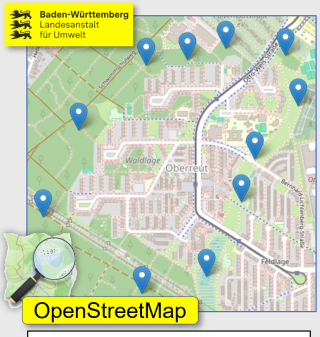
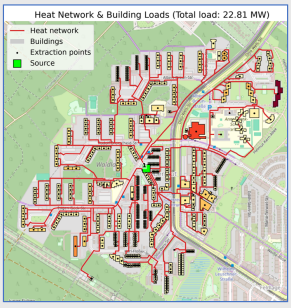
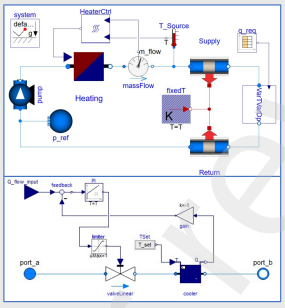
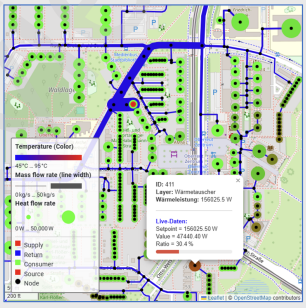
Graphical Abstract

Automated Generation and Multi-Level Validation of District Heating Network Models

Fatih Tüzün, Hüseyin K. Çakmak, Veit Hagenmeyer

Validated Architecture for Heat-Grid Generation and Simulation using Open Data

eHeatGen		eHeatSim	
Geospatial	Network	Simulation	Visualization
			
Open geospatial data are automatically processed to identify buildings and extract their heat demand from LUBW.	From the extraction points, the topology is automatically generated, hydraulically sized, and georeferenced.	The generated network is exported to Modelica. The new consumer model enables bottleneck identification.	Simulation results are visualized interactively on the map, showing temperatures, mass flows, and consumer loads per node.

Preprint not peer reviewed

This preprint research paper has not been peer reviewed. Electronic copy available at: https://ssrn.com/abstract=7674251

Highlights

Automated Generation and Multi-Level Validation of District Heating Network Models

Fatih Tüzün, Hüseyin K. Çakmak, Veit Hagenmeyer

- Automated GIS-to-Modelica workflow for rapid generation of district heating digital twins.
- Scalable framework and practical applicability demonstrated through a case study of an urban district encompassing roughly 450 buildings.
- Multi-level validation using real-world operational data for physical and topological fidelity.
- Proven diagnostic capability for identifying under-dimensioned pipes and supply deficits.

Preprint not peer reviewed

This preprint research paper has not been peer reviewed. Electronic copy available at: https://ssrn.com/abstract=7674251

Automated Generation and Multi-Level Validation of District Heating Network Models*

Fatih Tüzün[a], Hüseyin K. Çakmak[a] and Veit Hagenmeyer[a]

[a]Karlsruhe Institute of Technology, Institute for Automation and Applied Informatics, Hermann-von-Helmholtz-Platz 1, Eggenstein-Leopoldshafen 76344, Baden Württemberg, Germany

ARTICLE INFO

Keywords:
District Heating Networks
Digital Twin
GIS-based Modeling
Modelica
Control Engineering

ABSTRACT

This paper presents a novel automated workflow for the generation of validated digital models of district heating networks to support the decarbonization of urban heating systems. While high-resolution models are essential for optimizing urban heating, synthetic models can serve for planning of future heat grids. As such, synthetic network models constitute a central planning instrument for the heat transition, enabling the design and strategic assessment of future heat grid topologies before costly infrastructure investments are made. However, a significant gap remains between simplified synthetic models and computationally expensive full-scale simulations. To address this challenge, this paper presents an automated workflow that translates GIS data into Modelica models using comparative optimization algorithms for topology generation and automated hydraulic sizing. The framework's accuracy is proven through a rigorous three-level validation procedere using real-world data from the KIT Campus Nord (Germany), evaluating physical consumer behavior, topological fidelity, and computational efficiency. Through diagnostic simulations of network deficits, the present paper demonstrates that the proposed method achieves a critical balance between efficiency and physical accuracy. This framework provides a scalable tool for the rapid sensitivity analysis and optimization of urban heat networks, essential for the transition toward sustainable smart city infrastructures.

1. Introduction

1.1. Context

The transition to climate-neutral cities requires precise planning of district heating networks as a cornerstone of urban energy infrastructure. According to the European Commission's Green Deal and REPowerEU [1] initiatives, the electrification of industry, transportation, and buildings is essential to achieve carbon neutrality by 2050 [2]. District heating systems are expected to play a pivotal role in this transformation, particularly through integration of renewable heat sources, large-scale heat pumps, and thermal energy storage solutions.

However, smart energy systems demand high-resolution models to accurately assess the impact of these technologies on existing and future network infrastructures. The dynamic interaction between fluctuating renewable heat supply, time-varying consumer demand patterns, and hydraulic constraints necessitates simulation tools that capture both temporal and spatial dependencies within the network [3]. Simplified models that neglect pressure-dependent consumer behavior or thermal inertia of distribution pipes may lead to suboptimal design decisions, resulting in either over-dimensioned infrastructure with unnecessary capital expenditure or under-dimensioned networks prone to supply deficits during peak demand periods.

Furthermore, the increasing complexity of multi-source heating networks, in which industrial waste heat, geothermal sources, and conventional plants operate in parallel, requires modeling frameworks capable of handling distributed generation architectures. Traditional single-source assumptions no longer reflect the reality of modern urban heating systems, for which flexibility and redundancy are critical for ensuring supply security amid the energy transition [4].

Despite these requirements, a significant gap remains between simplified synthetic models used for rapid scenario analysis and computationally expensive full-scale simulations employed for detailed engineering studies. Urban planners and utility operators lack scalable tools that bridge this divide, enabling both rapid sensitivity analysis across multiple design alternatives and high-fidelity diagnostic simulations for identifying hydraulic bottlenecks. The present paper addresses this challenge by presenting an automated workflow that translates GIS data into validated Modelica models suitable for thermo-hydraulic simulation.

1.2. Literature Review

Current literature on district heating network modeling reveals a dualism between highly simplified approaches and computationally expensive high-fidelity simulations. This gap challenges urban energy planners who require both accuracy and efficiency in their decision-making tools.

Simplified Network Models. Traditional approaches often employ synthetic "box models" or aggregated demand representations that neglect the hydraulic and thermal dynamics of distribution networks [3, 5, 6]. While computationally efficient, these models fail to capture pressure-dependent consumer behavior, leading to systematic errors

*This work was conducted within the framework of the Helmholtz Programs "Materials and Technologies for the Energy Transition" (MTET) as well as "Energy System Design" (ESD) and funded by the KIT Energy Lab networking project E-Industry.

Fatih.Tuezuen@kit.edu (F. Tüzün)
ORCID(s): 0009-0004-1226-7842 (F. Tüzün); 0000-0002-1463-7606 (H.K. Çakmak); 0000-0002-3572-9083 (V. Hagenmeyer)

This preprint research paper has not been peer reviewed. Electronic copy available at: https://ssrn.com/abstract=7674251

Nomenclature

Acronyms/Abbreviations

AOI	Area of Interest
FMU	Functional Mock-up Unit
GIS	Geographic Information System
KIT	Karlsruhe Institute of Technology (Germany)
MAE	Mean Absolute Error
MST	Minimum Spanning Tree
MSSP	Multi-Source Shortest Path
OSM	OpenStreetMap
RMSE	Root Mean Square Error
SPT	Shortest Path Tree

Symbols/Variables

S_n	Combined source selection score
d_n	Normalized building density
$\hat{s}_n$	Normalized heat-weighted distance score
α, β	User-defined weights
L_{ring}	Maximum Euclidean distance for ring closing
$L_{\min}$	Minimum graph distance for ring closing
$\dot{Q}$	Thermal load / heat flow
$\dot{m}$	Mass flow rate
c_p	Specific heat capacity of water
ρ	Density of the fluid
$v_{\max}$	Maximum permissible flow velocity
d_{uv}	Inner diameter of pipe segment
$x(t)$	Internal state of the PI controller
$y(t)$	Valve opening signal $[0, 1]$

Superscripts/Subscripts

VL	Supply
RL	Return
in, out	Inlet, Outlet
set, req	Setpoint, Required
E, G	Euclidean, Graph
uv	Pipe segment from node u to node v

in component sizing. This results in either oversizing infrastructure or undersizing pipes. Such simplifications are particularly problematic when analyzing network resilience or diagnosing hydraulic bottlenecks.

High-Fidelity Simulation Tools. At the opposite end of the spectrum, detailed thermo-hydraulic models [7, 8] built in simulation environments such as Modelica [9] or TRNSYS [10] provide exceptional physical accuracy. However, these models typically require extensive manual effort to construct, limiting their applicability to small-scale case studies or well-documented existing networks. The manual translation from GIS data to simulation-ready models remains a significant bottleneck in large-scale urban energy system analysis.

Automated Graph-Based Approaches. Recent developments have sought to bridge this gap through automated graph-based model generation.

UESgraphs, developed at RWTH Aachen University's E.ON Energy Research Center, represents a notable advancement by extending the NetworkX graph class to represent buildings and energy networks [11, 12, 13]. Key features include physics-based hydraulic network sizing, Modelica template generation, quasi-static simulations using pandapipes, and demand estimation via TEASER integration [14]. However, while UESgraphs provides foundational graph-based representation and basic hydraulic sizing, its topological optimization capabilities remain limited. Specifically, the framework does not natively support comparative analysis of different topology generation strategies (MST [15] vs. SPT [15] vs. Steiner [16]) or automated redundancy modeling through ring-closing mechanisms.

Similarly, DistrictHeatingSim, developed at the University of Applied Sciences Zittau/Görlitz, provides a comprehensive GUI-based workflow for planning sustainable district heating networks [17, 18]. The tool integrates OSM data download, automatic heat network generation based on building and generator locations, thermohydraulic calculations using pandapipes, and economic analysis following VDI 2067 methodology. DistrictHeatingSim provides an accessible user interface with project management capabilities and supports cost-efficiency optimization. However, the network generation relies on standard shortest-path generic graphs without multi-method comparative analysis, and the pandapipes-based simulation focuses on steady-state hydraulic calculations rather than dynamic thermo-hydraulic behavior with pressure-driven consumers.

