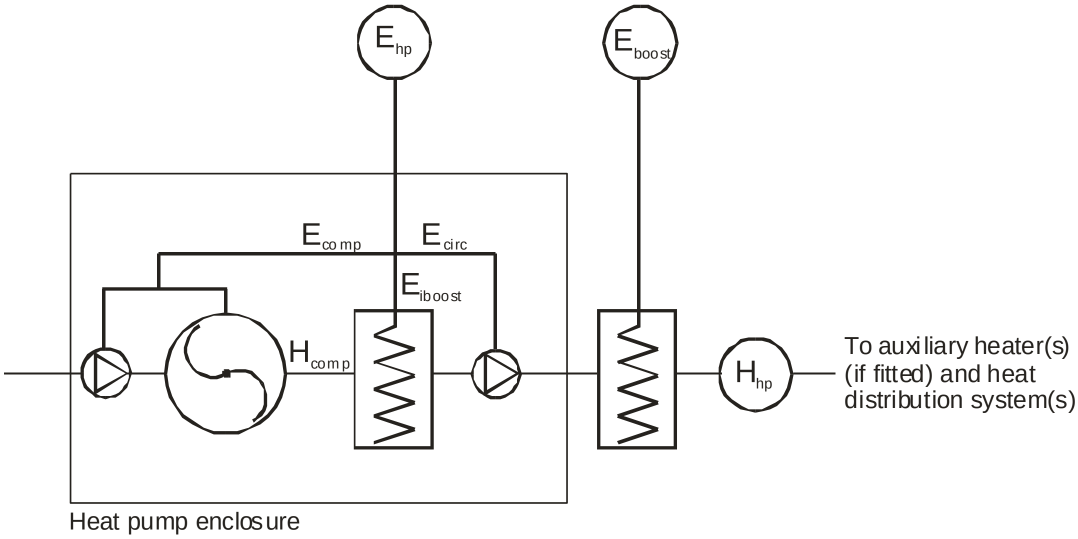


Figure 2.1: Model of a heat pump with internal boost heater

As the figure shows the energy metered as E_{hp} is assumed to be divided between the internal boost heater (E_{iboost}) and the heat pump compressor (E_{comp}). Throughout this document the value of E_{comp} includes the electricity used by the supply side circulation pump or fan: it is the SEPEMO H2 value [5].

The goal of the work described is to identify the value of E_{iboost}, using only measurements of E_{hp}, E_{boost} and H_{hp}. To this end we define the boost fraction of the heat meter enclosure, *bf,* as the ratio of internal boost energy to overall electricity consumption of the compressor and internal boost heater:

$$bf = \frac{E_{iboost}}{E_{iboost} + E_{comp}} = \frac{E_{iboost}}{E_{hp} - E_{circ}}$$

The energy used by the supply side circulation pump is included in the measured value of E_{hp}, and the energy it imparts to the flow (assumed to be equal to the power input to the pump) is included in H_{hp}. Both are subtracted back out. Because the separately metered boost heating (E_{boost}) is registered by the heat meter measuring H_{hp}, it is necessary to subtract it from the recorded value of H_{hp} to get the thermal output of the heat pump box. A unit SPF, which relates the thermal output of the compressor and internal boost heater to their total electrical input can then be defined:

$$SPF_u = \frac{H_{comp} + E_{iboost}}{E_{comp} + E_{iboost}} = \frac{H_{hp} - E_{circ} - E_{boost}}{E_{hp} - E_{circ}}$$

With a little algebraic manipulation, presented in full in Appendix A, it can be shown that:

$$bf = \frac{SPF_u - SPF_{H2}}{1 - SPF_{H2}}$$

As expected, when the unit is operating at an SPF_u equal to SPF_{H2} then no internal boost heating is in operation, and *bf* evaluates to zero. Correspondingly, when SPF_u is equal to one the system is using only the boost heater, and *bf* evaluates to one. The important result to come from this analysis is that between these extremes the boost fraction varies linearly. Thus if SPF_{H2} is known, the boost fraction and hence E_{iboost}, can be found simply from SPF_u.

3 The EQ plot

A simple plot of the amount of heat coming out of a heat pump (which may or may not have an internal boost heater) against the amount of electricity going in can be surprisingly informative. Figure 3.1 shows such a plot, generated from fabricated data. The data fabrication process has assumed a heat pump with thermal output 9kW, delivered at an SPF_{H2} of 3, corresponding to an electrical input of 3kW. The internal boost heater is assumed to have a capacity of 4kW. For the particular example generated the boost fraction is 10.2%. As described in the previous section, circulation pump energy and external boost heating applied before the heat measurement point have been assumed to be subtracted from the metered heat.

