*Data-Centric Engineering* (2024), 5: e39
doi:10.1017/dce.2024.44

RESEARCH ARTICLE

# Estimating the seasonal performance and electricity consumption of retrofitted heat pumps

Daniel René Bayer and Marco Pruckner

Modeling and Simulation, University of Würzburg, Würzburg, Germany
**Corresponding author:** Daniel René Bayer; Email daniel.bayer@uni-wuerzburg.de

**Received:** 09 April 2024; **Revised:** 09 August 2024; **Accepted:** 23 September 2024

**Keywords:** analysis of smart meter data; building simulation; heat demand estimation; heat pump retrofit; sector coupling

## Abstract

Gas furnaces are the prevalent heating systems in Europe, but efforts to decarbonize the energy sector advocate for their replacement with heat pumps. However, this transition poses challenges for power grids due to increased electricity consumption. Estimating this consumption relies on the seasonal performance factor (SPF) of heat pumps, a metric that is complex to model and hard to measure accurately. We propose using an unpaired dataset of smart meter data at the building level to model the heat consumption and the SPF. We compare the distributions of the annual gas and heat pump electricity consumption by applying either the Jensen–Shannon Divergence or the Kolmogorov–Smirnov test. Through evaluation of a real-world dataset, we prove the ability of the methodology to predict the electricity consumption of future heat pumps replacing existing gas furnaces with a focus on single- and two-family buildings. Our results indicate anticipated SPFs ranging between 2.8 and 3.4, based on the Kolmogorov–Smirnov test. However, it is essential to note that the analysis reveals challenges associated with interpreting results when there are single-sided shifts in the input data, such as those induced by external factors like the European gas crisis in 2022. In summary, this extended version of a conference paper shows the viability of utilizing smart meter data to model heat consumption and seasonal performance factor for future retrofitted heat pumps.

**Impact Statement**

Reducing carbon emissions in residential heating is a major concern until the mid-21st century. To estimate the annual electricity consumption of future retrofitted heat pumps that replace gas furnaces, we need to know the expected seasonal performance of these heat pumps. With the availability of more and more smart meter data, we propose and evaluate a new method to estimate the seasonal performance of future heat pumps. This method involves converting current annual heat consumption measures, based on annual gas consumption, to the electricity consumption that a retrofitted heat pump would require. The novelty of this method is the use of either the Jensen–Shannon Divergence (JSD) as an instance of a distance measure or the Kolmogorov–Smirnov (KS) test as an instance of a statistical test on an unpaired dataset of annual gas and heat pump electricity consumption values.

## 1. Introduction

In the following years, the number of installed heat pumps in the residential sector is expected to increase strongly, replacing existing gas furnaces, especially in Europe, in order to decarbonize the building sector (Rosenow et al., 2022). The European Union (EU) pursues this goal politically with vigor (European

Parliament, 2023). For instance, Germany aims to reach net climate neutrality by 2045, where heat pumps will be the main driver in the decarbonization of the residential heat sector (Federal Ministry for Economic Affairs and Climate Action (BMWK), 2022). As highlighted, for example, by Watson et al. (2023), these new heat pumps will substantially increase the electricity demand. This increased demand puts additional stress on all levels of the electricity system, including the distribution networks, causing voltage drops or overloads in the substations (Alpízar-Castillo et al., 2023). Thus, utility companies need to predict the impact of these retrofitted heat pumps on their distribution grids. Based on this prediction, possible grid bottlenecks and overloads can be identified at an early stage. As a result, grid expansion can be carried out early and in a targeted manner.

For the analysis of the effect of new heat pumps either on the building level or on the distribution grid level, utility operators need to know the heat consumption over a given period of time, and they need an assumption on the average efficiency of heat pumps in their supplied city. In this context, the most important metric is the (heating) seasonal performance factor (SPF). It is the amount of heat produced by a heat pump in a given year divided by the required electricity consumption. Existing analyses show that the SPF depends on the typical climatic conditions of a given city (Nouvel et al., 2015; Rossi di Schio et al., 2021) and changes from year to year due to varying weather conditions (Huchtemann and Müller, 2012). Knowing the SPF for an average heat pump in a given year allows for quick calculation of the electricity required by a heat pump to heat a building based on its annual heat consumption. However, measuring the SPF of existing heat pump installations is costly and effortful, for example, due to the installation of an additional heat meter in the heating circuit (Kelly and Cockroft, 2011). Moreover, the widely deployed electricity smart meters can only meter the electricity consumption of a heat pump, ignoring the heat output. On the other hand, the amount of gas consumed by a building can simply be converted to the heat consumption on the building level, assuming furnace efficiency and the gas heating value. Therefore, while smart meter data on heat pump electricity and gas consumption are readily available nowadays, the actual heat produced by heat pumps remains unknown.

In this article, we propose a novel method for estimating the average SPF of existing heat pumps on an annual level by utilizing two unpaired datasets—the heat pump electricity and the gas consumption smart meter data across all residential single- and two-family buildings in a city. We aim to find the average SPF value that maximizes the similarity between these two datasets. However, the challenge lies in the fact that these two datasets are not paired, as each building has only one heating system, either a heat pump or a gas furnace. Therefore, we need to determine if it is possible to find at least one SPF value where both distributions are similar in a statistically significant way.

In detail, we answer the following research questions in this paper:

1. How can we predict the mean SPF over all existing heat pump installations based on an unpaired heat pump electricity and gas consumption dataset using either the Jensen–Shannon Divergence (JSD) or the Kolmogorov–Smirnov (KS) test?
2. Which of the estimated SPF values result in a significant similarity of the given distributions using the KS test?
3. How much electricity would be consumed by heat pumps that replace existing gas furnaces based on the previously estimated SPF?

In order to address these research questions, we present a novel approach using either the JSD or the KS test to compare the heat pump electricity and gas furnace consumption distributions. A major contribution of the work is that the validity of the assumptions is also justified with the help of the KS test, which provides a statement about the statistical significance of the similarity of the two distributions.

The estimated SPF values can serve as a basis for estimating the future heat pump electricity consumption in order to analyze their effect on the local distribution grid. In contrast to existing works that must rely on assumptions about the actual SPF that are not aligned to existing data or weather realizations like Lechowicz et al. (2023) or Damianakis et al. (2023), our work can be used to simulate the heat pump retrofit in a given city where such data is available for multiple, recently past years. Thus, the

https://doi.org/10.1017/dce.2024.44 Published online by Cambridge University Press

variance in the SPF over different years caused by different weather realizations can be estimated properly.

The subsequent parts of this paper have the following structure. First, we will discuss the additions in this publication, which is an adapted version of a conference paper. Thereupon, we present related work dealing with similar research questions in Section 2. We follow with details on our methodology in Section 3 and present the results in Section 4. Subsequently, we discuss these results in Section 5, highlighting both the advantages and the limitations of this novel approach. Finally, we conclude the paper in Section 6.

### *Adaptations of the original conference paper*

This paper is an extended version of Bayer and Pruckner (2023b) published in the Proceedings of the 10th ACM International Conference on Systems for Energy-Efficient Buildings, Cities, and Transportation. In this version, we extend the existing methodology of estimating the SPF by using the KS test to get a statistical confidence interval for the estimated SPF values. Moreover, we include the data from 2022 into our analysis, showing a limitation of the presented methodology. The different methods for the data-driven estimation of the SPF are discussed at the end of the paper.

## 2. Related Work

Real measured gas consumption data is used by Kolter and Ferreira (2011) to predict and visualize the heat demand on the building level. Ko (2013) also utilizes heat data measured on the building level to provide an statistical analysis of heat and energy usage in the urban sector, including an overview of techniques for modeling urban energy usage. When hourly heat demand data for a building is available alongside weather data, the model proposed by Lumbreras et al. (2022) can be employed for long-term, hourly heat demand predictions.

Several models exist that estimate either the Coefficient of Performance (COP) (Shin and Cho, 2021; Priarone et al., 2020) or the SPF (Nouvel et al., 2015) based on theoretical computations or laboratory measures. In the latter, for instance, the authors present a model for estimating the SPF of a highly simplified sample building located in various regions of Europe, illustrating how location-specific climatic conditions can affect the SPF. This variation in SPF across different regions is also investigated by Rossi di Schio et al. (2021), who demonstrate that air-source heat pumps experience changes in performance due to the varying number of defrosting cycles required in different climates. Meanwhile, field studies on the electricity consumption and SPF of real-world heat pump installations, such as those by Huchtemann and Müller (2012), Miara et al. (2014) or Ruschenburg et al. (2014), reveal notable differences in the SPF from the theoretical computed values. Finally, Carroll et al. (2020) provide a detailed review of field studies on air-source heat pump efficiencies, highlighting notable variations in the resulting SPF values depending on the region under consideration.

