15TH IEA HEAT PUMP CONFERENCE
MAY 26–29, 2026 | HOFBURG
VIENNA 2026
DECARBONISATION THROUGH INNOVATION

The influence of frosting on typical day selection for the evaluation of heat pumps

Florian Will[a*], Matthias Mersch[a], Stephan Göbel[a], Christian Vering[a], Dirk Müller[a]

[a]*Institute for Energy Efficient Buildings and Indoor Climate, E.ON Energy Research Centre, RWTH Aachen University, Mathieustraße 10, 52074 Aachen, Germany*

Abstract

This study investigates the influence of frosting on the selection of typical days for heat pump performance testing. The objective is to reduce experimental time by using a typical-day method, where annual performance is extrapolated from measurements on a few typical days selected via time-series clustering of weather data. We analyze how the weighting of weather characteristics, ambient temperature, relative humidity, and solar irradiation, in the clustering algorithm impacts the accuracy of predicting the Seasonal Coefficient of Performance (SCOP).
Simulations of a building energy system with heat pump were conducted for the three German locations Potsdam, Garmisch and Sylt using a dynamic simulation model that includes a frosting and defrosting algorithm. The results demonstrate that ambient temperature is the dominant factor to accurately approximating the SCOP. For the building energy system model used in this study, high weights on the ambient temperature and low weights on the relative humidity and solar radiation yield SCOP approximation errors below 5% across the three locations. Furthermore, using four typical days proved to be an effective compromise, significantly reducing testing time while maintaining sufficient accuracy.

Keywords: Experimental measurements; Frosting; Hardware in the loop; Heat pump; Time series clustering; Typical days

1. Introduction

While dynamic modelling of heat pumps has improved significantly over the last years, exhaustive experimental testing remains an integral part in the evaluation of heat pumps, the development of new systems and the energy labelling of heat pumps. Energy labelling for heat pumps is done with the EN 14825 [1]. Experimental data is the foundation of model development, calibration, and validation. Furthermore, even the most advanced models are only an approximation of the real machines. Experimental testing is essential to investigate effects that cannot yet be captured by models. However, experiments are time-consuming and, thus, expensive, underlining the need to minimize testing time.

One key challenge for heat pump behaviour prediction with numerical models is the frosting of the evaporator. Frosting occurs when heat is extracted from cold moist air in winter, causing ice crystals to form on the evaporator. This frosting results in a reduced heat transfer in the evaporator and causes the heat pump performance to deteriorate significantly. Defrost cycles are required to remove the ice and restore the evaporator performance, but they are energy intensive. An accurate representation of frosting and defrost cycles is important for an accurate estimation of the heat pumps' performance over the year, i.e., the calculation of the seasonal coefficient of performance (SCOP).

While efforts have been made by researchers to model frosting, such as Zhu et al. [2] who developed frosting maps based on ambient temperature and relative humidity to indicate what degree of frosting will occur in general, and Popovac et al. [3] who conducted CFD simulations, frosting models generally rely on experimental data to calibrate them to a specific use case, such as the evaporator of a specific heat pump model. For example, Klingebiel et al. [4] used 30 experimental data points at different temperatures and relative

[*] Corresponding author. Tel.: +49-241-804-9840
E-mail address: florian.will@eonerc.rwth-aachen.de

humidities to calibrate their evaporator frosting model, based on which they optimised defrost cycles [4], [5]. Consequently, frosting effects are particularly important for the experimental characterisation of heat pumps.

Heat pump experiments are ideally an accurate representation of real-world operating conditions while also having defined conditions, being reproducible, and having a high measurement accuracy. This can be achieved by Hardware-in-the-loop (HiL) experiments, which emulate field tests in the lab [6]. However, such experiments require climate chambers to emulate varying ambient conditions, as well as hydraulic systems that can emulate the behaviour of buildings, which results in higher costs of experimental testing compared to standard steady-state testing time [7]. However, field tests in the lab are cheaper compared to field tests at the cost of neglecting some real-world effects that, e.g., occur due to user interaction. Therefore, there is interest in further developing field test in the lab procedures.

Ideally, experiments cover a whole year or even multiple years to achieve the most accurate representation of real-world annual performance. However, such long experimental campaigns are expensive and do not fit well into fast product development cycles that require a variety of different components, system designs, and controllers to be tested quickly to evaluate their performance.

To reduce the time required for experiments, a typical-day method can be applied. The idea is to perform the experimental measurements only for a few representative days on the HiL test bench. The annual performance of the heat pump has to be extrapolated from the measurements obtained for the typical days [8]. This approach reduces the required time significantly (a few days instead of a whole year).