Research Gap and Contribution. Table 1 summarizes the key differences between existing approaches and the proposed framework. While tools like DistrictHeatingSim and UESgraphs provide significant functionality for steady-state analysis and GIS integration, a critical gap remains in the automated generation of network graphs that preserve the street topology and are physically validated. Specifically, existing frameworks use connectivity approaches that produce direct line connections and lack automated redundancy modeling and systematic hyperparameter tuning.

To address these limitations, this present paper introduces an end-to-end automated workflow that bridges the gap between raw geospatial data and simulation-ready Modelica models. The primary contributions of this paper are:

- A multi-method optimization strategy (Shortest Path Tree, Minimum Spanning Tree, and Steiner Tree approximation) coupled with an automated ring-closing mechanism for redundancy modeling, with key hyperparameters tuned via Bayesian optimization using Optuna [19] to minimize structural deviation from real-world networks.

This preprint research paper has not been peer reviewed. Electronic copy available at: https://ssrn.com/abstract=7674251

Table 1
Comparison of automated district heating network modeling frameworks. Feature assessment based on the respective publications and source code repositories (accessed September 2026).

Feature	DistrictHeatingSim	UESgraphs	The Present Work
GIS-to-Model Automation	✓ (OSM)	✓ (via [12])	✓ (interactive AOI)
Topological Optimization	Multi (OSMnx, MST, Steiner)	Manual	Multi (SPT, MST, Steiner)
Multi-Source Support	✓	Partial	✓ (optimized placement)
Ring/Redundancy Modeling	✗	✗	Automated ring-closing
Hydraulic Sizing	✓ (pandapipes)	✓ (physics-based)	Recursive load aggregation
Pipe Model Differentiation	✗	✗	Adaptive (Static/Plug-Flow)
Simulation Engine	pandapipes (quasi-stationary)	Modelica/pandapipes/Python	Modelica
Pressure-Driven Consumers	✗	✓ (valve-based)	✓ (VarTVarDp)
Economic Analysis	✓ (VDI 2067)	✗	✗
Validation Framework	System-level plausibility	Verification vs. manual model	Multi-level (real-world data)
Hyperparameter Optimization	✗	✗	✓
GUI	✓ (PyQt6)	✗	✓

- Adaptive pipe model selection from the AixLib library [20], assigning computationally efficient static models to short segments while utilizing physically accurate plug-flow models for long distribution lines to capture thermal inertia and axial time delays.
- A new *VarTVarDp* pressure-driven consumer model enabling diagnostic simulations for the automatic identification of hydraulic bottlenecks and undersized components.
- A three-dimensional validation against operational data from the KIT Campus Nord network (Germany): (i) physical component behavior, (ii) topological fidelity, and (iii) hydraulic capacity.
- A demonstration of spatial scalability through application to an urban area in Karlsruhe-Oberreut (Germany) comprising approximately 450 buildings, covering automated plant placement, network generation, and diagnostic identification of undersupplied network regions.

The modular architecture allows for the rapid generation of networks for arbitrary urban areas via interactive polygon selection, reducing development time.

1.3. Aims of the Paper

The objective of this research is to demonstrate that an automated, GIS-based approach can achieve a level of topological and physical fidelity comparable to manually constructed models. Specifically, the present paper aims to: (i) Methodologically validate the efficiency of the end-to-end pipeline in reducing modeling effort while maintaining street topology. (ii) Diagnostically prove the capability of the pressure-driven consumer models to identify supply deficits, thereby supporting proactive network reinforcement planning. (iii) Practically provide a scalable tool for urban planners to perform rapid sensitivity analyses and optimize the transition toward sustainable smart city heating infrastructures.

2. Methodology, Optimization, and Validation

The introduced novel approach is divided into four main stages: geospatial data handling, network synthesis, simulation model generation, and result visualization. The following subsections will detail the modeling methodology and the validation framework.

2.1. Methodology

The overarching system architecture is designed as a modular pipeline as shown in Figure 1, ensuring that each step can be independently adjusted and validated. The data flow begins with geospatial data, where user-defined areas are captured via interactive maps. Subsequently, tabular processing retrieves geocoded address points and heat demand data from open data sources such as WärmeatlasBW [21], which provides information at building level or aggregated by street segment. In the *System Synthesis* stage, a hydraulic graph is generated, and pipe diameters are calculated based on recursive load aggregation. Finally, *Model Translation* automatically converts the network layout into Modelica classes, which are then compiled into Functional Mock-up Units for web-based co-simulation.

2.1.1. System Boundaries

The district heating area is defined using an interactive polygon-based extraction tool. By utilizing the *Folium* [22] library, users can define a precise area of interest (AOI) on a map. This polygon serves as the spatial filter for all subsequent data extraction steps, ensuring that only relevant building points and road segments from OpenStreetMap (OSM) [23] are processed.

2.1.2. Data Pipeline and Graph Generation

The construction of the hydraulic network follows a four-step process:

1. Spatial Layer Extraction. Address points are extracted from open data portals and filtered by the AOI polygon. These points are snapped to the nearest OSM road edges

This preprint research paper has not been peer reviewed. Electronic copy available at: https://ssrn.com/abstract=7674251

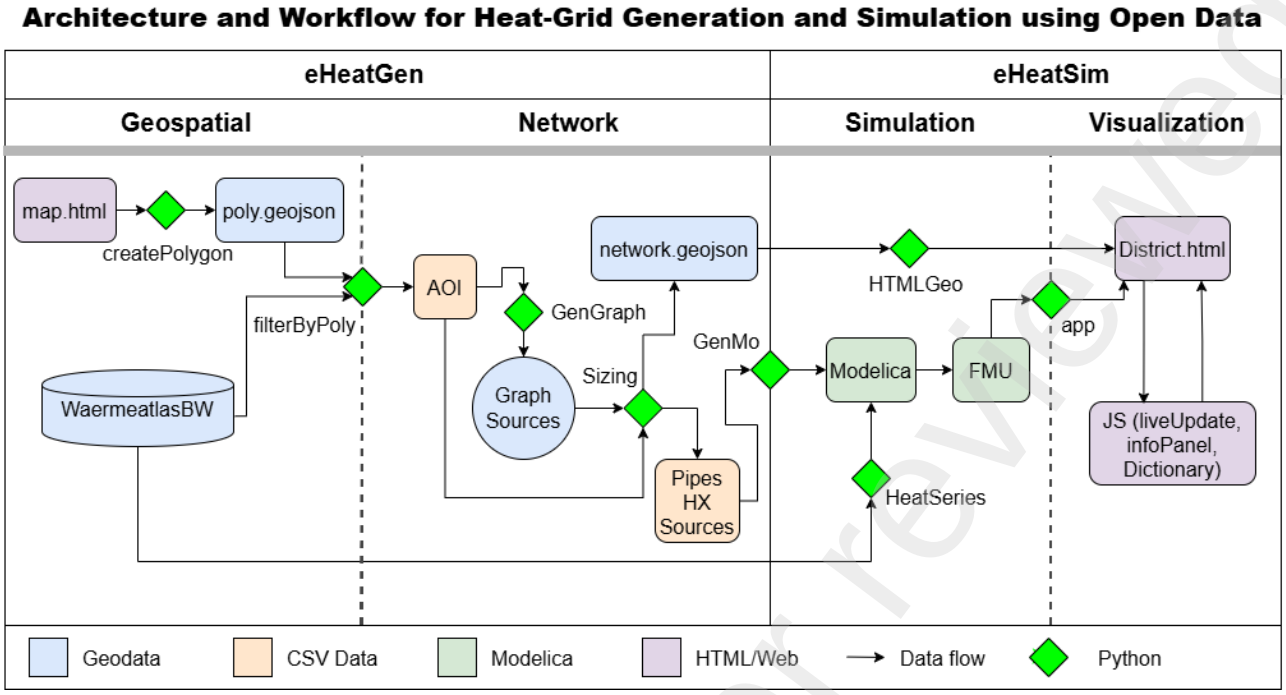


Figure 1: The automated workflow from geospatial data extraction to the generation of a validated Modelica-based digital twin.

by minimizing the Euclidean distance between the building point b_i and the orthogonal projection onto the edge line e_{uvk}. The modeling of the heat generation structure is differentiated according to the total heating load: Where the heating load in the selected area exceeds 20 MW, the heating loads are aggregated at road segment level and modeled as a single consumer; where the total heating load in the respective area reaches 50 MW or more, the implementation of several spatially distributed sources ensures system stability and load distribution.

2. Topology Generation and Network Optimization. The framework generates the network topology in a two-stage process: source selection and graph synthesis.

Stage 1: Source Selection. The proposed framework implements a multi-source optimization approach. This allows the system to handle multiple heating plants simultaneously. The number of sources is determined by a parameter that specifies the maximum load per source and the total annual heat demand. The urban area is then divided into branches of equal weight. The optimal location for each source is determined by a combined score S_n that balances building density d_n and the normalized heat-weighted distance $\hat{s}_n$ within each branch:

$$S_n = \alpha d_n + \beta(1 - \hat{s}_n) \tag{1}$$

where α and β are user-defined weights ($\alpha + \beta = 1$). While α governs the preference for peripheral locations with low building density, β drives the placement toward major load centers. By minimizing S_n, the framework places sources in areas of low building density (small d_n) while simultaneously maximizing the heat-weighted proximity $\hat{s}_n$ to major consumers.

Stage 2: Graph Synthesis. Once the sources are fixed, the framework provides three optimization algorithms to synthesize the network. To formally distinguish these strategies, their respective objective functions are defined based on a weighted graph $G = (V, E, w)$, where V represents the nodes, E the potential pipe segments, and w the edge weights.

(1) Shortest Path Tree (SPT): Utilizing a Multi-Source Dijkstra approach [24], the SPT minimizes the individual distance from each consumer node i to its nearest source s. The objective is to minimize the sum of shortest paths:

$$\min \sum_{i \in V} \text{dist}(s, i) \tag{2}$$

This strategy is prioritized when minimizing pressure drops and thermal losses, which are directly proportional to path length, is the primary objective.

(2) Minimum Spanning Tree (MST): The MST focuses on the global minimization of the total network length using Kruskal's algorithm [25]. The objective function minimizes the sum of all edge weights $w(e)$ included in the tree T:

$$\min \sum_{e \in E_T} w(e) \tag{3}$$

This represents a cost-driven approach, aiming to reduce the total investment in infrastructure and materials.