Given that the SPF is heavily influenced by both the geographic location and the specific year due to weather fluctuations, and real-world SPF values often differ notably from theoretical calculations, an accurate estimation of the SPF is crucial for effective energy planning. Computing the SPF in a field trial requires metering the actual heat produced by a heat pump, which can be labor-intensive and costly. Therefore, using smart meter data to estimate the SPF appears to be a promising approach for computing average SPF values at the city level, especially since the SPF can differ from one region to another, as mentioned in the previous paragraph. So far, only a few recent works have used real gas consumption data, assuming these gas furnaces should be replaced by heat pumps. For example, Lechowicz et al. (2023) use gas consumption data of a city to compute the residential energy demand on a household level using an assumed COP. They focus on selecting regions of the city where heat pumps should replace all existing gas furnaces to save maintenance costs at the gas grid level. Wamburu et al. (2022) present an analysis of gas consumption in a residential area based on real data. Additionally, they propose an optimization framework for reducing carbon dioxide emissions. Also, other publications focusing on the effects of a

https://doi.org/10.1017/dce.2024.44 Published online by Cambridge University Press

heat pump retrofit use assumed values for the COP or SPF that are not aligned to the concrete weather conditions of a year like Walker et al. (2022) or Jenkins et al. (2009). The disadvantage of using SPF values that are not aligned with the actual weather conditions occurs in cold winter. There, the actual performance is overestimated, leading to an underestimation of the required electricity consumption of the heat pumps (Petrović and Karlsson, 2016). Thus, the estimation of the SPF that is aligned with the actual annual weather and consumption behavior patterns is an important topic.

### *Contribution of this work*

Up to now, existing gas consumption data has not yet been combined with heat pump electricity consumption data to estimate the performance of current heat pump installations and assess the impact of heat pump retrofits in future energy system states. As heat pumps become increasingly popular, a data-driven approach for estimating their performance offers a promising avenue to foster the analysis of the potential effects of widespread heat pump retrofits. However, direct metering of the individual heat pump performance remains effortful, further emphasizing the need for efficient data-driven estimation methods. In this publication, we present the estimation of the seasonal performance of retrofitted heat pumps replacing existing gas furnaces based on an unpaired dataset of annual heat pump electricity and natural gas consumption within the same city and building class. We, therefore, propose two different methods to determine the SPF at which both distributions are most similar. Both methods are based on well-known statistical metrics.

The novelty of this work is that the selected statistical metrics have not yet been used for the analysis of two unpaired datasets on building heat consumption. Additionally, the methods involve additional methodological advancements. A distinct approach was developed around each applied test and utilized metric, particularly the JSD and KS tests. This encompasses the addition of a smoothing process and the minimization of the KS-test metric, which is not commonly done. This customized methodology enhances the accuracy and reliability of estimating the SPF, thereby providing a more robust analysis of heat pump performance in retrofitted systems.

### *COP and SPF: Measurement and calculation scope*

Whereas the COP typically measures the heating power produced by a heat pump at a given moment in time, the SPF measures the thermal energy, i.e., the heat, produced by a heat pump over a year divided by the electricity required to operate the heat pump (Hailu, 2021). The COP is a measure of the efficiency of a heat pump measured in a laboratory environment. The SPF, on the other hand, evaluates the actual physical work performed by a heat pump in a year. Gleeson and Lowe (2013) report different definitions of the SPF, depending on the scope of the measurement. As in most of the publications, we include the complete electricity consumption of a heat pump, especially including the reheat coil, for the SPF calculation. As a COP model typically does not include additional components like reheat coils, the conversion of a model-based COP time series to an annual SPF value is not directly possible.

## 3. Methodology

Let $H^{Gas}$ be the set of buildings with existing gas furnaces that should be replaced by heat pumps and $H^{HP}$ be the set of buildings with already existing heat pumps. Based on the existing gas consumption data of year $j$, we can compute the thermal energy $Q_j(h)$ produced by a gas furnace in a building $h \in H^{Gas}$ similar to Lechowicz et al. (2023) using

$$Q_j(h) = R_j(h) \cdot z \cdot v \cdot \eta_g \tag{1}$$

where $R_j(h)$ is the volume of gas consumed in year $j$, $z$ is the pressure factor, $v$ is the heating value of the gas, and $\eta_g$ is the efficiency of the gas furnace. In this paper, we expect that the heat consumption follows the heat demand.

https://doi.org/10.1017/dce.2024.44 Published online by Cambridge University Press

The heat consumption of a building is highly dependent on its individual insulation level. The insulation level between gas and heat pump heated buildings might be different, due to two reasons. Firstly, it is possible that a building's insulation level will be improved if a gas furnace is replaced by a heat pump, as it is reported in a field trial by Kelly and Cockroft (2011). Secondly, new buildings are generally better insulated than older buildings. However, it can be assumed that gas heating systems tend to be installed in older buildings, as heat pumps have started to become widespread in private buildings in the last decade. Therefore, we introduce the parameter $\gamma(h)$, which denotes the heat demand reduction that goes along with the replacement of the gas furnace in building $h \in H^{Gas}$. The heat consumption $Q_j^{\Delta}(h)$ after the retrofit is thus given by

$$Q_j^{\Delta}(h) = Q_j(h) \cdot (1 - \gamma(h)) \tag{2}$$

According to Hailu (2021), we define the SPF of a heat pump installed in building $h \in H^{HP}$ in the year $j$ as

$$\mathrm{SPF}_j(h) = \frac{Q_j^{HP}(h)}{E_j(h)} \tag{3}$$

denotes the heat (i.e., thermal energy) produced by this heat pump in year $j$. We note that the considered heat pumps produce domestic hot water (but no cooling loads) during summer in our considered location.

The first assumption for this methodology is that the annual heat consumption of a building currently heated by a gas furnace does not change upon a heat pump retrofit except for contemporaneous building insulation retrofits, i.e., $Q_j^{\Delta}(h) = Q_j^{HP}(h)$. We will discuss on this assumption in Section 5.2. Thus, we can compute the electrical energy that a retrofitted heat pump in building $h \in H^{Gas}$ would consume by combining Equations (1), (2), and (3):

$$E_j(h) = Q_j(h) \cdot \underbrace{\frac{1 - \gamma(h)}{\mathrm{SPF}_j(h)}}_{=: B_j^{-1}(h)} \tag{4}$$

As mentioned in the Section 1, measuring the heat produced by a heat pump can be very challenging. Nevertheless, the electricity consumption of an existing heat pump can be metered much more quickly if a separate smart meter is installed. As a result, calculating the SPF for individual buildings becomes impossible without dedicated meters to measure the annual heat output of their heat pumps. However, as we have data on the electricity consumption of heat pumps on one hand and, on the other, data on the annual heat consumption of buildings from the same class from gas smart meter data, we compute the average SPF in the following by shifting our focus from the individual building level to all buildings. In the further course, the distribution of the annual heat consumption over all single- and two-family buildings might be denoted as $Q_j$. The heat consumption distribution of those only heated by a heat pump is denoted as $Q_j^{HP}$, with $Q_j^{HP} \subset Q_j$, and those only heated by a gas furnace are denoted by $Q_j^{Gas}$, with $Q_j^{Gas} \subset Q_j$.

Over the complete building stock of a city, the individual annual heat consumption values typically show a high degree of variation, as the annual heat consumption of a residential building depends on multiple, often hidden parameters (Lim and Zhai, 2017). The most important parameters are as follows:

- the weather conditions,
- the building size (Kolter and Ferreira, 2011),
- the household size and income (Sardianou, 2008),
- the user behavior and their preferred temperature (Delzendeh et al., 2017),
- misconfigurations of the heating system (Weigert et al., 2022), and
- the insulation standard.

