Novel validated method for GIS based automated dynamic urban building energy simulations

P. Nageler a,*, G. Zahrer b, R. Heimrath a, T. Mach a, F. Mauthner c, I. Leusbrock c, H. Schranzhofer a, C. Hochenauer a

* Institute of Thermal Engineering, Graz University of Technology, Inffeldgasse 25/B, A-8010 Graz, Austria
b Institute of Geodesy, Graz University of Technology, Steyrergasse 30, A-8010 Graz, Austria
c AEE – Institute for Sustainable Technologies, Feldgasse 19, 8200 Gleisdorf, Austria

Abstract

The modelling of whole urban districts requires an automated process to parameterize simulation tools. This paper presents a validated methodology for fully automated building modelling within urban districts based on publicly available data. Dynamic building models with detailed heating systems are created in the simulation environment IDA ICE. The method of data collecting and processing and result visualization in a geographical information system (GIS) and the data storage procedure in a PostgreSQL database is described in detail. The building simulation model is validated with consumption data available from 69 buildings of the town Gleisdorf (Austria). The results of the annual heating and domestic hot water demand show a good approximation to the measurement data with a mean deviation of -0.98 %. The urban simulation process was then extended to the whole community with its 1,945 buildings. This method helps to model and quantitatively describe current building stock in an efficient and timesaving way and enables to develop future smart energy systems, in which the buildings interact with the district heating networks, with limited effort.

Keywords: dynamic urban building energy simulation; automation; GIS; IDA ICE; PostgreSQL database

* Corresponding author. Tel.: +43 316873 7310
email: peter.nageler@tugraz.at

Highlights:

- A methodology for automated dynamic urban building energy simulation is developed.
- The automation process is based on publicly available data.
- Both building model and heating system are automated modelled and dimensioned.
- A PostgreSQL database structure is created for building and heating system models.
- The urban simulation results are visualized in QGIS.
## NOMENCLATURE

### Abbreviations

<table>
<thead>
<tr>
<th>Abbreviation</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>ACH</td>
<td>air change per hour</td>
</tr>
<tr>
<td>ASL</td>
<td>above sea level</td>
</tr>
<tr>
<td>BCVTB</td>
<td>building control virtual test bed</td>
</tr>
<tr>
<td>BIM</td>
<td>building information modelling</td>
</tr>
<tr>
<td>DB</td>
<td>database</td>
</tr>
<tr>
<td>DHW</td>
<td>domestic hot water</td>
</tr>
<tr>
<td>ERD</td>
<td>entity relationship diagram</td>
</tr>
<tr>
<td>FK</td>
<td>foreign key</td>
</tr>
<tr>
<td>GIS</td>
<td>geographic information system</td>
</tr>
<tr>
<td>HS</td>
<td>heating system</td>
</tr>
<tr>
<td>IDA ICE</td>
<td>simulation tool: IDA Indoor Climate and Energy</td>
</tr>
<tr>
<td>MFH</td>
<td>multi family house</td>
</tr>
<tr>
<td>NSB</td>
<td>night setback of the heating system</td>
</tr>
<tr>
<td>PK</td>
<td>primary key</td>
</tr>
<tr>
<td>SFH</td>
<td>single family house</td>
</tr>
<tr>
<td>RC model</td>
<td>resistance and capacity model</td>
</tr>
</tbody>
</table>

### Symbols, unit

<table>
<thead>
<tr>
<th>Symbol</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>A</td>
<td>area, m²</td>
</tr>
<tr>
<td>ACH</td>
<td>air change per hour, hr⁻¹</td>
</tr>
<tr>
<td>cp</td>
<td>specific heat capacity, J (kg K)⁻¹</td>
</tr>
<tr>
<td>COP</td>
<td>coefficient of performance</td>
</tr>
<tr>
<td>d_insulation</td>
<td>insulation thickness, m</td>
</tr>
<tr>
<td>dT</td>
<td>temperature difference, K</td>
</tr>
<tr>
<td>GFA</td>
<td>gross floor area, m²</td>
</tr>
<tr>
<td>h</td>
<td>height, m</td>
</tr>
<tr>
<td>k_value</td>
<td>power law coefficient; emitted power per heat distribution system length and degree (raised to ( n-value )), W (m Kⁿ⁻¹)</td>
</tr>
<tr>
<td>m</td>
<td>mass flow, kg s⁻¹</td>
</tr>
<tr>
<td>m_sp</td>
<td>specific mass flow of the solar plant, kg (hr m²_collector)⁻¹</td>
</tr>
<tr>
<td>n_value</td>
<td>power law exponent</td>
</tr>
<tr>
<td>P</td>
<td>designed power of the heat supply unit, W</td>
</tr>
<tr>
<td>P_peak</td>
<td>required peak power for space heating, W m⁻²</td>
</tr>
<tr>
<td>P_specific_geothermal_probe</td>
<td>specific heat transfer of the geothermal probe, W m⁻¹</td>
</tr>
<tr>
<td>P_sp_dist</td>
<td>specific radiator power, W m⁻²</td>
</tr>
<tr>
<td>Q_dhw_person</td>
<td>domestic hot water heat demand per year and person, kWh (person a⁻¹)</td>
</tr>
<tr>
<td>Q_utilisation</td>
<td>heat per year and collector area, kWh (a m²_collector)⁻¹</td>
</tr>
<tr>
<td>r</td>
<td>radius, m</td>
</tr>
<tr>
<td>RMSE</td>
<td>root mean square error, kW</td>
</tr>
<tr>
<td>s</td>
<td>standard deviation, kW</td>
</tr>
<tr>
<td>SC</td>
<td>solar coverage, %</td>
</tr>
<tr>
<td>T</td>
<td>temperature, °C</td>
</tr>
<tr>
<td>UA_heat_exchanger</td>
<td>UA value of the heat exchanger, W K⁻¹</td>
</tr>
<tr>
<td>V</td>
<td>volume, m³</td>
</tr>
<tr>
<td>y</td>
<td>experimental data</td>
</tr>
<tr>
<td>ŷ</td>
<td>model output</td>
</tr>
<tr>
<td>ε</td>
<td>error or residual, kW</td>
</tr>
<tr>
<td>¯ε</td>
<td>mean error, kW</td>
</tr>
<tr>
<td>% error</td>
<td>percent error, %</td>
</tr>
</tbody>
</table>

