Energy Conversion and Management: X 25 (2025) 100860

Contents lists available at ScienceDirect

Energy Conversion and Management: X

journal homepage: www.sciencedirect.com/journal/energy-conversion-and-management-x

K-means and agglomerative clustering for source-load mapping in distributed district heating planning

Amir Shahcheraghian [a], Adrian Ilinca [a,*], Nelson Sommerfeldt [b]

[a] Department of Mechanical Engineering, École de Technologie Supérieure, Montreal, QC, Canada
[b] KTH Royal Institute of Technology, Stockholm, Sweden

A R T I C L E I N F O

Keywords:
Distributed district heating
Clustering
Energy performance certificates
Economic viability
Urban energy planning
Heat source allocation
Sustainability
Marginal cost of heat
Policy implications

A B S T R A C T

This study introduces a high-resolution, data-driven approach for optimizing district heating networks using source-load mapping, focusing on Stockholm as a case study. The methodology integrates detailed building energy performance data (2014–2022) with geographic data from the Swedish Survey Agency, employing advanced clustering techniques such as K-means Clustering, Agglomerative Clustering, DBSCAN, Spectral Clustering, and Gaussian Mixture Model (GMM) Clustering to identify optimal locations for distributed heat sources, including data centers, supermarkets, and water bodies. Quantitative results show that these environmentally friendly sources could supply 54 % of Stockholm's total annual heat demand of 7.7 TWh/year, equating to 4.2 TWh from residual heat sources. Data centers contribute 0.48 TWh, water bodies provide 3.4 TWh, and supermarkets contribute 0.3 TWh annually. Economic analysis further reveals that 98 % of residual heat sources are economically viable, with marginal costs of heat (MCOH) for data centers, supermarkets, and water bodies estimated at 12.7 EUR/MWh, 16.0 EUR/MWh, and 20.0 EUR/MWh, respectively—well below the Open District Heating (ODH) market price of 22.0 EUR/MWh. The policy implications of these findings are profound. Policymakers can leverage this methodology to identify economically viable heat sources, enabling the creation of regulations that incentivize the integration of distributed heat sources into existing district heating networks. This can lead to reduced energy costs, enhanced sustainability, and more resilient energy systems. Practically, urban planners and energy utilities can use clustering insights to optimize the placement of new infrastructure, such as data centers, ensuring they are strategically located in high-demand zones. Furthermore, the study's methodology can be replicated in other urban contexts, offering cities worldwide a scalable tool for improving the efficiency and sustainability of their heating networks. These findings support the transition to low-carbon energy solutions and provide actionable recommendations for the long-term development of urban energy systems.

Introduction

District heating networks offer significant opportunities to rapidly transform energy supply in buildings and cities. Changing the fuel sources for heating plants makes it possible to increase the proportion of renewable energy or decarbonize the heating supply with minimal intervention from building owners. Sweden is a global leader in this area, with approximately 500 cities or communities utilizing district heating networks that supply heat to 55 % of the building area [1]. Most heat is generated by combustion fueled by residuals from the forestry industry and municipal solid waste, with peaks often covered by fossil fuels [2].

Traditionally, heating networks are centered around large, centralized plants. However, Stockholm's Open District Heating market illustrates how distributed heat sources—such as supermarkets, ice rinks, and data centers—can recover and contribute heat that would otherwise be wasted [3]. This system allows facilities with substantial cooling demands to sell their excess heat to the network, creating a unique prosumer market for heat similar to the prosumer model in the electricity sector. While heavy industry has long contributed to heating networks through specialized agreements, this marketplace opens participation to smaller, distributed stakeholders.

High spatial and temporal resolution data are essential to effectively integrate distributed heat sources into district heating networks. The

* Corresponding author.
E-mail address: adrian.ilinca@etsmtl.ca (A. Ilinca).

https://doi.org/10.1016/j.ecmx.2024.100860
Received 14 October 2024; Received in revised form 24 December 2024; Accepted 26 December 2024
Available online 27 December 2024

growing availability of high-resolution geographic information systems (GIS) data has led to numerous district heating planning studies. On the demand side, large cities such as London [4], Berlin [5], and Helsinki [6] have mapped out their building energy demands. Studies in Switzerland [7] and China [8] have mapped the potential of ground-source heat pumps.

So far, high-resolution mapping has primarily focused on buildings and heat demand. However, source-load mapping, which identifies potential heat sources, is often conducted at a minimum resolution of 1 km, which is insufficient for urban environments. Renewable heat sources, which are seasonally variable and typically produce low temperatures, require enhancements via heat pumps. The electrification of heat production demands hourly time resolution to account for fluctuations in supply profiles and temperature levels, both of which impact system efficiency. Additionally, electricity prices, which have become more volatile since the withdrawal of Russian natural gas from the European market, vary by the hour. High temporal resolution is also crucial for accurately modeling storage systems, especially in such a volatile pricing environment.

This study uses high-resolution source-load mapping to develop a novel data-driven approach to district heating network planning during the pre-feasibility stages. Unlike existing methods, this approach integrates detailed spatial data for both distributed heat demand and supply sources, identifying optimal locations for new heat sources and potential areas for heat recovery. This leads to more efficient and cost-effective district heating infrastructure designs.

Stockholm is used as a case study due to the availability of data, particularly the distributed heat-source dataset from Su et al. [9], on which this research is based. The study's objectives include analyzing how existing distributed resources can be optimally allocated to meet the city's heat demand, identifying the best locations for new data centers that can serve as significant heat sources, assessing the economic feasibility of utilizing heat from these distributed sources, and developing a comprehensive spatial database that includes detailed building energy performance metrics across multiple years (2014–2022) along with extensive geographical data. By addressing these issues, the study aims to provide actionable insights that enhance the sustainability and cost-effectiveness of urban heating solutions.

This study addresses several significant research gaps in the district heating field, particularly in integrating distributed heat sources into urban energy systems. First, traditional district heating studies have primarily focused on centralized heat production, overlooking the potential of distributed sources such as data centers, supermarkets, and water bodies. As a result, the significant untapped potential of residual heat from these sources has often been underutilized. Another critical gap is the reliance on coarse-resolution spatial and temporal data in previous studies, which limits the ability to capture urban heat supply and demand variability and dynamics. The lack of high-resolution data hampers the accuracy of supply–demand matching and the identification of optimal resource allocation.

Additionally, while the technical potential of distributed heat sources has been well-documented, insufficient research has explored their economic viability in the context of current market conditions. This gap leaves policymakers without the necessary insights to make informed decisions about integrating these sources into existing infrastructure. Furthermore, previous studies have not fully leveraged advanced analytical approaches like machine learning algorithms. In particular, unsupervised learning methods, including K-means, DBSCAN, and Gaussian Mixture Models, have been underutilized in analyzing the complex spatial and temporal patterns of heat supply and demand. The lack of these advanced analytical techniques has constrained the ability to handle noise, non-linear distributions, and the scalability needed in urban energy analysis.

Another overlooked aspect is the role of data centers, which, despite their potential as significant heat sources, have often been underrepresented in district heating studies due to outdated or insufficient data. The limited use of localized analysis tailored to specific urban contexts has further hindered the recognition of their potential. Moreover, many studies fail to provide methodologies that can be replicated across different urban contexts, limiting the broader applicability of their findings. This lack of replicability impedes the widespread adoption of sustainable district heating solutions. Finally, a critical gap exists between academic insights and the practical needs of urban planners and policymakers. Previous research has often failed to offer actionable recommendations, hindering the translation of academic findings into real-world applications.

By addressing these gaps, this study provides a high-resolution, economically grounded, and methodologically robust framework for integrating distributed heat sources into district heating networks, offering valuable insights for both researchers and urban planners.

Literature review and background

Blanco et al. [10] introduced a novel data-driven methodology for urban energy analysis by classifying urban areas into sixteen distinct morphological units called Urban Energy Units (UEUs). This approach utilizes open-source data and machine learning techniques, including a random forest model to address missing building data and a decision tree model for UEU classification. This data-driven method facilitates the creation of modular energy districts within cities, leading to more targeted and effective energy planning. Applied in Oldenburg, Germany, the methodology demonstrates the practical benefits of their classification system and underscores the importance of data-driven approaches in urban energy analysis for informed decision-making. Their study not only tackles the issue of missing data but also lays the groundwork for future research to validate and expand the classification framework to other urban areas. This work aligns with our studies, particularly in its application to district heating systems, highlighting the importance of data-driven approaches for informed energy planning.

Zhang et al. [11] tackle the critical issue of rising energy consumption in data centers, which has surged dramatically in recent years. The paper comprehensively reviews the latest thermal management techniques and evaluation metrics for data centers, discussing energy conservation strategies and safe operation practices. It covers critical advancements in cooling technologies, innovative energy-saving methods such as free cooling and heat recovery, and optimization techniques to enhance system efficiency. The paper emphasizes the significance of thermal evaluation metrics in maintaining equipment safety and energy efficiency. It concludes by calling for further research and strategic development in the thermal management of data centers.

Wahlroos et al. [12] explore the potential of reusing waste heat from data centers in Nordic countries, highlighting both the challenges and opportunities. The paper addresses issues such as the lack of transparency in business models between district heating and data center operators and waste heat reuse's economic and environmental implications. Their proposed 8-step change process provides a structured approach to overcoming barriers, aligning closely with our project's focus on waste heat management in data centers and district heating systems. The review emphasizes the importance of understanding energy efficiency metrics and economic considerations directly relevant to our project's objectives.

Davies et al. [13] investigate the potential for reusing waste heat from data centers in London, focusing on using heat pumps to upgrade waste heat for district heating networks. The study reveals significant energy, carbon, and cost savings potential, especially when data center waste heat is integrated into district heating systems with eligibility for Renewable Heat Incentive (RHI) payments. It aligns potential waste heat sources with the heat demands of various London districts, demonstrating substantial energy and carbon savings benefits.

Cichowicz et al. [14] compared Energy Performance Certificate (EPC) data calculation methods with actual energy consumption to evaluate energy efficiency in multi-family residential buildings. Their findings revealed a 14.5 % difference in Practical Energy Consumption (PEC) indices and a 14.7 % difference in Final Energy (FE) indices between the two methods. While both methods showed comparable performance, the reliance on user behavior in the consumption method raised concerns about its reliability across different building types. This study highlights the need for standardized procedures considering building-specific characteristics in EPC-based energy assessments.