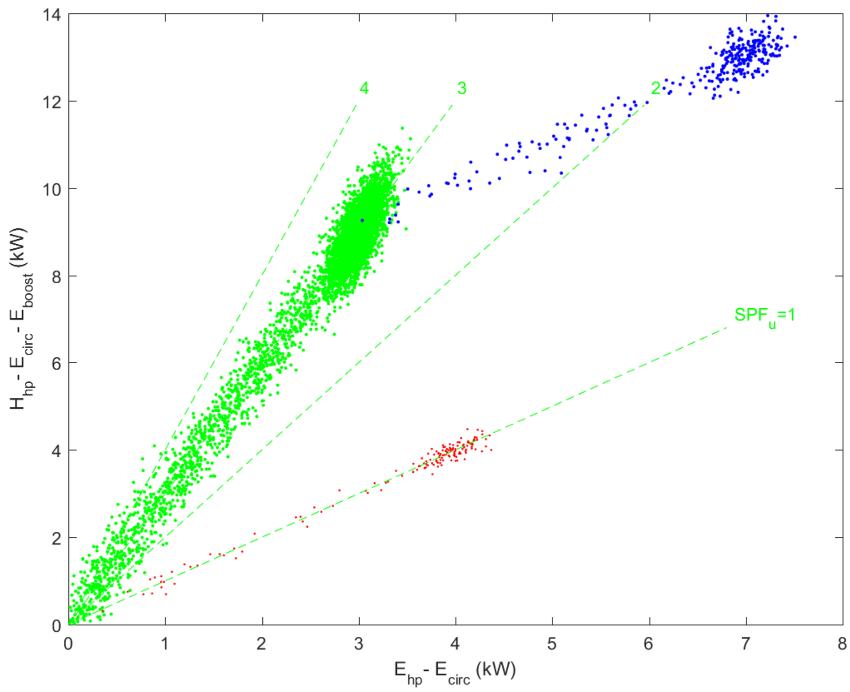


Figure 3.1: EQ plot of fabricated test data

The green points on the plot represent periods during which the heat pump operated on its own. The thermal output is greater than the electrical input by a factor of SPF_{H2}. The bulk of the points are centred around the heat pump capacity, 9kW thermal and 3kW electrical. There are some points on the way up to this centre and, unless the heat pump uses variable speed control on its compressor, these will correspond to the times at which the system turned on or off, at some point during the data recording interval. The slope of the line followed by these points is equal to the assumed SPF_{H2}.

The blue points correspond to the situation in which the heat pump does not have the required output capacity, and uses its internal boost heater to top up. At this point the electrical consumption of the boost heater is added to that of the heat pump, and its thermal output (assumed equal to the electrical input) is added to the thermal output of the heat pump. As a result, the slope of the line followed changes to one.

Finally, the red points represent periods when the internal boost heater of the heat pump runs, but the compressor does not. In this situation the unit SPF is one.

As part of the UCL investigations some cases have been identified where there are data points along the x-axis. This implies that electrical energy is being used by the heat pump, but no corresponding thermal output is seen. It is suspected that this may be due to the heat pump supplying power to a remote immersion heater for a periodic pasteurisation cycle. Once again, the problem leads to under-reporting of SPF_{H2}. This has not been examined in this report, but it is clear that once the EQ plot has been prepared it could be easily identified.

4 Simple analysis of the EQ plot

Some points on the EQ plot are highly likely to correspond to operation of internal boost heating. In particular:

- points at which the electrical input is greater than the rated consumption of the compressor and its ancillaries;
- points at which the thermal output is equal to the electrical input ($SPF_u = 1$). Whilst this could just indicate a very poorly performing heat pump, it is more likely to imply that only the boost heater is running.

Given these criteria it should, in principle at least, be possible to generate a lower bound on the amount of internal boost heating being used by a heat pump by identifying points with an E_{hp} greater than the compressor rating or an SPF_u equal to one.

4.1 The simple algorithm

As is so often the case, this very simple algorithm needs some refinement to make it robust in the face of real data:

- in practice the rated consumption of the compressor is unlikely to be known in advance, and deriving it from the rated thermal output requires prior knowledge of the value of SPF_{H2}. However it can be derived from the measured data. If it is assumed that the heat pump spends most of the time operating without boost heating then the majority of the data points should be at or below this threshold. In this work the 90th percentile of the points when the compressor is believed to be running has been used to determine this value. If this result was used directly, then some of the points scattered slightly above the threshold would be incorrectly classified as internal boost heating, and to avoid this a safety zone of 20% is added to the estimated peak compressor power. In practice it is highly unlikely that the capacity of a boost heater would be less than 20% of the heat pump capacity and so this should not create any errors.

- a similar issue arises with the second condition. If the threshold for attributing internal boost to points with a low SPF_u is set exactly to one then points which scatter into the region slightly above this value will not be correctly classified. To avoid this the criterion for boost operation is set to $SPF_u < 1.2$.

- a final pathology arises at very low input powers. These may correspond to periods when circulation pumps are running but the compressor is not, or simply to power used by controls when the system is idle. Although the individual energies are small, there may be a large number of them. To avoid incorrectly classifying them as boost heating periods with power inputs less than 250W are excluded from the analysis.

Applying the resulting algorithm to the fabricated data shown on Figure 3.1 gives a boundary, outside which boost heating is assumed to be operating, shown on Figure 4.1 below.