### Subscript

<table>
<thead>
<tr>
<th>Subscript</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>solar</td>
<td>solar collector</td>
</tr>
<tr>
<td>Abbreviation</td>
<td>Description</td>
</tr>
<tr>
<td>--------------</td>
<td>------------------------</td>
</tr>
<tr>
<td>hdist</td>
<td>heat distribution system</td>
</tr>
<tr>
<td>sp</td>
<td>specific</td>
</tr>
<tr>
<td>ret</td>
<td>return</td>
</tr>
<tr>
<td>sup</td>
<td>supply</td>
</tr>
<tr>
<td>w</td>
<td>water</td>
</tr>
</tbody>
</table>
1 INTRODUCTION

Cities are currently responsible for over two-thirds of the world’s energy consumption and account for more than 70% of global CO₂ emissions [1]. European buildings account for approximately 37% of the total final energy use, 26% is taken by residential and 11% by commercial buildings [2]. Driven by growing urbanization (from 54% in 2014 to 66% in 2050 [3]), total urban energy use will further increase from 240 EJ in 2005 to 740 EJ in 2050 [4]. The energy consumption of buildings thus plays a central role in climate change. To reach the climate targets of the European Commission by 2030, a reduction of our current greenhouse gas emissions by 40% (from 1990 levels) is necessary [5]. Additionally, the share of renewables have to be increased to 27% in the same time span [5]. The usage of highly dynamic fluctuating renewable energy sources, such as wind or solar, results in a temporal mismatch between supply and demand. The main future challenge here is to match the available energy from renewables with the energy demand in space and time. Consequently, a transformation of the old existing energy systems into a new smart energy system is necessary, which integrates the high fluctuation of renewable energy into the system. In this context, the 4th generation district heating network [6] provides a promising approach.

To enable and structure this transformation process, an analysis of the actual building stock is necessary to identify and quantify, for example, the energy demand, the refurbishment potential of buildings, the usage of decentral heating supply options or the expansion potential of district heating networks. The underlying approaches, challenges and opportunities for urban energy system models are presented by Keirstead et al.[7] and Kavgic et al. [8]. Reinhart et al. [9] published a review of urban energy modelling tools. The tools differ in level of detail regarding the spatial- and temporal resolution and the modelling approach. In literature, three main groups of tools exist for urban modelling:

1. The first group uses model simplification procedures, which are for example physical simplifications such as simplified thermal zones [10] [11] or resistance and capacity (RC) models for thermal zones [12] [13] [14] [15] [16], model order reduction method [17], a linear regression analysis [18] [19], statistical method [20] or a probabilistic method [21].
2. The second group uses simplifications regarding to their temporal resolution for example monthly heat balance models such as SimStadt [22] or annual heat balance models such as BREDEM [23].
3. Finally, the last group uses detailed building simulation tools, which provide multi-zone analysis with any temporal resolution. These tools consider also mutual building shading and the heating system of the buildings. The building simulation tools EnergyPlus [24] [25] [26] and Dymola [27] were already used on the urban scale.

Furthermore, it is difficult to collect all the necessary data for a building energy analysis due to data protection for private people and corporate secrets. The data of the building stock—which e.g. includes the material layers and their respective thicknesses of the building envelope, the year of construction, the type of use and the behavior of the building users—are mostly unknown. Approaches to close the data gaps were presented in the TABULA project [37] and by Aksoezen et al. [38], which use the building age as an indicator for the building energy consumption. The TABULA project assigns age-related building
envelopes to building types. On the other hand, various data sources are available for the building geometry. Here, CityGML provides a promising approach for a standardized 3D data format [28]. This standardization can be used to exchange data from different tools [29] [30]. Another standardized building model is BIM, in which the whole building information can be stored (building envelope with its materials and thicknesses), but often there are errors during data exchange between the architect and the simulation expert. The 3D geometry can also be obtained through 3D laser scanning point clouds [31]. Mahdjoubi et al. [32] combine the BIM and the laser scanning technology.

GIS are software tools to collect, manipulate, organize and visualize geographical data and are already widely used for potential analysis in urban areas. In literature, for instance potential studies for solar [39] [40] [41], wind [42] [43] and geothermal probes [44] exist. The usefulness of a combination of GIS and energy simulation programs has already been recognized [27] [45] and will become even more important in the future as the amounts of data and project sizes will grow and thereby increasing the complexity of the problem. The collected and manipulated data is stored in a database and an interface produces a simulation model for the simulation tool. Zucker et al. [24] use a Co-Simulation framework (BCVTB [46]) to couple automated generated buildings (EnergyPlus) with zone controllers (Ptolemy [47]).

Building modelling of whole urban districts or towns and cities is a time consuming effort. An automated process is required for this task in order to reduce the enormous modelling time of manual input and error rate due to typos. Urban energy simulation tools are essential to evaluate and simulate actual performance or renovation scenarios. Fuchs et al. [27] developed an automated workflow for combined modelling of buildings (RC model) and district energy systems. Eicker et al. [48] use a standardized geometrical format (CityGML) for the building geometry. Other tools [49] [50] [51] are used for automated complex thermal auto-zoning. Dogan and Reinhart present an automated multi-zone method to perform urban simulations [11].