This leads to a broad distribution of the actual heat consumption over the complete residential building stock in a given city (see for example Figure 1). Even though early publications reported less impact of the

https://doi.org/10.1017/dce.2024.44 Published online by Cambridge University Press

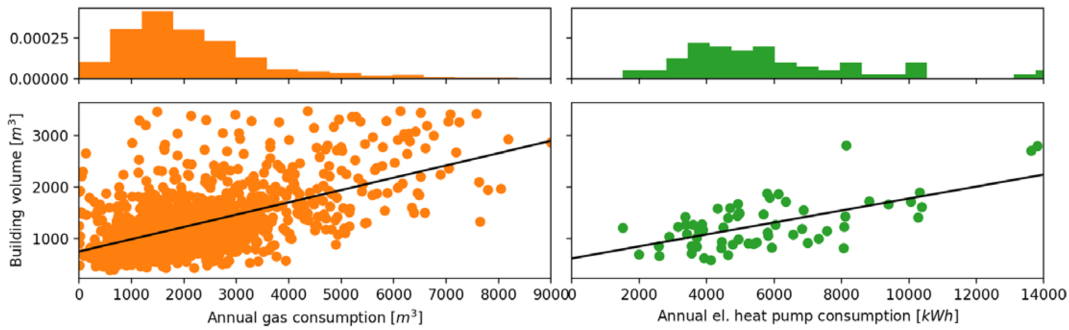


***Figure 1.*** *Histogram of the annual gas (orange) and heat pump electricity consumption (green) from our dataset in 2020 in the upper row. Lower row: The scatter plot of the annual consumption values across all single- and two-family buildings plotted against the corresponding building size. The solid black line is the result of the linear regression.*

insulation on the building heat consumption, especially in the case of an increase of the insulation level during a retrofit (Capper and Scott, 1982), more recent approaches report notable impacts dependent on the climatic conditions (Pan et al., 2012).

The main idea of this work is to assume that the influence exerted by the above-listed parameters on the annual heat consumption of an individual building is independent of the specific heating system type. This assumption implies a uniform impact of these factors on energy consumption, regardless of the underlying heating system. In other words, the distribution of the annual heat consumption over all buildings heated with gas furnaces $Q_j^{Gas}$ and heated with heat pumps $Q_j^{HP}$ is assumed to be drawn from the same underlying distribution $Q_j$ of all heat consumptions that only depends on the above-listed parameters but is independent of the heating system type, i.e.,

$$\forall j \in J:\ Q_j \sim Q_j^{HP} \sim Q_j^{Gas} \tag{5}$$

where $J$ denotes the set of considered years and the tilde denotes the equality of two distributions. We note that we only consider heat pumps and gas furnaces as heating systems in this paper. Nevertheless, the main assumption does not imply that the above-listed parameters would not have any impact on the residential heat consumption or the efficiency of heat pumps itself. As reported in related works like Huchtemann and Müller (2012), the concrete weather conditions in a given year impact the seasonal performance of heat pumps notably. Since the other parameters mentioned above that affect heat consumption in residential buildings do not change frequently between years for a given building, it can be assumed that the main cause of varying heat pump performance is the changing weather conditions.

Let us further assume that the SPF divided by $1-\gamma(h)$ (i.e., $B_j^{-1}(h)$ in Eq. (4)) is independent of the actual heat consumption $Q_j(h)$ of an arbitrary building $h$. Finally, it can be assumed that the influence of the weather on the SPF affects all heat pumps equally. This implies that $Q_j^{HP} = E_j \cdot B_j$ holds with $E_j$ denoting the distribution of the metered annual heat pump electricity consumption and $B_j$ denotes the average value of $B_j(h)$ overall existing heat pumps in year $j$ with $B_j$ only influenced by the changing weather conditions. Then, Eq. (6) follows from Eq. (5) for the distributions computed over the obtained dataset:

$$\forall j \in J:\ Q_j^{HP} = E_j \cdot B_j \sim Q_j^{Gas} \tag{6}$$

Obtaining statistically significant similarities between $E_j \cdot B_j$ and $Q_j^{Gas}$ for at least one $B_j$ for every considered year $j$ provides evidence of the assumptions' validity (see Section 4 and Table 1).

***Table 1.*** *Results (first section), changes of the mean value of the input data for different years (center section) and exogenous variables (last section) for the four evaluated years*

| Year | $B_j^*$ with minimal JSD | KS stat. | Range[1] of $B_j^*$ $p<0.05$ | Relative change in mean annual consumption[2] natural gas | heat pump el. | Average daily outdoor temperature in winter[3] Mean | Min |
|---|---|---|---|---|---|---|---|
| 2019 | 3.3 | 3.1 | [3.0, 3.1] | −6.0% | 3.3% | 3.6 °C | −4.7 °C |
| 2020 | 3.5 | 3.4 | [3.2, 3.8] | 0.0% | 0.0% | 4.1 °C | −2.1 °C |
| 2021 | 3.0 | 2.9 | [2.8, 3.1] | 3.6% | 16.3% | 2.4 °C | −10.3 °C |
| 2022 | 3.1 | 2.8 | [2.8, 2.8] | −14.2% | −1.7% | 4.1 °C | −5.2 °C |

[1] $p$-value of KS test
[2] Compared to the reference year 2020
[3] The winter is defined as the period from 01.01. until 20.02. and 01.11. until 31.12. of each year

### *3.1. Find optimal B*

To get the value of $B$ where both distributions of Eq. (6) overlap best, we propose two different methods. First, we select the value for $B_j$ minimizing the JSD between the two distributions $E_j \cdot B_j$ and $Q_j^{Gas}$. In the second step, we use the KS test statistic to find the value of $B_j$ where both distributions are most similar. In both methods, we calculate the statistics over a predefined and sufficiently large discrete set of values for $B_j$, within which we assume the minimum value is located. If the minimum value is not included in this predefined discrete set, we would notice that the minimum value is either at the lower or upper boundary of the set, indicating that the range for exploration was chosen too narrowly.

We explicitly note that this minimization should be applied for every year $j$ individually, as the SPF can differ in different years. Suppose we add an assumption about the mean reduced heat demand $\gamma(h)$ over all buildings with a heat pump $h \in H^{HP}$, e.g., based on a heat cadaster. In this case, we can directly estimate the mean SPF over all existing heat pumps. We validate our approach by comparing our results to field trials for SPF computation.

#### *3.1.1. Find* B *minimizing the distance using the JSD*

The first approach to find the $B_j^*$ where both distributions are most similar is to compute the JSD between $E_j \cdot B_j$ and $Q_j^{Gas}$ for a high resolved, discrete set of values of $B_j$, and take the $B_j$ minimizing the JSD, which we denote as $B_j^*$. The JSD is a symmetric measure for the similarity of two distributions (see Lin (1991) for theoretical details)

$$\mathrm{JSD}(P \parallel R) = \frac{1}{2} D_{\mathrm{KL}}(P \parallel R) + \frac{1}{2} D_{\mathrm{KL}}(R \parallel P) \tag{7}$$

where $P$ and $R$ are two distributions and $D_{\mathrm{KL}}$ denotes the Kullback–Leibler Divergence. It takes smaller values the more similar the distributions are and reaches 0 if and only if both distributions are equal. The upper bound of the JSD is given by $\log_x 2$ with $x$ being the basis of the logarithm used in the computation of the JSD (Lin, 1991; Nielsen, 2020). In this paper, we use $x = e$. To smooth the histogram of both distributions, we apply a kernel density estimation using an exponential kernel. We define the optimal value $B_j^*$ for $B_j$ based on the JSD as

$$B_j^* := \arg\min_{B_j} \mathrm{JSD}\left(E_j \cdot B_j, Q_j^{Gas}\right). \tag{8}$$

#### *3.1.2. Find* B *minimizing the KS test statistic*

A disadvantage of finding $B$ by minimizing the JSD between $E_j \cdot B_j$ and $Q_j^{Gas}$ is that this method does not provide a range of values for $B$ where both distributions are similar in a statistically significant way.

https://doi.org/10.1017/dce.2024.44 Published online by Cambridge University Press