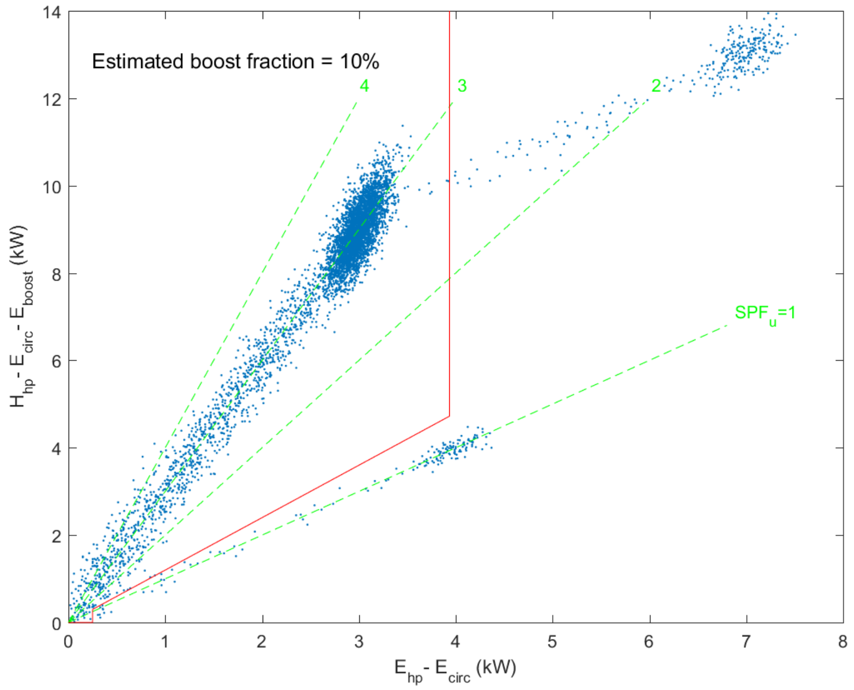


Figure 4.1: Application of the simple algorithm to the fabricated data

To determine the total boost electricity used, and hence the boost fraction, it is necessary to consider the area outside the boundary in two parts:

- for points in the part of the boundary defined by the $SPF_u < 1.2$ criterion the compressor is assumed not to be operating and so all of the input electricity is assumed to be boost, and

- for the remaining points it is assumed that the compressor is also running, and thus the boost energy is given by the input energy minus the previously estimated compressor capacity.

Once these energies have been totalled, the boost fraction can be easily evaluated. For this idealised dataset the algorithm correctly reports it as 10%.

5 Distribution based analysis of the EQ plot

This analysis aims to detect the operation of internal boost heating in the presence of more complicated (although not necessarily more optimal) control of the boost heater. In particular it should identify situations where both compressor and internal boost heating are running below their maximum outputs.

The analysis works by looking at the amount of electrical energy consumed at different values of the unit SPF.

5.1 The distribution algorithm

Figure 5.1 shows the test data described above, broken down by the unit SPF. The figure shows the amount of electrical energy consumed at each SPF value.

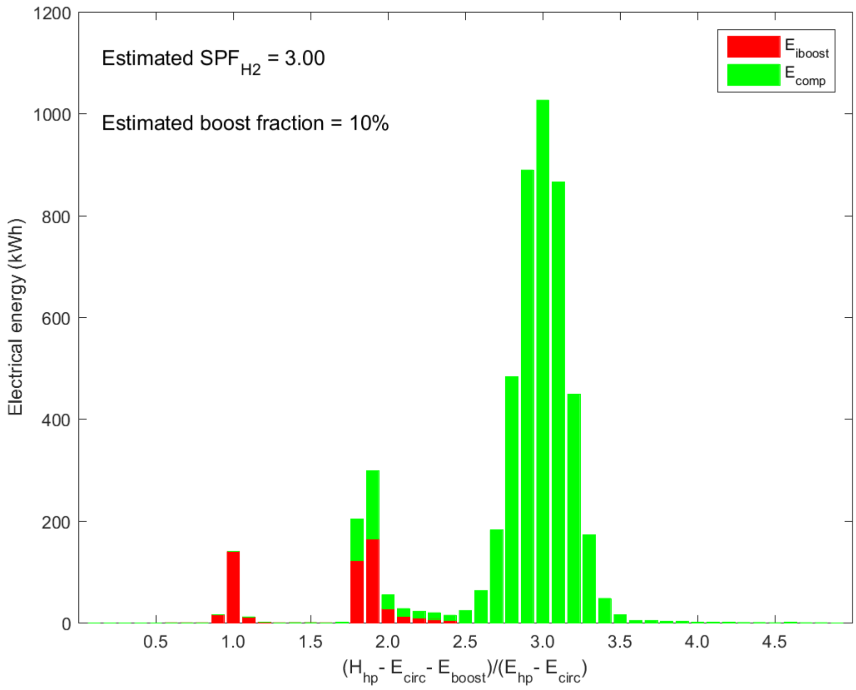


Figure 5.1: Analysis of the distribution of SPF_u

There is a clear maximum at an SPF_u of 3. This corresponds to the green points on Figure 3.1, where no internal boost heating is in use, and the unit runs at SPF_{H2}. There is a second maximum at around 1.9. This corresponds to periods where both the compressor and the internal boost heating are running at capacity. In this situation the electrical input is 3 + 4 = 7kW, and the thermal output is 9 + 4 = 13kW, giving a unit SPF of 1.86.