This study presents a validated methodology for fully automated building modelling within urban districts based on publicly available data. The proposed method combines statistical and engineering approaches. The method of data collecting and processing in a geographical information system (GIS) and the data storage in a PostgreSQL database is described in detail. In an urban development area, the database frequently changes in course of investor announcements. For this reason, the proposed simulation model is flexible enough to add new buildings to the model and change the data of the old building stock in order to refine the model easily to any level of detail. Further, the building heating systems connect the buildings with the district heating network and the decentralized renewable energy sources. Consequently, modelling of heating systems is essential to grasp the dynamic interactions with smart energy systems or renewables. Such information is important for e.g. energy suppliers, urban planners and investors. In this paper, an urban building model is equipped with a detailed heating system. This includes water tanks, boilers, ambient air and ground heat pumps, solar thermal power, photovoltaic and wind turbine power, district heating connection and a heat distribution system. Further, it is important to consider mutual building shading in urban areas. The model is thus simulated with the dynamic building energy simulation tool IDA ICE [52], which fits into the third group of simulation tools. The town Gleisdorf (Austria) was chosen as the area of
application. First, the method is validated with measurement data of 69 buildings with a district heating connection. Afterwards the whole community with its 1,945 buildings is simulated. This method helps to model and quantitatively describe current building stock in an efficient and time-saving way and enables to develop future smart energy systems, in which the buildings interact with the district heating networks, with limited effort. Further, the presentation of results is difficult in urban areas. A brief and clear presentation of the results is introduced in QGIS, which are essential for clients.

2 CASE STUDY

2.1 Study area

The methodology is tested on the basis of a case study. The investigated area is the building stock of the community Gleisdorf (47°6' N 15°42' E 365 m ASL) in Austria. Gleisdorf is a typical Central European small town with terraced houses in the town center and many detached houses on the outskirts. The examined building stock consists of 1,945 buildings, which are spread over an area of approximately five square kilometers. The community can be characterized by the following mean values: the average building age is 49 years, the mean building area is 253 m², the medium-sized gross floor area is 403 m² and the mean building height is 5.12 m. The buildings are typically aligned to the southwest. Fig. 1 shows the investigated area. A district heating network supplies 69 buildings, which have an average building age of 32 years and a medium-sized gross floor area of 574 m², with heat for domestic hot water (DHW) and space heating. The district heat is produced by two biomass woodchip boilers (each 850 kW), a gas boiler (only in the winter period) and three solar thermal plants with (300 m², 285 m² and 495 m² collector area) [53]. The district heating network water supply temperature at the substations is set to 80 °C and the temperature decrease at the heat exchanger unit is set to 30 K.
2.2 Measurement data, validation and applied metrics

Data on building stock and district heating infrastructure for this case study were supplied by the utility services of Gleisdorf. The case study is split into two cases. The first case refers to the district heating network. In this case, the methodology is validated with real measurement data derived from the utilities’ database of the city Gleisdorf [53]. The buildings as well as the heating system, which is connected via heat exchanger with the district heating network, are modelled. The utilities of the community Gleisdorf operate the district heating network and collect the measurement data for heat demand at the substations. The heat is measured with gauged heat meters. In this case, the annual heat demand for space heating and DHW of all 69 buildings and the temporal power course of space heating and DHW of five buildings are taken to validate the methodology. The buildings consist of 46 single family houses (SFH), 15 multi family houses (MFH), five schools, one office and two commercial business buildings. Furthermore, the building levels, the building type and the year of construction are given from the utilities of the community. In this article the simulation results comparison with real measurement data relies on the evaluation of heat demand at the substations. The evaluation interval is from 1.1.2015 to 31.12.2015 with hourly resolution. The second case refers to a larger case area. The building energy demand for space heating is calculated for the whole building stock of Gleisdorf. The simulated buildings are shown in Fig. 1 with blue and orange. In this case, the building height is obtained from a 10x10 m elevation model by airborne laser scanning [54]. The building type and category is assigned with the proposed methodology, see section Methodology. In both cases the weather data from ZAMG are used [55].
The error \( \varepsilon \) or residual at instant \( t \) is defined according to eq. (1), compare with Ljung [56]. To compare a specific simulation result \( \hat{y}(t) \) a reference value \( y(t) \) is needed. This error may be calculated for any simulation result, e.g. the district heating power consumption.

\[
\varepsilon(t) = y(t) - \hat{y}(t)
\] (1)

Based on eq. (1) a number of evaluation metrics is possible to indicate the deviation. The mean error \( \bar{\varepsilon} \) is suitable to indicate a bias, that is a constant under- or overestimation. The root mean square error (RMSE) is used to assess the average unsigned error of a certain simulation tool. The empirical standard deviation \( s \) is used to quantify the common spread of an error. Another indication of the spread of the model error is given by the maximum absolute error \( \max(\vert \varepsilon \vert) \), which indicates the worst outlier for the evaluated interval. Finally, the percent error \( \% \text{error} \) shows the differences between the simulated and measured values, as a percentage of the measured value.

3 METHODOLOGY

This paper presents a validated methodology for fully automated building modelling within urban districts based on publicly available data. This section describes the workflow from the raw data to the visualization of the simulation results. First, a database (DB) system is created, which consists of a project-related DB and a reusable typology DB. Afterwards, a five-step workflow is used to fill the project-related database and to visualize the dynamic simulation results. The data flow and their direction between the DB and each process step is depicted with arrows. The workflow is divided into the following process steps (see Fig. 2):

1) data acquisition,
2) categorization/dimensioning,
3) building modelling,
4) dynamic simulation and
5) result visualization.
**DB system**

The DB system allows the user to manage multiple data sources and make the same data accessible to several simulation tools. The DB system stores all the necessary simulation input data and the relevant simulation results. The methodology contains two object-relational PostgreSQL databases [63], see Fig. 2. The structure of these databases is represented in an entity relationship diagram (ERD). The ERD is a graphical representation of a DB that shows the relationships between entities (tables). Primary- (PK) and foreign keys (FK) define the relationships between the entities. The PK of a table is a unique identifier for the table rows and can then be used as a FK in another table. By matching key values in the two tables, the relationships between the rows of the one and the rows of the other table are then recognized. Each entity is represented as a rectangle, in which the table name is represented first, followed by the primary key and the attributes (columns) with their data types.

Separately to the DB a file system exists, in which the automated generated thermal city model are stored as text files. All assumptions are combined in these text files and the results can be reproduced.