This preprint research paper has not been peer reviewed. Electronic copy available at: https://ssrn.com/abstract=7674251

(3) Steiner Tree Approximation: To minimize overall network infrastructure cost while connecting only designated sources and high-load buildings, a Steiner Tree formulation is employed [16, 26]. This approach finds a minimum-weight connected subgraph $T = (V_{ST}, E_{ST})$ that spans a subset of terminal nodes $K \subseteq V$, introducing intermediate junction nodes ($V_{ST} \setminus K$, or Steiner points) where beneficial:

$$\min_{E_{ST} \subseteq E} \sum_{e \in E_{ST}} w(e) \quad \text{subject to } K \subseteq V_{ST}, \; T = (V_{ST}, E_{ST}) \text{ is connected.} \tag{4}$$

Classic approximation heuristics [26] solve this NP-hard problem by evaluating shortest-path distances between terminals to build a distance network, upon which a minimum spanning tree is constructed.

Buildings are sorted by their distance to the road and connected iteratively. This allows the formation of connection chains. Each building is linked to the nearest network element either a street segment or an existing house connection. In each iteration, the identified edge is split at the point of orthogonal projection to insert a junction node, to which the building is then connected via a terminal edge. The computational complexity of this process is approximately $O(n \log(m + n))$, where n denotes the number of buildings and m represents the number of edges in the initial street network.

3. Redundancy and Ring Closing. Since real-world urban networks are rarely pure trees and often incorporate rings for reliability and pressure stabilization, the framework includes a *Ring Closing* mechanism. After the initial tree (SPT, MST, or Steiner) is generated, the algorithm searches for pairs of nodes that are spatially close but topologically distant. If the Euclidean distance between two nodes is less than a threshold L_{ring} and their distance within the graph exceeds a minimum path length L_{min}, a redundant edge is added. L_{ring} is scaled with the system's structural size. The minimum path length L_{min} is set in direct proportion to L_{ring}.

$$\text{Add Edge}(u, v) \quad \text{if} \quad (d_E < L_{ring}) \wedge (d_G > L_{min}) \tag{5}$$

This ensures that a redundant edge is exclusively added between nodes that are physically close in Euclidean space but currently separated by a long routing path within the existing topology. The insertion of this redundant edge connects two separate path segments, thereby closing a loop and establishing a meshed network topology.

4. Hydraulic Sizing. Once the topology is fixed, the pipe diameters are determined through a recursive load aggregation. Starting from the leaf nodes (consumers), the heat loads are summed upstream toward the sources. A Shortest Path Tree (SPT) is calculated. This means that the path along which heat flows from the building to the source during normal operation is defined. Pipes that are not part of this shortest path (such as the redundancy rings) would, according to this logic, have a load of 0.0. To prevent redundancy lines from being designed with a minimum diameter, if a line is marked as a redundancy ring and carries no load during normal operation, it is assigned the greater of the two node loads at its adjacent nodes. For each edge (u, v), the required inner diameter d_{uv} is derived from the cumulative heat flow Q_{uv}:

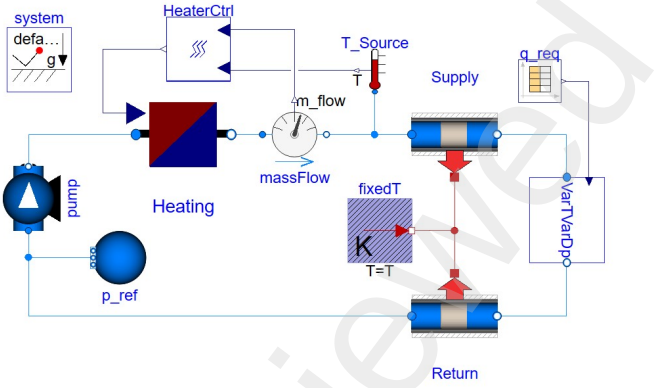


Figure 2: Core model applied to generate the district heating network.

$$\dot{m} = \frac{Q_{uv} \times 10^3}{c_p \Delta T}, \quad A = \frac{\dot{m}}{\rho v_{max}}, \tag{6}$$

$$d_{uv} = \max\left(\sqrt{\frac{4A \cdot SF}{\pi}}, d_{min}\right) \tag{7}$$

where $\dot{m}$ is the mass flow rate, Q_{uv} is the thermal heat flow of the pipe branch to be designed, c_p is the specific heat capacity of water, ΔT is the design temperature difference, A is the cross-sectional area, ρ is the fluid density, v_{max} is the maximum permissible flow velocity, SF is the safety factor, d_{min} is the minimum allowable diameter, and d_{uv} is the resulting pipe diameter.

2.1.3. Model Translation and Core Physics

The final stage is the automated translation of the graph into a Modelica-based simulation model. To optimize the balance between simulation speed and physical fidelity, the framework introduces an adaptive pipe selection logic.

Model Parameterization. The heat source and pump are automatically dimensioned from the nominal heat loads of all exchangers. The total heating capacity is calculated as the sum of all individual nominal heat flows.

The required mass flow is derived from the energy balance with a fixed temperature spread of $\Delta T = 35K$. The pipe network pressure drop Δp_{supply} is computed using Darcy-Weisbach along the hydraulically critical path (maximum loss among all consumer branches). Assuming a symmetric return network, the total system pressure becomes:

$$\Delta p_{\text{nominal}} = 2 \cdot \Delta p_{\text{supply}} + \Delta p_{\text{exchanger}} \tag{8}$$

Preprint not peer reviewed
This preprint research paper has not been peer reviewed. Electronic copy available at: https://ssrn.com/abstract=7674251

with $\Delta p_{\text{exchanger}} = 15000 Pa$ for a single heat exchanger.

The `ControlledPump` is parameterized as with $\dot{V} = \{0, 1.2 \cdot \dot{m}_{\text{nominal}}/\rho\}$ and $H = \{1.2 \cdot \Delta p_{\text{nominal}}/(\rho \cdot g), 0\}$. The pump characteristic uses a linear interpolation between shut-off head (at zero flow) and the design point. All parameters are computed during model generation and written directly to the Modelica file.

Pipe Model Differentiation. The system distinguishes between two types of pipe model based on the physical characteristics of the segment:

- **Static Pipes:** Used for short segments (Length < Threshold) or small diameters (Diameter < Threshold). These models ignore axial time delays to increase the solver speed.
- **Plug Flow Pipes:** Used for wide or long distribution lines. These utilize the *AixLib* [20] plug-flow formulation to account for thermal inertia and axial delay $\tau = L/v$, where τ is the transport delay, L is the pipe length, and v is the flow velocity.

This selection can be performed automatically by the script or overridden manually by the user for specific critical sections of the network.

2.1.4. Core Model of the Heating Network

The generated Modelica model represents the district heating network as a closed-loop system (see Figure 2) as a minimal core model. The left-hand side represents a source element and consists of a heating element and a pump. The right-hand side contains the piping, which consists of a supply and return line. The right-hand side is scaled up according to the graph. The new consumer model itself is covered in detail later.

The New Pressure-Driven Consumer Model. A new reduced consumer model `VarTVarDp` (see Figure 3) represents heat consumption by taking the required heat flow rate as setpoint. Unlike simplified models that prescribe a fixed mass flow, this model is pressure-driven, meaning the actual heat delivery is a function of the available pressure difference Δp across the consumer node.

System Topology and Pressure Control. The network's hydraulic head is provided by a centrifugal pump model (`ControlledPump`) from the Modelica Standard Library [9]. The pump connects the return line back to the supply line, maintaining a nominal pressure setpoint p_{set} at the pump outlet. This closed-loop configuration allows for the simulation of complex hydraulic interactions, such as pressure drops across long-distance distribution pipes and the impact of flow reversals in ring-structured networks.

The consumer model is composed of three coupled subcomponents:

(1) Linear Control Valve: Acts as the primary flow regulator. The mass flow rate $\dot{m}$ is determined by the valve

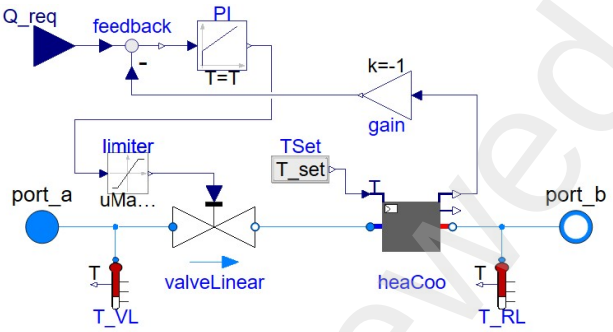


Figure 3: Closed-loop consumer model with controller.

opening $y(t)$ and the pressure difference Δp between the supply and return pipes:

$$\dot{m} = f(y(t), \Delta p) \tag{9}$$

(2) Sensible Cooler: Simulates the heat extraction process. The energy balance is governed by the sensible heat relation, where the heat flow Q_{flow} depends on the mass flow and the temperature drop:

$$Q_{flow} = \dot{m} c_p (T_{in} - T_{out}) \tag{10}$$

Either the outlet temperature T_{out} or the temperature difference $T_{in} - T_{out}$ can be prescribed.

To prevent numerical instabilities and physically unrealistic behavior at very low mass flow rates (e.g., when heat demand is low), the temperature difference is scaled by a smoothing function $\sigma(\dot{m})$. The actual temperature difference ΔT_{act} is calculated as:

$$\Delta T_{act} = \Delta T_{set} \cdot \sigma(\dot{m}) \tag{11}$$

where $\sigma(\dot{m})$ is a sigmoid function based on the hyperbolic tangent, ensuring a smooth transition to zero as the mass flow $\dot{m}$ drops below a minimum threshold $\dot{m}_{min}$:

$$\sigma(\dot{m}) = \frac{1}{2}\left[1 + \tanh\left(k_{smooth} \cdot (\dot{m} - \dot{m}_{min})\right)\right] \tag{12}$$

where, k_{smooth} determines the steepness of the transition. For the dynamic case, the outlet temperature T_{out} is governed by a first-order differential equation:

$$\frac{dT_{out}}{dt} = \frac{1}{k \cdot \tau}\left(T_{in} - \Delta T_{act} - T_{out}\right) \tag{13}$$

where τ is the thermal time constant and k is a correction factor that adjusts the dynamics based on the normalized mass flow to ensure stability during steady-state transitions.