In its simplest form the algorithm first seeks the highest local maximum. This is assumed to correspond to the ideal situation in which the unit is running without boost heating. In practice there is always some fluctuation around this peak, and the data above the maximum is combined with an equal number of points from below the maximum in an attempt to generate an unbiased estimate of SPF_{H2}. Operation at values below this is assumed to be due to the fact that the internal boost heater is being used. The amount of boost heating can be inferred using the relationship developed in

Section 2, where it was shown that this varies linearly between zero and one as the unit SPF varies between SPF_{H2} and one.

As before, this initially straightforward algorithm requires refinement to make it robust in the face of real data.

- the largest local maximum is found by working down from the maximum value of SPF_u. Used on its own, this process would be susceptible to finding small local maxima in the noise of the upper tail of the distribution. For this data there are such maxima, barely discernible on the figure, at SPFs of 4.6, 3.9 and 3.7. This problem is avoided by requiring that at least 20% of the data points have been traversed before a maximum can be declared;

- once the algorithm has reached the first peak of the distribution, it may classify a small undulation in the distribution as the maximum. This is avoided by requiring that the value falls by at least 10% before the preceding point is declared the maximum;

- further complications arise when the boost fraction is attributed to periods of operation at an SPF_u below the estimated SPF_{H2}. The analysis presented in Section 2 revealed that this fraction varies linearly as boost heating causes the unit SPF to fall from SPF_{H2} to one. This analysis assumed that the value of SPF_{H2} was a single number. However the result returned by the analysis is actually a distribution of SPF_{H2} values. Applying the simple linear rule to the example shown on Figure 5.1 therefore starts to indicate boost heating at an SPF of 2.9, whereas this data actually corresponds to part of the distribution of performance without any boost. When calculating the average SPF it is, of course, balanced by the periods when the observed SPF is 3.1. As a result it is necessary to evaluate the boost energy across the whole range of SPFs, including the negative values when the observed SPF is greater than SPF_{H2}. In order to produce a meaningful graphical representation on Figure 5.1 these negative values have been removed, along with a balancing number of positive values. This therefore has no overall effect on the calculated boost input.

With these refinements in place the method correctly estimates both SPF_{H2} and the boost factor for the highly idealised test data set.

6 Application to real data

98 sites were identified which are

- in Sample B2 (cropped)
- known to contain an internal resistance heater within the HP unit

These have been analysed using the algorithms described. Some yielded very high apparent boost fractions, and these sites have been examined manually, and the more comprehensive distribution algorithm and space heating/hot water separation used to clarify issues. As expected, the inspection process weeded out a lot of the cases where unfeasibly large amounts of boost input were originally diagnosed.

Some example sites where manual investigation yields false alarms are as follows:

6.1 A false alarm: poor heat pump performance (RHPP5620)

For this site the simple algorithm reports a boost fraction of 69%.

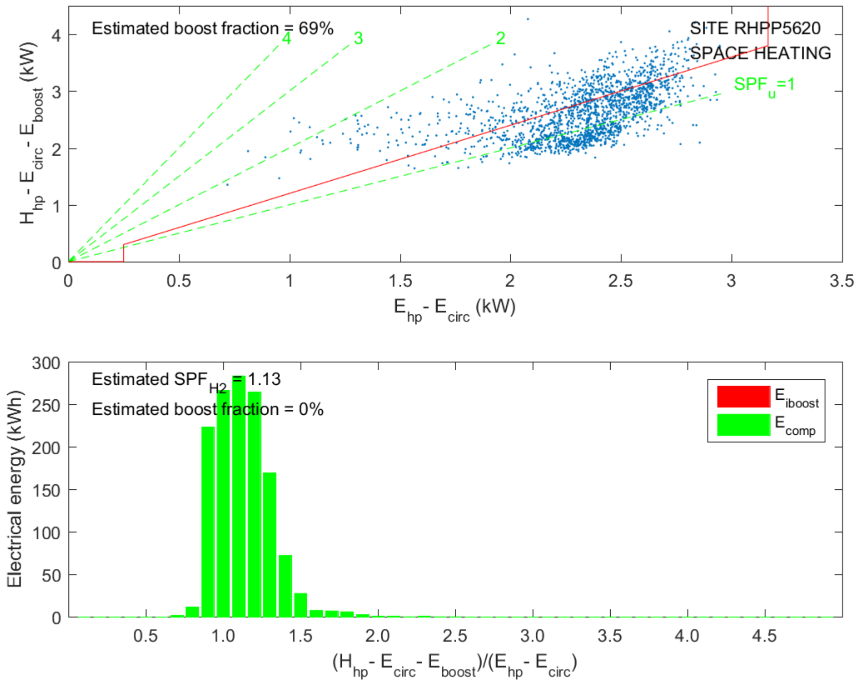


Figure 6.1: Analysis of site with poor heat pump performance

However, the more complex distribution analysis concludes that the boost energy is zero, due to the very poor performance of the heat pump. Even the occasions when the SPF of the unit falls below one are attributed to the fact that the distribution of SPF without boost extends down to that value. It therefore seems that the poor performance of this system may be due to factors other than the excessive use of internal boost heating.