Zhang et al. [15] explore the evolution of district heating and cooling (DHC) systems in Sweden, focusing on adapting these systems to future changes in demand profiles and renewable energy supplies. Using a generalized methodology framework that integrates future changes, various operational scenarios, and system design optimizations, the study concludes that a fifth-generation district heating and cooling (5GDHC) system is the most economically viable option in a future scenario with low-energy building stock and increased cooling demand [16]. The study also highlights the significant impact of electricity prices on the cost-efficiency of 5GDHC and ultra-low temperature district heating and cooling (ULTDHC) systems, particularly with a 50 % share of wind power in the national grid. The methodology presented can be applied to similar systems to enhance understanding of system transitions.

Huang et al. [17] investigate the dual role of data centers in district energy systems, where they act as both consumers and producers of energy. The study reviews the integration of renewable energy and the recovery and reuse of waste heat in data centers, emphasizing the importance of optimizing both upstream and downstream processes. It discusses how upstream integration involves procuring and managing renewable energy sources, such as solar and wind, for data center operations. At the same time, downstream waste heat utilization refers to managing and repurposing waste heat for district heating systems. The findings indicate that while integrated global controls for managing energy production, operation, and waste heat generation are still developing, regional climate studies are crucial for optimizing these integrations. The study also highlights the development of global energy metrics as essential for quantifying data center performance and providing a comprehensive approach to sustainable energy use within data centers.

Su et al. [9] focus on decarbonizing Stockholm's district heating sector to achieve net-zero emissions by 2040. They use an integrative GIS-based analysis to map clean, non-fossil fuel heat sources within Stockholm, achieving high-resolution mapping and addressing data availability challenges. Their results show that the city has abundant clean heat sources, including water bodies and data centers, capable of covering 100 % of the district's heating net annual energy needs. The study identifies clusters of heat sources for prioritized exploitation. It provides a method pipeline applicable to other cities transitioning to clean district heating, emphasizing the importance of strategic planning based on local heat source availability.

Ogliari et al. [18] proposed a methodology integrating clustering and neural networks to enhance day-ahead thermal load forecasting for a District Heating (DH) system in Northern Italy. The study comprehensively addressed data preprocessing, variable correlation analysis, clustering, and forecasting. Three clustering techniques—k-means, Hierarchical Agglomerative Clustering (HAC), and DBSCAN—were evaluated using the Caliński-Harabasz and Silhouette indexes, identifying three optimal clusters while excluding outliers, but no single superior method emerged. Forecasting was explored using three strategies: training a single neural network for all utilities, per cluster (based on HAC), and per substation. Results indicated that training a neural network for each substation yielded the highest accuracy. However, a cluster-based approach was recommended for larger systems to balance accuracy and computational efficiency. The authors emphasized the need for further research with expanded datasets and additional substations to validate and generalize their findings.

Lumbreras et al. [19] highlight the critical role of the building sector, consuming approximately 40 % of primary energy in the European Union, and the increasing focus on energy efficiency driven by EU directives. District Heating (DH) networks, covering 13 % of EU building heat loads, are evolving towards 4th Generation District Heating (4GDH), characterized by low-temperature heat distribution and renewable energy integration. The study identifies a gap in the literature regarding the application of unsupervised clustering techniques to heating demand data, unlike the extensive application in electricity demand analysis. The authors propose a multistep clustering framework using density-based clustering for outlier detection and k-means for identifying heating consumption patterns. The methodology reveals the interpretative challenges of clustering, as optimal classifications vary based on cluster validation indices (e.g., K = 3). While the framework offers valuable insights into demand patterns, its replicability remains limited. Future work aims to explore correlations between demand patterns and climatic or calendar variables and extend the framework's application to additional buildings within the DH network to enhance generalizability and applicability.

Methodology

The proposed method involves aggregating spatially labelled heat energy demand data with distributed heat sources, clustering these demands to map supply, and identifying optimal locations for new heat sources. Additionally, it includes calculating levelized heating costs by cluster.

Energy demand data is sourced from the Swedish Housing Authority's (Boverket) Energy Performance Certificates (EPC) database and geographically located using data from the Swedish Survey Agency (Lantmäteriet) [20,21]. This database is merged with the locations of existing data centers, supermarkets, and water bodies, providing a comprehensive, multi-dimensional perspective for practical spatial analysis [9].

The integration process synchronizes building-specific energy performance metrics with their geographical coordinates, ensuring that each data point accurately reflects both location and energy profile. Advanced clustering techniques, including k-means and agglomerative methods, are then employed to segment the region based on heat demand characteristics [22,23]. These clusters are analyzed to identify central nodes representing optimal locations for new data centers, considering their potential contributions to district heating systems.

The methodology ensures the optimal placement of new data centers within these identified clusters using quantitative metrics derived from k-means and agglomerative clustering analyses. These analyses leverage heat demand data from all buildings in Stockholm City. This approach allows for evaluating data center locations' proximity to high-heat demand areas and their potential for integration into heat recovery systems. It aligns with existing urban elements, such as supermarkets and water bodies, to enhance sustainable energy management. Ultimately, this method provides a data-driven framework for strategic urban planning and resource management, facilitating the development of district heating solutions that are both economically viable and environmentally sustainable.

The methodology for this analysis involves using Python for data processing and machine learning, with key libraries such as Pandas, NumPy, Scikit-learn, and Matplotlib. The simulation applies various clustering models, including K-means Clustering, Agglomerative Clustering, DBSCAN, Gaussian Mixture Models (GMM), and Spectral Clustering, to group the data based on features like heat demand, energy usage, and geographical locations. Visualizations are enhanced by overlaying geographical scatter plots with a background image, providing context for the clustering results. The dataset for Stockholm spans multiple years, with EPC (Energy Performance Certificate) data sizes for each year as follows: 10,368 data points for 2014, 39,320 for 2018, 42,566 for 2019, 28,830 for 2020, 14,304 for 2021, and 7,896 for 2022. The total dataset size across these years amounts to 143,284 data points. This dataset includes detailed information on energy performance, property designation, and geographical coordinates, which are essential for clustering and analysis in the simulation. K-means Clustering serves as the primary solver in this environment, along with other unsupervised learning algorithms such as Agglomerative Clustering, DBSCAN, Gaussian Mixture Model (GMM), and Spectral Clustering. These clustering methods are employed to categorize the data based on key attributes, including heat demand, energy usage, and geographical locations.

The dataset is comprised of both categorical and numerical attributes. To handle missing values, imputation and interpolation methods are applied. Additionally, the data is normalized to ensure it is properly scaled for effective clustering analysis.

Processing of EPC data with Swedish survey agency data

The EPC data was integrated with geographic information from the Swedish Survey Agency (Lantmäteriet) to identify optimal locations for future data centers based on heat demand characteristics [20,21].

The data collection encompassed a wide array of energy metrics, categorized by energy types used for heating and domestic hot water (e.g., district heating, heating oil, natural gas, firewood, and various forms of electricity, including several types of heat pumps). These metrics were complemented by primary and normalized energy use figures for buildings, yielding a year-corrected value known as the Energy Index. Additional details included each building's address, complexity, category, type, construction year, and structural specifics, such as the number of heated basement floors, above-ground floors, stairwells, and residential apartments.

During integration, the energy performance data for each building was geocoded using primary and postal addresses, ensuring precise mapping. This accuracy is critical for practical spatial analysis, which underpins energy demand mapping and site selection for data centers.

In the data processing phase, efforts were made to standardize and normalize the data across different energy types and measurement units, facilitating uniform analysis. Rigorous data cleansing was conducted to correct inconsistencies, impute or exclude missing values based on their significance, and eliminate duplicate records. This involved removing whitespace from string fields for consistency, extracting non-numeric prefixes from property designations and address identifiers to standardize these fields, and handling missing values by removing rows with null entries. Additionally, coordinates were added based on property designations, and categorical variables were encoded using one-hot encoding. These steps ensured the dataset's integrity and reliability for subsequent analysis.

Through these processes, a robust and detailed spatial database was constructed to identify high-heat demand areas and determine the most strategic locations for future data center installations. This database not only supports detailed spatial analysis but also forms the foundation for the advanced clustering techniques used in the later stages of the study.

Heat demand clustering using k-means and agglomerative clustering

In this study phase, the geographical area was segmented into clusters based on buildings' heat demand characteristics using k-means and agglomerative clustering techniques.

The Elbow Method was utilized to determine the optimal number of clusters for k-means clustering. This method plots the sum of squared distances (inertia) from each point to its assigned cluster center against the number of clusters. As the number of clusters increases, inertia decreases. However, there is a point where the rate of decrease sharply slows, forming an "elbow" shape on the plot. This "elbow" indicates the optimal number of clusters, balancing the trade-off between underfitting and overfitting. Employing the Elbow Method ensured that the segmentation was meaningful and efficient for further analysis [24].

Subsequently, four centroids were selected within the dataset for the k-means clustering process. Each data point, representing the heat data of a building, was assigned to the nearest centroid based on the Euclidean distance between the data point's features and the centroid's features [25]. This assignment and recalculation of centroids were iteratively performed until the centroids stabilized, marking the algorithm's convergence. As a result, the dataset was divided into four distinct clusters, each representing an area with homogeneous heat demand characteristics.

Concurrently, agglomerative clustering was employed as a hierarchical method, starting with each data point as its own cluster. Clusters were merged based on their similarity and assessed using Ward's method, which aims to minimize the variance within a cluster. This merging process continued until all data points were consolidated into four major clusters [26].

Allocation of heat sources to clusters

In the next step, k-means and agglomerative clustering were employed again, given their effectiveness in grouping data based on similarities, to identify clusters of Stockholm's heat demands. To allocate existing renewable energy sources, Stockholm's heat demands were segmented into ten districts using the k-means clustering algorithm, accessed via the sci-kit-learn library [25].

The methodology begins by verifying and preparing the dataset, which includes critical analysis parameters. Data anomalies, such as infinite values or missing data points within the heat demand metric, are replaced with a statistically representative figure—typically the mean or median—to maintain data integrity. The heat demand data undergo a normalization process, ensuring a uniform scale for comparative visualization. The size of the visual markers is proportionally scaled to reflect the heat demand, enabling an accurate spatial representation of heat intensities.

For the clustering analysis, geographic coordinates are extracted to delineate the urban landscape into ten distinct zones. This segmentation is achieved through an iterative clustering technique that organizes the space based on proximity and similarity in heat demand, ensuring that the identified clusters are stable and replicable in subsequent analyses.