**Typology database**

The typology database stores non-geometrical data, which is project independent. This requires an example for clarification. The wall constructions for typical buildings are stored in the typology database, thus they are reusable for the next project (D) and data that has not yet been implemented can be extended in the database (E), see Fig. 2. Fig. 3 shows the ERD of the typology database. The centre of the typology database is the building type (table: buildingtype), which connects all the relevant non-geometric simulation data. A building type combines all data regarding a special type of building. This contains data of the wall- and window constructions, the building category, the percentage of glazing, the specific heat demand and peak...
power for space heating. The building type has relationships to the user profile (table: userprofile) with time-resolved load profiles for internal gains (table: loadprofile) [64], the building category (table: buildingcategory) and the building envelope (table: buildingtype_hull).

Building envelopes (table: buildingtype_hull) consists of wall constructions (table: hull) with their wall layers (table: layer) and materials (table: material) [37] [65]. For example, a building envelope consists of wall constructions (e.g. 25 cm brick with 10 cm insulation) of each hull type (table: hulltype), which are walls, roofs and floors. The construction of these hull types are made of layers (table: layer) (e.g. 25 cm brick) and their materials (table: material). Data sources for building envelopes are for example the TABULA project [37] and the ECR project [65]. The TABULA project presents 96 building envelopes for residential buildings, which are split into two refurbishment classes (usual- and advanced refurbishment).

A building category defines the kind of use, for example a SFH, a terraced house, a MFH or an apartment block are classified as residential buildings, and a trade post, an office, a school, a hospital and a hotel are classified as non-residential buildings. However, the user profile can differ from the building category in that a building category can have several user profiles, which consist of square meter specific hourly resolved profiles for internal gains regarding occupancy, equipment and lighting. User profiles in a single-family house include e.g.: sleeping rooms and living room, laundry room and kitchen. Furthermore, the user profile differs with respect to the geographic location. The choice of the building category generally limits the selection of the usage profile. For example, the user profile single/group office or large office is assigned to an office building. This allocation can then be made by means of the building size.

The year of construction and the building category is an indicator for the energy demand in buildings. The building stock is divided in these categories. The building age (table: building_agegroups) is divided into eight building year classes regarding their historical background, which are for example in Austria: until 1919; 1919-1944; 1945-1960; 1961-1980; 1981-1990; 1991-2000; 2001-2009; 2010-today [37]. If the year of construction of individual buildings is not available, a statistical distribution (table: building_agegroups_distribution) [66] is used to assign the missing year of construction to the buildings. The statistical distribution of the heating systems (table: heatingsys_distribution) according to age-groups and types is applied in the same way for the allocation of the missing heating system types (table: heatingtype) of the respective buildings. Each heating system type is assigned an efficiency (table: heatingtype_utilizationdegree) according to age-groups. Heating system types are boilers, a district heating connection, a geothermal probe, a solar thermal plant, an ambient air to water heat pump, a photovoltaic plant, a wind turbine and a water tank connection for DHW. Finally, the DHW load profiles (table: dhw_profile) are calculated with the DHWcalc [67], which considers the simultaneity in the user behavior.
Fig. 3. Typology database structure.

**Project database**

The project database stores the relevant project data with a spatial reference, which is filled in three steps (see Fig. 2). In the first step (A), the required input data are stored. After a categorization/dimensioning process (2), the assigned data (B) are appended. Finally, the simulation results (C) are stored in the database after simulation. This means that the same typology database is reusable and the project database is created for each new project or project version.

Fig. 4 shows the ERD of the project database. The building (table: buildings) is the central part of the project database, which connects with its building id (attribute gid) all the relevant geometric and non-geometric simulation input data and the simulation results. The geometric input data are the ground plan (attribute: the_geom) and building height (attribute: h_mean) and derived attributes such as the building floor area (attribute: building_area_m2), the gross floor area (GFA) (attribute: gross_floor_area_m2), detached/joining (attribute: detached), the number of floors (attribute: floors), the number of dwellings (attribute: dwelling_units) and the number of persons (attribute: persons). The non-geometrical data are the building type (attribute: buildingtypeid), the year of construction (attribute: build_year) and if there is more
than one usage per building, two additional building categories with percentage building use can be specified.

The heatingsys_buildings table lists the selected heating type or a combination of heating types for the respective building. The selected heating systems are then dimensioned. In order to make the heating system scalable, attributes are scheduled for each heating system type. These are e.g. (table: heatingsys): the nominal mass flow (attribute: mnominal), the design power (attribute: qrating) or the $COP$.

Only selected simulation results are stored in the database to reduce the memory consumption. The results are split into groups. Those, which can be assigned to the building zones (table: result_zone) and those to the heating system (table: result_plant). The reason for this is that several building zones can be assigned to a heating system. The table result_zone stores the required results, which are necessary for the energy balance of the thermal zone. Furthermore, the room air temperature as well as the heating and cooling load per square meter are stored. Whereas the power of the power supply units is stored in the table result_plant.

The land utilization plan (table: utilization_plan) is not connected with the other tables and is used for the assignment of the building category.
1) Data acquisition

The data acquisition is a major problem in urban energy analysis [8] [57]. This project thus uses publicly available data. The new developed methodology is based on only three kinds of data:

1. The first piece of information is the building geometry, which consists of the building ground plan and the building height. The ground plan can be achieved through Open Data such as OpenStreetMap [58] or basemap.at [59] or through cadastral plans. Normalized digital surface models provide an opportunity to obtain the building height, which is available in two spatial resolutions for the case study area [60]. 1x1 m grid laser scanning data allows an accurate illustration of the building height. The second option is a free 10x10 m resampled model, but it does not always represent the real building height correctly because the lower spatial resolution causes problems with small buildings. The laser scanning data consists of terrain-, surface- and difference model data and offers one elevation point within the grid.
2. The second piece of information is the land utilization plan by which the buildings are assigned to a building category [61].

3. The third piece of information is the year of construction. Possible data sources are city databases or archives, on-site inspections, aerial photos and expert opinions of regional city planners. The building age are used as an indicator for the building energy consumption. Age-related building envelopes can then be assigned to buildings.