6.2 Classic boost heating behaviour (RHPP5460)

This system shows all the characteristics of internal boost heating outlines in Section 3.

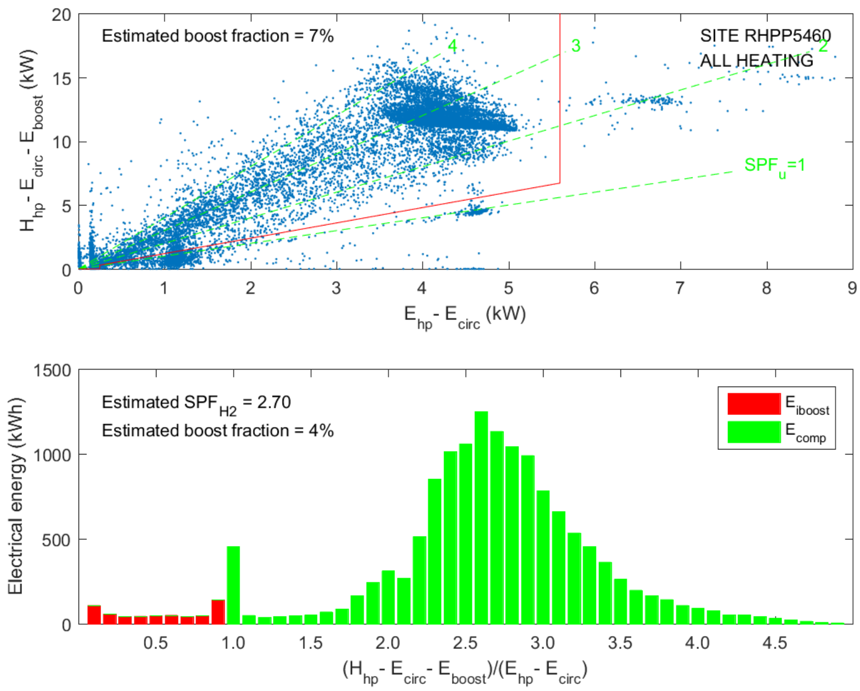


Figure 6.2: Analysis of a system showing classic boost behaviour

By contrast to the previous system this site shows the symptoms of a case with internal boost heating. There are two clusters of points which take off from the capacity of the heat pump with a slope of one, suggesting that the boost heating is being applied in two stages, each corresponding to 2kW. Finally, there is a cluster of points at an SPF_u of one, corresponding to a power input of just over 4kW, suggesting that both boost heaters were operated whilst the compressor was not running.

In this case the distribution algorithm underestimates the amount of boost heating because the distribution of SPF_u is significantly skewed to the right, violating a key assumption.

6.3 A false alarm: water heating performance mistaken for boost (RHPP5707)

In this example the distribution algorithm fails, spectacularly over predicting internal boost fraction. When the data is separated into space heating and hot water production periods, the reason becomes clear.