Furthermore, the algorithm calculates the central points of these clusters, pinpointing critical areas of heat concentration. The final visualization includes a legend correlating to the clusters, with the geographic scope of the map meticulously adjusted to encompass the Stockholm region of interest. This comprehensive display of the city's heat demand landscape is an indispensable tool for urban planning and the efficient allocation of energy resources.

Marginal cost of heat

The economic impact of waste heat recovery is evaluated using the marginal cost of heat (MCOH), primarily driven by electricity prices and limited to operational costs. This focus on operational costs is due to the relatively minor contribution of capital costs to the overall life cycle cost. The capital costs for heat recovery equipment are conservatively estimated at €1.5 million per MW of cooling capacity [27–29], representing about 3 % of the operational cost for a single year. Over a 20-year lifespan with a 10 % discount rate, capital costs account for only 0.3 % of the total life cycle cost, which falls well within the range of economic uncertainty.

Although capital costs are not insignificant, they are less critical in this context. Previous studies have indicated that economic returns for data center [27] and supermarket [30] owners are uncertain, partly due to the variable value of heat throughout the year and the network owner's willingness to pay for it [25]. This study adopts the perspective of the district and city, viewing waste and environmental heat as part of a resource portfolio where the lowest marginal costs determine the merit order.

For cooling devices and heat pumps, the MCOH is entirely influenced by the coefficient of performance (COP) and the price of electricity (p_{el}), as represented by Eq. (1). Maintenance costs, like capital costs, are also considered negligible and are therefore excluded from the analysis.

$$MCOH = \frac{p_{el}}{COP} [€/MWh] \tag{1}$$

Electricity prices and COPs fluctuate throughout the year, as does heat value, making detailed hourly simulations the most effective method for assessing cost-effectiveness at any given time [30,31]. However, high-resolution time simulations are beyond the scope of this study. Instead, an economic performance map is created to capture a range of COPs and electricity prices. This map includes cost ranges for data centers, supermarkets, and water bodies, comparing them to Stockholm's district heating prices at retail and wholesale levels. Wholesale prices are derived from Stockholm Exergi's 2024 base retail prices [32] and their Open District Heating (ODH) market [33].

Equation (2) provides the formula for calculating the weighted MCOH for each district or cluster. Each cluster is denoted as i, and the MCOH of each heat source (with each technology denoted as j) is multiplied by the heat supply for each source in the cluster portfolio ($Q_{j,i}$), then divided by the total heat demand of the cluster (Q_i). This analysis reveals spatial differences in heating costs across the city, which can significantly influence urban development, particularly in cities lacking district heating networks.

$$MCOH_i = \frac{\sum \left(p_{el,j}/SCOP_j\right) Q_{j,i}}{Q_i} [€/MWh] \tag{2}$$

Hyperparameter selection and justification

The hyperparameters listed in Table 1 are critical for selecting the optimal values for various clustering algorithms used in the study. In the case of KMeans (n_clusters = 4), this value was chosen based on the assumption that the data can be divided into distinct groups, such as high-demand versus low-demand heat zones. Selecting four clusters enables the identification of several types of demand and supply locations, representing different heating demand profiles across the city. The Elbow Method was employed to determine the ideal number of clusters by evaluating the sum of squared distances within clusters. A smaller number of clusters, such as two or three, would oversimplify the data, while more clusters could overfit the model and introduce noise, making four clusters an ideal balance for spatial clustering in district heating. Similarly, Agglomerative Clustering (n_clusters = 4) was selected to partition the dataset into distinct heat supply and demand regions, with the number of clusters adjusted based on the dendrogram, which visually represents hierarchical relationships. Choosing fewer clusters could obscure subtle spatial patterns, while more clusters might increase complexity and risk overfitting, making four clusters a reasonable choice.

For DBSCAN (eps = [0.1, 0.3, 0.5, 0.7, 1.0], min_samples = [3,5,10,15]), the parameter eps controls the maximum distance between two points to be considered part of the same neighborhood. A range of values from 0.1 to 1.0 was selected to explore different neighborhood sizes, which is crucial for identifying tightly packed clusters (small eps values) or more spread-out ones (larger eps values). The min_samples parameter determines the minimum number of points required to form a "core" point in a neighborhood. Smaller values (e.g., 3) lead to the identification of more clusters, while larger values (e.g., 10 or 15) focus on denser, more meaningful clusters. This approach ensures the algorithm can handle spatial data with varying densities, such as urban heat sources, and avoid noise. SpectralClustering (n_clusters = 4, affinity = 'nearest_neighbors', random_state = 42) was chosen to focus on local spatial structures by using the 'nearest_neighbors' affinity, which is critical for clustering heat demand or supply based on geographical proximity. Fixing the random state ensures reproducibility of the results, which is essential for research purposes. The number of clusters affects the granularity of the identified patterns, influencing the resolution of heat demand–supply matching.

Table 1
Hyperparameter Selection for Clustering Algorithms.

Algorithm	Hyperparameter	Value/Range	Description
KMeans	n_clusters	4	Number of clusters to form.
Agglomerative Clustering	n_clusters	4	Number of clusters to form.
DBSCAN Clustering	eps	[0.1, 0.3, 0.5, 0.7, 1.0]	The maximum distance between two samples for them to be considered as in the same neighborhood.
	min_samples	[3,5,10,15]	The number of samples in a neighborhood for a point to be considered as a core point.
Spectral Clustering	n_clusters	4	Number of clusters to form.
	random_state	42	Seed used by the random number generator for reproducibility.
	affinity	'nearest_neighbors'	Metric used to compute the similarity matrix.
Gaussian Mixture Clustering	n_components	4	The number of mixture components (clusters).
	covariance_type	'full'	Type of covariance matrix. Options: 'full', 'tied', 'diag', 'spherical'.
	reg_covar	1.00E-04	Regularization to ensure covariance matrices are positive semi-definite.
	max_iter	500	Maximum number of iterations.
Data Preprocessing	imputer_strategy	'mean'	Strategy to use for imputing missing values.
	scaler	StandardScaler	Standardizes the data by removing the mean and scaling to unit variance.

For GaussianMixture (n_components = 4, covariance_type = 'full', reg_covar = 1e-4, max_iter = 500), the number of components was set to 4 to model the heat sources in Stockholm as a combination of four distinct distributions, each representing different heating profiles. The 'full' covariance type was selected to allow each component to have its own covariance matrix, offering maximum data modeling flexibility. Regularization (reg_covar = 1e−4) ensures the covariance matrices are positive semi-definite, which enhances model stability, especially with sparse data. The max_iter parameter was set to 500 to allow the algorithm sufficient time to converge to an optimal model. Fewer components may simplify the model, losing important detail, while more components could overfit the data.

Data preprocessing choices included imputing missing values using the mean strategy (imputer_strategy = 'mean') and standardizing the data with StandardScaler. The mean imputation was chosen because it is a simple and effective strategy for handling missing data, particularly when the data is missing at random. Standardizing the data ensures that all features contribute equally to the clustering process by eliminating bias from features with more extensive ranges, such as geographical coordinates or temperature values. Other imputation strategies or scaling methods could be considered, but standard scaling is typically ideal for clustering tasks, as it normalizes the influence of each feature.

Overall, the choice of hyperparameters is crucial for accurately mapping and clustering Stockholm's heat supply and demand sources. Variations in cluster numbers, neighborhood sizes, and covariance models impact the clustering techniques' flexibility, sensitivity, and scalability. By selecting appropriate values, the study ensures that the results reflect meaningful patterns while avoiding overfitting or underfitting. The use of multiple clustering techniques, KMeans, AgglomerativeClustering, DBSCAN, and GaussianMixture, enables the adaptation to different data characteristics, such as noise, density, and distribution shapes, providing a robust framework for integrating distributed heat sources into district heating networks.

Methods limitation

While the proposed methodology offers a comprehensive framework for optimizing heat resource allocation in urban environments, it has certain limitations. Any inaccuracies, missing entries, or outdated information can directly impact the clustering results and the overall accuracy of heat demand mapping. Additionally, the methodology assumes standardization across various energy types and units, which may not fully capture regional variations or temporal fluctuations in heat demand and supply. The clustering techniques used, such as k-means and agglomerative clustering, depend on the selection of hyperparameters, like the number of clusters, which can be somewhat subjective and may not fully reflect the complex spatial dynamics of urban heat demand.

Another significant limitation is the simplified economic model used for evaluating the marginal cost of heat (MCOH). By focusing primarily on operational costs and assuming that capital and maintenance costs are negligible, the methodology may overlook important economic factors influencing the feasibility of integrating new heat sources. Moreover, the method does not account for practical challenges such as policy constraints, stakeholder engagement, and the adaptability to real-time changes in heat demand or supply. These factors can significantly impact the implementation of the proposed solutions in real-world urban settings. Addressing these limitations is essential for enhancing the robustness and applicability of the methodology in diverse urban contexts.

Results

In this research, the marginal cost of environmental heat was assessed locally to determine its impact on the merit order of heat deployment and the resulting cost to the city. The heat demand across Stockholm was mapped using data from the Energy Performance Certificate (EPC) database, which was integrated with property locations provided by the Swedish Land Survey (Lantmäteriet). After removing duplicate property description entries (known as 'fastighetsbeteckning' in Swedish), the number of unique properties in the EPC and Lantmäteriet datasets are 31,389 and 61,952, respectively. When the EPCs are filtered for those connected to district heating (approximately 52 % of all buildings and 92 % of the floor area in Stockholm) and cross-referenced for matching property descriptions, there are 14,103 entries. This is 86 % of all DH connected properties. However, the total demand is found to be 7.7 TWh/year and is in agreement with prior studies [34]. The breakdown by year of record for EPCs and matching properties with coordinates are given in Table 2 and shown spatially in Fig. 1. Table 2 shows Unique total property counts for EPCs and those with coordinates by year.

A heat map was created to further illustrate the spatial distribution and intensity of heat demand, as shown in Fig. 2. This heat map visualizes the density of heat demand data points across Stockholm, using the x and y coordinates to represent longitude and latitude. Lighter, more saturated areas on the map indicate regions with greater heat demand, clearly highlighting areas where heat resources are most needed. This visualization is crucial for identifying optimal locations for future data centers, enabling a more strategic and data-driven approach to district heating planning.