The typical days are determined through clustering of meteorological time series [9], but the weighting of the different parameters (e.g., temperature, solar irradiation, wind speed, humidity) in the clustering algorithm is complex. Therefore, in this study, we analyse the impact of different weighting factors on the selection of typical days and the accuracy of the heat pump performance prediction. We place a particular focus on frosting, as it has a considerable impact on the heat pump performance but is rarely considered in modelling and labelling. To achieve a more robust analysis, the analysis is conducted for a variety of locations in Germany with different climatic conditions.

2. Method

Fig. 1 gives an overview of the methods used in this study. We generate typical days from different weather data using clustering. We evaluate the influence of different weights for the weather characteristics on the quality of the clustering. The typical days are used for a prediction of the SCOP of the heat pump. The quality of the clusters and the typical days is evaluated with a comparison of the approximated SCOP with the actual SCOP. The approximated SCOP is determined by using the SCOP of a building energy system with heat pump on the typical days. The actual SCOP results of a yearly simulation of the same building energy system with heat pump. Low deviations between approximated SCOP and actual SCOP mean a high quality of the cluster.

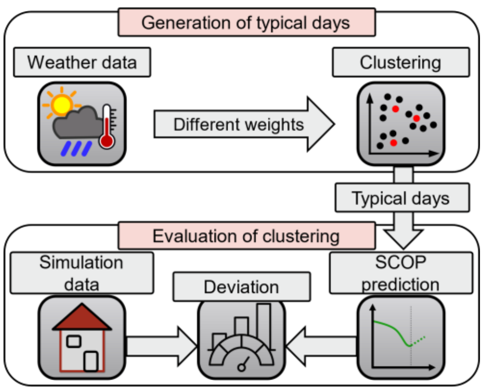

Figure 1. Method to generate typical days and evaluate the quality of the generated clusters and typical days.

2.1. Modelling of building and heat pump

The analysis is performed based on dynamic heat pump and building models implemented in Modelica, an object-oriented programming language designed to model and simulate physical systems. The model builds on the open-source BESMod library [10], a modular Modelica model library for building energy systems.

The building model is based on the high-order building model in the open-source library AixLib [11], which is described and validated in [12] and [13]. The modelled building is a typical newly built single-family home with a floor area of 158 m^2. The building is insulated according to the German energy conservation standard for new builds from 1995 (WSVO95) with a nominal heating load of 53 W/m^2 at an ambient temperature of -14 °C according to EN 12831. The building is heated by an underfloor heating system with a nominal supply temperature of 38 °C, which is connected to the heat pump via a 900 L buffer tank. Additionally, the heat pump supplies domestic hot water (DHW) via a heat exchanger where fresh tap water is heated to the required temperature of 60 °C. The daily hot water demand is assumed to be 200 L, which is supplied over 24 individual draw-off events.

The heat pump itself is modelled using the approach described in Wüllhorst et al. [14]. Two polynomials provide data for performance maps to estimate compressor electricity demand (P_{el}) as well as the supplied heat in the condenser ($\dot{Q}_{\mathrm{con}}$) based on ambient temperature (T_{a}), supply temperature (T_{sup}), and compressor speed (n_{c}). The polynomial for the compressor electricity demand is shown in Eq. (1), the polynomial for the supplied heat is shown in Eq. (2).

$$P_{\mathrm{el}} = \left(c_1 \cdot T_{\mathrm{sup}}^2 + c_2 \cdot T_{\mathrm{sup}} + c_3\right) \cdot n_{\mathrm{c}}^2 + \left(c_4 \cdot T_{\mathrm{sup}}^2 + c_5 \cdot T_{\mathrm{sup}} + c_6\right) \cdot n_{\mathrm{c}} \tag{1}$$

$$\dot{Q}_{\mathrm{con}} = \left(c_7 \cdot T_{\mathrm{a}} - T_{\mathrm{sup}} + c_8\right) \cdot n_{\mathrm{c}} \tag{2}$$

In our case, an inverter-controlled air-source heat pump using propane (R290) as refrigerant is modelled. The coefficients c_1 to c_8 are fitted by a heat pump manufacturer to measurement data. Fig. 2 shows the Coefficient of Performance (COP) of the heat pump model dependent on T_{sup} and T_{a} for $n_{\mathrm{c}} = 60$ Hz.