Therefore, we introduce another method for finding an optimal $B$ based on a goodness-of-fit test. These tests can be divided into two groups, namely parametric and nonparametric tests. As we do not want to make an assumption about the underlying distributions of $E_j \cdot B_j$ and $Q_j^{Gas}$, we employ an nonparametric goodness-of-fit test. According to Arnold and Emerson (2011), the most popular one in this group is the KS test. Compared to other nonparametric goodness-of-fit tests, the KS test is also robust to outliers (Arnold and Emerson, 2011). In our case, we utilize the two-sample variant of the KS test. This means, that an empirical distribution, here $E_j \cdot B_j$, is compared against another empirical one, in this case $Q_j^{Gas}$. Thus, the tested null hypothesis has the form

$$H_0 : F_{E_j \cdot B_j} = F_{Q_j^{Gas}} \tag{9}$$

where $F_X$ denotes the empirical cumulative distribution function of a variable $X$. The KS test statistic $d_K$ is defined as

$$d_K\left(F_{E_j \cdot B_j}, F_{Q_j^{Gas}}\right) = \sup_{t \in \mathbb{R}} \left\| F_{E_j \cdot B_j}(t) - F_{Q_j^{Gas}}(t) \right\| \tag{10}$$

following the literature (Lehmann and Romano, 2022, Chapter 16.3). Consequently, we define the optimal value $B_j^*$ for $B_j$ based on the KS test as

$$B_j^* := \arg\min_{B_j} d_K\left(F_{E_j \cdot B_j}, F_{Q_j^{Gas}}\right). \tag{11}$$

The advantage of using the KS test as an instance of a statistical test compared to the JSD is that we get a range of values for $B_j$ in which the two distributions overlap significantly at a given significance level. Therefore, we compute the KS test statistic $d_K\left(F_{E_j \cdot B_j}, F_{Q_j^{Gas}}\right)$ from Equation 10 for every value of $B_j$ in the predefined range of values to query and check, if the null hypothesis holds by checking

$$\frac{m \cdot n}{d} \cdot d_K\left(F_{E_j \cdot B_j}, F_{Q_j^{Gas}}\right) \leq j_\alpha \tag{12}$$

denotes the greatest common divisor for $m$ and $n$, and $j_\alpha$ denotes the critical value for rejecting the hypothesis given a significance level $\alpha$. A detailed discussion on the computation of the $j_\alpha$ can be found for example in Comment 38 in Hollander et al. (2014) or in Hodges (1958). The most important aspect from Equation (12) is that the distribution of $j_\alpha$ is independent from the empirical distributions $E_j \cdot B_j$ and $Q_j^{Gas}$, and the number of samples $m$ and $n$.

### *3.2. Dataset*

Our analysis is based on about 1400 smart meter time series of gas meters with a daily resolution and 73 time series of electricity meters exclusively metering a heat pump with an hourly resolution from 2019 until 2021. The data were measured in Haßfurt, a small city in Southern Germany, including most of the residential single- and two-family buildings there. For all of the obtained time series, we know the location of the meter and surplus information about the building, like its volume. The annual consumption of natural gas or electricity is computed by summing over the time series values measured by the smart meters. The histogram of the total gas and heat pump electricity consumption in 2020 is depicted in the upper row of Figure 1. A correlation analysis between the building volume as main parameter of the building size and the annual natural gas or heat pump electricity consumption results in $r(\text{gas cons.}, \text{volume}) = 0.58$ and $r(\text{heat pump el. cons.}, \text{volume}) = 0.66$, respectively, for 2020 (see the lower row in Figure 1). Such low correlations highlight the impact of the other parameters as listed above. Moreover, related works show similar correlations between the building size and its energy consumption for heating. For example, Kolter and Ferreira (2011) report a correlation of $r = 0.611$ between the annual heating energy consumption and the living area. The reasons for the high variation of annual heat consumption have been discussed multiple times in the literature (Lim and Zhai, 2017), which are mainly caused by the parameters as discussed in Section 3.

https://doi.org/10.1017/dce.2024.44 Published online by Cambridge University Press

Additionally, we conducted a survey among all customers of the utility company, which we validated against the building volume of all buildings ($t(5824) = -0.36$, $p = 0.72$, $d = -0.01$). The response rate was around 17%. For the buildings, that are already heated by a heat pump, we additionally asked about the type of the heat pump. Thus, we are able to identify the share of installed heat pump types, showing that 76% are air-source ($n = 70$) and 24% ($n = 22$) are ground-source heat pumps.

The metered gas consumption of our dataset might include the consumption of additional gas-powered appliances like gas stoves used for cooking. For this analysis, we treat the complete metered gas consumption as caused by the heating system, neglecting especially cooking demands, as cooking is only responsible for 6.3% of the final energy consumption in households in Europe in 2022 (European Statistical Office, 2022) and only a small fraction of around 6% of all households in Germany use gas stoves in 2018 according to the Federal Statistical Office of Germany (2018). Moreover, related works like Kolter and Ferreira (2011) also ignore possible additional gas consumers in the residential buildings.

#### *3.2.1. Dataset validation*

In order to validate the dataset, we first analyze the individual heat pump as gas consumption profiles on a daily level. Secondly, we present the criterion for a profile to be included in our analysis. Our dataset shows a correlation between the mean daily heat pump electricity consumption and the mean daily outside temperature of $r = -0.95$ for 2021 and $r = -0.92$ for 2019 and 2020. When comparing the mean daily gas consumption with the mean daily outside temperature, we still get a correlation of $r = -0.92$ for 2020 and 2021 and $r = -0.93$ for 2019. These results go along with related publications. For example, Wamburu et al. (2022) show a correlation of $r = -0.90$ between the daily gas consumption and the temperature time series. In terms of data quality, we ensured the robustness of our dataset by only including time series from smart meters that had measurements for at least 98% of all hours. This stringent criterion helps to mitigate the impact of data gaps and ensures the reliability of our analysis. Any missing data points were handled using linear interpolation, which is a common method for filling gaps in time series data. This approach ensures that the continuity and trends in the data are preserved, minimizing any potential biases due to missing values. After this preprocessing and validation step, there remain between 62 and 66 smart meter time series on the heat pump electricity consumption per year and between 992 and 1493 time series on the gas consumption.

#### *3.2.2. Daily analysis*

Using the data from 2019, the comparison between the average daily heat consumption from all gas furnaces and the average electricity consumption of all installed heat pumps shows a strong correlation of $r = 0.99$. For 2020 and 2021, the correlation remained high at $r = 0.98$ and $r = 0.97$, respectively. These consistently high correlations across the years are beneficial for simulating the replacement of gas furnaces with heat pumps in the future. They indicate that the daily energy consumption patterns for both systems are very similar and that outdoor temperature has a comparable effect on both systems.

### ***3.3. Determination of the minimal amount of required data points***

The presented methodology requires a sufficiently large amount of data points for the annual gas and heat pump electricity consumption. Our dataset shows that gas furnaces are the predominant heating system at least in Europe, but the number of heat pump installations is still at a low level. Thus, we need to define the minimal amount of required data points on the annual heat pump electricity consumption are required to apply our methodology properly.

Therefore, we sample a random subsample of size $m$ from the original universe, i.e., the set of all annual heat pump electricity consumption values in our obtained dataset, for every considered year and apply our methodology as presented in Section 3.1.2 on the dataset with the reduced size. We gradually reduce the subsample size $m$, starting at $m = 60$ until reaching a lower limit of $m = 5$. Since the sampling process is random, we need to repeat this sampling process multiple times depending on $m$, the size of the

https://doi.org/10.1017/dce.2024.44 Published online by Cambridge University Press

subsample. For an arbitrary data point $x$ from our universe, the probability of being selected in one subsample of size $m$ is $\frac{m}{n}$, with $n$ being the size of the universe, as this sampling process can be described using the geometric distribution.

For this analysis, we aim to sample enough subsamples so that we can expect to have an arbitrary data point $x$ at least in $r = 3$ subsamples. Thus, the number of required subsamples for a given subsample size $m$ follows the expected value of the negative binomial distribution with $r = 3$ and a success probability of $p = \frac{m}{n}$. Finally, the number of required subsamples $N_m$ is $N_m = \frac{r}{p} = \frac{r \cdot n}{m}$. Additionally, we always sample at least  subsamples, even though $N_m$ could be lower.

## 4. Results

Depending on the year of the analysis, different optimal values of $B_j^* = \frac{\mathrm{SPF}_j}{1-\gamma}$ are found by computing the JSD and the KS test for a discrete set of values for $B_j$, ranging from 1.5 to 4.0 in steps of 0.1. Table 1 shows the values of $B_j$ that achieve a minimal value of the JSD or the KS test for different years, along with relevant metrics of the input data helping to explain the results, i.e., the mean outdoor temperature in the winter (from January 1st to February 20th and from November 1st to December 31st) and the relative change in the average annual consumption of the used input data. The concrete optimal values $B_j^*$ range from 3.0 to 3.5 using the JSD and 2.8 to 3.4 using the KS test. A plot of the JSD and KS test statistic values for all evaluated values for $B_j$ is depicted in Figure 2.