GIS has the purpose to manipulate, visualize and analyze the collected data. QGIS [62] is chosen as an open source GIS tool. GIS provides the option to visualize the 2D building geometry and to display their attributes, hence offering the option to either alter or update the collected building data. The 2.5D geometry consists of the 2D geometry and the height of the laser scanning. The building height is assigned to the buildings directly in QGIS. QGIS stores then the collected data in the DB, see Fig. 2.

2) Categorization/dimensioning

In this methodology, statistical and empirical assignments are made on various levels in order to make basic assumptions about the refurbishment rate or the heating type. It is important to note, that every assumption can be overridden manually for a single housing unit, if more information is available. Fig. 2 shows how the data flow works between the building categorization/dimensioning procedure and the rest of the method. First, the requested input data are collected, which are the building geometry, the land utilization plan and the year of construction from the project DB (A) (connection 2.1 in Fig. 2) and statistics from the typology DB (connection 2.2 in Fig. 2). Necessary data that has not yet been created in the DB is extended in step 2.3 (see Fig. 2) and can thus be re-used in the next project. For example, new building types can be generated by using new materials or load profiles. After all necessary data are available, the categorization/dimensioning procedure takes place and the assigned data are stored in the DB (B) in step 2.4 (see Fig. 2).

Fig. 5. Categorization/dimensioning procedure: data inputs dark grey and outputs white.
Fig. 5 shows schematically the three steps of the building categorization/dimensioning procedure. In the first step, a building category is assigned to each building in the database. The building category may be obtained from the geometry and the land utilization plan. The attributes of the building height, the floor area and detached/joining buildings have an especially strong influence on the assignment. For instance, a building in the residential area with 7 m height, 80 m$^2$ floor area and a detached building is assigned to a SFH, but with a joining building to a terraced house.

In the second step, the building type is assigned by information from building category, year of construction and refurbishment rate. The refurbishment state is assigned via available statistics. A compound interest calculation is thus used with 1 % refurbishment rate per year [66]. It is assumed that buildings, which were built after 2000, are not being renovated. For example, a SFH built in 1996 is assigned to a building type SFH in the construction year class 1991 – 2000 with a refurbishment probability of 22 %. Further, the heating system type is assigned to a building of the stock. The heating system is estimated by means of the year of construction of the building and statistical data of the heating system types distribution depending on the year of construction [68]. Heating system types are used for the heat-, electric energy- and the DHW supply.

Predefined and scalable heating systems are used for the energy supply. In the third step, the heating systems are dimensioned in such a way that the heat is provided by one heat generation unit or with a solar thermal plant combination. The solar thermal plant is dimensioned according to eqs. (2), (3), (4) and (5). The heating utilization $Q_{\text{utilisation}}$ is calculated in eq. (2) and is defined as the heating energy demand per year and collector area. The equation is obtained by a best-fit curve of several solar thermal plant calculations with changing solar coverage $SC$ in Polysun [69]. In the methodology the $SC$ is assumed to be 20 % [70]. The collector area $A_{\text{solar}}$ is calculated from the annual heat demand $Q_{\text{heating} + \text{DHW}}$ for space heating and DHW, see eq. (3). The space heating demand is used from the corresponding building category and the DHW use is set to 850 kWh per year and person with the assumption of 50 l per day water use and a temperature difference of 40 K between cold and hot water. The pumps in the solar thermal plant are dimensioned according to eq. (4). The specific solar mass flow is set to 15 kg (hr m$^2$ collector)$^{-1}$ [71]. The heat exchanger is dimensioned by using eq. (5). The design power of the heat exchanger is calculated with a UA – value (overall heat transfer coefficient multiplied by the contact area) and the temperature difference between supply and return liquid, which is set to 10 K.

$$Q_{\text{utilisation}} = 211845.82 \cdot SC^{-1.6002}$$  \hspace{1cm} (2)

$$A_{\text{solar}} = \frac{Q_{\text{heating} + \text{DHW}}}{Q_{\text{utilisation}}}$$ \hspace{1cm} (3)

$$\dot{m}_{\text{solar}} = A_{\text{solar}} \cdot \dot{m}_{\text{sp}}$$ \hspace{1cm} (4)

$$UA_{\text{heat exchanger}} = 88.561 \cdot A_{\text{solar}} + 328.19$$ \hspace{1cm} [71] \hspace{1cm} (5)
Eq. (6) and (7) sizes the boiler, the ambient air to water heat pump, the district heating connection and the geothermal probe. The design power \( P \) (eq. (6)), which is used for the dimensioning of the heat generators and heat exchangers, is calculated from the specific peak power for space heating \( P_{\text{peak}} \), the GFA and a safety factor. The specific peak power for space heating (W m\(^{-2}\)) is calculated for each building type and is stored in the typology database. The dimensioned mass flow is determined with the design power, the specific heat capacity of water \( c_{pw} \) and the temperature difference \( dT \) between supply and return water. The temperature difference \( dT \) is set to 10 K for solar thermal plants, district heating connection and boilers and to 5 K for ambient air to water heat pumps and geothermal probes. Furthermore, the boiler efficiency depends on the boiler type and the year of construction [68].

\[
P = P_{\text{peak}} \cdot 1.2 \cdot \text{GFA} \quad [72]
\]

\[
\dot{m} = \frac{P}{c_{pw} \cdot dT} \quad [7]
\]

In addition, the necessary length of the geothermal probe \( l_{\text{borehole}} \) is determined by eq. (8). The required peak power demand is composed of the space heating- and the \( \text{DHW} \) power demand, which is set to 250 W per person [73]. The geothermal probe is modelled with a heat pump. The condenser power must thus be converted to the evaporator power with the Coefficient of Performance (COP). The COP of the heat pump is set to 3.5 [74]. The specific geothermal probe is assumed to be 50 W m\(^{-1}\) [73]. The number of ground probes are calculated with a maximum borehole depth of 150 m [73].

\[
l_{\text{borehole}} = \frac{(P+250 \text{ persons}) \left(1 - \frac{1}{\text{COP}}\right)}{P_{\text{specific geothermal probe}}} \quad [73]
\]