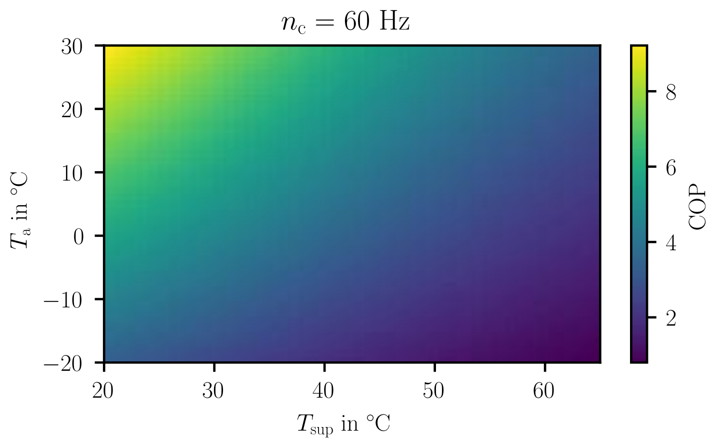

Figure 2. COP of the heat pump model dependent on T_{a} on the y-axis and T_{sup} on the x-axis for a constant n_{c} of 60 Hz.

To account for frosting, the heat pump model is enhanced by using a frosting factor that corresponds to the mass of ice accumulated on the evaporator. It affects the heat gained in the evaporator and thus the available heating capacity as well as the COP. The mass of ice is determined based on the frost growth rate, which is calculated from the ambient temperature and relative humidity using a frosting model calibrated to experimental data [5]. Defrosting is initiated regularly in defined time intervals, which represents the currently most commonly used algorithm, even though research has shown the potential of adaptive defrost controllers [5].

The heat pump controller is also modelled in Modelica. The space heating supply temperature setpoint is determined from a heating curve. A hysteresis element determines whether the heat pump is switched on or off, while another one determines whether there is DHW demand and switches to the higher DHW supply temperature setpoint if necessary. While the heat pump is running, a PI controller sets the compressor speed to control the supply temperature. At the same time, another PI controller sets the speed of the heating circuit pump to control the water mass flowrate.

The heat pump model is connected to the building model such that the dynamic operation of the building energy system can be simulated. Using weather data for different locations as well as the DHW demand profile mentioned above, the energy demand of the building is simulated. From the thermal building model, the heating water return temperature can be determined, which is an input to the heat pump controller and thus affects the heat pump operation. The heat pump model calculates the actual supply temperature, which is an input to the building simulation in the next time step. Thus, the two models are directly coupled, and the operation of the entire building energy system can be realistically modelled.

2.2. Generating representative timeseries

To reduce the experimental duration required for an accurate annual performance assessment, a typical-day methodology is employed. This approach is based on the premise that the annual performance of a heat pump system can be approximated by conducting experiments for a limited number of representative days, the results of which are then weighted and aggregated [7]. The selection of these representative, or "typical," days is achieved through the application of time-series clustering algorithms to annual meteorological data.

The clustering process in this study is performed using the Python package tsam in conjunction with the Gurobi solver. The hierarchical clustering algorithm is selected, as it has been demonstrated by Kotzur et al. [8] to yield results comparable to k-medoids clustering while offering superior computational efficiency. A key advantage of both hierarchical and k-medoids clustering is that the resulting cluster centroids are actual days from the original dataset (medoids), ensuring that the typical days represent physically plausible weather sequences.

The input data for the clustering consists of Test Reference Year (TRY) weather data for various locations. For the clustering, the weather data is divided into 365 time periods of 24 hours with an hourly resolution. These time periods always start at hour 0 on the respective day. The goal of the clustering is to group the 365 time periods into clusters.

The relevant weather characteristics considered are dry-bulb temperature (T_a), relative humidity (φ), and global horizontal solar irradiation ($\dot{q}_\mathrm{s}$). To investigate the influence of these parameters on the clustering outcome and the subsequent accuracy of the annual performance prediction, the weighting factors assigned to each characteristic are systematically varied. Each weight (w_{T_a}, w_{φ}, $w_{\dot{q}_\mathrm{s}}$) is varied within the range [0, 1] in steps of 0.05, under the constraint that the sum of the weights is normalized to one.