(3) PI Controller (*Heater_Ctrl*): To track the required heat flow $\dot{Q}_{req}$, a proportional-integral (PI) controller dynamically adjusts the valve opening $y(t) \in [0, 1]$. The controller processes the error $e(t) = \dot{Q}_{req} - \dot{Q}_{actual}$ between demanded and delivered heat flow, ensuring that the substation reacts to transient thermal loads. When the valve saturates at $y = 1$ while $e(t) > 0$ persists, the substation is

This preprint research paper has not been peer reviewed. Electronic copy available at: https://ssrn.com/abstract=7674251

hydraulically constrained — this constitutes the diagnostic indicator for undersizing discussed below. The controller is defined as follows:

$$\dot{x}(t) = \frac{1}{T}\, e(t) - a\, x(t) \tag{14}$$

$$y(t) = k\,(x(t) + e(t)) \tag{15}$$

where $e(t) = \dot{Q}_{req} - \dot{Q}_{actual}$ denotes the control error, $x(t)$ is the (normalized) integral state of the controller, and $y(t)$ is the valve opening signal. For $a = 0$, the controller equations reduce to a classical PI controller with transfer function $k\,(1 + 1/(Ts))$. where T is the integrator time constant and $a \geq 0$ is an optional decay rate of the integral state, which gradually forgets accumulated control error and thereby mitigates integrator windup after prolonged actuator saturation. In the present paper, $a = 0$ is used.

Diagnostic Utility for Network Sizing. The coupling of a closed-loop hydraulic system with pressure-driven consumers allows the framework to act as a diagnostic tool. In an optimally sized network, the pump provides sufficient pressure to allow the valve to reach the required $\dot{m}$ to satisfy Q_{req}. However, in an undersized network, the excessive pressure drop Δp_{loss} in the pipes reduces the available pressure at the consumer node. Consequently, the valve may reach a fully open state ($y = 1$) without achieving the necessary mass flow, resulting in a failure to meet the heating demand. This behavior allows for the precise identification of hydraulic bottlenecks within the urban grid.

2.2. Validation Framework

The validation is based on a high-resolution dataset from the 'Campus Nord' of the Karlsruhe Institute of Technology (KIT) in Germany. Campus Nord supplies more than 300 buildings via a second-generation district heating network with a total single pipe length of 17 km. The meshed network is fed at a single central injection point into three main supply lines ('north', 'east', 'south'), with supply temperatures of up to 110 °C at 6–7 bar. Annual heat delivery amounts to 70–80 GWh [27].

The primary performance indicators for the physical validation are the Mean Absolute Error (MAE) and the Root Mean Square Error (RMSE). The MAE provides a linear representation of the average error, defined as:

$$\mathrm{MAE} = \frac{1}{n}\sum_{i=1}^{n} |y_i - \hat{y}_i| \tag{16}$$

where y_i represents the measured value and $\hat{y}_i$ denotes the simulated value. To penalize larger outliers the RMSE is utilized:

$$\mathrm{RMSE} = \sqrt{\frac{1}{n}\sum_{i=1}^{n}(y_i - \hat{y}_i)^2} \tag{17}$$

2.2.1. Model Validation (Bottom-Up)

The physical accuracy of the system is validated through a bottom-up approach, ensuring that the building blocks of the network behave according to observed physics.

Consumer Model. The consumer model is validated using time series from the Campus Nord. The comparison focuses on the thermal load $\dot{Q}$, the supply and return temperatures (T_{VL}, T_{RL}), and the mass flow rates $\dot{m}$. The model's ability to replicate the dynamic thermal response of buildings is assessed by comparing the simulated heat demand against the measured energy consumption over a period of 115.7 days. For validation, the measured temperature is used as the input instead of the heating system.

Pipe Model. To minimize computational overhead and ensure reliability, the pipe models are based on the established *AixLib* library [20]. This library provides pre-validated thermal-hydraulic components specifically designed for district heating simulation, covering pressure drop calculations and heat loss characteristics.

2.2.2. Topological Validation (Graph Fidelity)

Beyond physical parameters, the structural accuracy of the automatically generated network is compared against the As-Built technical drawings of the Campus Nord. To ensure a fair comparison despite differences in granularity and oversegmentation in the legacy data, a topological alignment process was implemented. This involved filtering out intermediate transit nodes (degree = 2), thereby focusing the validation on the "network skeleton" consisting of critical nodes [28]. The actual locations of the heat exchangers, along with the measured annual heat loads, are used for generic model generation.

Spatial Fidelity. To quantify the geometric accuracy of the GIS-based extraction, a mapping is performed between the generic nodes N_{gen} and the physical nodes N_{real} using nearest-neighbor assignment with a distance threshold [29]. The spatial fidelity is measured by the Mean Absolute Error (MAE) of the Euclidean distance d:

$$\mathrm{MAE}_{dist} = \frac{1}{|N_{gen}|}\sum_{i=1}^{|N_{gen}|} \min_{j \in N_{real}} \|\mathbf{p}_i - \mathbf{p}_j\| \tag{18}$$

where $\mathbf{p}$ denotes the coordinate vector of the respective nodes. This metric is analogous to the Hausdorff distance used in shape matching applications [30].

Structural and Functional Connectivity. The global robustness and the strategic organization of the network are validated through three complementary metrics. First, the *algebraic connectivity* is analyzed, defined as the second smallest eigenvalue λ_2 of the graph Laplacian matrix $L = D - A$ (where D is the degree matrix and A is the adjacency matrix). This metric, introduced by Fiedler [31], quantifies how well-connected a graph is as a whole.

The relative error in algebraic connectivity (Fiedler value λ_2) is calculated as:

$$\mathrm{RE}_{\lambda_2} = \frac{|\lambda_{2,\mathrm{gen}} - \lambda_{2,\mathrm{real}}|}{\lambda_{2,\mathrm{real}}} \tag{19}$$

This preprint research paper has not been peer reviewed. Electronic copy available at: https://ssrn.com/abstract=7674251

Second, the *betweenness centrality* C_B is compared via the Pearson correlation coefficient r to verify if the so-called "hubs" (nodes with high traffic importance) are positioned identically in both models. Betweenness centrality, defined by Freeman [32, 33], measures the fraction of shortest paths passing through each node:

$$C_B(v) = \sum_{s \neq v \neq t} \frac{\sigma_{st}(v)}{\sigma_{st}} \tag{20}$$

where σ_{st} is the total number of shortest paths from node s to node t, and $\sigma_{st}(v)$ is the number of those paths passing through v.

Finally, the *cycle space dimension* $k = |E| - |V| + 1$ is compared to verify that the number of independent loops is preserved, complemented by the average cycle length derived from a fundamental cycle basis obtained via depth-first search [28]. While the fundamental basis is not unique, the cycle count is invariant, and the average length provides an indicative measure of whether redundancy is implemented via short local rings or large peripheral loops.

Transport Efficiency. To evaluate the functional performance of the network, the global transport efficiency is analyzed by comparing the graph diameter (the longest shortest path) and the average shortest path length $\bar{d}$. These metrics are standard measures in network science for characterizing navigation efficiency [34, 35]. The relative error (RE) is calculated to ensure that the synthetic network replicates the real-world travel distances:

$$\mathrm{RE}_{\mathrm{path}} = \frac{|\bar{d}_{gen} - \bar{d}_{real}|}{\bar{d}_{real}} \tag{21}$$

Local Topology and Degree Distribution. To verify the correct ratio of dead-ends, transit lines, and junctions, the node degree distributions are compared. The structural difference is quantified using the Mean Squared Error (MSE) between the normalized histograms (probability mass functions) of the generic distribution P_{gen} and the real distribution P_{real} [36]:

$$\mathrm{MSE} = \frac{1}{k} \sum_{i=1}^{k} (P_{\mathrm{gen}}(i) - P_{\mathrm{real}}(i))^2 \tag{22}$$

where k represents the maximum node degree considered. Degree distribution analysis is a fundamental tool in complex network characterization [37].

Physical Parameter and Hydraulic Capacity. Finally, a direct comparison of pipe attributes is conducted. The relative error is calculated for the total pipe length L. To validate the hydraulic capacity, the distribution of diameters D is compared using the same MSE metric as defined in Eq. (22), where $P(i)$ now represents the normalized frequency of diameter bin i.

The total volumetric capacity V is evaluated using the sum of length-weighted squared diameters [38]:

$$V = \sum_{e \in E} L_e \cdot D_e^2 \tag{23}$$

where L_e and D_e represent the length and diameter of edge e, respectively. For pressurized pipe networks, this metric correlates with pumping energy requirements and capital investment costs [39].

2.3. Iterative Network Optimization

To ensure the topological and physical accuracy of the synthetic district heating networks, an iterative optimization framework was implemented. The generation process employs a multi-stage pipeline consisting of network synthesis, hydraulic sizing, and rigorous validation.

The synthesis phase evaluates the three topological strategies introduced in Section 2 (SPT, MST, and Steiner Tree approximation), supplemented by the ring-closing procedure to introduce structural redundancy. Following topological generation, pipe diameters are calculated based on heat load requirements using recursive load aggregation.

The resulting synthetic networks are compared against a ground-truth network using the multi-metric validation suite. This suite evaluates the metrics explained in 2.2.2.

The calibration process employs Bayesian optimization via the *Optuna* framework [19], which uses Tree-structured Parzen Estimator (TPE) algorithms for efficient hyperparameter search [40]. Three key hyperparameters are optimized: L_{ring}: Maximum Euclidean distance between two nodes for adding redundant ring edges; L_{min}: Minimum graph path length between two nodes required before ring closing is triggered; Method: Categorical selection between {MST, SPT, Steiner}.

This loop iteratively adjusts key hyperparameters to minimize a weighted loss function defined as:

$$\mathcal{L} = \sum_{i=1}^{n} w_i \cdot \frac{\mathrm{Metric}_i}{\mathrm{Normalizer}_i} \tag{24}$$

where w_i represents the relative importance of the i-th validation metric. This iterative approach explores the function space of graph generation using various graph generation methods and parameters.

3. Results

This section first presents a comparative evaluation of the different network generation methods, then details the thermo-hydraulic and physical validation of the resulting topologies, and finally discusses the overall performance and operational trade-offs of the optimized district heating networks.

3.1. Consumer Model Validation

To validate the proposed framework under real-world conditions, a case study is conducted using Building 574 at KIT Campus Nord. Unlike standard office buildings, Building 574 functions as a multi-use facility that accommodates both administrative spaces and specialized laboratory and technical infrastructure, including energy storage research facilities. For this validation, the measured supply temperature was used as a boundary condition instead of simulating

This preprint research paper has not been peer reviewed. Electronic copy available at: https://ssrn.com/abstract=7674251

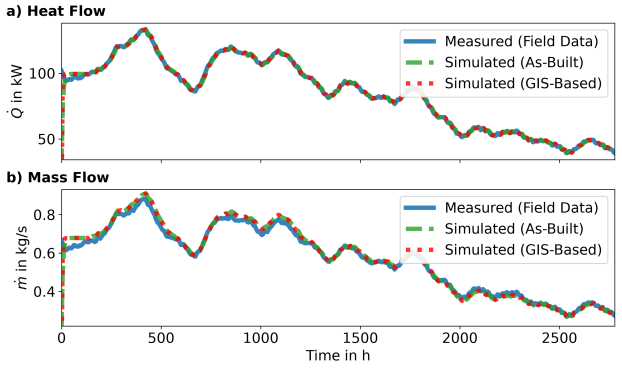


Figure 4: Consumer model validation: Measured vs. simulated heat flow $\dot{Q}$ and mass flow $\dot{m}$ for building 574 of KIT Campus Nord. Top: Heat demand profile. Bottom: Mass flow response. The As-Built network simulation (dashed green) and GIS-based network simulation (dotted orange) both closely track the measured field data (solid blue).