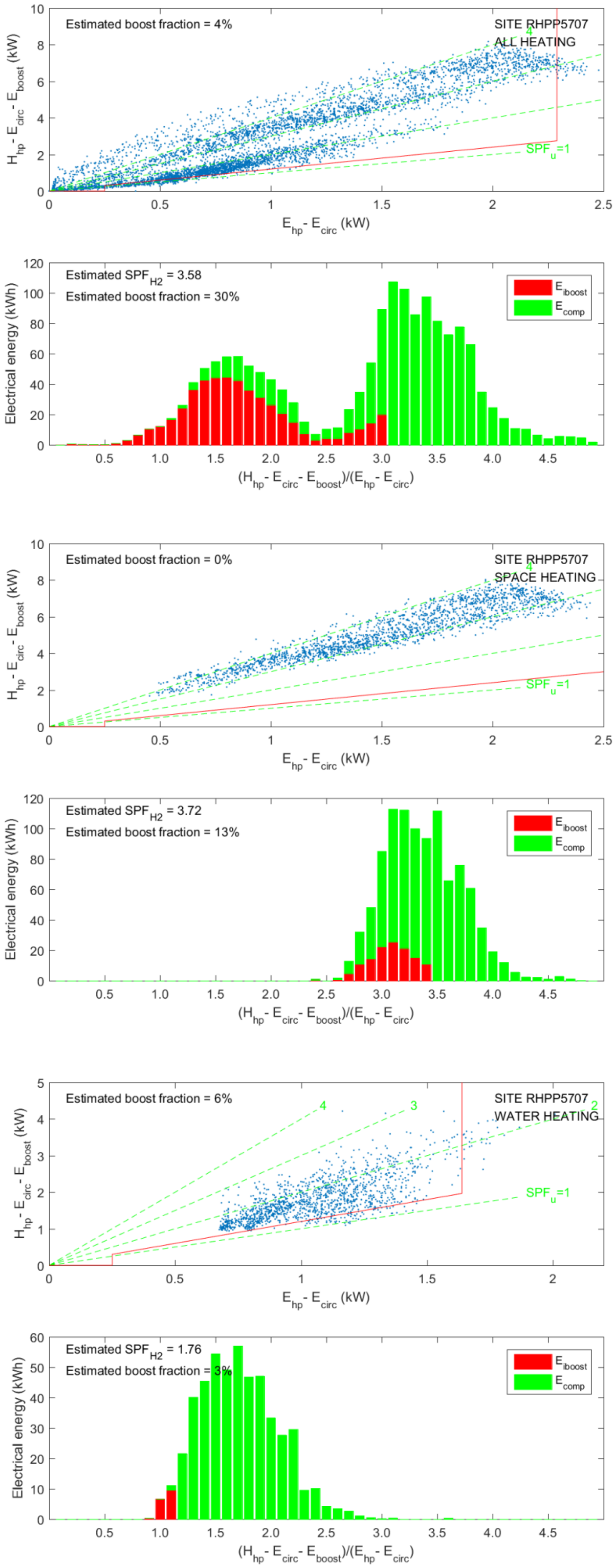


Figure 6.3: Distribution algorithm falsely identifies hot water production as boost heating

When the whole dataset (space heating and hot water) is analysed the simple algorithm indicates only a small amount of boost. However it is clear from the distribution plot that the system spends a significant amount of time operating at a reduced SPF, and the distribution algorithm attributes this to the operation of internal boost heating, as a result estimating a boost fraction of 30%. When operation is separated into space heating and hot water production it becomes apparent that the second peak corresponds to hot water production and is quite possibly not related to boost heating at all. This demonstrates another shortcoming of the distribution algorithm: because it assumes operation at a constant SPF_{H2} it may fail in situations where other factors cause this parameter to vary.

6.4 Analysis across the sample

In general, the simple algorithm has been found to be the more robust, and for this reason the discussion of the whole sample is led by those results. Figure 6.4 shows the distribution of boost fractions estimated by the simple algorithm, using the complete data sets (space heating and hot water production).

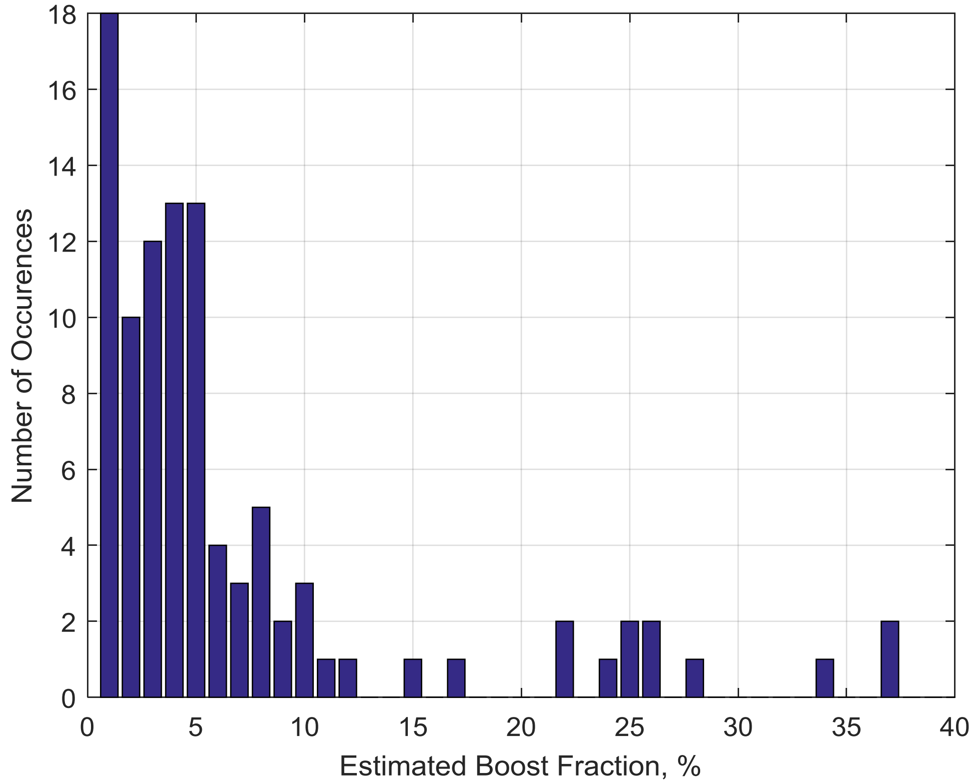


Figure 6.4: Estimated boost fractions across whole sample (space heating and hot water)

There are 15 sites where the estimated boost is greater than 10%, and 83 sites (85%) in which the boost is less than 10% of the total electricity at the nominal H2 bound[14]. The median boost fraction is 3.8%. The expression presented in Appendix D shows that, for a system with a true SPF_{H2} of 2.5 (the RED threshold), this would result in an under estimation of SPF_{H2} of 0.055, or approximately 2%.