The clustering algorithm assigns each day of the year to a specific cluster and returns the corresponding cluster medoid, which serves as the typical day for that cluster. The annual performance is then approximated by calculating a weighted average of the Seasonal Coefficient of Performance (SCOP) measured on each typical day. For a clustering result with *k* clusters, the approximation is calculated as shown in Eq. (3):

$$\mathrm{SCOP_{app}} = \sum_i^k \mathrm{SCOP}_{\mathrm{day},i} \cdot \frac{n_i}{365} \tag{3}$$

where n_i is the number of days assigned to cluster i and $\mathrm{SCOP}_{\mathrm{day},i}$ is the SCOP for the typical day of cluster i. Since a full-year simulation is performed for each system configuration, the approximated $\mathrm{SCOP_{app}}$ can be directly compared to the simulated annual $\mathrm{SCOP_{year}}$. The accuracy of the typical-day method is evaluated by the relative deviation ($\varepsilon_{\mathrm{SCOP}}$) defined as in Eq. (4):

$$\varepsilon_{\mathrm{SCOP}} = \frac{(\mathrm{SCOP_{year}} - \mathrm{SCOP_{app}})}{\mathrm{SCOP_{year}}} \tag{4}$$

3. Results

Annual simulations of the coupled building-heat pump system are conducted for three distinct locations in Germany: Potsdam (continental climate), Garmisch-Partenkirchen (alpine climate), and Sylt (maritime climate). These locations are selected to represent a range of climatic conditions. The TRY of Sylt has mild T_a between 2 °C and 17 °C with high relative humidities. Potsdam has higher T_a in summer and lower T_a in winter compared to Sylt because of the more continental climate. Garmisch has the lowest T_a of the locations and high relative humidities.

For all locations, the heat pump model includes the frosting and defrosting dynamics described in Section 2.1. An additional simulation for Potsdam is performed with a simplified heat pump model that excludes frosting effects, to isolate the impact of frosting on the typical-day selection process. The results compare the accuracy of the SCOP approximation for different weighting combinations of the weather characteristics and for different numbers of clusters: 2, 4, 8, and 12.

3.1. Sensitivity analysis

A linear sensitivity analysis is conducted to establish a baseline for the relative importance of the weather characteristics (T_a, φ and $\dot{q}_s$) on the annual SCOP. For each characteristic, three full-year simulations are performed per location: one with the original TRY data, one with the characteristic multiplied by (1+κ), and one with it multiplied by (1-κ), where κ is a variation coefficient. The resulting SCOP values are used to perform a linear regression. The slope of this regression is taken as the weight for the respective characteristic. These weights are normalized to sum to one.

The calculated normalized weights are presented in Table 1. The results indicate that the weights are consistent across the different locations. The ambient temperature carries the highest weight, reflecting its dominant influence on building heat load and heat pump efficiency. The weights for relative humidity and solar irradiation are significantly lower. Notably, for the Potsdam simulation without frosting, the weight for relative humidity approaches zero, which is expected since φ has no effect on the heat pump performance in the absence of a frosting model.

Table 1: Sensitivities (w) for the locations Potsdam, Sylt and Garmisch-Partenkirchen. The last line displays the deviation of the SCOP of the yearly simulation to the $SCOP_{app}$.

Parameter	Potsdam (no Frosting)	Potsdam	Sylt	Garmisch-Partenkirchen
w_{T_a}	0.9957	0.956	0.9444	0.9537
w_{φ}	0	0.0405	0.0522	0.0404
$w_{\dot{q}_s}$	0.0043	0.0035	0.0034	0.0059
ε_{SCOP} (%)	3.5	4.5	-3.4	7.5

Fig. 3 shows an example calendar of generated data clusters for Potsdam, with each cluster represented by a distinct color. The calendar-plot shows seasonal patterns: a winter cluster (blue) is prominent in January, February, November, and December, while a summer cluster (red) dominates the months from June to September. The remaining two clusters represent the transitional seasons, with one cluster (yellow) appearing mostly in spring (April-May) and the other (orange) occurring in both spring and fall. The cluster medoids, which represent the typical day for each group, are marked with an 'x'. These medoids are the typical days used to calculate the $SCOP_{app}$.

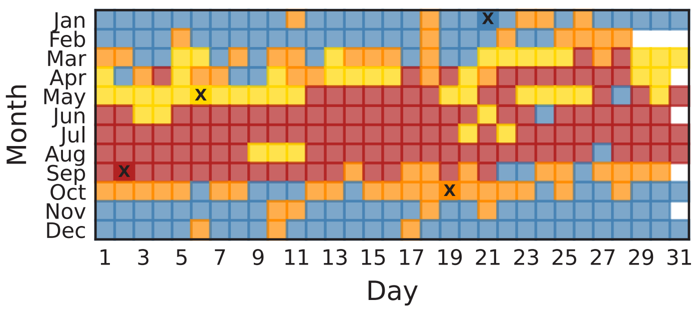

Figure 3. The year divided into four clusters for the location Potsdam.
The cluster medoid is marked with an x.