Table 2
Validation metrics: Consumer model performance for building 574

Metric	As-Built	GIS-Based
MAE $\dot{Q}$ [kW]	0.89	1.20
RMSE $\dot{Q}$ [kW]	3.55	4.73
MAE $\dot{Q}$ [% of peak]	0.7	0.9
DTW $\dot{Q}$ [kW]	77.4	103.0
Norm. DTW $\dot{Q}$ [–]	0.191	0.254
MAE $\dot{m}$ [kg/s]	0.015	0.016
RMSE $\dot{m}$ [kg/s]	0.028	0.035
MAE $\dot{m}$ [% of max]	1.7	1.9
DTW $\dot{m}$ [kg/s]	0.507	0.680
Norm. DTW $\dot{m}$ [–]	0.186	0.249
MAE ΔT [K]	1.35	1.42
RMSE ΔT [K]	1.51	2.24
MAE ΔT [% of mean]	3.8	4.0
DTW ΔT [K]	34.0	50.4
Norm. DTW ΔT [–]	0.002	0.003

Norm. DTW: DTW distance normalized by the total variation of the measured signal, i.e. $\widetilde{\mathrm{DTW}}_{\dot{Q}} = \mathrm{DTW} / \sum_{k=2}^{N} |\dot{Q}_k^{\mathrm{meas}} - \dot{Q}_{k-1}^{\mathrm{meas}}|$, analogously for $\dot{m}$; for ΔT: $\widetilde{\mathrm{DTW}}_{\Delta T} = \mathrm{DTW}/(N \cdot \overline{\Delta T}^{\mathrm{meas}})$. Raw DTW distances carry the unit of the signal multiplied by the number of samples $N = 504$.
DTW computed after the dynamic programming approach of Sakoe and Chiba [41].

the heating plant dynamics. In this case, the newly developed component is operated at a constant temperature difference of 35 K. The measured heat demand time series for each building $\dot{Q}_{set}$ was prescribed as input to the PI controller. This pressure-driven control strategy ensures that the actual heat delivery $\dot{Q}_{actual}$ tracks the setpoint while respecting hydraulic constraints. Both measured and simulated datasets were pre-processed with identical low-pass filtering (cut-off frequency: 5e-7 Hz) to eliminate measurement noise.

Heat Flow and Mass Flow Accuracy. Figure 4 presents the time-series comparison between measured and simulated values over a representative heating period of 115.7 days (January–April 2024). In this case, the measured values are compared with the manually created and the generically created simulation models. The results demonstrate excellent agreement for both thermal load and hydraulic behavior across both network topologies (As-Built technical drawing vs. GIS-based synthetic network).

The PI controller successfully regulates the heat flow $\dot{Q}$ by adjusting the valve opening $y(t)$ based on the error between $\dot{Q}_{set}$ and $\dot{Q}_{actual}$.

Table 2 summarizes the quantitative validation metrics. The As-Built network model achieved a Mean Absolute Error (MAE) of 0.89 kW (0.7% of peak load) for heat flow and 0.015 kg/s (1.7% of maximum) for mass flow. The GIS-based synthetic network showed comparable performance with slightly higher deviations: MAE of 1.20 kW (0.9%) for heat flow and 0.016 kg/s (1.9%) for mass flow. These marginal differences confirm that the topological simplifications introduced during automated GIS extraction do not significantly compromise local hydraulic behavior at consumer nodes.

In addition to point-wise error metrics, Dynamic Time Warping (DTW) was applied to assess shape similarity independently of small phase shifts between measurement and simulation, which typically arise from the thermal inertia of the building and the control dynamics of the consumer valve. The normalized DTW distances confirm the same ordering of model quality: 0.191 (As-Built) versus 0.254 (GIS-Based) for heat flow and 0.186 versus 0.249 for mass flow. Since the DTW-based deviations are of the same order as the point-wise MAE, the residual error can be attributed predominantly to amplitude discrepancies rather than to timing offsets, i.e., the models reproduce both the magnitude and the temporal evolution of the consumer load profiles.

Temperature Difference Dynamics. The temperature spread $\Delta T = T_{VL} - T_{RL}$ is a critical indicator of thermal energy transfer efficiency at the consumer station. Figure 5 compares the temperature difference between measurement and both simulation models. The mean ΔT values show strong agreement: 35.52 °C (measured), 35.00 °C (As-Built), and 34.93 °C (GIS-based). The MAE for temperature difference remains below 1.5 K for both models (3.8–4.0% relative error).

Minor transient deviations during rapid load changes reflect uncertainties in thermal capacitance parameters, which cannot be fully identified from available operational data. Nevertheless, the mean temperature levels remain within acceptable bounds (±1.5 K), validating the pressure-driven consumer model formulation.

Conclusion on Consumer Model Validation. The multi-metric validation demonstrates that the proposed consumer model achieves engineering-grade accuracy suitable for district heating network analysis. Key findings include:

This preprint research paper has not been peer reviewed. Electronic copy available at: https://ssrn.com/abstract=7674251

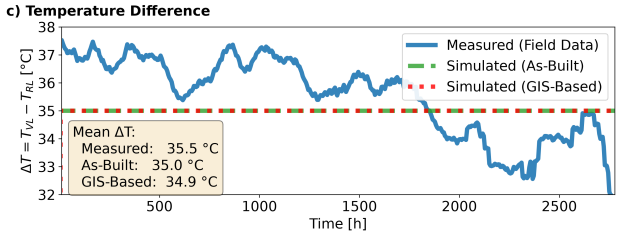


Figure 5: Temperature difference comparison: $\Delta T = T_{VL} - T_{RL}$ for building 574 of KIT Campus Nord. The simulated profiles closely match the measured field data, with mean deviations below 1.5 K. The statistical summary inset shows the mean ΔT for all three datasets. This justifies the assumption of a constant temperature difference for the active heating period.

- Heat flow prediction error below 1% of peak load (MAE: 0.89–1.20 kW)
- Mass flow prediction error below 2% of maximum (MAE: 0.015–0.016 kg/s)
- Temperature difference error below 1.5 K (MAE: 1.35–1.42 K)
- Comparable performance between As-Built and GIS-based network topologies

These results validate the bottom-up modeling approach and justify the use of the automated GIS-to-Modelica workflow for large-scale urban heating network studies.

3.2. Optimization Results

The Bayesian optimization framework was executed for 50 trials within the search space defined in Table 3, identifying optimal hyperparameters that minimize the weighted loss function between synthetic and real network characteristics. The loss function weights are specified in Table 4, prioritizing hydraulic capacity. Figure 6 provides a comprehensive analysis of the optimization process.

As illustrated in Figure 6(a), the loss function is a weighted metric incorporating physical capacity, topological fidelity, and centrality, and it shows a decline, converging toward a minimum. The optimization reached its optimum at trial 12, with a minimum loss value of approximately 8.98. The interaction between the L_{ring} and the detour-factor ($= L_{min}/L_{ring}$) is visualized in the contour plot in Figure 6(b). The region of minimal loss (dark purple) is concentrated at lower values of the detour-factor and moderate closing distances.

Figure 6(c)–(e) illustrate the parameter sampling history of the Bayesian optimizer. The scatter of the ring closing distance over the trials (Figure 6(c)) shows that the sampler explores the entire search range (50–500 m) with an initially uniform coverage, before increasingly concentrating on the lower and mid-range values after the optimum at 140.06 m has been identified. For the detour-factor (Figure 6(d)), the sampling density shifts progressively toward the lower bound (1.04), indicating that the optimizer exploits a region

Table 3
Hyperparameter Search Space and Optimal Values

Parameter	Search Range	Optimal Value
L_{ring}	50–500 m	140.06 m
Factor	1.0–3.0	1.04
Method	{MST, SPT, Steiner}	SPT

Table 4
Loss function weights by metric category

Category	Metric	Weight
Geometric	Node MAE	1.0
Hydraulic	Volume RE	5.0
	Diameter Dist Error	10.0
Structural	Cycle Diff	1.0
	Algebraic Connectivity	0.1
Functional	Centrality Correlation	3.0
	Param RE %	0.5

Table 5
Top 5 Optimization Trials by Loss Score

Trial	Method	Ring [m]	Min Graph [m]	Vol RE	Loss
12	SPT	140.06	146.22	0.672	**8.98**
16	MST	189.42	274.43	0.535	9.27
13	SPT	161.24	163.60	0.729	9.30
39	Steiner	227.72	254.13	0.456	9.47
11	SPT	134.68	167.30	0.731	9.66

in which direct ring closing is preferred over strongly detoured topologies. The categorical parameter in Figure 6(e) demonstrates that all three topology-generation methods (*mst*, *spt*, *steiner*) are sampled throughout the search, confirming a balanced exploration of the discrete design space; the Bayesian surrogate nevertheless returns the *spt* configuration as the optimum, which is consistent with the low-loss cluster observed for this method in Figure 6(h).

The sensitivity plot in Figure 6(f) reveals a basin-like structure, suggesting that an optimal window exists around 140.06 m. Values significantly higher or lower than this range lead to an increase in loss. A monotonic trend is observed in Figure 6(g), where the lowest loss is consistently achieved at the lower bound (Opt: 1.04). Comparing the categorical performance in Figure 6(h), the *spt* (Shortest Path Tree) method demonstrated a denser cluster of low-loss trials compared to the *steiner* or *mst* approaches.