The tank volume \( V_{\text{tank}} \) is composed of the tank volume for the \( \text{DHW} \) use and for the solar plant, see eq. (9). The \( \text{DHW} \) tank volume is dimensioned for two days and the temperature difference between hot and cold water is assumed to be 40 K. The tank volume for the solar plant depends on the solar coverage and the collector area. The tank height and radius is calculated according to eqs. (10) and (11). The insulation dimensioned according to eq. (12).

\[
V_{\text{tank}} = \frac{\text{persons} \cdot Q_{\text{dhw person}}}{c_{pw} \cdot \frac{3600}{dT}} + \frac{55.33792 \cdot e^{0.0131 SC} \cdot A_{\text{solar}}}{1000} \quad [9]
\]

\[
h_{\text{tank}} = 1.8442 \cdot V_{\text{tank}}^{0.373} \quad [71]
\]

\[
r_{\text{tank}} = \sqrt[3]{\frac{V_{\text{tank}}}{h_{\text{tank}} \cdot \pi}} \quad [11]
\]
\[ d_{\text{insulation}} = 0.131 \cdot V_{\text{tank}}^{0.2899} \]  

[71]  

The heat distribution system is modelled as a radiator. The power-law exponent changes regarding to the type of the heat distribution system and is set for radiators to 1.28 and for floor heating to 1.1 [72]. The radiator length is calculated with the water peak power for space heating, the gross floor area, the specific radiator power \( P_{sp \, dist} \) and the radiator height, see eq. (13). The radiator height is assumed as 0.5 m. The \( P_{sp \, dist} \) depends on the year of construction, because of a number of technological advances in the heating systems the water supply temperature has decreased. This leads to a smaller temperature difference between radiator and room air. The dimensioned mass flow \( \dot{m}_{hdist} \) is determined with the design power, the specific heat capacity of water \( c_{pw} \) and the temperature difference between supply and return water, see eq. (14). The temperature difference depends on year of construction and is between 5 and 20 K. The \( k \) – value (heat distribution system to room air) is calculated according to eq. (15) and uses a logarithmic temperature difference.

\[ l_{hdist} = \frac{P_{\text{peak}} \cdot GFA}{P_{sp \, dist} \cdot h_{hdist}} \]  

(13)  

\[ \dot{m}_{hdist} = \frac{P_{\text{peak}} \cdot GFA}{c_{pw} \cdot dT} \]  

(14)  

\[ k \text{ – value} = \frac{P_{\text{peak}} \cdot GFA}{\left( \frac{T_{w \, sup} - T_{w \, ret}}{\ln\left(\frac{T_{w \, sup} - T_{room \, air}}{T_{w \, ret} - T_{room \, air}}\right)} \right)^n \cdot l_{hdist}} \]  

(15)  

3) Building modelling

A generation tool for automated thermal city modelling is applied within this research [78]. The tool is used to generate an automated building model for IDA ICE and is written in the programming language Java. The tool adapts the source code from the IDA ICE model directly and executes lisp-based scripts within the IDA ICE simulation environment. The required data for each building is stored in the typology and project DB (see Fig. 2) and is imported through an interface [79]. The inputs from the project DB are the building geometry and the assigned data (connection 3.1 in Fig. 2). The assigned building types in the project DB refer to the stored buildings types (connection 3.2 in Fig. 2) in the typology database. The output (connection 3.3 in Fig. 2) of the modelling process is the thermal city model in the form of a text file.

In this methodology, one building is represented by vertical layered thermal zones. The number of zones per building depends on the number of different uses, but is currently limited by three zones. The reason for this is that buildings in cities are often used according to the same scheme: business areas in the
ground floor, office usage in the first floor and residential use in the upper floors. The separation of the zones with different user profiles has the advantage that different set points for the room air temperature can be considered. A mixed user profile with a mean set point temperature would under-predict the heating demand since the heating periods would be shortened or extended regarding the set points. A further decrease of the heating demand occurs, because loads, solar gains and thermal mass effects may cancel each other out in the one-zone model. The zone geometry consists of the ground plan, the building height from the laser scanning data and the percentage of glazing, which depends on the façade orientation and the building type. Fig. 6 shows the 2.5D geometry of the building model. IDA ICE takes into account the mutual shading between buildings. Further, relevant shading elements with their degree of transparency can be added. The urban green land with evaporation effects is not taken into account, but the ground reflectance can be altered for each façade element. The building envelope is composed of the geometry and the building constructions. Furthermore, internal masses are inserted into the thermal zone to consider the storage effect of the intermediate ceilings. The occupant model of IDA ICE [52] is used instead of the occupancy profiles. The occupancy is also used to allocate the air exchange of the thermal zone, which is set to 30 m³ per person and hour. The window ventilation increases during the summer months due to the behavior of the residents. To consider the increased window ventilation an ideal cooler is implemented to limit the room air temperature to 26 °C. Without this limitation, the room air temperature becomes too high in the summer and consequently the heating period starts too late. Further, the heating system is modelled with a night setback. The set points are 21 °C between 4 am and 6 pm and 17°C between 6 pm and 4 am.

Fig. 6. Automated generated 2.5D geometry model.

The building model generation tool also generates the heating system and parametrizes the heating system with the automated dimensioned values, which are stored in the project database. Fig. 7 shows the graphical representation of all the currently implemented heating system types, which are described with components containing equations, variables and parameters. The connection between two components is represented by lines. Each building is equipped with a heating system, which is composed of a combination of heating types of Fig. 7. The IDA Early Stage Building Optimization (ESBO) plant is used as basis for the heating systems [80]. The models are adapted to make the requirements in the present case. The ESBO-plant is not usable for more buildings, because IDA ICE is developed to model only one building with one heating
system. Thus, the heating system is encapsulated in a macro object and each macro is connected to the corresponding building, see Fig. 7. A stratified water tank with heat losses always connects the heating system with the heat distribution system. The number of water layers in the tank is set to eight. The DHW is heated either by a centralized heating system or by a decentralized, electric unit. The solar thermal plant is mostly built in combination with another heating system type. A standard boiler is used for each boiler type by adapting the utilization factor. The district heating network is connected with the tank in the heating system via a heat exchanger.