Table 2
Unique total property counts for EPCs and those with coordinates by year.

Year	Total EPCs	EPC and Coordinates	Match Rate
2014–2017	5583	1999	36 %
2018	4258	3324	78 %
2019	4931	2651	54 %
2020	4867	2715	56 %
2021	5696	1886	33 %
2022	6054	1528	25 %
Totals	31,389	14,103	45 %

Allocation of future data centers using k-means and agglomerative clustering

Fig. 3 offers a comprehensive visualization of the spatial distribution of heat demands across Stockholm, with intensity levels represented by the size and colour of the circles. Larger and darker circles indicate areas of higher heat demand, helping to identify critical zones with concentrated energy requirements. Five advanced clustering techniques—K-Means, Agglomerative Clustering, DBSCAN, Spectral Clustering, and Gaussian Mixture Model (GMM)—were applied to determine the optimal locations for future data centers. Each clustering method is represented on the map, with K-Means cluster centers marked by red triangles, Agglomerative cluster centers by blue stars, DBSCAN cluster centers by blue cross, Spectral Clustering centers by purple circles, and Gaussian Mixture Model centers by green squares. This multi-method approach ensures a robust and reliable analysis, enhancing confidence in the recommended locations for data centers.

The clustering analysis reveals that Stockholm's highest heat demand concentrations are located primarily in its central and northern regions, as evident from the denser and darker circles in these areas. Each clustering method provided unique insights into the spatial distribution of heat demand. K-means and Agglomerative Clustering consistently overlapped, demonstrating their reliability in identifying high-demand zones. DBSCAN excelled in identifying smaller, denser clusters, highlighting micro-level variations in heat demand. Spectral Clustering and GMM offered additional perspectives by identifying flexible cluster shapes and providing probabilistic clusters, respectively. These nuances help refine the decision-making process and broaden the scope of infrastructure planning.

Each of the five clustering methods brings unique strengths and perspectives to the analysis, offering a comprehensive understanding of heat demand patterns in Stockholm. K-Means Clustering provides a simple and efficient solution for partitioning data into well-defined clusters, though its assumption of spherical and similarly sized clusters may not always reflect the actual spatial distribution. Agglomerative Clustering effectively captures hierarchical relationships and yields flexible, interpretable cluster shapes, but it can become computationally intensive and less scalable than K-Means. DBSCAN excels at identifying dense regions of heat demand and uncovering smaller clusters that other methods may overlook, yet it struggles with sparse data and is highly sensitive to parameter choices. Spectral Clustering accommodates complex, non-linear cluster shapes, making it ideal for intricate spatial patterns, although it is computationally expensive and sensitive to the prescribed number of clusters. Finally, the Gaussian Mixture Model (GMM) offers probabilistic clusters with nuanced boundaries that facilitate uncertainty quantification, but its performance may hinge on initial conditions and extensive parameter tuning. Together, these methods provide a robust toolkit for exploring, analyzing, and interpreting the spatial variability of heat demand in Stockholm.

No single method is universally superior. Instead, using multiple

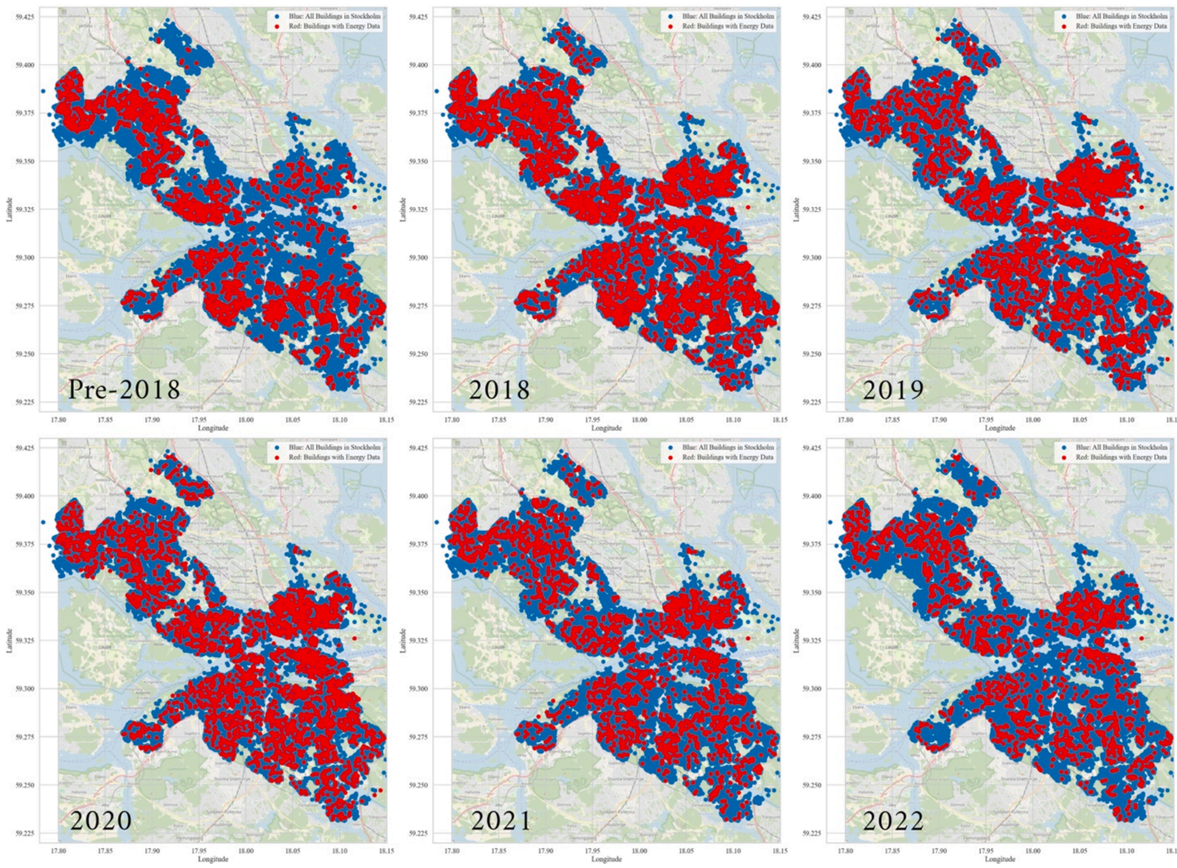

Fig. 1. Unique buildings/properties in Stockholm (blue dots) compared with properties that have energy data (Red dots). (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

methods enriches the analysis by capturing different aspects of the data. For instance, DBSCAN effectively detects localized high-demand clusters, while GMM offers a probabilistic perspective that accounts for variations within the data. The overlaps between K-Means and Agglomerative Clustering provide consistent insights, reinforcing the reliability of their suggested locations. With its ability to handle non-linear patterns, Spectral Clustering complements the other methods by uncovering additional potential sites.

Table 3 provides a numerical comparison of the cluster center locations identified by each method.

The clustering results indicate that while K-Means and Agglomerative Clustering tend to converge on similar cluster centers, DBSCAN, Spectral Clustering, and GMM identify additional or alternative locations. This divergence underscores the value of employing multiple clustering techniques to capture global and local data patterns. DBSCAN highlighted micro-level heat demand variations that could be critical for localized planning. Spectral Clustering uncovered complex patterns, suggesting potential sites near existing infrastructures. GMM, with its probabilistic approach, provided insights into areas of uncertainty, enabling more informed decision-making.

The strategic placement of data centers based on these clustering methods offers significant potential for optimizing Stockholm's district heating network. Data centers located near high-demand areas can efficiently supply heat, leveraging their excess heat for sustainable urban energy solutions. Moreover, integrating supermarkets and other existing infrastructures into the analysis further highlights opportunities for collaborative heat recovery initiatives, enhancing the overall efficiency and sustainability of the heat distribution network.

However, clustering-based approaches are inherently limited by their reliance on historical data and static assumptions. They do not account for future urban development, demographic shifts, or evolving heat demand patterns. Smaller but strategically important pockets of demand might not significantly influence the cluster centroids and could be overlooked. To address these limitations, clustering results should be complemented with predictive models, urban planning considerations, stakeholder engagement, and adaptive strategies to accommodate future changes.

By leveraging five advanced clustering methods—K-Means, Agglomerative Clustering, DBSCAN, Spectral Clustering, and GMM—this analysis provides a comprehensive understanding of heat demand patterns in Stockholm. The multi-method approach not only identifies optimal locations for future data centers but also highlights the strengths and limitations of each technique. Integrating these insights with urban planning strategies ensures a balanced, data-driven, and adaptive approach to sustainable energy solutions. This methodology demonstrates the potential for clustering techniques to guide the development of efficient district heating systems, supporting Stockholm's broader sustainability goals.

Allocation of heat sources to clusters

Fig. 4 presents a map segmented into 10 districts using a k-means clustering algorithm accessed via the sci-kit-learn library. Each ten district is assigned the heat sources within its cluster and water body sources, which are connected based on the minimum connection distance. The heat supply from each source is calculated, and the total cost is determined using the marginal cost of heat (MCOH). In Fig. 4, each bubble represents a heat load, with the bubble size corresponding to the annual energy demand. Table 4 provides each cluster's total heat energy demand and the allocated residual heat supply by type. The total heating demand of 7.7 TWh/year matches the reported district heating supply of the city [34], indicating that the EPC dataset offers comprehensive