For all evaluated years, the fit of both distributions using the optimal value $B_j^*$ based on the KS test statistic falls within a significant range using a significance level of 5% (see the second and third column in Table 1). This similarity of both distributions is visualized for all evaluated years in a histogram in Figure 3 using the optimal values based on the KS test. From 2019 until 2021, the analysis reveals an almost linear relationship between the optimal value $B_j^*$ and the mean outdoor temperature in the winter months. This

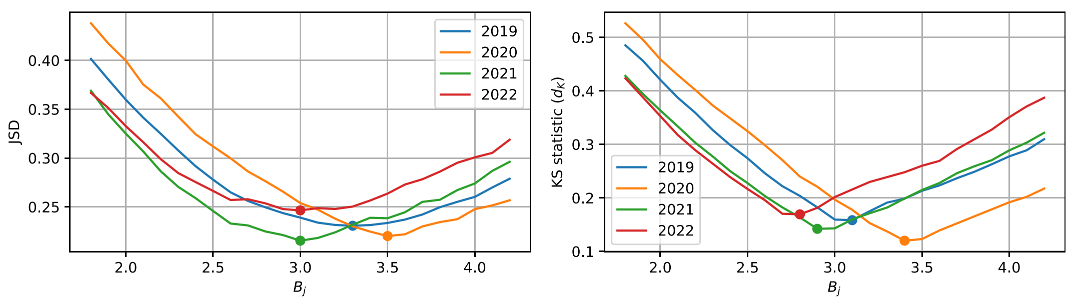


***Figure 2.*** *Plot of the JSD (left) and the KS test statistic (right) for different values of $B_j$ for the different years. The point of each line marks the minimal value.*

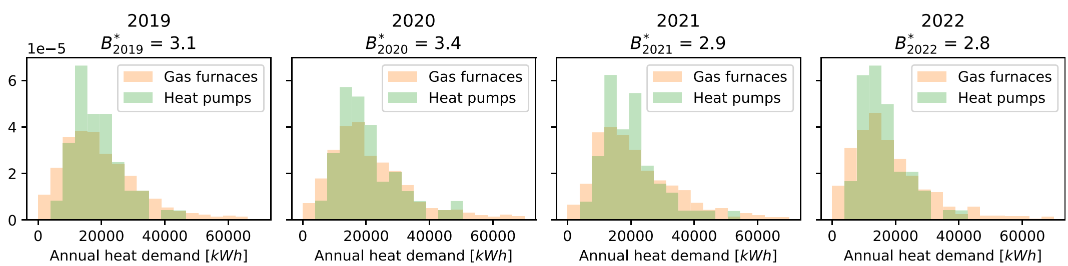


***Figure 3.*** *Histograms of the annual heat demand computed from the gas consumption data (orange) and using the described methodology for the heat pumps (green) annually grouped. The used values for $B_j^*$ from Equation (4) are the result minimizing the KS test statistics.*

https://doi.org/10.1017/dce.2024.44 Published online by Cambridge University Press

relationship is plausible since our data contains lots of air-source heat pumps that show a reduced efficiency during colder winters (Swardt and Meyer, 2001; Vocale et al., 2014). However, for 2022, the projected value of $B_j^*$ does not align with the outdoor temperature. This is caused by the fact that on the one hand the natural gas consumption of residential buildings dropped notably during the European gas crisis in 2022 (Ruhnau et al., 2023). But on the other hand, we see no strong reduction in the average annual heat pump electricity consumptions between 2020 and 2022. Thus, this one-sided reduction obscures the effects of climatic conditions, leading to a reduced value of $B_j^*$.

The town council created a heat cadaster to evaluate the current state of building insulation within the municipal area. Therefore, they randomly selected more than 1000 residential buildings and used a questionnaire to assess both the heat demand and insulation level of each building. Based on this cadaster, we can assume that buildings heated with a heat pump have a heat requirement of around $\overline{\gamma} = 10.5\%$ lower than buildings currently heated with a gas furnace. Utilizing this information, we can compute the average SPF over all buildings to 3.0 for 2019 and 3.1 or 2.7 for 2020 or 2021 by multiplying $B_j^*$ with $1 - \overline{\gamma}$. These results fall within a reasonable and expected range, as field trials conducted in Germany report mean SPF values ranging from 2.3 to 3.1 for air-source heat pumps and from 2.9 to 4.0 for ground-source heat pumps (Huchtemann and Müller, 2012; Miara et al., 2014).

### *4.1. Annual demand*

The estimation of the SPF is crucial for the estimation of the future electricity consumption of heat pumps replacing existing gas furnaces. Let us assume that all existing gas furnaces in the residential buildings were replaced by heat pumps, and the insulation level of all gas-heated buildings was equivalent to those with a heat pump. Then, they would have consumed 9.0 GWh in 2019 and 8.0 or 9.5 GWh in 2020 or 2021 in total using the estimated values for the SPF or $B_j^*$, respectively, based on the KS test. This corresponds to an increase in electricity consumption over all gas-heated buildings between 140% in 2020 and 166% in 2021. Only in 2022, the summed electricity consumption of all retrofitted heat pumps would be at a lower level of 5.9 GWh. The ratio of the heat pump to building electricity consumption is visualized in Figure 4 for all evaluated years. We might note that the building electricity consumption does not contain the electricity consumption caused by the heat pump. Thus, a ratio smaller than one can appear. The median of this household-to-heat-pump electricity demand ratio is higher for the future systems replacing gas furnaces (center and right plot) than for the already installed heat pumps (left plot). Additionally, this ratio shows a higher variance for future systems than for existing ones. The different variance magnitudes are already present in the initial dataset. Assuming an insulation retrofit, the median of the heat pump to residential electricity consumption ratio remains almost stable over the different evaluated years ranging

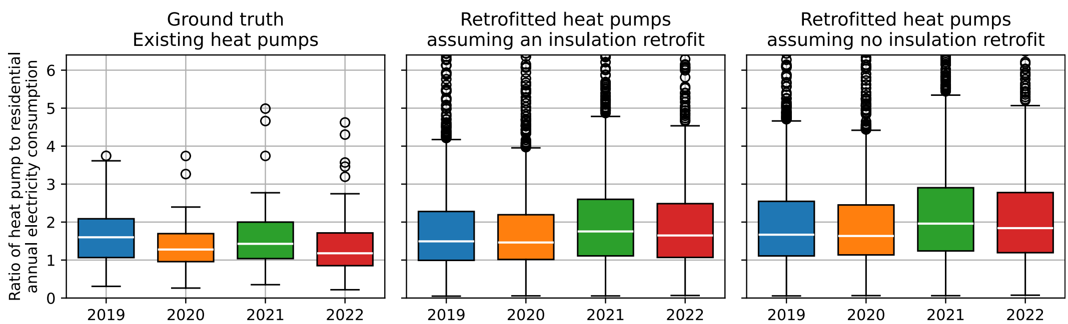


***Figure 4.*** *Boxplots of the share of heat pump electricity demand divided by the residential electricity demand (excluding the heat pump) annually grouped. The left plot visualizes the share for existing heat pumps. The center and the right plot visualizes the distribution of the share using the estimation of the SPF with (center) and without (right) insulation retrofit. For the sake of clarity, the y-axis is clipped at a level of 6.4 and buildings with a electricity consumption lower than 500 kWh/a are excluded.*

https://doi.org/10.1017/dce.2024.44 Published online by Cambridge University Press

between 1.46 for 2020 and 1.75 for 2021. Assuming no insulation retrofit, the median of the ratio increases to a range between 1.63 for 2020 and 1.96 for 2021.

### 4.2. Determination of the minimal amount of required data points

To determine the minimal number of heat pump measurements per year, we apply the method as described in Section 3.3. We accept a deviation of 0.1 below or above the value for $B_j^*$ computed over the entire data set to be acceptable as a result of applying the methodology with each subsample with reduced size (see the gray band in Figure 5). When repeating the sampling and computing procedure $N_m$ times given the subsample size $m$, we can report that for a subsample size of 55, at least 90% of the subsamples result in a value for $B_j^*$ within the given limits for the years 2019, 2021, and 2022. Only for the year 2020, even with a subsample size of $m = 60$, the estimates for $B_j^*$ are only in 80% of the cases inside of the acceptable range. A visualization of the distributions of the resulting estimates for $B_j^*$ can be found in Figure 5.