Table 1 shows the $\varepsilon_{\mathrm{SCOP}}$ for four clusters for all locations. We see low deviations of 3.5% for Potsdam (no frosting) and -3.4% for Sylt. For Garmisch we have the largest deviation of 7.5%.
This sensitivity analysis, while informative, requires a full system model and multiple annual simulations, which is computationally intensive and requires high modeling effort. A key objective of this paper is to determine whether robust ranges of weights exist that yield accurate results across different locations, thereby potentially reducing the need for such a preliminary analysis.

3.2. Influence of frosting on the typical days

The impact of including the frosting model on the typical-day method is first analysed for Potsdam using four clusters, a number considered a reasonable compromise for experimental testing duration. Fig. 4 illustrates the deviation $\varepsilon_{\mathrm{SCOP}}$ as a function of $w_{T_{\mathrm{a}}}$, w_{φ} and $w_{\dot{q}_{\mathrm{s}}}$. $w_{\dot{q}_{\mathrm{s}}}$ is constant along the white dotted lines and decreases from bottom-left ($w_{\dot{q}_{\mathrm{s}}} = 1$) to top-right in steps of 0.2. Areas where $\varepsilon_{\mathrm{SCOP}}$ is higher than 10% are highlighted in red, areas with $\varepsilon_{\mathrm{SCOP}}$ lower than 10% are highlighted in pink.

For both system configurations, high weights assigned to solar radiation (corresponds to low $w_{T_{\mathrm{a}}}$ and w_{φ}) generally lead to higher errors, which is consistent with its low sensitivity weight. Accurate approximations are achievable for a wide range of weights when $w_{\dot{q}_{\mathrm{s}}}$ is below approximately 0.4.

For the system without frosting (Fig. 4, left), high accuracy is maintained across a broad plateau of $w_{T_{\mathrm{a}}}$(0.4 to 1.0) and w_{φ}(0 to 1). The system with frosting (Fig. 4, right) shows similar tendencies, with $w_{T_{\mathrm{a}}} < 0.4$ leading to high approximation errors of the SCOP. The region with $w_{T_{\mathrm{a}}} > 0.4$ shows more variability for the system with frosting model compared to the one without, with several weight combinations resulting in errors exceeding 5%. The most consistent and accurate results for the frosting model are achieved with a high weight on temperature ($w_{T_{\mathrm{a}}}$ between 0.6 and 1) and a low weight on solar radiation ($w_{\dot{q}_{\mathrm{s}}} < 0.2$).

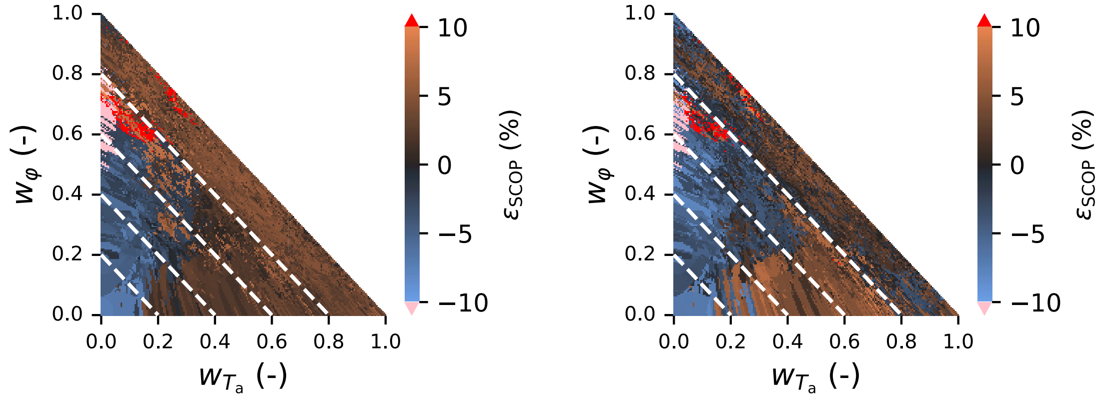

Figure 4. Heatmaps of $\varepsilon_{\mathrm{SCOP}}$ for the location Potsdam dependent on $w_{T_{\mathrm{a}}}$, w_{φ} and $w_{\dot{q}_{\mathrm{s}}}$. On the white dotted lines, $w_{\dot{q}_{\mathrm{s}}}$ is constant. On the left, the results for the heat pump model without frosting are shown, the results on the right include frosting.