Table 5 summarizes the top 5 performing trials by total loss score. The weighted loss as defined in Equation (24) decreased from an initial value of 10.25 to 8.98, representing a 12.3% improvement. The Shortest Path Tree (SPT) method achieved the best overall performance in Trial 12. MST-based networks showed moderate performance with slightly lower volumetric capacity, while Steiner tree approximations

This preprint research paper has not been peer reviewed. Electronic copy available at: https://ssrn.com/abstract=7674251

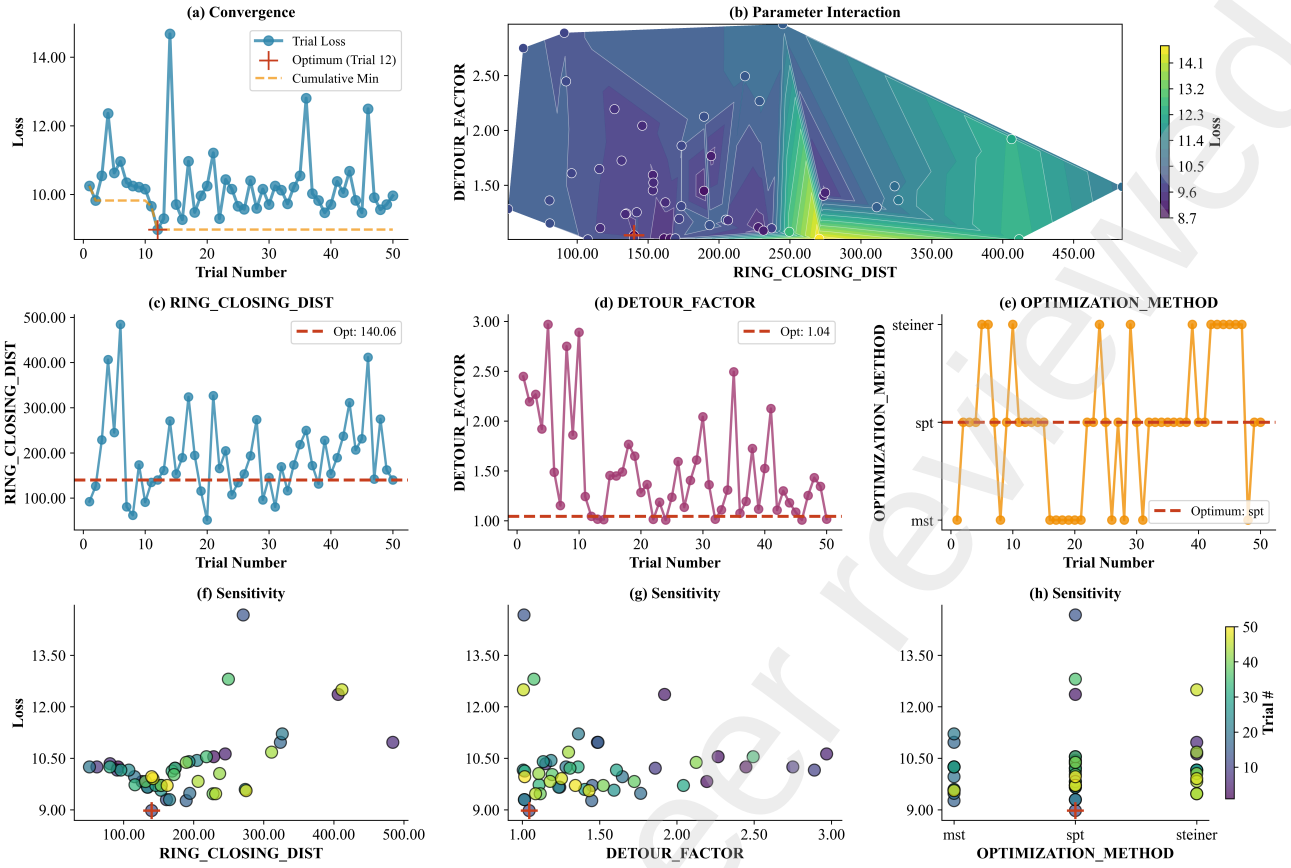


Figure 6: Bayesian optimization results for the district heating network design. (a) Convergence of the loss function over 50 trials, highlighting the global optimum at trial 12. (b) Parameter interaction contour plot between the ring closing distance and the detour factor. (c–e) Evolution of the hyperparameters over the optimization process. (f–h) Sensitivity analysis showing the relationship between individual parameters and the objective loss function.

yielded more conservative networks with lower ring density but higher path length deviations.

3.3. Topological Validation Results

The best-performing synthetic network (SPT with ring-closing; Trial 12) is compared against the As-Built Campus Nord district heating network across all validation metrics in detail. Table 6 presents the complete validation metric suite for the optimized network. Figure 7 visualizes the spatial alignment between the generated and real networks.

Spatial Fidelity (Node MAE). The mean absolute error between critical nodes (degree $\neq 2$) was $\text{MAE}_{dist} = 25.23m$ (Equation (18)), with a median distance of 20.64 m. 89.3% of the 345 generated nodes were successfully mapped after filtering (2.2.2) within the 75 m threshold. The remaining 10.7% (28 nodes) represent local deviations.

Node Degree Distribution. The degree distribution shows strong alignment with the ground truth network (MSE = 0.058, Equation (18)). Both networks exhibit similar proportions of dead-ends (degree 1), transit nodes (degree 2), and junctions (degree 3–4). Minor deviations occur at higher degrees (5+), which are rare in physical networks but occasionally emerge in the synthetic model due to ring-closing operations.

Table 6
Topological Validation Metrics: Optimized SPT Network vs. Ground Truth (Trial 12)

Metric	Value	Target	Assessment
Node MAE [m]	25.23	0	Acceptable
Param RE [%]	7.88	0	Good
Algebraic Connectivity RE	14.89	0	Poor
Centrality Correlation (r)	0.45	1.0	Limited
Cycle Count RE	0.00	0	Excellent
Avg Cycle Len Diff	0.97	0	Moderate
Diameter RE	0.26	0	Good
Avg Path RE	0.16	0	Good
Degree Dist MSE	0.058	0	Good
Diameter Dist MSE	0.042	0	Good
Volume RE	0.672	0	Poor

Edge Parameters and Hydraulic Capacity. The total pipe length deviation of 7.88% indicates that the synthetic network achieves comparable coverage with slightly more efficient routing. The volumetric capacity deviation (Volume

This preprint research paper has not been peer reviewed. Electronic copy available at: https://ssrn.com/abstract=7674251

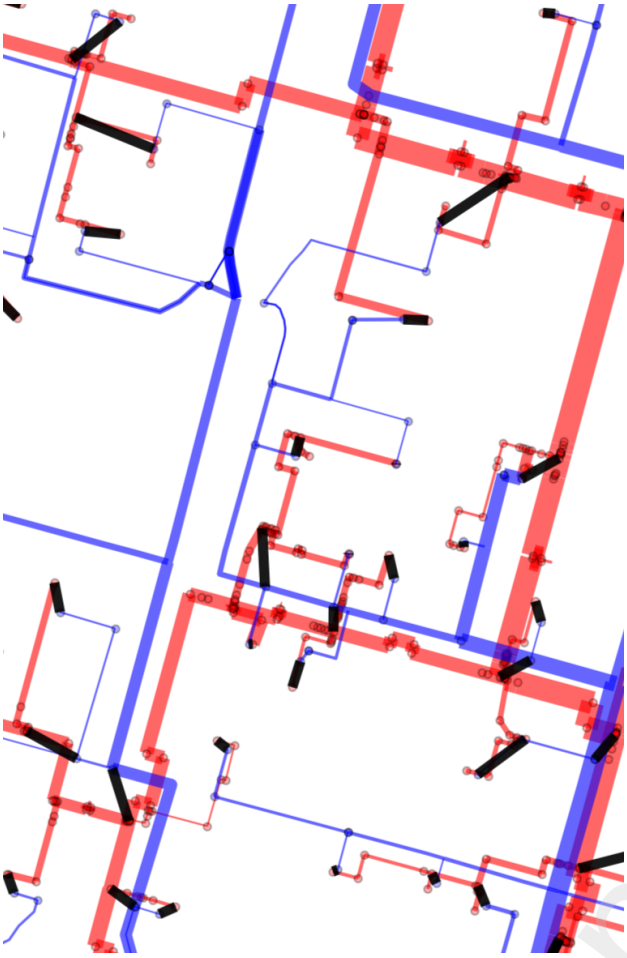


Figure 7: Spatial comparison between generic (blue) and real (red) district heating networks of KIT Campus Nord. Black lines indicate node mappings for the calculation of Node MAE between the generated und real network. Line widths of red and blue lines represent pipe diameters.

RE = 0.672, Equation (23)) reflects differences in diameter assignment strategies: the sizing algorithm prioritizes monotonic diameter reduction from sources to leaves, while the As-Built network contains legacy infrastructure with non-optimal dimensioning.

The diameter distribution MSE of 0.042 confirms that the relative frequencies of standard pipe sizes (DN 50–400) are well-replicated, ensuring hydraulic plausibility of the synthetic model.

Discussion of Hydraulic Capacity Deviation. It must be acknowledged that the volumetric capacity deviation of 67.2% (Volume RE) represents the largest remaining discrepancy in the validation suite. This deviation reflects a systematic difference in sizing philosophy rather than a model error: the recursive load aggregation produces a strictly demand-driven, monotonically decreasing diameter profile, whereas the As-Built network comprises legacy pipe segments that were dimensioned for historically higher loads, parallel trunk lines, or strategic reserves. While the diameter *distribution* (MSE = 0.042) and thus the plausibility of standard pipe sizes are well reproduced, absolute volumetric capacity should therefore not be interpreted as a calibrated quantity. For applications where absolute pumping energy or water inventory matters, the safety factor SF introduced in the hydraulic sizing step provides a calibration handle, which was not tuned in this present paper.

Transport Efficiency. Path length metrics demonstrate strong functional equivalence: the average shortest path deviation of 15.7% and diameter deviation of 25.9% (Equation (21)) indicate that media transport times through the synthetic network closely approximate real-world conditions.

Algebraic Connectivity and Robustness. The relative error in algebraic connectivity ($RE_{\lambda_2} = 14.89$, Equation (19)) reflects fundamental topological differences: the real network evolved organically over decades with multiple reinforcement phases, while the synthetic network follows algorithmic construction rules. Despite this deviation, the generated network maintains global connectedness and exhibits comparable vulnerability patterns to edge failures.

Centrality Correlation Limitations. The betweenness centrality correlation of $r = 0.45$ (Equation (20)) represents the weakest metric in the validation suite. This indicates that high-traffic hub nodes do not perfectly align between models. The deviation is attributed to historical expansion patterns in the real network that are not captured by single-phase optimization.

3.4. Analysis A: Performance Benchmark (Model Hierarchy)

To quantify the trade-off between computational efficiency and physical fidelity, two modeling abstraction levels were compared in isolation under identical operating conditions (115-day winter scenario). The first level employs a prescribed mass flow model, which decouples network hydraulics from thermal transport. In contrast, the second level implements the proposed coupled thermo-hydraulic model, where the mass flow rate is determined as a function of the local pressure differences across the network.