If we want to have at least 50% of the subsamples to produce estimates of $B_j^*$ within the defined band, a subsample size between $m = 30$ and $m = 45$ is required. For lower sample sizes, the percentage of subsamples that result in estimates of $B_j^*$ within the acceptable range drops notably.

## 5. Discussion

First of all, the presented results reveal some general insights for the modeling of future retrofitted heat pumps. Our heat consumption analysis reveals a notable variation in the annual heat consumption among buildings of the same volume (see Figure 1). Consequently, any investigation of the annual electricity consumption of a potentially retrofitted heat pump should be based on the specific heat consumption of each individual building, when such data is available, as demonstrated by Lechowicz et al. (2023), rather than relying solely on building parameters like its volume. Additionally, the strong daily correlation between gas consumption and the electricity consumption of heat pumps suggests that gas consumption

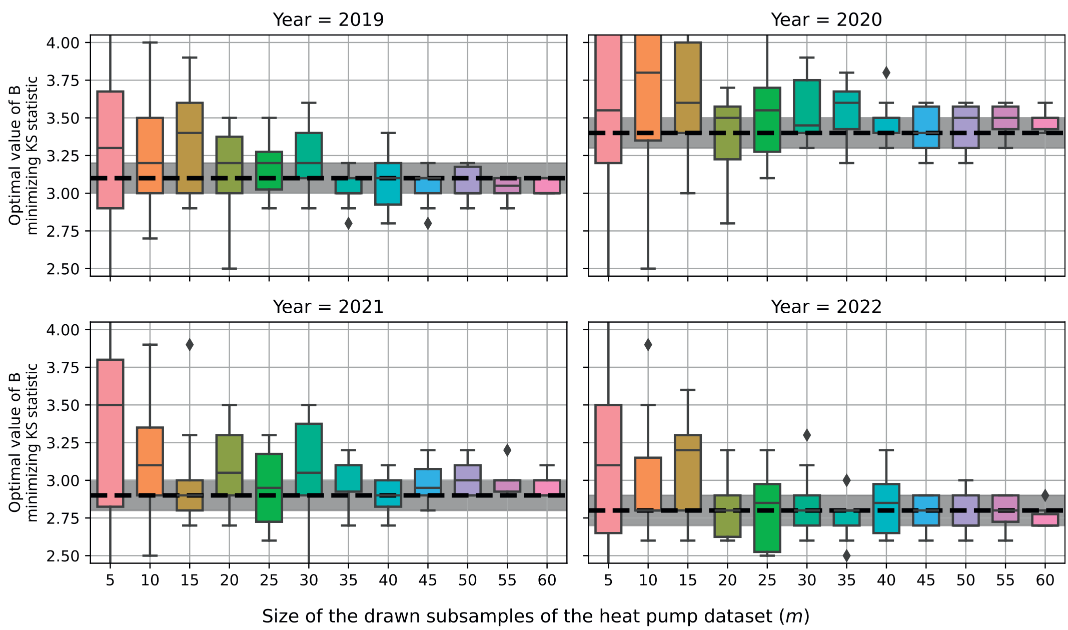


***Figure 5.*** *Boxplots of estimated value for $B_j^*$ minimizing the KS statistic over all subsamples grouped by their size m. The result of $B_j^*$ over the complete data set minimizing the KS statistic is visualized as the dashed line. The gray band around the dashed line is the $B_j^* +/- 0.1$.*

profiles can be effectively translated into equivalent heat pump profiles in the case of modeling a heat pump retrofit replacing a gas furnace.

Regarding the SPF estimates, our results for different years lie in a valid range when comparing our results to field trials in Germany. However, the notable annual deviations of the average SPF argue against using constant values for the SPF. Furthermore, the notable annual deviations that appeared within the same city highlight the fact that analysis considering the effect of a heat pump retrofit on the electrical distribution system should always consider multiple years with different weather realizations.

The results of this paper can be used to predict the electricity consumption of future retrofitted heat pumps. On the one hand, such a prediction is helpful for analyses on the building level like Lechowicz et al. (2023); Wamburu et al. (2022) or for developing city-scale digital twins like those presented in Bayer and Pruckner (2023a). On the other hand, utility companies can use this methodology to estimate the electricity demand produced by retrofitted heat pumps. This knowledge can help to optimally dimension future distribution systems and thus help to save resources on our planet as companies can avoid wrong energy system dimensioning. Generally, this paper shows that smart meter data on gas and electricity consumption becomes very valuable when predicting the electricity demand of retrofitted heat pumps.

### 5.1. Differences of $B^*_j$ using KS test and JSD

Using the two different methods to estimate the optimal $B_j^*$ results in slightly different values, where $B_j^*$ is generally higher using the JSD than using the KS test in our case. The results using the KS test seem to be more plausible, as all of them are significant (using a significance level of 5%) meaning that we decide in favor of the null hypothesis of cumulative distribution equality (Equation 9). In contrast, the estimates for $B_j^*$ using the JSD are not inside the range of significant values using the KS test for two out of four years (2019 and 2022). The reason for this deviation of the results based on the KS test and the JSD is that the first relies on the cumulative distribution functions of the given distributions. In contrast, the latter relies on the smoothed histograms of the distributions, paying more attention to outliers and sampling deviations.

### 5.2. Justification of the assumptions and advantages of the KS test

This paper requires four major assumptions, as stated in Section 3. We discuss these assumptions and their implications in this section.

*Independence of the heat consumption and the heating system type*

At first, we assume that the heat consumption of a building is independent of its heating system type. Thus, the heat consumption does not change if the heating system is changed during a retrofit. This assumption seems valid as existing literature dealing with influencing factors of building heat consumption does not report notable correlations of the heating system on building heat consumption. From a practical point of view, the heating system type influences the demand side, not the consumption side. Based on the obtained dataset, we note that buildings heated with an heat pump are currently better insulated than those heated by gas furnaces, so we introduce a factor in Equation (2) to account for this. Nevertheless, the insulation standard is not dependent on the heating system type itself.

*Independence of heating system type and heat consumption influencing parameters*

Second, we assume that the impact of multiple parameters influencing the annual heat consumption of a residential building, like its size or the local weather conditions, is independent of the heating system type. This is a specialization of the first assumption: If the heating system type is independent of the heat consumption, it must also be independent of the factors and parameters that mainly influence the heat consumption.

Without this one, the Equations (4) and (6) would not be necessarily valid. As a result, we find a $B_j^*$ for every year for which we can decide in favor of the null hypothesis from Equation 9 on the equality of both

https://doi.org/10.1017/dce.2024.44 Published online by Cambridge University Press

distributions. This provides empirical evidence for this assumption and strengthens the understanding of the underlying factors that impact the heat consumption of residential buildings. If we did not find a value for the parameter $B_j$ for which we obtain a significant similarity between both distributions from Equation (6), this assumption would be questionable, meaning that the choice of the heating system would impact the building's annual heat consumption.

Moreover, this fact highlights an advantage of the KS test compared to the JSD. If we could not find a value of $B_j^*$ for which we get significant results, we would have evidence that at least one of the assumptions, as defined in Section 3, would be violated. Using the JSD, we always find a minimal value over a discrete set of tested values for $B_j$ without getting evidence of the significance of the result.

*Independence of the SPF of the actual heat consumption*

Third, we assume that the SPF and a potential insulation enhancement are independent of the actual heat consumption. Without this assumption, Equation (6) might be invalid. The first part of the assumption, i.e., the independence of the SPF from the heat consumption, seems justified by the fact that the size of every newly installed heat pump is dimensions based on the (assumed or actual) heat consumption of a building. For example, in Germany, this process is normed by the German Institute for Standardization (2017). Nevertheless, an accidentally wrong sizing of a heat pump can reduce the seasonal performance notably (Madonna and Bazzocchi, 2013). The second part of the assumption, i.e., the independence of the insulation enhancement of the heat consumption, is based on the idea that the insulation increase during a complete retrofit of a building should reach the best possible outcome at the moment and not depend on the current heat consumption that is highly driven, e.g., by the current user behavior.