Fig. 7. Predefined and scalable heating system types [78].
4) Dynamic simulation
The building simulation tool IDA ICE is used, because the software allows a detailed dynamic validated building simulation [75], the modelling processes can be automated through lisp scripts and the simulation process is parallelizable. Thus, the limiting factor regarding the investigation area is the hardware, namely the number of CPU cores. The future trend is towards cloud computing—for example Microsoft Azure [76]—so that huge models with multiple thousands of thermal zones can be simulated [77].

The simulation tool requires as input the automated generated thermal city model (4.1 in Fig. 2). IDA ICE writes the previously defined results into text files and these are written into the database by means of a Java script (4.2 in Fig. 2).

5) Result visualization
The presentation of the results is an important part of introduced methodology, because the number of results increases exponentially with the size of the investigation area. It is now a difficult task to filter the main results for the simulation expert or the client. The simulation experts are primarily interested in the correctness of the results. The method uses predefined and scalable heating systems for each heating type. This avoids errors during the modelling process. A check of the room air temperature ensures the right dimensioning of the heating system. The heating system is oversized when the room air temperature overshoots the set point and is undersized when it reaches too late or not the set point. The room air temperature can easily be plotted for all buildings. Clients are e.g. energy suppliers, urban planners and investors, which are interested in a brief and clear presentation of the results. In this context, Diagrams and tables are not enough to display results. GIS provides thus an opportunity to present the results graphically in the form of heat maps, 3D visualizations or temporal courses of the actual building power consumption, see section 4.

4 RESULTS AND DISCUSSION

This section presents the validation results of the methodology introduced on the basis of 69 buildings. The SFH have the greatest impact, because of 46 samples, whereas office, school and commercial buildings have less influence. Finally, the results of the town Gleisdorf are presented.

Fig. 8 shows the heating and DHW energy consumption, which is cumulated for each building category. The SFH show a good approximation with a mean deviation of -0.98 % of the annual heating and DHW demand. The simulated energy consumption is 789.6 MWh and is 7.8 MWH less than the measured energy consumption. The results indicate that assignments of building types to buildings is correct. The MFH show a greater deviation of the annual energy demand with a difference of 850 MWh (~36.5 %). The reason is that the simulation model only uses a maximum of three vertical layered thermal zones for a building. This approximation leads to an underestimation of energy demand, because the simulation tool treats a zone as a node with perfectly mixed air volumes. The thermal solar gains of south facing rooms are
thus mixed with north facing rooms. This insight is in agreement with Smith et al. [51] and becomes more important with larger buildings. This effect also leads to an underestimation in schools (-14.08 %), offices (-1.51 %) and commercial buildings (-60.55 %). Furthermore, the user behavior is difficult to estimate for commercial buildings, due to differences in energy usage [65]. For example, a bakery or a hairdressing salon use much equipment leading to higher internal gains, thus the spatial heating demand decreases. By contrast a shop has much less internal gains through electrical equipment, but the number of people increases. In addition, laser scanning data with a spatial resolution of 10x10 m leads to a slightly mismatch of the height of some buildings.

![image](image.png)

**Fig. 8.** Cumulated building category specific annual heating and DHW energy consumption, in MWh.

Fig. 9 shows the deviation of the simulation results to the measurement data in dependence on the annual heating and DHW energy consumption. The spread of the SFH is between -64 % and 60 %. A histogram of SFH is shown in a subplot of the figure. The maximal energy consumption of a SFH is 34 MWh. The MFH have a spread between -12.8 and 74.9 %. The figure shows the underestimation explained above. The refurbishment rate of the buildings explains the relatively large spread of the simulation results. In the proposed methodology, the refurbishment rate is assumed with 1 % per year. Thus, some buildings are simulated with modelled refurbishment, but they are in reality not refurbished and vice versa. This effect causes two frequency peaks in the histogram. Furthermore, the extent of a refurbishment is always different. This can be for example, a window change, a new insulation or a new heating system. Consequently, an advanced refurbishment of a building with a year of construction before 1919 causes a decrease of the overall heat transfer coefficient of walls from 1.4 to 0.12 W (m² K)⁻¹ [37].
Fig. 9. Building category specific annual heating and DHW energy consumption, in MWh; histogram of the SFH in the subplot.

Fig. 10 shows the dynamic behavior of two SFH. The main plot shows a typical building (gid=127) with an annual heating demand of 15 MWh and a small mean error; compare Fig. 9. Other SFH show a similar behavior pattern. The upper main plot of Fig. 10 indicates that the trend of the district heating power consumption over the whole year is well represented. The results show a bias of -0.06 kW, which decreases in the summer period without heating, a standard deviation $s$ of 1.163 kW, a RMSE of 1.164 kW $\pm$ 4 kW and a maximum error of 6.174 kW. The trajectory of the district heating power indicates that the assumption with an ideal cooler in the summer period is correct. The ideal cooler limits the room air temperature to 26 °C. In reality, the ventilation increases during the summer due to window openings. Consequently, the heating period starts correct in autumn. In addition, a SFH (gid=114) is shown in a side plot, which clearly underestimates the energy consumption (compare Fig. 9), which occurs because of the problem with the renovation rate. Nevertheless, the figure shows that the trend is correct represented, but a bias is present.
Fig. 10. District heating power consumption of two SFH.

Fig. 11 shows a one-week interval of a SFH, MFH, office, school and a commercial building. The time interval is during a typical week in winter from 15.1.2015 to 21.1.2015. The night setback of the heating system NSB is visible in the plots for all building categories (highlighted in Fig. 11). The results of the SFH in particular indicate a good approximation of the reality. The results of the office building can also follow the trend. The energy consumption in SFH and MFH decreases at the weekend, because the use of electrical equipment and the number of people at home in dwellings rises. The small peaks during one day can be attributed to the DHW consumption. The small peaks occur in the morning and the evening (exemplary highlighted in Fig. 11). The duration and height of the peaks are statistically distributed [67]. The results of the school building indicate an underestimation during school days. Internal gains during the day increase due to people in classrooms, but the infiltration rises because of window openings during breaks. The window openings have to be implemented in the simulation model with a variable infiltration or a window-opening model, which means high computational costs. The problems in commercial buildings have already been above mentioned.
Fig. 11. One-week interval of the district heating power consumption of each building category.