Metrics. The prediction error is evaluated with respect to measured field data as ground truth. In both reduced-order consumer models, the supply-line temperature T_{VL} is imposed as a prescribed boundary condition, so the actual model output to be validated is the return-line temperature T_{RL}. Root-mean-square errors (RMSE) of T_{RL} is given in Kelvin. For the mass flow rate $\dot{m}$ and the heat flow rate $\dot{Q}$, relative errors are used; because the analyzed period includes substation standstill ($\dot{m} = 0$, $\dot{Q} = 0$), the relative errors are normalized to the respective *mean operating value* rather than to the instantaneous measurement, which would degenerate at zero crossings.

This preprint research paper has not been peer reviewed. Electronic copy available at: https://ssrn.com/abstract=7674251

Table 7
Model hierarchy comparison, simulated over the 115.7-day winter scenario with DASSL in Dymola. T_{VL} is prescribed in both models; return-line temperature errors are given as RMSE in K, and mass flow and heat flow errors are given relative to the mean operating value.

Model	# Eq.	RMSE T_{RL} [K]	Rel. Err. $\dot{m}/\dot{Q}$ [%]	CPU [s]	Acc. Steps
SupplyDp	78	5.67	11.9 / 0.0	0.16	38,230
VarDp	65	3.96	8.2 / 2.2	0.31	77,630

Table 7 summarizes the performance metrics for the reduced-order consumer modeling strategies, computed over $N = 501$ synchronized time samples of the winter scenario.

Discussion. The `VarTSupplyDp` [20] variant is a *prescribed* (open-loop) consumer model. The heat flow rate is imposed as a fixed input, which is why the reported heat flow deviation is 0.0 % by construction. Its meaningful, independent error therefore lies in the return-line temperature, for which it yields an RMSE of 5.67 K (relative error 11.0 %). The mass flow error of 11.9 % is correspondingly higher than that of `VarTVarDp`.

The `VarTVarDp` model, in contrast, is a fully *closed-loop* formulation in which the mass flow is computed from the thermal and hydraulic state of the substation. The heat flow is the setpoint. It reproduces the measured return-line temperature with an RMSE of 3.96 K (relative error 7.3 %), while limiting the mass flow error to 8.2 % and the heat flow error to 2.2 % relative to the respective mean operating values.

The higher predictive quality of `VarTVarDp`, which estimates the free boundary quantities from the substation state rather than imposing them as inputs, confirms the hypothesis that the secondary heating circuit in residential buildings is negligible for district heating network simulation: the proposed closed-loop model reproduces the hydraulic and thermal state at the substation boundary, with return-temperature errors below 4 K, at lower computational cost.

Computational Cost. Despite a 17% lower equation count (65 vs. 78), the closed-loop `VarTVarDp` model requires roughly twice the integration effort of the prescribed `VarTSupplyDp` variant (77,630 vs. 38,230 accepted DASSL steps; 0.31 s vs. 0.16 s CPU time for the 115.7-day scenario). The overhead stems from the feedback loop: the PI-controlled valve introduces state events and a stiffer system behavior, reflected in the model being translated as a DAE rather than an ODE. However, in absolute terms both models simulate a full heating season in well under one second on a standard desktop machine, so the factor-of-two overhead is negligible for the intended use in large-network scenario studies. We therefore consider the accuracy gain (RMSE T_{RL} reduced from 5.67 K to 3.96 K, mass flow error reduced from 11.9% to 8.2%) a favorable trade-off.

Table 8
Statistical evaluation of exchanger fulfillment under undersizing stress conditions.

Metric	Total Exchangers	Mean	Median
Value	428	90.59 %	99.98 %
Metric	**Min**	**Max**	**Std. Dev.**
Value	0.00075 %	100.00 %	20.73

3.5. Analysis B: Diagnostic Capability (Stress Test)

The pressure-driven consumer architecture enables diagnostic simulations that identify hydraulic bottlenecks and supply deficits before physical implementation.

Urban Integration: Karlsruhe City Model. To illustrate spatial scalability, the framework was applied to a 2.5 km^2 urban area in Karlsruhe-Oberreut (Germany). Oberreut is a residential district in southern Karlsruhe, located approximately 2.5 km southwest of the city center. The case study encompasses approximately 450 buildings with an aggregated annual heat demand of 18.3 GWh. The multi-source optimization identified the optimal plant location based on density-weighted scoring (Equation (1)). Network generation completed in 52 s, producing a model with 646 nodes and 1,290 edges. Figure 8 illustrates component-level behavior during supply shortfall conditions. When the required heat output $\dot{Q}_{req}$ cannot be met due to insufficient differential pressure Δp, the PI controller drives the valve to maximum opening, yet actual heat delivery remains below setpoint. The ratio $\dot{Q}_{actual}/\dot{Q}_{req}$ serves as a quantitative undersupply indicator, enabling automated identification of vulnerable network regions.

Undersizing Detection Feature. A deliberate undersizing scenario was created by reducing the pump pressure by 2 bar compared to the hydraulically optimized baseline. This stress test revealed critical insights into the robustness of the district heating network:

- In the demonstrated scenario, 5 out of 428 exchangers (1.2%) experienced sustained undersupply, with fulfillment ratios below 10% of the required heat load during the winter peak period. All affected substations are located in peripheral or high-rise building clusters, as shown in Figure 9.
- Insufficient supply disproportionately affected large consumers at the peripheral edges of the district, where cumulative heat demand and network branching exacerbate pressure losses.
- The most critical case (Exchanger 155) exhibits near-complete supply failure, with a fulfillment ratio of $\leq$ 5% of the required heat flow while the median fulfillment across all exchangers remains at 99.98%.

This preprint research paper has not been peer reviewed. Electronic copy available at: https://ssrn.com/abstract=7674251

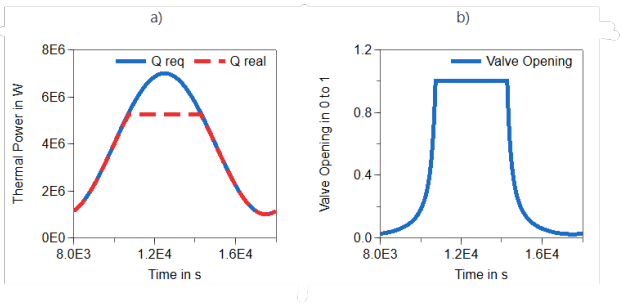


Figure 8: Demonstration of an insufficiently dimensioned heat exchanger: (a) Heat demand profile, and (b) Valve opening dynamics. The PI controller reaches maximum valve opening (100%), yet actual heat delivery remains insufficient due to hydraulic limitations.

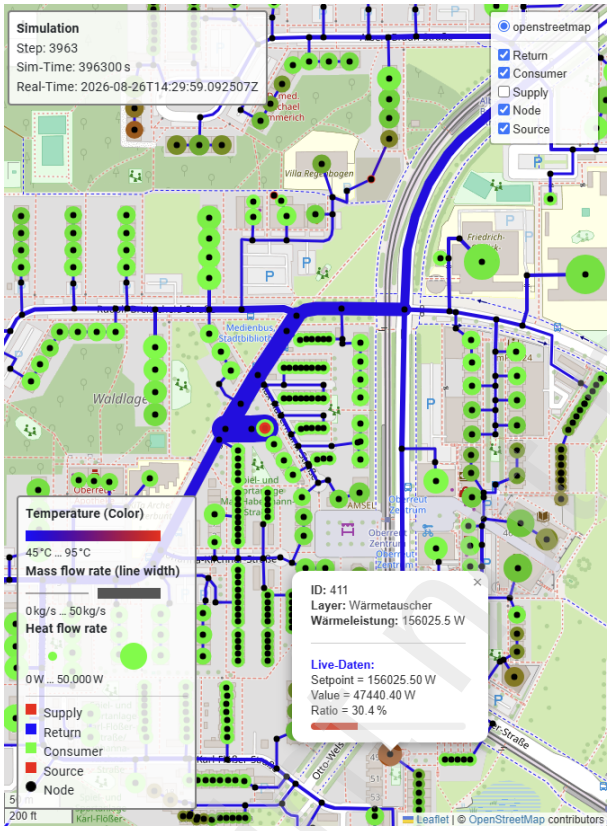


Figure 9: Spatial distribution of exchanger fulfillment under undersizing stress conditions in the Karlsruhe City Model. Green circles indicate adequate heat supply (radius scales with heat consumption); red markers highlight substations with sustained undersupply ($\dot{Q}_{actual}/\dot{Q}_{req}$).

4. Conclusions

This paper presents an end-to-end automated workflow for generating high-fidelity digital models of district heating networks from GIS data. The framework successfully bridges the gap between simplified synthetic models and computationally expensive manual modeling approaches, while maintaining engineering-grade accuracy.

The application to a 2.5 km^2 urban area in Karlsruhe-Wolfartsweier demonstrates the spatial scalability of the workflow: automated plant placement and network generation produced a model with 646 nodes and 1,290 edges in 52 s. The diagnostic capability of the *VarTVarDp* consumer model was further exploited to identify vulnerable network regions via the quantitative undersupply indicator.

Key Achievements. The multi-level validation against the KIT Campus Nord reference network demonstrates several key achievements. Regarding physical fidelity, consumer model heat flow predictions achieves an MAE of less than 1% of peak load (0.89–1.20 kW), with temperature differences accurate to within 1.5 K. Topological accuracy is evidenced by a node positioning MAE of 25.23 m, a pipe length deviation of 7.88%, and a diameter distribution MSE of 0.042. Moreover, the diagnostic utility of the pressure-driven consumer models successfully identifies hydraulic bottlenecks and undersupply conditions in stress test scenarios, with 5 out of 428 exchangers exhibiting sustained undersupply under a deliberately weakened pump configuration.

Limitations of GIS Automation. Observed deviations between synthetic and As-Built networks arise primarily from simplified geometry, for which building connection points are approximated from footprint centroids rather than actual trench entry locations, typically resulting in an offset of 5 to 30 m. Historical growth patterns also play a significant role, as real networks contain decommissioned segments, temporary bypasses, and phased expansions that violate single-phase optimization assumptions. Because the selected optimization parameters did not show particularly high sensitivity to the loss, no optimization method could truly replicate the historically grown network. While some error metrics were well optimized, imitating arbitrarily grown infrastructure using rule-based algorithms remains challenging. Additionally, data granularity issues persist since legacy technical drawings include valve chambers and maintenance access points not visible in cadastral GIS layers. Finally, mapping threshold sensitivity plays a role, as the 75 m threshold excludes peripheral nodes where network extensions follow property boundaries rather than straight-line connections.