This demonstrates that the performance of the typical-day method is sensitive to the system's physical characteristics. Since frosting is a real-world phenomenon that significantly impacts performance, it is essential to account for relative humidity in the clustering process when the goal is to test and evaluate actual heat pumps. In the next section, we look at the influence of the location on the clustering.

3.3. Influence of the location on the typical days

The results for the locations Garmisch and Sylt, using the heat pump model with frosting and four clusters, are shown in Fig. 5. The general pattern observed in Potsdam is confirmed: high weights on solar radiation, which corresponds to low $w_{T_{\mathrm{a}}}$ and low w_{φ} lead to significant approximation errors.

Location-specific differences are apparent. For instance, at low $w_{T_{\mathrm{a}}}$ (<0.2), Garmisch shows a tendency towards overestimation ($\varepsilon_{\mathrm{SCOP}}$ < -10%), while Sylt shows underestimation ($\varepsilon_{\mathrm{SCOP}}$> +10%) of the $\mathrm{SCOP}_{\mathrm{app}}$. This underscores the influence of local climate on the optimal weight selection.

For all three locations, weight combinations with $w_{T_\mathrm{a}} > 0.4$ yield to mostly low approximation errors ($|\varepsilon_\mathrm{SCOP}| < 5\%$). This suggests that for air-source heat pumps susceptible to frosting, a weighting scheme that prioritizes ambient temperature is likely to produce reliable typical days across various climatic conditions in Germany. The weights for solar radiation and relative humidity should be set based on the requirements of the study to conduct: if the relative humidity should be representative of the yearly weather data (e.g. for experiments that study frosting), a high w_φ should be set. Accordingly, for high representativity of the solar radiation, a high weight w_{T_a} should be set.

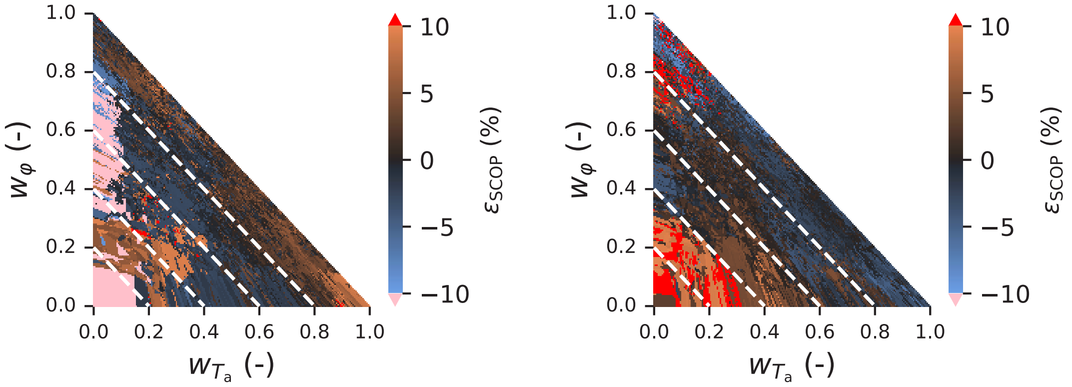

Figure 5. Heatmaps of $\varepsilon_\mathrm{SCOP}$ for the locations Garmisch-Partenkirchen (left) and Sylt (right) dependent on w_{T_a}, w_φ and $w_{\dot{q}_\mathrm{s}}$. On the white dotted lines, $w_{\dot{q}_\mathrm{s}}$ is constant. The results are generated with the heat pump model with frosting.

3.4. *Influence of the number of clusters on the typical days*

The analysis is extended to evaluate the impact of the number of selected typical days on the accuracy of the annual performance approximation. Simulations are conducted for all locations using the heat pump model with frosting, considering cluster numbers of 2, 4, 8, and 12. The results are summarized in Fig. 6, which depicts the distribution of the approximation error ($\varepsilon_\mathrm{SCOP}$), across all weighting combinations for each cluster size.

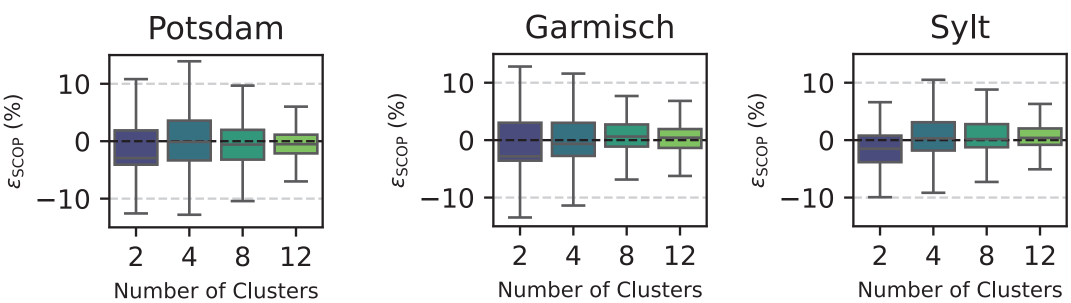

Figure 6. Approximation error of the SCOP for the locations Potsdam, Garmisch-Partenkirchen and Sylt for different number of clusters. Included are the results for all combinations of weights as boxplots.