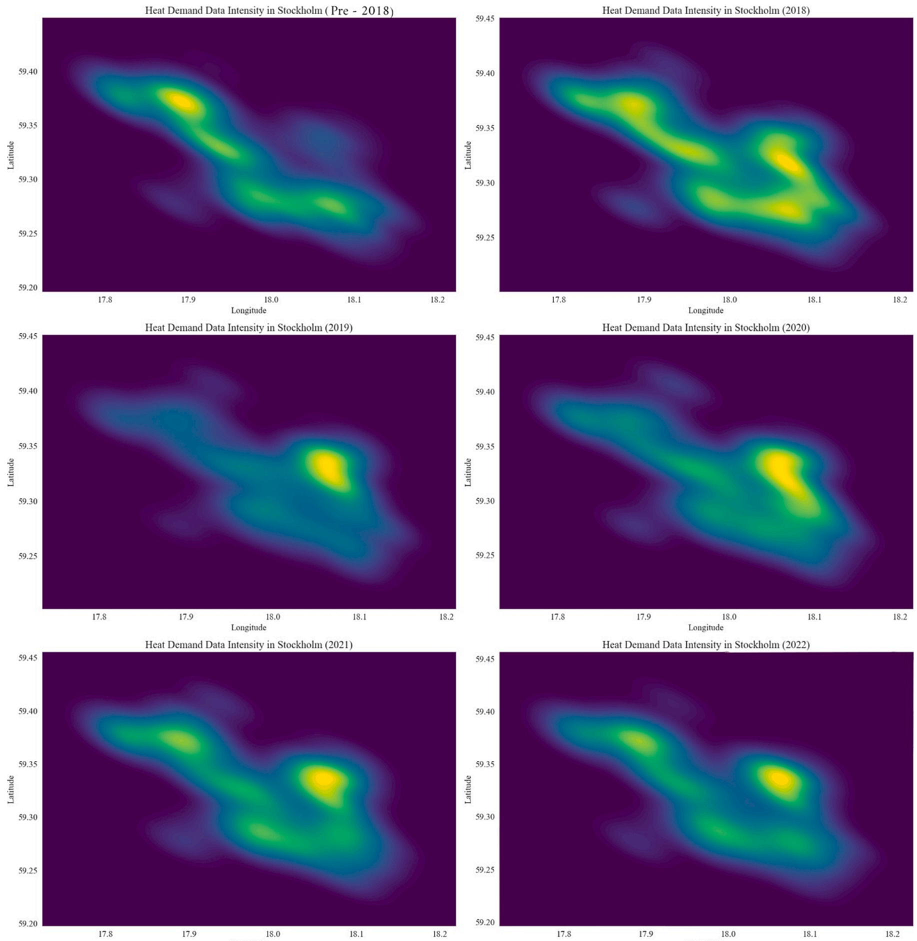

Fig. 2. Heat demand intensity map in Stockholm.

coverage.

The heating supply from environmental sources is 4.2 TWh/year, accounting for approximately 54 % of the city's total annual demand. This figure contrasts with previous mapping by Su et al. [9], where data centers contributed 3.2 TWh/year; in this study, their contribution is reduced to 0.48 TWh/year. This discrepancy arises from updated GIS data, which shows fewer data centers within Stockholm's administrative boundaries (the focus of this study) and indicates that the heat output from the average existing data center is lower than previously reported by Su et al., who based their findings on relatively large data centers by Nordic standards [12,35]. Non-viable heat generation during warmer periods (12 °C and greater) are also removed here, consistent with how the Open DH market in Stockholm functions.

A distinct color scheme is employed to differentiate the clusters on the visual map, resulting in an illustrative scatter plot that overlays normalized heat demand data onto the city's geographic layout. Each cluster is marked with unique hues, with geographic points scaled to represent their respective heat demands.

The clustering results presented in Fig. 4 provide valuable insights into the spatial distribution of heat demand across Stockholm. By segmenting the city into distinct clusters, urban planners and policymakers can better understand the regions with the highest and most consistent heat demand, enabling targeted investments in district heating infrastructure. For example, areas with large clusters of high heat demand, such as those shown in red, yellow, and pink clusters, could be prioritized to install additional heat recovery systems, including integrating data centers, supermarkets, and water bodies as heat sources. This strategic allocation of heat sources helps to optimize energy use and minimize the need for conventional, fossil-fuel-based heating solutions. Furthermore, the bubble size on the map representing heat load can help policymakers allocate resources more efficiently, ensuring that areas with higher demand are adequately supplied.

In addition to these infrastructural recommendations, the heat source allocation strategy has significant economic implications. By linking data centers and supermarkets to the district heating system, cities can reduce operational costs through heat recovery and reliance on external energy sources. For instance, data centers in Cluster 2 (blue) can be incentivized to supply heat to the surrounding areas, especially in regions like Cluster 1 (red), where heating demand is high. This strategy could reduce the marginal cost of heat (MCOH), ensuring that heat produced from environmental sources remains cost-competitive compared to traditional heating methods. Additionally, understanding how environmental heat sources contribute to the total demand helps shape policies that promote renewable heat generation while balancing the economic feasibility of these systems. By incorporating such policies, local governments can align their heating strategies with broader sustainability goals, reducing carbon emissions and fostering the growth of green technologies in urban environments.

Fig. 5 is a heatmap that shows the distances between various cluster centers, calculated based on their geographic coordinates and displayed in kilometers. The colours range from dark red (indicating smaller distances) to light red (indicating larger distances), helping to identify clusters that are close to each other and those that are farther apart. This visualization highlights the spatial distribution of clusters, with closer centers suggesting areas of high heat demand and more distant centers indicating regions with dispersed or unpredictable demand. These insights are useful for optimizing district heating systems.

The study's clustering results and data center allocation provide an

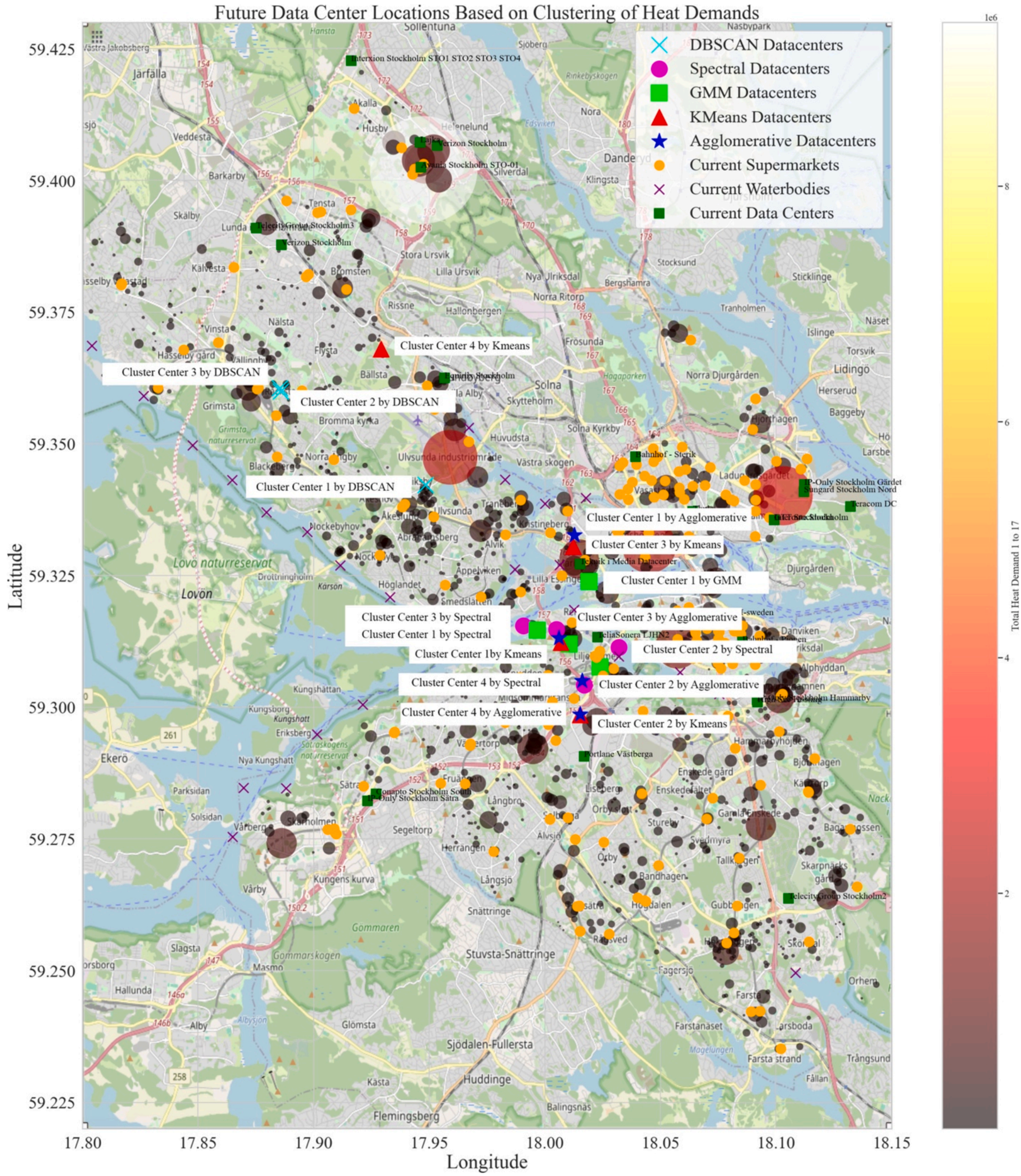

Fig. 3. Total heat demand mapping and the locations of future data centers based on five clustering methods.

in-depth analysis of the distance variation between different cluster centers, offering valuable insights for optimizing district heating systems. These clusters represent areas with varying heat demand across Stockholm, and clustering techniques help identify strategic locations for placing distributed heat sources, such as data centers, supermarkets, and water bodies. The variation in distances across clusters and the different clustering methods applied significantly affect the overall system.

The centers of the clusters identified using different methods (DBSCAN, Spectral Clustering, GMM, KMeans, and Agglomerative Clustering) exhibit considerable variability. This is to be expected, as each algorithm has distinct characteristics that influence the placement

Table 3
Future Datacenter's location in Stockholm based on 5 clustering methods.

Data Centers	K-means (Lat, Long)	Agglomerative (Lat, Long)	DBSCAN (Lat, Long)	Spectral (Lat, Long)	GMM (Lat, Long)
1	59.3123, 18.0070	59.3326, 18.0128	59.3422, 17.9483	59.3148, 18.0050	59.3238, 18.0191
2	59.2986, 18.0153	59.3050, 18.0162	59.3600, 17.8858	59.3042, 18.0171	59.3120, 18.0104
3	59.3305, 18.0123	59.3131, 18.0061	59.3606, 17.8853	59.3154, 17.9908	59.3076, 18.0239
4	59.3680, 17.9292	59.2986, 18.0153	59.3543, 17.9065	59.3113, 18.0321	59.3146, 17.9968

of centers. DBSCAN, for example, detects denser and more localized clusters, making it ideal for pinpointing smaller hotspots of heat demand that other methods might miss. On the other hand, Spectral Clustering identifies flexible and complex shapes, which is particularly useful for regions with irregular or non-linear heat demand patterns. The Gaussian Mixture Model (GMM) provides a probabilistic view of clusters, offering a nuanced understanding of uncertainty in heat demand, which can benefit urban planners considering variability in demand patterns.