[14] According to the SEPEMO definition, boost electricity is included in SPFH3 not SPFH2; but for these sites, the measured heat pump electricity has included this boost and therefore, for these sites, the SPFH2 value in this report actually should be labelled as SPFH3.

7 Conclusions and suggestions for further investigations

The RHPP monitoring specification provides temperature sensors which were to be used to detect operation of an internal boost heater. The resulting data was used in the preliminary data analysis carried out by BEIS. Unfortunately, due mainly to installation, constraints this approach was subsequently decided to be unreliable and was eventually discontinued.

The work described has demonstrated that it is possible at least to estimate how much internal boost energy is being used with an algorithmic approach. An artificial dataset has been produced, mainly to provide graphical demonstrations of the algorithms in action. For this (highly idealised) data both algorithms correctly recover the amount of internal boost energy being used, and hence the correct value of SPF_{H2}.

Of the two algorithms, the simpler appears to give more robust results when confronted with real data. The second, more sophisticated, algorithm can, in principle, identify internal boost heating in a wider range of circumstances, but when faced with real data it fails more frequently. In when, as in this report, data from space heating and hot water production are treated together it overestimates the amount of internal boost heating.

A further issue with the data has been identified by UCL. Some sites may be taking the power required to pasteurise the hot water cylinder from the metered heat pump supply, and not registering the heat produced. The analysis presented here could readily be used to identify this.

In view of these comments the following three areas for further investigation are suggested:

- revisit the original approach of using T_{co} in conjunction with T_{sf} and T_{wf} to identify periods of internal boost operation. Determine whether it is possible to refine the way in which this data is used to obtain a third estimate of when internal boost heating is being used, and whether that estimate is consistent with the methods developed here. This would mean that it could be incorporated directly into the data analysis process. Although this will have little impact on the results presented, the fact that internal boost is seen to be explicitly treated will increase the credibility of those results;

- explore whether better agreement between the simple algorithm and the distribution algorithm if space heating and hot water are treated separately;

- detect points along the x-axis and tabulate the number of sites where they appear, and the amount of energy consumed. Depending on the significance of this effect, determine from installers and/or manufacturers whether it really is due to auxiliary heating powered from within the heat pump.

Additional notes: the relationship between boost fraction and SPF_u

In Section 2 the boost fraction, *bf*, was defined as the ratio of internal boost energy to overall heat pump box electricity consumption, which is equal to the combined internal boost energy and compressor energy:

$$bf = \frac{E_{iboost}}{E_{iboost} + E_{comp}}$$

With a little re-arrangement this is equivalent to:

$$E_{iboost} = \frac{bf}{1 - bf} E_{comp}$$

Because the separately metered boost heating (E_{boost}) is registered by the heat meter measuring H_{hp}, it is necessary to subtract it from the recorded value of H_{hp} to get the thermal output of the heat pump box. The unit SPF, which relates the combined thermal output of the compressor and internal boost heater to the total electrical input is defined as:

$$SPF_u = \frac{H_{comp} + E_{iboost}}{E_{comp} + E_{iboost}} = \frac{H_{hp} - E_{circ} - E_{boost}}{E_{hp} - E_{circ}}$$

For the compressor (and associated source side pump) alone it must also be true that:

$$H_{comp} = SPF_{H2}\, E_{comp}$$

Substituting into the definition of the unit SPF gives:

$$SPF_u = \frac{SPF_{H2}\, E_{comp} + \frac{bf}{1 - bf} E_{comp}}{E_{comp} + \frac{bf}{1 - bf} E_{comp}}$$

Multiplying top and bottom by *(1 - bf)* and dividing by E_{comp} then gives:

$$SPF_u = (1 - bf)\, SPF_{H2} + bf$$

Finally, the result below, already presented in Section 2, is obtained:

$$bf = \frac{SPF_u - SPF_{H2}}{1 - SPF_{H2}}$$

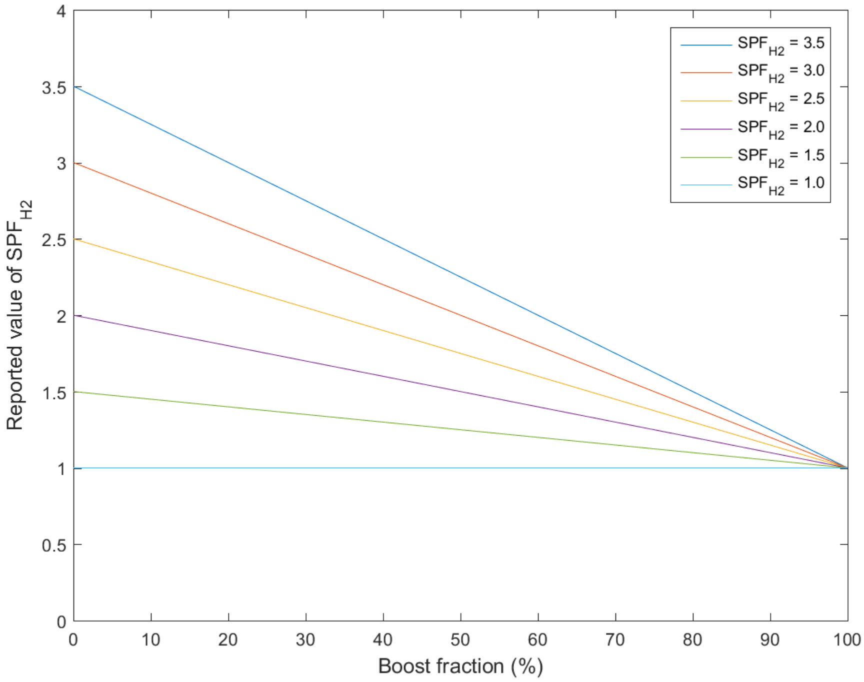
Additional notes: sensitivity of reported SPF_{H2} to boost fraction

The impact of the decision to ignore internal boost heating is that the value reported for SPF_{H2} is actually SPF_u. This Appendix evaluates the magnitude of this error in terms of the boost fraction.

From previous section:

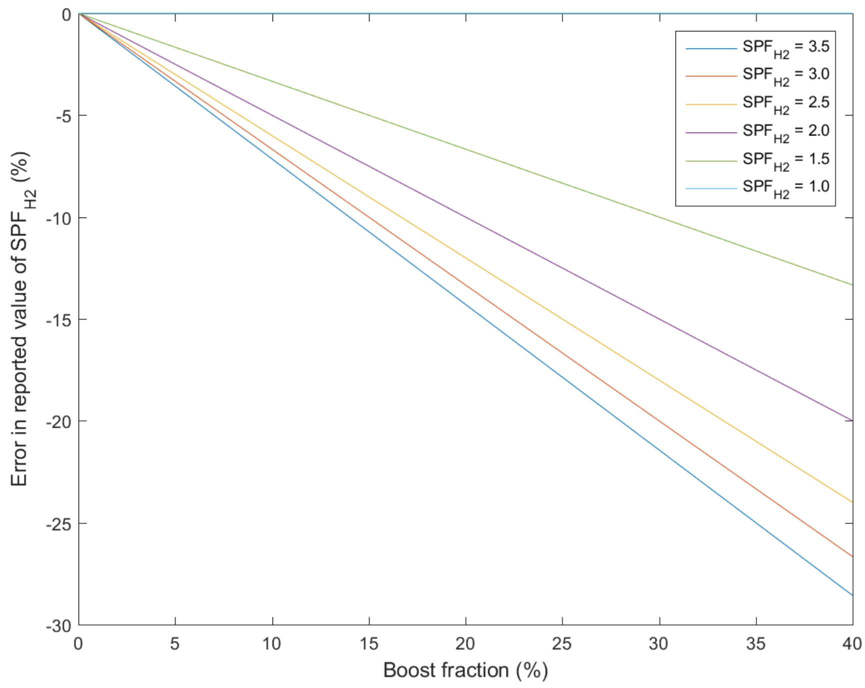
$$SPF_u = (1 - bf)\, SPF_{H2} + bf$$

The figure below shows how ignored internal boost operation impacts on the reported SPF.



As expected, the figure shows that the reported SPF varies linearly between SPF_{H2} when the system is using no boost energy, and one when it is using 100% boost.

In practice it is more useful to translate the boost fraction into a percentage error in reported SPF. The result is shown on the next figure.


Error in reported value of SPF_{H2} (%)
Boost fraction (%)
0
-5
-10
-15
-20
-25
-30
0
5
10
15
20
25
30
35
40
SPF_{H2} = 3.5
SPF_{H2} = 3.0
SPF_{H2} = 2.5
SPF_{H2} = 2.0
SPF_{H2} = 1.5
SPF_{H2} = 1.0

REFERENCES

[1] BEIS RHPP heat pump monitoring: Review of data analysis software developed by UCL. Chris Martin. Energy Monitoring Company. December 2015.

[2] BEIS RHPP heat pump monitoring:
Developing a consistent approach to monitoring heat pump installations
Chris Martin. Energy Monitoring Company. Version 3: October 2011

[3] Overview of RHPP analysis framework
Denis Fan. BEIS.

[4] RHPP metering site schematics
Prepared by: Chris Wickins and Andrew Thorne
Version 2.1: September 2012

[5] Concept for evaluation of SPF: A defined methodology for calculation of the seasonal performance factor and a definition which devices of the system have to be included in this calculation.
Andreas Zottl and Roger Nordman Version 2.0: August 2011

UCL ENERGY
INSTITUTE

BSRIA

UNIVERSITY OF
WESTMINSTER

SP