*Equal impact of the weather on the performance over all heat pumps*

Fourth, we assume that the influence of the weather on the SPF affects all heat pumps equally. This assumption is grounded on several considerations. All heat pumps operate on the same thermodynamic principles, which means that changes in outdoor temperature similarly affect their efficiency. Especially the decrease in the COP that happens when the temperature differential between the heat source and sink increases is a universal phenomenon, further indicating that weather affects the SPF of all heat pumps in a similar way. Since we are considering heat pumps within a single city, we can assume that the weather conditions are generally consistent across the entire city.

### 5.3. Generalization of the results and limitations

Our methodology is transferable to any other town or city in the world, requiring the presence of a sufficient number of annual natural gas and heat pump consumption measurements. The building classes of the buildings currently heated with heat pumps or gas furnaces should be the same to ensure the assumptions as defined in Section 3 hold. The presented methodology can be used to compute the actual heat pump performance for retrofitted heat pumps, which is especially required to estimate the heat pump electricity consumption for future retrofitted heat pumps. Nevertheless, the results also show theoretical limitations of the methodology. If there is a strong change in the consumption only affecting either the gas furnace or the heat pump usage, but not both, the interpretation of $B_j^*/(1-\overline{y})$ as an exact estimation of the average SPF of the already installed heat pumps might be misleading. This situation happened for example in 2022 due to the energy crisis in Europe, where the users reduced their natural gas consumption in the mean by 14.2% compared to 2020, but the occupants in already heat pump-heated buildings did not notably change their heat consumption behavior (see Table 1). Thus, the estimation of $B_j^*$ is 2.8 for 2022, while the estimation is 3.4 for 2020 using the results based on the KS test. Following the weather conditions of these two years, we would expect a result for $B_j^*$ in 2022 that is in a similar region as in 2020.

However, it is essential to note that assumptions about the electricity consumption of future retrofitted heat pumps are not affected, assuming that there is no change in the heating energy consumption of the buildings and all other parameters remain constant in the future. This means that for the year 2022, with

https://doi.org/10.1017/dce.2024.44 Published online by Cambridge University Press

the gas crisis, we assume that buildings currently having gas heating systems would not have reduced the energy used for heating if they already had installed a heat pump. This fact is also visible in the boxplots of the heat pump to residential electricity consumption ratio in Figure 4, where the distributions of the ratios in 2022 (red boxes in the center and right plot) follow those from the previous years (blue, orange and green boxes in the center and right plot). Moreover, when trying to estimate the actual SPF of the already installed heat pumps out of an unpaired natural gas and heat pump electricity consumption dataset, we require knowledge of different average insulation levels between the buildings with gas furnaces on the one hand and heat pumps on the other hand.

The existing literature shows that the SPF differs notably between air- and ground-source heat pumps (Wu et al., 2009; Huchtemann and Müller, 2012). It is, therefore, advisable to use the method presented separately for air- and ground-source heat pumps. Nevertheless, we do not consider these two heat pump types separately in this paper as we lack a sufficient amount of labeled data. From the 92 heat pumps with known types based on our conducted survey (see Section 3.2), we can only attribute 15 labels per group to separately metered heat pump profiles. As presented in Section 4.2, 15 samples per heat pump type are not sufficient data to use the presented methodology properly.

## 6. Conclusion and outlook

The SPF is an important metric to estimate the annual electricity consumption of retrofitted heat pumps replacing existing gas furnaces. This work presents a novel method to estimate the SPF based on an unpaired dataset of energy consumption data of buildings within the same building class and located within the same city, where one group is formed by the buildings currently heated by heat pumps and the other group heated by gas furnaces. The SPF can be computed by minimizing either the JSD of these two unpaired distributions or the KS test statistic. We apply our method to an exemplary, recent dataset and validate the results with results from field trials. The resulting values using the KS test statistic for estimating the average SPF or $B$, i.e., the SPF divided by the insulation difference between the buildings with heat pump and those with gas furnace, are all statistically significant. The large differences in the results of the estimated values for SPF or $B$ for the various years under consideration emphasize the initial assumption that different weather conditions have a massive impact on the performance of heat pumps. Additionally, our analysis shows that shifts in consumption patterns as caused by the European gas crisis in 2022 only affecting the heating patterns in gas-heating buildings reduce the interpretability of the resulting SPF estimates. Moreover, we estimate the annual electricity consumption of future retrofitted heat pumps that impact the local distribution system.

As discussed in the introduction, heat pumps will continuously replace gas furnaces in the future. Whereas it is effortful to get measurements of the SPF of existing heat pump installations in a city, this work helps researchers and utility companies to estimate the average SPF comparably easily based on a unpaired dataset of gas and heat pump electricity consumption that is fast to obtain by reading the smart meter values. Based on the SPF estimation, the additional electricity consumption caused by heat pumps can be estimated for recently passed years and projected into the future, assuming that the weather conditions seen in the years with data available will happen in the future again. Our results highlight that using one average estimate for the SPF will not adequately reflect the strong year-to-year variability in heat pump performance that our analysis has shown. Consequently, relying on an average year would fail to capture the full diversity of conditions, potentially leading to inaccurate energy consumption estimates and less reliable simulations. Thus, it is essential to consider the annual variations in SPF for more precise and comprehensive modeling.

Therefore, the data-driven estimation of the SPF and electricity consumption of retrofitted heat pumps can serve as a foundation for analyzing the potential carbon emission reductions when replacing gas furnaces with heat pumps. For example, such an analysis as presented in Bayer and Pruckner (2024) is based on such a data-driven SPF estimation. Another technical question that remains open is how the average SPF will change with the ongoing development of smart grids. More sophisticated control strategies for heat pumps, such as those aiming to maximize photovoltaic (PV) self-consumption as presented by Yousefi et al. (2021), or those trying to minimize the overall heat demand as showcased by Bayer and Pruckner (2022), might change the SPF of an individual heat pump, as emphasized by Pospíšil et al. (2019). Another aspect to consider is categorizing heat pumps based on their energy source, which can also affect the SPF. As we do not know the heat pump type for most of the obtained electricity profiles, we plan to develop a detection mechanism based on pattern recognition to identify the different heat pump types based on their hourly electricity consumption profile.

**Acknowledgments.** We would like to thank our research partner Stadtwerk Hassfurt GmbH for providing the data set.

**Data availability statement.** The data are supplied by our research partner Stadtwerk Hassfurt GmbH. The raw data cannot be published as we do not have the necessary authorization to do so, and the data contains personal information.

**Author contribution.** Conceptualization: DB. Methodology: DB. Data curation: DB. Data visualisation: DB. Writing original draft: DB. Writing review: MP. Supervision: MP. Funding acquisition: MP. All authors approved the final submitted draft.

**Funding statement.** This paper is an outcome of the research project *DigiSWM* (DIK0298/02) founded by the Bavarian State Ministry of Economic Affairs, Regional Development and Energy.

**Competing interest.** The authors declare no competing interests.

**Ethical standard.** The research meets all ethical guidelines, including adherence to the legal requirements of the study country.

## References

**Alpízar-Castillo J**, **Ramírez-Elizondo L and Bauer P** (2023) The effect of non-coordinated heating electrification alternatives on a low-voltage distribution network with high PV penetration. In *2023 IEEE 17th International Conference on Compatibility, Power Electronics and Power Engineering (CPE-POWERENG)*. IEEE.

**Arnold TB and Emerson JW** (2011) Nonparametric goodness-of-fit tests for discrete null distributions. *The R Journal 3*(2), 34.

**Bayer D and Pruckner M** (2022) Enhancing the performance of Multi-Agent reinforcement learning for controlling HVAC systems. In *2022 IEEE Conference on Technologies for Sustainability (SusTech)*, virtual, USA. IEEE, pp. 187–194.

**Bayer D and Pruckner M** (2023a) A digital twin of a local energy system based on real smart meter data. *Energy Informatics 6*(1).

**Bayer DR and Pruckner M** (2023b) Modeling of annual and daily electricity demand of retrofitted heat pumps based on gas smart meter data. In *Proceedings of the 10th ACM International Conference on Systems for Energy-Efficient Buildings, Cities, and Transportation*. ACM, pp. 248–251.

**Bayer DR and Pruckner M** (2024) Data-driven heat pump retrofit analysis in residential buildings: Carbon emission reductions and economic viability. *Applied Energy 373*, 123823.

**Capper G and Scott A** (1982) The economics of house heating: Further findings. *Energy Economics 4*(2), 134–138.

**Carroll P**, **Chesser M and Lyons P** (2020) Air source heat pumps field studies: A systematic literature review. *Renewable and Sustainable Energy Reviews 134*, 110275.