The distances between the cluster centers identified by each method provide important insights into the spatial distribution of heat demand. As seen with K-means and Agglomerative Clustering, smaller distances between centers suggest areas with well-defined high heat demand, which can be reliably targeted for infrastructure development, such as placing new data centers or optimizing existing ones for heat recovery. Larger distances, identified by methods like DBSCAN and GMM, point to regions where heat demand is more unpredictable, indicating that localized solutions, such as smaller-scale heat sources, might be needed. These methods also reveal potential gaps or underutilized areas that could benefit from heat recovery solutions. The variation in distances highlights the differing granularity of each algorithm's approach, with GMM's probabilistic nature identifying uncertain or boundary areas that could play a key role in future district heating expansions.

The variation in cluster distances directly impacts the optimization of district heating infrastructure. Clusters with tightly packed centers, as identified by methods like K-means and Agglomerative Clustering, suggest high, concentrated heat demand areas. These regions can be prioritized for expanding or upgrading existing district heating infrastructure, facilitating the integration of heat recovery systems. Conversely, as revealed by DBSCAN, larger distances imply more dispersed heat demand, which may benefit from decentralized heat solutions, such as localized heat pumps or smaller heat recovery units. Spectral Clustering, with its ability to capture complex clusters, may also identify areas that require tailored heat supply solutions. The economic feasibility of heat recovery is influenced by the alignment of these clusters with existing infrastructure. Areas with a higher concentration of heat demand and residual heat sources, such as data centers or water bodies, tend to have a lower marginal cost of heat (MCOH), making them more economically viable for district heating investments.

Temporal variations in heat demand, such as seasonal changes, could significantly alter the ideal cluster center locations. In colder months, heat demand is likely to be higher, necessitating the placement of heat recovery systems to meet these increased needs. During warmer months, when data centers and supermarkets generate more residual heat, clustering methods that account for dynamic heat generation, such as GMM, can provide better insights into when and where to optimize heat recovery systems. Integrating time-based data would enhance the accuracy of clustering and the strategic placement of heat recovery units, ensuring efficiency year-round.

From a strategic urban planning perspective, the clustering analysis provides crucial data for optimizing energy use and reducing reliance on fossil fuels. By identifying clusters with high heat demand and their proximity to residual heat sources, urban planners can strategically place heat-generating facilities, such as data centers, in areas with high demand, minimizing infrastructure costs and ensuring that heating systems are aligned with local needs. Collaborative heat recovery initiatives, integrating sources like data centers and supermarkets into district heating networks, can reduce operational costs and carbon emissions.

The variation in cluster distances across different methods offers a comprehensive view of Stockholm's heat demand landscape. By combining these insights with considerations of economic feasibility and temporal variations, urban planners can make informed decisions that optimize both system performance and cost-effectiveness. The use of multiple clustering methods adds robustness to the analysis, ensuring that no significant heat demand areas are overlooked and that infrastructure investments align with sustainable, low-carbon goals.

Calculation of levelized cost of heat per cluster and city

Fig. 6 presents a map of the marginal cost of heat (MCOH) for a range of seasonal coefficient of performances (SCOP) and electricity prices, with several key values. The zones of MCOH values (in different shades of blue) are formed by Eq. (1) and the three solid price lines represent three different values of heat; in yellow is the weighted average sale price over the year for the Open District Heating (ODH) market (i.e. a wholesale price) at 22.0 EUR/MWh[1], in orange is the ODH price when warm weather sales over 12 °C are removed at 27.4 EUR/MWh, and the final price in red is the 2024 seasonally weighted retail price of heat at 39.4 EUR/MWh [32].

Seasonal COP values for water bodies, data centers, and supermarkets are shown in Fig. 6, with dashed boxes noting the ranges found in the literature. Water bodies are assumed to have an average SCOP of 2.8, which is relatively conservative compared to existing heat pumps in Stockholm, reaching above 3.0 [34]. Data centers can have many designs, including advanced water cooling, without needing a heat pump. However, the example here relies on typical air-cooled servers with an average SCOP of 4.4 [31]. Supermarkets can use their waste heat directly to help offset retail purchases, which offers greater value than selling to ODH and means the realized SCOP for supermarket sales can vary dramatically depending on the sales strategy [33,36]. It is assumed here that self-consumption is the preferred strategy; therefore, the SCOP for the district heating supply will be in the lower range, with an average of 3.5.

Using the electricity price and SCOPs described above, the resulting MCOH values for data centers, supermarkets, and water bodies are 12.7, 16.0, and 20.0 EUR/MWh, respectively. All prices are below the 22.0 EUR/MWh price for ODH, corroborating the results from the recent Nordic Energy Outlook [35], which concluded that 98 % of all residual heat sources in Stockholm are economically viable.

The total cost and weighted price for heating by cluster are given in Table 5, which shows a reduction in heating price in all clusters and is most significant in those with the least heat coming from the existing district heating supply. In an advanced and well-developed district heating system like Stockholm, cluster analysis may not be relevant, but the method highlights how cities without district heating may prioritize development around environmental heat sources to minimize costs.

LCOH is a convenient metric for comparison but does not capture the complexities of investment planning or operational strategies. Swedish DH networks have become primarily based on boilers and/or co-heat and power (CHP) plants fueled by biomass or municipal solid waste, electric boilers, and heat pumps [1]. CHP plants and heat pumps have complex optimization challenges because they interact with two independent markets: heat and electricity. Higher electricity prices favor CHP plants, which earn higher revenues, whereas heat pump supply

[1] All prices are converted from SEK to EUR using the 2023 average conversion rate of 11.5 SEK/EUR.

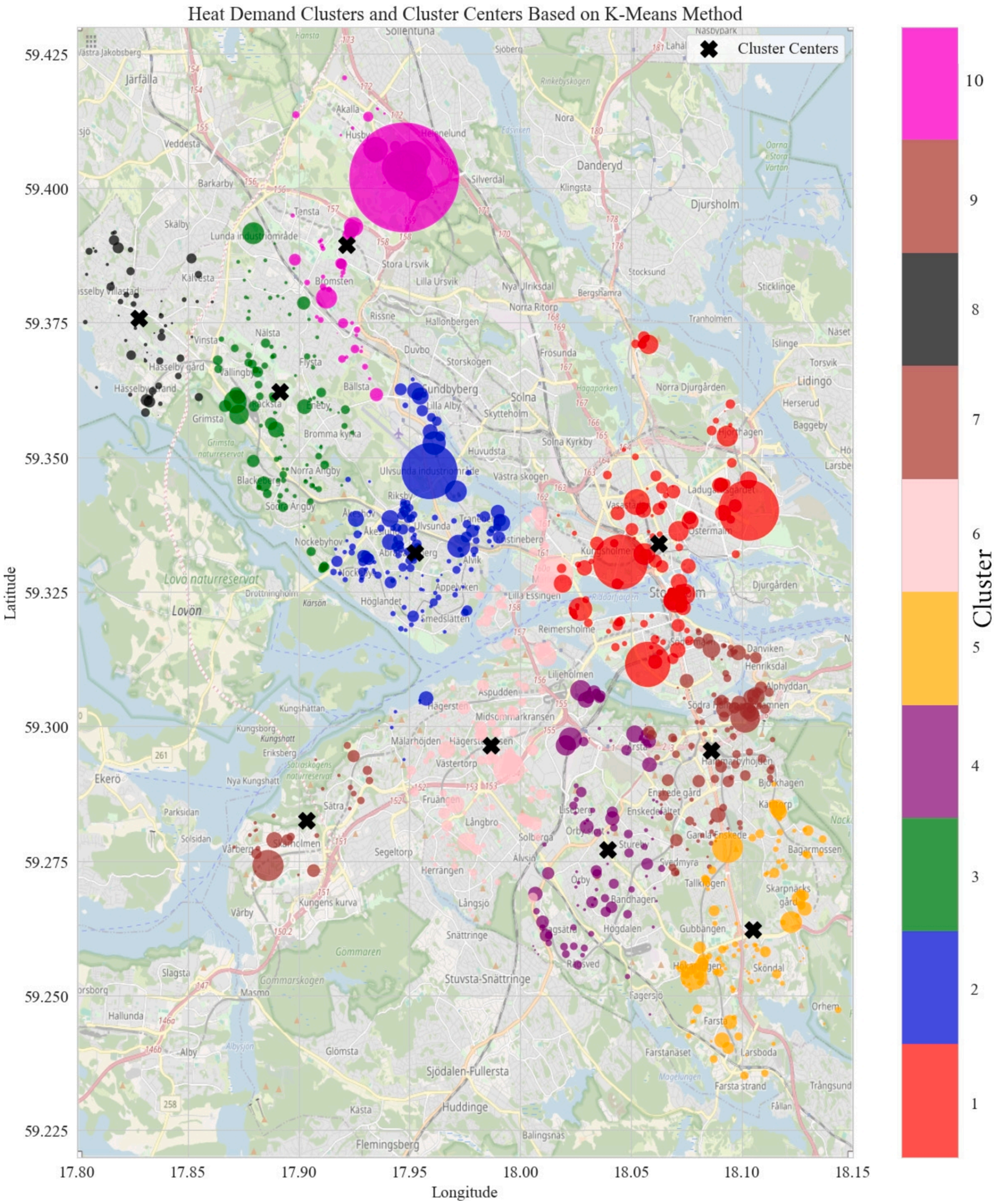

Fig. 4. Heat demand mapping of Stockholm with 10k-means clustered districts.

becomes more expensive [37], as shown in Fig. 6. Electricity price volatility has also increased in Sweden, increasing the opportunity for arbitrage and/or ancillary services, which can moderately increase CHP revenues but are limited by the need to supply heat [37]. Likewise, heat pumps can react to price volatility to minimize operational costs [34].

From an investment perspective, the spatial differences in electricity and heating markets is a point of conflict for district heating supply technologies. Heating demand is local (i.e., within a city), and the electricity market is at a minimum regional (i.e., prices set by regional zones) but influenced by neighboring price zones, often in other countries. National electricity portfolios and transmission capacity will impact the viability of heat pump and CHP investments [38] and require deeper scenario and risk analysis than provided in these results. If electricity price trends continue towards moderate average increases and higher volatility, then investments in thermal storage will be a key enabler of greater heat pump adoption in district heating [39] and help maintain the cost competitiveness found in this study.