First, the methodology has been successfully validated in small-scale. Subsequently the method is used to calculate the thermal energy demand of the whole community of Gleisdorf with its 1,945 buildings. The whole town is divided into 28 sections, whereas areas with adjacent buildings are simulated in one model to consider mutual shading. One model contains 70 buildings with its heating system.

QGIS is used for the representation of the simulation results, because GIS provides the opportunity to present the results graphically in the form of heat maps, 3D visualizations or temporal courses of the actual building power consumption. Fig. 12 shows two heat maps of the town Gleisdorf with different grid cells. The simulated buildings are depicted in the map through their ground plan. The power consumption of the HS is represented as colors in a matrix of 100x100 m (left heat map) and in a matrix of 50x50 m (right heat map). A heat map shows the areas with high and less energy consumption. A heat map of this kind can be used to identify suitable areas for a district heating network or an extension. Different resolutions of the grid cells play a central role in the planning of district heating networks. Larger raster cells can be used to determine the areas in which district heating should be applied. Whereas smaller grid cells help energy suppliers to plan the exact pipe laying. The scale of the legend differs with the grid size because the undeveloped area changes.
Fig. 12. Heat map (left/right 100/50 m grid) of Gleisdorf; energy demand for space heating and DHW use.

The building-related representation offers a further possibility to represent the annual energy consumption. Fig. 13 shows the building-related annual energy consumption of Gleisdorf. A further opportunity would be to consider the energy consumption per area. QGIS offers thus with the Qgis2threejs plugin a possibility to represent 3D building data [82]. The QGIS Qgis2threejs plugin exports terrain data and vector data to a web browser. The 2.5D buildings and the terrain data can be viewed in the web browser. In addition, the essential attributes regarding each building can be transferred to the browser. For example, this may be information about the heating system or simulation results. Urban planners can use this feature to display the actual situation or development scenarios. This information can thus improve negotiations with investors regarding energy planning.
It is essential to grasp the dynamic interaction of the buildings with smart energy systems or renewables. The representation of temporal courses thus plays an important role for energy suppliers. The TimeManager plugin in QGIS provides an opportunity to animate vector features based on a time attribute [81]. This allows building-related representations of temporal courses. Animations can be created directly in QGIS or image series can be exported. Fig. 14 shows a selection of an images series of the heating system power consumption of Gleisdorf during one day. In addition to the heat map, additional attributes such as the total requirements of hot water, room heat, $HS$ and electrics of all Gleisdorf buildings and relevant climatic data such as ambient air temperature and the direct radiation are displayed. The first image shows the $HS$ power consumption per square meter $GFA$ at midnight, which is low because of the night setback. The next image shows the buildings with heating systems switched on. The heating power drops with increased direct radiation. The last image shows the $HS$ demand after activation of the night setback.

Fig. 13. 3D visualization of Gleisdorf.
The methodology is only valid within a wider observation horizon, because the method uses statistics and probabilities, which are refurbishment rates, statistics of heating types and assignment of building types to buildings. This means that a single building may not be correctly reproduced due to the assumptions in building use and construction. If the data is more accurate, the automatically created model can be flexible refined or expanded to any level of detail. The simulation environment IDA ICE provides the opportunity to model the basic building with reasonable computational costs up to relatively complex models such as geothermal probe fields or double facades [52]. It is also possible to integrate new districts to the automatically generated simulation model. This is very useful in practice, because investors or urban planners often wish to expand or renovate an urban development area.

5 CONCLUSION

This paper presents a new and validated methodology for fully automated building modelling within urban districts based on publicly available data. Each building is connected with a detailed heating system. This includes, for example, water tanks, boilers, ambient air and ground heat pumps, solar thermal power, photovoltaic and wind turbine power, district heating connection and a heat distribution system. It was discovered that a building represented by a maximum of three vertical layered thermal zones can adequately model the reality of the situation. The simulation of detached single family houses in particular shows that adequate results with a mean deviation of -0.98 % have been achieved. The spread of the 46 detached single family houses is between -64 % and 60 %. The refurbishment rate of the buildings explains the relatively
large spread of the simulation results. In the proposed methodology, the refurbishment rate is assumed to be 1% per year. In multi family houses or other big buildings the approach with a maximum of three thermal zones per building leads to an underestimation of the energy demand. Thus, a compass direction zoning with a core zone may improve the methodology.

The proposed method is able to predict the energy demand of the building stock and to examine the dynamic interactions between buildings and a district heating network, which can be modelled also in IDA ICE or in another dynamic simulation tool and the required data is exchanged via co-simulation. Furthermore, the simulation model is flexibly expandable at any level of detail. This means that the simulation model is flexible enough for the addition of new buildings and for changing the old building stock data in order to refine the model easily at any level of detail.

QGIS provides an adequate opportunity to manipulate, visualize and analyze the collect input data. Further, a brief and clear representation of the simulation results are import for investors, urban planners and energy suppliers. QGIS provides thus a possibility to present the results graphically in form of heat maps, 3D visualizations or temporal courses.

The limitation regarding model size with approximately 300 buildings with simple geometry or 70 buildings with detailed heating system lead to the necessity of model split. IDA ICE provides a parallelizable simulation process and a co-simulation interface between these models. The limiting factor in the context of model size is thus the hardware, specifically the number of CPU cores.

ACKNOWLEDGEMENTS
The reported research has been conducted within the framework of the Research Studio Austria “EnergySimCity - Ganzheitliche Analyse und Simulation von Energiesystemen und Ressourcenverbünden in Städten und Stadtquartieren” (Project-No.: 844732), funded by the Federal Ministry of Science, Research and Economics (bmfw). Further, the utilities of the city of Gleisdorf have enabled the validation process by providing measurement data.
REFERENCES