**Damianakis N**, **Mouli GRC**, **Bauer P and Yu Y** (2023) Assessing the grid impact of electric vehicles, heat pumps & PV generation in Dutch LV distribution grids. *Applied Energy 352*, 121878.

**Delzendeh E**, **Wu S**, **Lee A and Zhou Y** (2017) The impact of occupants' behaviours on building energy analysis: a research review. *Renewable and Sustainable Energy Reviews 80*, 1061–1071.

**European Parliament** (2023) Press release: 'MEPs back plans for a climate neutral building sector by 2050'. Released on 2023 March 14th.

**European Statistical Office** (2022) Energy Consumption in Households by Type of End-Use. Available at https://ec.europa.eu/eurostat/statistics-explained/index.php?title=Energy_consumption_in_households#Energy_products_used_in_the_residential_sector (accessed 10 July 2024)

**Federal Ministry for Economic Affairs and Climate Action (BMWK)** (2022) Climate action in figures.

**Federal Statistical Office of Germany** (2018) Equipment of private households with household and other appliances *(in German)*. Available at https://www.destatis.de/DE/Themen/Gesellschaft-Umwelt/Einkommen-Konsum-Lebensbedingungen/Ausstattung-Gebrauchsgueter/Tabellen/a-evs-haushaltsgeraete-d.html (accessed 10 July 2024).

**German Institute for Standardization** (2017) DIN EN 12831-1:2017-09.

**Gleeson C P and Lowe R** (2013) Meta-analysis of european heat pump field trial efficiencies. *Energy and Buildings 66*, 637–647.

**Hailu G** (2021) 8—energy systems in buildings. In Borge-Diez D and Rosales-Asensio E (eds), *Energy Services Fundamentals and Financing*, Energy Services and Management. Academic Press, pp. 181–209.

**Hodges JL** (1958) The significance probability of the smirnov two-sample test. *Arkiv für Matematik 3*(5), 469–486.

**Hollander M**, **Wolfe DA and Chicken E.** (2014) *Nonparametric Statistical Methods*. Wiley.

**Huchtemann K and Müller D** (2012) Evaluation of a field test with retrofit heat pumps. *Building and Environment 53*, 100–106.

**Jenkins D**, **Tucker R and Rawlings R.** (2009) Modelling the carbon-saving performance of domestic ground-source heat pumps. *Energy and Buildings 41*(6), 587–595.

**Kelly N and Cockroft J** (2011) Analysis of retrofit air source heat pump performance: Results from detailed simulations and comparison to field trial data. *Energy and Buildings 43*(1), 239–245.

**Ko Y** (2013) Urban form and residential energy use. *Journal of Planning Literature 28*(4), 327–351.

**Kolter J and Ferreira J** (2011) A large-scale study on predicting and contextualizing building energy usage. *Proceedings of the AAAI Conference on Artificial Intelligence 25*(1), 1349–1356.

**Lechowicz A**, **Bashir N**, **Wamburu J**, **Hajiesmaili M, and Sheno P** (2023) Equitable network-aware decarbonization of residentialheating at city scale. In *The 14th ACM International Conference on Future Energy Systems (e-Energy '23)*. ACM.

**Lehmann E and Romano JP** (2022) *Testing Statistical Hypotheses*. Springer International Publishing.

**Lim H and Zhai ZJ** (2017) Review on stochastic modeling methods for building stock energy prediction. *Building Simulation 10* (5), 607–624.

**Lin J** (1991) Divergence measures based on the Shannon entropy. *IEEE Transactions on Information Theory 37*(1), 145–151.

**Lumbreras M**, **Garay-Martinez R**, **Arregi B**, **Martin-Escudero K**, **Diarce G**, **Raud M and Hagu I** (2022) Data driven model for heat load prediction in buildings connected to district heating by using smart heat meters. *Energy 239*, 122318.

**Madonna F and Bazzocchi F** (2013) Annual performances of reversible air-to-water heat pumps in small residential buildings. *Energy and Buildings 65*, 299–309.

**Miara M**, **Russ C**, **Günther D**, **Kramer T and Henning H** (2014) Efficiency of heat pumps in real operating conditions. *Rehva Journal 2014*, 7–12.

**Nielsen F** (2020) On a generalization of the Jensen–Shannon divergence and the Jensen–Shannon centroid. *Entropy 22*(2), 221.

**Nouvel R**, **Cotrado M and Pietruschka D** (2015) European mapping of seasonal performances of air-source and geothermal heat pumps for residential applications. In *Proceedings of International Conference CISBAT 2015 Future Buildings and Districts Sustainability from Nano to Urban Scale*, number CONF. LESO-PB, EPFL, pp. 543–548.

**Pan D**, **Chan M**, **Deng S and Lin Z** (2012) The effects of external wall insulation thickness on annual cooling and heating energy uses under different climates. *Applied Energy 97*, 313–318. Energy Solutions for a Sustainable World—Proceedings of the Third International Conference on Applied Energy, May 16–18, 2011—Perugia, Italy.

**Petrović SN and Karlsson KB** (2016) Residential heat pumps in the future danish energy system. *Energy 114*, 787–797.

**Pospšil J**, **Špiláček M and Charvát P** (2019) Seasonal COP of an air-to-water heat pump when using predictive control preferring power production from renewable sources in the czech republic. *Energies 12*(17), 3236.

**Priarone A**, **Silenzi F and Fossa M** (2020) Modelling heat pumps with variable eer and cop in energyplus: A case study applied to ground source and heat recovery heat pump systems. *Energies 13*(4), 794.

**Rosenow J**, **Gibb D**, **Nowak T and Lowes R** (2022) Heating up the global heat pump market. *Nature Energy 7*(10), 901–904.

**Rossi di Schio E**, **Ballerini V**, **Dongellini M and Valdiserri P** (2021) Defrosting of air-source heat pumps: Effect of real temperature data on seasonal energy performance for different locations in italy. *Applied Sciences 11*(17), 8003.

**Ruhnau O**, **Stiewe C**, **Muessel J and Hirth L** (2023) Natural gas savings in Germany during the 2022 energy crisis. *Nature Energy 8*(6), 621–628.

**Ruschenburg J**, **Ćutić T and Herkel S.** (2014) Validation of a black-box heat pump simulation model by means of field test results from five installations. *Energy and Buildings 84*, 506–515.

**Sardianou E** (2008) Estimating space heating determinants: An analysis of greek households. *Energy and Buildings 40*(6), 1084–1093.

**Shin J-H and Cho Y-H** (2021) Machine-learning-based coefficient of performance prediction model for heat pump systems. *Applied Sciences 12*(1), 362.

**Swardt CAD and Meyer JP** (2001) A performance comparison between an air-source and a ground-source reversible heat pump. *International Journal of Energy Research 25*(10), 899–910.

**Vocale P**, **Morini GL and Spiga M** (2014) Influence of outdoor air conditions on the air source heat pumps performance. *Energy Procedia 45*, 653–662.

**Walker IS**, **Less BD and Casquero-Modrego N** (2022) Carbon and energy cost impacts of electrification of space heating with heat pumps in the US. *Energy and Buildings 259*, 111910.

**Wamburu J**, **Bashir N**, **Irwin D and Shenoy P** (2022) Data-driven decarbonization of residential heating systems. In *Proceedings of the 9th ACM International Conference on Systems for Energy-Efficient Buildings, Cities, and Transportation*. ACM.

**Watson S**, **Crawley J**, **Lomas K and Buswell R** (2023) Predicting future GB heat pump electricity demand. *Energy and Buildings 286*, 112917.

**Weigert A**, **Hopf K**, **Günther SA and Staake T** (2022) Heat pump inspections result in large energy savings when a pre-selection of households is performed: a promising use case of smart meter data. *Energy Policy 169*, 113156.

**Wu R** *et al.* (2009) Energy efficiency technologies–air source heat pump vs. ground source heat pump. *Journal of Sustainable Development 2*(2), 14–23.

**Yousefi M**, **Hajizadeh A**, **Soltani MN and Hredzak B** (2021) Predictive home energy management system with photovoltaic array, heat pump, and plug-in electric vehicle. *IEEE Transactions on Industrial Informatics 17*(1), 430–440.

**Cite this article:** Bayer DR and Pruckner M (2024). Estimating the seasonal performance and electricity consumption of retrofitted heat pumps. *Data-Centric Engineering*, 5, e39. doi:10.1017/dce.2024.44