Temporal changes in heat demand, supply, and electricity prices can

Table 4
Heat demand and allocation per cluster (in GWh/year).

Cluster	Heating Demand	Data Centers	Super-markets	Water Bodies	Total Allocation
1 (red)	1,672	0	24	100	124
2 (blue)	1,006	60	26	100	186
3 (green)	464	80	18	0	98
4 (purple)	557	20	23	0	43
5 (yellow)	558	60	50	800	910
6 (pink)	959	160	104	300	564
7 (mauve)	170	40	12	500	552
8 (grey)	155	0	9	700	709
9 (mauve)	635	20	16	500	536
10 (magenta)	1548	40	17	400	457
Totals	7,754	480	299	3,400	4,179

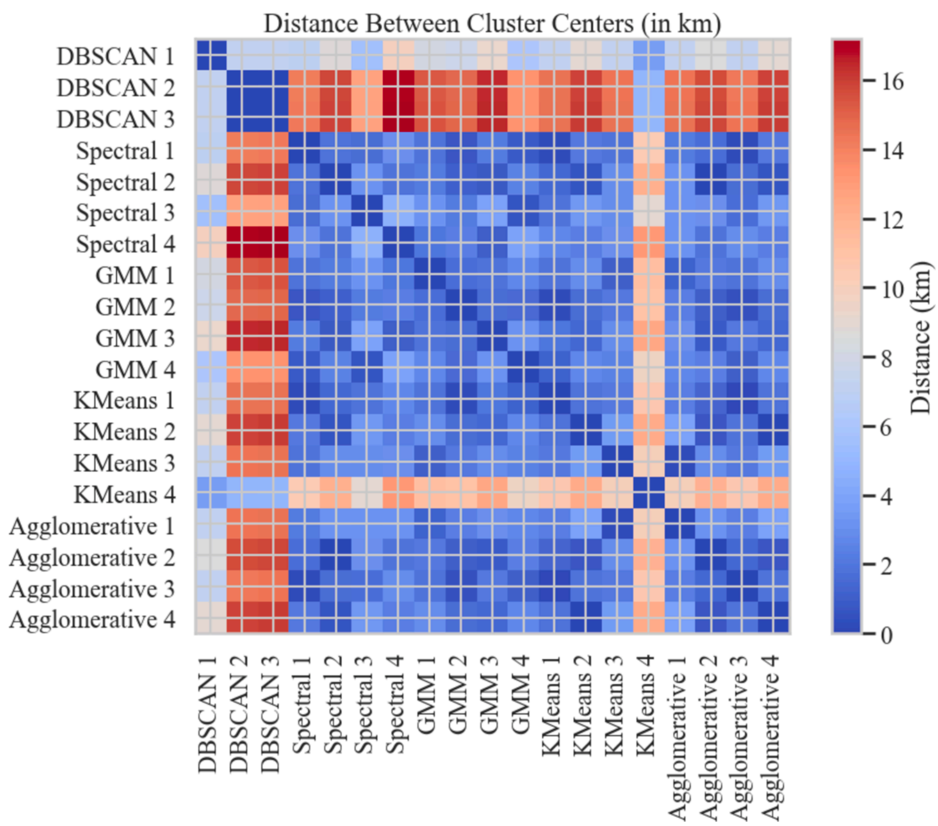

Fig. 5. Distances between clusters Centers.

significantly impact clustering results and the economic outcomes of district heating systems. Heat demand fluctuates seasonally and daily, affecting the allocation of resources if not considered during clustering. Residual heat sources, such as data centers, supermarkets, and water bodies, also experience temporal variability in heat availability, with data centers' cooling demands and supermarket refrigeration systems varying across seasons. Additionally, electricity prices fluctuate, influencing these sources' marginal cost of heat (MCOH), making their economic viability sensitive to market conditions. Failure to incorporate temporal data in clustering could lead to suboptimal resource allocation, inefficiencies in heat supply–demand matching, and inaccurate economic evaluations. Integrating time-based data and adjusting for seasonal and hourly variations would improve the precision of clustering results and provide more sustainable, cost-effective planning for district heating networks, benefiting both urban energy planners and policymakers.

Conclusion

This research introduces a novel data-driven approach to district heating network planning by leveraging high-resolution source-load mapping to address the spatial and temporal variability of distributed heat sources. Unlike traditional methods focusing on centralized heat production, this study integrates detailed spatial data for distributed heat supply sources, such as data centers, supermarkets, and water bodies, alongside demand points within Stockholm. Employing advanced clustering techniques, including K-means, agglomerative clustering, DBSCAN (Density-Based Spatial Clustering of Applications with Noise), and Gaussian Mixture Model clustering, Spectral Clustering, the study robustly analyzes and identifies optimal locations for resource allocation. This multi-method approach ensures adaptability to varying data characteristics, such as noise and non-linear distributions.

Additionally, the study pioneers an in-depth techno-economic assessment of distributed heat sources, revealing that 98 % of

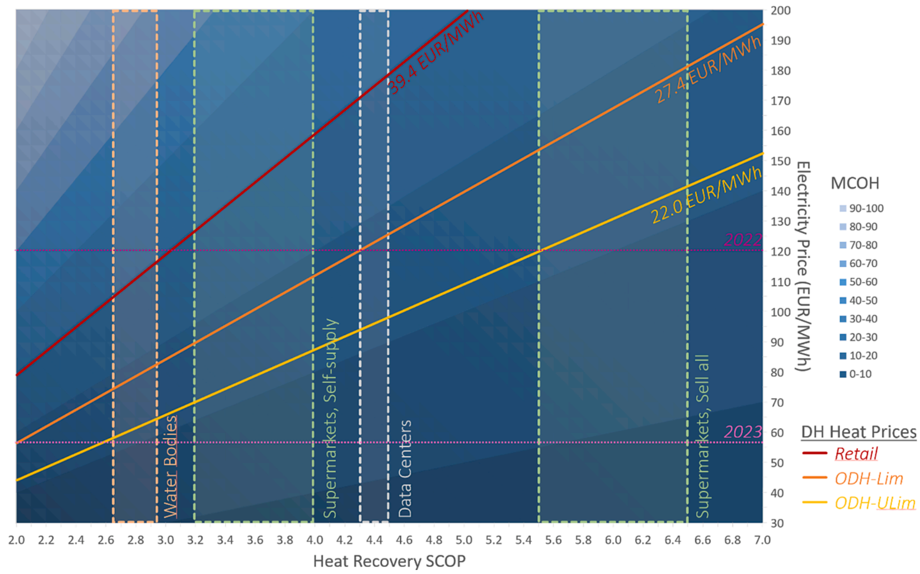

Fig. 6. Marginal cost of heat (in EUR/MWh) by source and electricity price.

Table 5
Total annual costs and weighted prices for heat supplied by residual source and cluster.

Cluster	Existing DH *MEUR/yr*	Data Centers *MEUR/yr*	Super-markets *MEUR/yr*	Water Bodies *MEUR/yr*	Total Cost *MEUR/yr*	Weighted Price *EUR/MWh*
1 (red)	34.1	0.0	0.4	2.0	36.4	21.8
2 (blue)	18.0	0.8	0.4	2.0	21.2	21.1
3 (green)	8.0	1.0	0.3	0.0	9.4	20.2
4 (purple)	11.3	0.3	0.4	0.0	11.9	21.4
5 (yellow)	0.0	0.8	0.8	16.0	17.6	29.9
6 (pink)	8.7	2.0	1.7	6.0	18.4	19.2
7 (mauve)	0.0	0.5	0.2	2.4	3.1	18.0
8 (grey)	0.0	0.0	0.1	2.9	3.1	19.8
9 (mauve)	2.2	0.3	0.3	10.0	12.7	20.0
10 (magenta)	24.0	0.5	0.3	8.0	32.8	21.2
Totals	106.3	6.1	4.8	49.4	166.6	21.5

Stockholm's residual heat sources are economically viable under current market conditions. High spatial resolution (sub-kilometer scale) and temporal granularity are utilized, surpassing previous studies that rely on coarser datasets. These advancements enable precise identification of heat source locations and improved modeling of supply–demand dynamics. By focusing on Stockholm, a global leader in district heating, the research provides actionable insights and establishes a replicable methodology for other urban centers aiming to adopt sustainable heating systems.

The study's contributions include the development of a comprehensive spatial database encapsulating detailed building energy performance metrics for Stockholm (2014–2022), enriched with extensive geographic data to support informed urban energy planning. Quantitative analysis demonstrates that environmentally friendly sources can supply 54 % of Stockholm's annual heat demand of 7.7 TWh/year, with significant contributions from data centers (0.48 TWh), supermarkets (0.3 TWh), and water bodies (3.4 TWh). Strategic clustering methods identify optimal sites for future data centers, maximizing their role as heat sources while aligning with Stockholm's urban energy goals. Economic analysis further confirms that distributed heat sources are cost-effective, with marginal heat costs below the current Open District Heating market price of 22.0 EUR/MWh. Specifically, the marginal heat (MCOH) costs for data centers, supermarkets, and water bodies are estimated at 12.7 EUR/MWh, 16.0 EUR/MWh, and 20.0 EUR/MWh, respectively.

This study underscores the importance of localized, high-resolution data by bridging gaps in existing literature, particularly in source-load mapping resolution and the role of residual heat sources. The methodology, combining GIS data, unsupervised machine learning techniques, and economic analysis, offers a replicable framework for optimizing district heating networks in other cities. This research highlights the dynamic potential of distributed heat sources and provides strategic insights for policymakers and urban planners, setting a benchmark for sustainable and cost-effective district heating solutions.

Future studies in district heating network planning should focus on several key areas to improve and expand upon the findings of this study. First, **data collection and accuracy** should be enhanced by integrating high-resolution, real-time data on heat demand and supply. This would allow for dynamic modelling that accounts for fluctuations in weather, energy prices, and demand, improving system efficiency and responsiveness. Second, **exploring new heat sources** is essential. While this study focused on data centers, supermarkets, and water bodies, future research should investigate other renewable sources, such as geothermal energy, waste heat from industrial processes, or heat recovery from electric vehicles. This would diversify the heat supply and improve the sustainability of district heating systems. **Scalability** should also be a major focus. The methodology used in this study was applied to Stockholm, but future research should test the approach in other cities with varying climate conditions and energy systems. Comparative studies would help refine the models and assess their generalizability in different urban contexts.