The results exhibit a consistent trend across all locations. The median of the $\varepsilon_\mathrm{SCOP}$ distribution is negatively biased for two clusters, indicating a systematic underestimation of the annual SCOP. As the number of clusters increases, the median of the distribution converges towards zero. Furthermore, the variability of the results decreases. The interquartile range (the range between the 25th and 75th percentiles) narrows with an increasing number of clusters. For 4 clusters, the interquartile range predominantly lies within $\varepsilon_\mathrm{SCOP} = \pm 5\%$, while for 12 clusters, it is contained within $\varepsilon_\mathrm{SCOP} = \pm 3\%$.

This demonstrates that a higher number of clusters enhances the representativeness of the typical days, leading to a more accurate and less biased approximation of the annual performance. However, a greater number of clusters directly translates into a longer required experimental or simulative duration. Therefore, a compromise must be struck between the desired accuracy and the practical constraints of testing time. The optimal number of clusters strongly depends on the boundary conditions of the study to conduct. For simulative studies, a higher number, like 12 clusters, might be optimal, because it still reduces the computational time significantly but provides high accuracies. For experimental studies, lower number of clusters might be optimal, because experimental time is expensive. For instance, 4 clusters appear to offer a reasonable balance, providing a robust approximation with errors generally within an acceptable ±5% range for a wide range of weighting factors. This result is in line with studies by Huchtemann et al. [15].

4. Conclusions

While dynamic simulations are essential tools for evaluating heat pump performance, they cannot capture all physical effects. Therefore, experimental testing remains critical for validating performance and developing heat pumps. To make extensive experimental campaigns feasible within practical development cycles, methods to reduce testing time are necessary. This study addresses this need by investigating the use of typical days.

This study demonstrates that the selection of typical days for heat pump testing is highly sensitive to the weighting of weather characteristics in the clustering process. Ambient temperature is the dominant factor for accurate prediction of the annual performance (SCOP) for frosting-sensitive heat pumps. For weights of ambient temperature between 0.4 and 1 (the sum of all weights is 1), we find approximations errors of the SCOP below 5% across the locations Potsdam, Sylt and Garmisch. The weighting for relative humidity and solar radiation should be chosen based on the requirements of the study to conduct.

Using four typical days was shown to be an effective compromise, reducing experimental time while maintaining sufficient accuracy. For higher accuracy of the SCOP approximation, a higher number of clusters is needed. This provides a practical basis for designing efficient test campaigns that account for real-world phenomena like frosting.

There are limitations in this study. The findings are based on a specific building and heat pump model with a fixed control strategy. Different system configurations or control algorithms may alter the optimal weights. We also only include a fixed time period of 24 hours. Considering different time periods may result in different results.

Future work should include experimental validation on a test bench and extend the analysis to other building types and heat pump systems. In future work, the representative time periods will be used for the development and evaluation of heat pump controllers. For this use-case, we might need to change the method to determine the typical days. One approach could be to generate demanding sequences of events (defrosting, start mode, DHW storage loading, …) to evaluate the controller on.

Acknowledgements

We gratefully acknowledge the financial support by the German Federal Ministry for Economic Affairs and Energy (BMWE), promotional references 03EN1096B.

This study was realized together with the project partners Dresden University of Technology with Martin Knorr, who supported the building energy system modelling, Viessmann Holding International GmbH and Kermi GmbH.

Parts of the results were generated in the bachelor thesis of Leon Jungkurth.

References

[1] 'EN 14825:2018, Air conditioners, liquid chilling packages and heat pumps, with electrically driven compressors, for space heating and cooling - Testing and rating at part load conditions and calculation of seasonal performance', CEN/TC 113, Bruxelles, Belgium, Dec. 2018. [Online]. Available: https://standards.cen.eu/dyn/www/f?p=204:110:0::::FSP_PROJECT,FSP_ORG_ID:63007,6095&cs=1F46F6ECBABB4348B281EABF58FFAA34B

[2] J. H. Zhu, Y. Y. Sun, W. Wang, S. M. Deng, Y. J. Ge, and L. T. Li, 'Developing a new frosting map to guide defrosting control for air-source heat pump units', *Appl. Therm. Eng.*, vol. 90, pp. 782–791, Nov. 2015, doi: 10.1016/j.applthermaleng.2015.06.076.