Practical Implications. Despite these limitations, the validated synthetic networks achieve sufficient fidelity for hydraulic simulation and pressure drop analysis, expansion planning and scenario comparison, vulnerability assessment and resilience studies, as well as rapid sensitivity analysis for decarbonization pathways.

Future Work. Ongoing research addresses several key directions for extending this framework. Multi-energy coupling will expand the methodology to integrated electricity-heat systems for comprehensive sector-coupling analysis with the eASiMOV platform [42]. City-scale application involves validation on networks exceeding 10,000 buildings

This preprint research paper has not been peer reviewed. Electronic copy available at: https://ssrn.com/abstract=7674251

using reduced-order models. Additionally, graph neural networks will be explored for automated graph generation from geospatial data.

The proposed framework provides urban planners and utility operators with a scalable tool for evidence-based decision-making in the transition toward climate-neutral heating infrastructures. By automating the translation from geospatial data to validated simulation models, the present paper removes a critical bottleneck in large-scale sector-coupled urban energy system analysis.

CRediT authorship contribution statement

Fatih Tüzün: Conceptualization, Methodology, Software, Formal analysis, Investigation, Data curation, Visualization, Writing - original draft. **Hüseyin K. Çakmak:** Conceptualization, Validation, Supervision, Writing - review & editing, Project administration. **Veit Hagenmeyer:** Conceptualization, Supervision, Funding acquisition, Resources, Writing - review & editing.

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.

Acknowledgments

This work was conducted within the framework of the Helmholtz Programs “Materials and Technologies for the Energy Transition” (MTET) as well as “Energy System Design” (ESD) and funded by the KIT Energy Lab networking project E-Industry.

References

[1] European Commission. Repowereu plan. COM(2022) 230 final, 2022.

[2] European Commission. The european green deal. COM(2019) 640 final, 2019.

[3] H. Lund, S. Werner, R. Wiltshire, S. Svendsen, J. E. Thorsen, F. Hvelplund, and B. V. Mathiesen. 4th generation district heating (4GDH): Integrating smart thermal grids into future sustainable energy systems. *Energy*, 68:1–11, 2014.

[4] X.J. Lin, N. Zhang, Y.H. Mao, J.Y. Chen, X.T. Tian, and W. Zhong. A review of the transformation from urban centralized heating system to integrated energy system in smart city. *Applied Thermal Engineering*, 240:122272, 2024.

[5] S. Lu, Z. Gao, Y. Sun, S. Zhang, B. Li, C. Hao, Y. Xu, and W. Gu. Aggregate model of district heating network for integrated energy dispatch: A physically informed data-driven approach. *IEEE Transactions on Sustainable Energy*, 15(3):1859–1871, 2024.

[6] H. V. Larsen, B. Bøhm, and M. Wigbels. A comparison of aggregated models for simulation and operational optimisation of district heating networks. *Energy Conversion and Management*, 45(7):1119–1139, 2004.

[7] B. van der Heijde, M. Fuchs, C. Ribas Tugores, G. Schweiger, K. Sartor, D. Basciotti, D. Müller, C. Nytsch-Geusen, M. Wetter, and L. Helsen. Dynamic equation-based thermo-hydraulic pipe model for district heating and cooling systems. *Energy Conversion and Management*, 151:158–169, 2017.

[8] A. Dénarié, M. Aprile, and M. Motta. The importance of an accurate numerical model for the simulations of new generation district heating systems. *Journal of Physics: Conference Series*, 2893:012030, 11 2024.

[9] Modelica Association. Modelica language specification version 3.6. https://specification.modelica.org/, 2023.

[10] S. A. Klein et al. *TRNSYS 18: A Transient System Simulation Program, Volume 4: Mathematical Reference (Standard Component Library, Type 604)*. Solar Energy Laboratory, University of Wisconsin-Madison, 2017.

[11] M. Fuchs, J. Teichmann, M. Lauster, P. Remmen, R. Streblow, and D. Müller. Workflow automation for combined modeling of buildings and district energy systems. *Energy*, 117:478–484, 2016.

[12] M. Fuchs and D. Müller. Automated design and model generation for a district heating network from OpenStreetMap data. In *Proceedings of Building Simulation 2017: 15th Conference of International Building Performance Simulation Association*, pages 2050 – 2059, San Francisco, USA, 2017.

[13] M. Mans, T. Blacha, T. Schreiber, and D. Müller. Development and application of an open-source framework for automated thermal network generation and simulations in Modelica. *Energies*, 15(12):4372, 2022.

[14] RWTH Aachen University, E.ON Energy Research Center, Institute for Energy Efficient Buildings and Indoor Climate (EBC). UESgraphs: Urban energy system graphs framework documentation. https://github.com/RWTH-EBC/UESgraphs, 2023. Accessed: 2026-08-27.

[15] T. H. Cormen, C. E. Leiserson, R. L. Rivest, and C. Stein. *Introduction to Algorithms*. The MIT Press, 4th edition, 2022.

[16] F. K. Hwang, D. S. Richards, and P. Winter. *The Steiner Tree Problem*, volume 53 of *Annals of Discrete Mathematics*. North-Holland, Amsterdam, 1992.

[17] J Pfeiffer and M. Kunick. Development of an integrated software workflow for district heating network planning: A structured methodological approach. *ACC Journal*, 31(1):48–63, 2025.

[18] J. Pfeiffer et al. DistrictHeatingSim: GUI-based tool for sustainable district heating network planning. https://github.com/JonasPfeiffer123/DistrictHeatingSim, 2025. Accessed: 2026-08-24.

[19] T. Akiba, S. Sano, T. Yanase, T. Ohta, and M. Koyama. Optuna: A next-generation hyperparameter optimization framework. In *Proceedings of the 25th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining*, pages 2623–2631, 2019.

[20] Laura Maier, David Jansen, Fabian Wüllhorst, Martin Kremer, Alexander Kümpel, Tobias Blacha, and Dirk Müller. Aixlib: an open-source modelica library for compound building energy systems from component to district level with automated quality management. *Journal of Building Performance Simulation*, 17(2):196–219, 2023. Accessed: 2025-10-29.

[21] KEA Klimaschutz- und Energieagentur Baden-Württemberg and LUBW Landesanstalt für Umwelt Baden-Württemberg. Wärmeatlas baden-württemberg: Gebäudescharfe wärmebedarfsdaten und interaktive online-karte im energieatlas bw. Online-Datenportal, 2026. Accessed: 2026-08-28.

[22] Rob Story and folium contributors. folium: Python data visualization on leaflet.js maps. https://github.com/python-visualization/folium, 2026. Accessed: 2026-08-25.

[23] OpenStreetMap contributors. OpenStreetMap database. https://openstreetmap.org, 2026. Accessed: 2026-09-28.

[24] E. W. Dijkstra. A note on two problems in connexion with graphs. *Numerische Mathematik*, 1:269–271, 1959.

[25] J. B. Kruskal. On the shortest spanning subtree of a graph and the traveling salesman problem. *Proceedings of the American Mathematical Society*, 7:48–50, 1956.

[26] L. Kou, G. Markowsky, and L. Berman. A fast algorithm for steiner trees. *Acta Informatica*, 15(2):141–145, 1981.

[27] P. Zwickel, F. Bauer, K. Schätzler, U. Steiner, S. Waczowicz, V. Hagenmeyer, and E. Schill. Integrated concept of local heat transition using geothermal heat production, aquifer storage and optimization

This preprint research paper has not been peer reviewed. Electronic copy available at: https://ssrn.com/abstract=7674251

of distribution network. In *European Geothermal Congress 2022 : Berlin, Germany, 17-21 October 2022 : proceedings*. European Geothermal Energy Council (EGEC), 2022. 37.12.03; LK 01.

[28] M. Newman. *Networks*. Oxford University Press, 2nd edition, 2018.

[29] D. Conte, P. Foggia, C. Sansone, and M. Vento. Thirty years of graph matching in pattern recognition. *International Journal of Pattern Recognition and Artificial Intelligence*, 18(3):265–298, 2004.

[30] D. P. Huttenlocher, G. A. Klanderman, and W. J. Rucklidge. Comparing images using the hausdorff distance. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 15(9):850–863, 1993.

[31] M Fiedler. Algebraic connectivity of graphs. *Czechoslovak Mathematical Journal*, 23(2):298–305, 1973.

[32] L. C. Freeman. A set of measures of centrality based on betweenness. *Sociometry*, 40(1):35–41, 1977.

[33] L. C. Freeman. Centrality in social networks conceptual clarification. *Social Networks*, 1(3):215–239, 1978.

[34] M. E. J. Newman. The structure and function of complex networks. *SIAM Review*, 45(2):167–256, 2003.

[35] D. J. Watts and S. H. Strogatz. Collective dynamics of 'small-world' networks. *Nature*, 393:440–442, 1998.

[36] A. Clauset, C. R. Shalizi, and M. E. J. Newman. Power-law distributions in empirical data. *SIAM Review*, 51(4):661–703, 2009.

[37] R. Albert and A.-L. Barabási. Statistical mechanics of complex networks. *Reviews of Modern Physics*, 74:47–97, 2002.

[38] O. M. Awe, S. T. A. Okolie, and O. S. I. Fayomi. Optimization of water distribution systems: A review. *Journal of Physics: Conference Series*, (1378), 2019.

[39] T. M. Walski, D. V. Chase, D. A. Savic, W. Grayman, S. Backwith, and E. Koelle. *Advanced Water Distribution Modeling and Management*. Haestad Press, 2004.

[40] J. Bergstra, R. Bardenet, Y. Bengio, and B. Kégl. Algorithms for hyper-parameter optimization. In *Proceedings of the 25th International Conference on Neural Information Processing Systems*, pages 2546–2554, 2011.

[41] H. Sakoe and S. Chiba. Dynamic programming algorithm optimization for spoken word recognition. *IEEE Transactions on Acoustics, Speech, and Signal Processing*, 26(1):159–165, 1978.

[42] H. K. Çakmak, A. Kocher, Cheng H., J. Vucinic, and V. Hagenmeyer. Collaborative modelling and co-simulation for sector-coupled multi-energy system analysis with eASiMOV-eCoSim, 2025. doi:10.5445/IR/1000182805.

Preprint not peer reviewed

This preprint research paper has not been peer reviewed. Electronic copy available at: https://ssrn.com/abstract=7674251