Further, **advancing clustering and machine learning techniques** could provide deeper insights into heat demand patterns. Techniques like deep learning or hybrid models that combine clustering with predictive analytics could improve the accuracy of heat distribution planning. Additionally, **economic and policy analysis** will be crucial to evaluate the long-term viability of distributed heat sources and assess the impact of different regulatory frameworks on their deployment. Future research should also explore **stakeholder engagement and understand** the challenges building owners, utility providers, and policymakers face. Social and political factors play a key role in adopting new technologies and should be studied to facilitate smoother integration of district heating networks. Finally, integrating district heating with other urban infrastructure, such as smart grids and waste management systems, could enhance energy efficiency. Studies focusing on **long-term sustainability** and **climate adaptation** are also needed to ensure that district heating networks remain resilient in the face of climate change and energy market fluctuations.

CRediT authorship contribution statement

Amir Shahcheraghian: Writing – original draft, Validation, Software, Methodology, Investigation, Formal analysis, Data curation, Conceptualization. **Adrian Ilinca:** Writing – review & editing, Supervision, Methodology, Funding acquisition, Conceptualization. **Nelson Sommerfeldt:** Writing – review & editing, Writing – original draft, Resources, Methodology, Investigation, Funding acquisition, Conceptualization.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

The authors acknowledge the financial support from the Swedish Energy Agency (Energimyndigheten), project number (51538-1), within the Termo program (for Nelson Sommerfeldt) and from NSERC Discovery Grant – RGPIN/4220-2019 (for Adrian Ilinca).

The energy performance certificate (EPC) data used in this article is provided by The Swedish Housing Authority (Boverket). The authors do not have permission to share the data, but requests can be made to Boverket. The cadastral building data is provided by The Swedish National Land Survey agency (Lantmäteriet) and the authors do not have permission to share the data.

Data availability

The authors do not have permission to share data.

References

[1] Werner S. District heating and cooling in Sweden. Energy 2017;126:419–29.
[2] Agency SE. Energy in Sweden – facts and figures 2023. https://www.energimyndigheten.se/en/news/2023/energy-in-sweden—facts-and-figures-2023/ (accessed).
[3] Exergi S. Heat Recovery. https://www.stockholmexergi.se/en/heat-recovery/ (accessed April 11, 2024).
[4] London Mo. London Heat Map; 2024. [Online]. Available: https://apps.london.gov.uk/heatmap/.
[5] IFAF. Visualisierung von Heizenergieverschwendungen in ̈offentlichen Geb ̈auden durch eine Heatmap; 2024. [Online]. Available: https://www.ifaf-berlin.de/projekte/heatmap/.
[6] Helsinki Co. Energy and climate atlas. https://kartta.hel.fi/3d/atlas/#/ (accessed April 11, 2024).
[7] Unternährer J, Moret S, Joost S, Maréchal F. Spatial clustering for district heating integration in urban energy systems: application to geothermal energy. Appl Energy 2017;190:749–63.
[8] Su C, Madani H, Palm B. Building heating solutions in China: a spatial techno-economic and environmental analysis. Energ Conver Manage 2019;179:201–18.
[9] Su C, Dalgren J, Palm B. High-resolution mapping of the clean heat sources for district heating in Stockholm City. Energ Conver Manage 2021;235:113983.
[10] Blanco Bohorquez LA, Alhamwi A, Schiricke B, Hoffschmidt B. Data-driven classification of Urban Energy Units for district-level heating and electricity demand analysis. Sustain Cities Soc 2023;101.
[11] Zhang K, Zhang Y, Liu J, Niu X. Recent advancements on thermal management and evaluation for data centers. Appl Therm Eng 2018;142:215–31.
[12] Wahlroos M, Pärssinen M, Rinne S, Syri S, Manner J. Future views on waste heat utilization – case of data centers in Northern Europe. Renew Sustain Energy Rev 2018;82:1749–64.
[13] Davies G, Maidment G, Tozer R. Using data centres for combined heating and cooling: an investigation for London. Appl Therm Eng 2016;94:296–304.
[14] Cichowicz R, Jerominko T. Comparison of calculation and consumption methods for determining Energy Performance Certificates (EPC) in the case of multi-family residential buildings in Poland (Central-Eastern Europe). Energy 2023;282:128393.
[15] Zhang Y, Johansson P, Kalagasidis AS. Assessment of district heating and cooling systems transition with respect to future changes in demand profiles and renewable energy supplies. Energ Conver Manage 2022;268:116038.
[16] Stef Boesten WI, Dekker Stefan C, Eijdems Herman. 5th generation district heating and cooling systems as a solution for renewable urban thermal energy supply. https://adgeo.copernicus.org/articles/49/129/2019/ (accessed May 8th, 2024).
[17] Huang P, et al. A review of data centers as prosumers in district energy systems: renewable energy integration and waste heat reuse for district heating. Appl Energy 2020;258:114109.
[18] Ogliari E, Eleftheriadis P, Nespoli A, Polenghi M, Leva S. Machine Learning methods for clustering and day-ahead thermal load forecasting of an existing District Heating. In: 2022 2nd International Conference on Energy Transition in the Mediterranean Area (SyNERGY MED). IEEE; 2022. p. 1–6.
[19] Lumbreras M, Martin-Escudero K, Diarce G, Garay-Martinez R, Mulero R. Unsupervised clustering for pattern recognition of heating energy demand in buildings connected to district-heating network. In: 2021 6th International Conference on Smart and Sustainable Technologies (SpliTech)IEEE; 2021. p. 1–5.
[20] The Swedish Survey Agency. https://www.lantmateriet.se/en (accessed April 16th, 2024).
[21] Boverket – the Swedish National Board of Housing, Building and Planning. https://www.boverket.se/en/start/ (accessed April 16th, 2025).
[22] JdB. et al. sklearn.cluster.KMeans. https://scikit-learn.org/stable/modules/generated/sklearn.cluster.KMeans.html (accessed April 16th, 2024).
[23] JdB. et al. sklearn.cluster.AgglomerativeClustering. https://scikit-learn.org/stable/modules/generated/sklearn.cluster.AgglomerativeClustering.html (accessed April 16th, 2024).
[24] Rushirajsinh Z. The elbow method: finding the optimal number of clusters. https://medium.com/@zalarushirajsinh07/the-elbow-method-finding-the-optimal-number-of-clusters-d297f5aeb189 (accessed May, 23th, 2024).
[25] www.geeksforgeeks.org. K means clustering – introduction. https://www.geeksforgeeks.org/k-means-clustering-introduction/ (accessed April 16th, 2024).
[26] Wijaya CY. Breaking down the agglomerative clustering process. https://towardsdatascience.com/breaking-down-the-agglomerative-clustering-process-1c367f74c7c2 (accessed April 16th, 2024).
[27] Oró E, Taddeo P, Salom J. Waste heat recovery from urban air cooled data centres to increase energy efficiency of district heating networks. Sustain Cities Soc 2019;45:522–42.
[28] Murphy AR, Fung AS. Techno-economic study of an energy sharing network comprised of a data centre and multi-unit residential buildings for cold climate. Energ Build 2019;186:261–75.
[29] Antal M, et al. Reuse of data center waste heat in nearby neighborhoods: a neural networks-based prediction model. Energies 2019;12(5):814.
[30] Giunta F, Sawalha S. Techno-economic analysis of heat recovery from supermarket's CO2 refrigeration systems to district heating networks. Appl Therm Eng 2021;193:117000.
[31] Sintong JE. Data centres as prosumers: a techno-economic analysis; 2023.
[32] Exergi S. Prices for condominium associations. https://www.stockholmexergi.se/bostadsrattsforening/vadkostardetbostadsrattsforening/ (accessed June 6th, 2024).
[33] Raka Adrianto L, Grandjean P-A, Sawalha S. Heat recovery from CO2 refrigeration system in supermarkets to district heating network. 13th IIR Gustav Lorentzen Conference. 2018.
[34] Levihn F. CHP and heat pumps to balance renewable power production: lessons from the district heating network in Stockholm. Energy 2017;137:670–8.
[35] Arizton. Arizton. Nordic Data Center Construction Market – Industry Outlook & Forecast 2023-2028; 2023.
[36] Steuer AJD, Magnius R, Arias J, Sawalha S. Techno-economic analysis of heat export from supermarket refrigeration systems: field measurements analysis of three case studies. Presented at the 26th IIR International Congress of Refrigeration, Paris; 2023.
[37] Gao S, Jurasz J, Li H, Corsetti E, Yan J. Potential benefits from participating in day-ahead and regulation markets for CHPs. Appl Energy 2022;306:117974.
[38] Beiron J, Göransson L, Normann F, Johnsson F. A multiple system level modeling approach to coupled energy markets: Incentives for combined heat and power generation at the plant, city and regional energy system levels. Energy 2022;254:124337.
[39] Holmér P, Ullmark J, Göransson L, Walter V, Johnsson F. Impacts of thermal energy storage on the management of variable demand and production in electricity and district heating systems: a Swedish case study. Int J Sustain Energ 2020;39(5):446–64.

Nomenclature

Nomenclatures and abbreviations

EPC: Energy Performance Certificates
GIS: Geographic Information Systems
ODH: Open District Heating
TWh: Terawatt-hour
COP: Coefficient of Performance
MCOH: Marginal Cost of Heat
SCOP: Seasonal Coefficient of Performance
DH: District Heating
DBSCAN: Density-Based Spatial Clustering of Applications with Noise
GMM: Gaussian Mixture Model
KMeans: K-means clustering algorithm
HAC: Hierarchical Agglomerative Clustering
UEUs: Urban Energy Units
RHI: Renewable Heat Incentive

Parameters and Variables

n_clusters: Number of clusters for clustering algorithms
eps: Maximum distance between two samples for DBSCAN
min_samples: Minimum number of points required to form a "core" point in DBSCAN
affinity: Metric used to compute similarity in Spectral Clustering
reg_covar: Regularization for ensuring covariance matrices are positive semi-definite in GMM
max_iter: Maximum number of iterations for GMM
imputer_strategy: Strategy for imputing missing values (mean strategy)
scaler: Standardization method (StandardScaler)
Qj,i: Heat supply for each source in the cluster
Qi: Total heat demand of the cluster
KMeans Centroids: Centroids of clusters identified by KMeans
Ward's Method: Method used in Agglomerative Clustering to minimize the variance within a cluster
Cluster Center Locations: Locations of cluster centers as identified by different clustering methods