[3] M. Popovac, J. Emhofer, and Ch. Reichl, 'Frosting in a heat pump evaporator part B: Numerical analysis', *Appl. Therm. Eng.*, vol. 199, p. 117488, Nov. 2021, doi: 10.1016/j.applthermaleng.2021.117488.

[4] J. Klingebiel, M. Salamon, P. Bogdanov, V. Venzik, C. Vering, and D. Müller, 'Towards maximum efficiency in heat pump operation: Self-optimizing defrost initiation control using deep reinforcement learning', *Energy Build.*, vol. 297, p. 113397, Oct. 2023, doi: 10.1016/j.enbuild.2023.113397.

[5] J. Klingebiel *et al.*, 'Optimal defrost initiation for air-source heat pumps: Evaluating the improvement potential of common defrosting controllers', *Energy*, vol. 324, p. 135871, Jun. 2025, doi: 10.1016/j.energy.2025.135871.

[6] C. Vering *et al.*, 'Towards a defossilized building sector with field tests in the lab: Review, development, and evaluation', *Appl. Energy*, vol. 365, p. 123225, Jul. 2024, doi: 10.1016/j.apenergy.2024.123225.

[7] S. Goebel, F. Will, P. Stoffel, C. Vering, and D. Müller, 'Optimizing energy efficiency in buildings: Experimental findings on the role of heat pump control interfaces in model predictive control', *Energy Convers. Manag. X*, p. 101306, Oct. 2025, doi: 10.1016/j.ecmx.2025.101306.

[8] P. Mehrfeld *et al.*, 'Dynamic evaluations of heat pump and micro combined heat and power systems using the hardware-in-the-loop approach', *J. Build. Eng.*, vol. 28, p. 101032, Mar. 2020, doi: 10.1016/j.jobe.2019.101032.

[9] L. Kotzur, P. Markewitz, M. Robinius, and D. Stolten, 'Impact of different time series aggregation methods on optimal energy system design', *Renew. Energy*, vol. 117, pp. 474–487, Mar. 2018, doi: 10.1016/j.renene.2017.10.017.

[10] F. Wüllhorst, L. Maier, D. Jansen, L. Kühn, D. Hering, and D. Müller, 'BESMod - A Modelica Library providing Building Energy System Modules', presented at the American Modelica Conference 2022, Dallas, October 26-28, Feb. 2023, pp. 9–18. doi: 10.3384/ECP211869.

[11] L. Maier *et al.*, 'AixLib: an open-source Modelica library for compound building energy systems from component to district level with automated quality management', *J. Build. Perform. Simul.*, vol. 17, no. 2, pp. 196–219, Mar. 2024, doi: 10.1080/19401493.2023.2250521.

[12] A. Constantin, R. Streblow, and D. Müller, 'The Modelica HouseModels Library: Presentation and Evaluation of a Room Model with the ASHRAE Standard 140', in *Proceedings of the 10th International Modelica Conference, March 10-12, 2014, Lund, Sweden*, in Linköping Electronic Conference Proceedings. Linköping University Electronic Press, Jan. 2014, pp. 293–299. doi: 10.3384/ECP14096293.

[13] K. Xanthopoulou, P. Mehrfeld, P. Kangowski, F. Lanzerath, and D. Müller, 'Validation of a building model as part of the AixLib Modelica library for dynamic plant and building performance simulations', *Energy Build.*, vol. 250, p. 111248, Nov. 2021, doi: 10.1016/j.enbuild.2021.111248.

[14] F. Wüllhorst, D. Jansen, P. Mehrfeld, and D. Müller, 'A Modular Model of Reversible Heat Pumps and Chillers for System Applications', in *Proceedings of 14th Modelica Conference 2021, Linköping, Sweden, September 20-24, 2021*, Sep. 2021, pp. 561–568. doi: 10.3384/ecp21181561.

[15] K. Huchtemann, H. Engel, P. Mehrfeld, M. Nürenberg, and D. Müller, 'Testing method for evaluation of realistic seasonal performance of heat pump heating systems', in *CLIMA 2016 proceedings of the 12th REHVA World Congress*, Aalborg, Denmark, May 2016, p. 21. [Online]. Available: http://vbn.aau.dk/files/233716892/paper_257.pdf