BauSIM 2024 in Wien
23. September – 26. September

Revolutionizing Heating Planning: A Data-Driven Approach for Accurate and Scalable Energy Consumption Predictions

Stefanie Segbers, Alexander Küster-Inderfurth, Christoph Nytsch-Geusen
Universität der Künste Berlin, Deutschland, E-Mail: s.segbers@udk-berlin.de

Abstract

Accurately predicting heating energy consumption for buildings is crucial for optimizing communal heating planning. However, the lack of detailed building information hinders precise estimations necessary for city decarbonization and efficient use of local renewable or waste energy. To address this challenge, this study utilizes 3D geometries and open data of Berlin's half a million buildings to perform building-specific simulations and trains AI-models to predict heating consumption.

A physics library was developed using statistical data from 16 building types and 6 refurbishment scenarios, providing building model components with physical properties. Additionally, a usage library comprising 24 different usage types and a weather database spanning 13 recent years, and 3 future climate scenarios for the next 8 decades were created. Simulations on a subset of Berlin's buildings were conducted using these libraries.

To extend results city-wide, an Artificial Neural Network (ANN) was trained using input data and simulation outcomes. This approach replaces traditional simulation software, offering a more efficient alternative. An intuitive web-application with a fast AI surrogate model at its core allows individuals to access and modify results for their buildings as well as correcting values to reflect actual demands. Corrected data is incorporated into neural network training, improving accuracy for various climate scenarios and potential refurbishments.

This iterative process not only streamlines communal heating planning in Berlin but also establishes a flexible framework for continuous improvement. This emphasizes the adaptability and scalability of our proposed methodology, contributing to enhanced predictions and a more efficient approach to city-wide heating planning.

Introduction

For decarbonization concepts of cities, communities and neighborhoods, a building-specific database of the building’s properties and energy demands is essential. Since German legislators have passed a law on communal heat planning (Bundesregierung 2023), concerns on how to approach the obligations of the law are growing. A first step must be a thorough compilation of building data that represents the status quo. Often however, energy cadastres provide only block-level or 100m x 100m grid consumption data (Paardekooper, et al. 2018), (Dochev, Seller und Peters 2019), (Behörde für Stadtentwicklung und Wohnen, Stadt Hamburg 2019). Yet, to undertake precise planning and to locally utilize renewable energies or waste heat from sources such as industry, data centers, or wastewater, building-specific data is indispensable.

In addition to the current heating consumption of buildings, future scenarios for refurbishments and the impacts of climate change (Kahlenborn, et al. 2021), such as increasingly mild winters or warmer summers, are crucial for long-term heat planning. Unfortunately, there is a scarcity of comprehensive and freely available information on buildings for those scenarios and in general in Germany.

Consequently, in practice energy concepts for neighborhoods, districts, communities, or cities often rely on rough assumptions regarding heating and domestic hot water consumption. Typically, the demands are extrapolated based on the building's area and educated guesses of specific values. However, more accurate predictions are imperative for long-term and precise heat planning.

Drawing from insights gained in a previous project, Open eQuarter (Nytsch-Geusen et al., 2016) which utilized a GIS-based statistical approach to assess the specific energy demands of buildings in Berlin, this new method aims to further enhance accuracy and efficiency. To generate precise forecasts of the heating energy consumption for tens of thousands of buildings and different refurbishment and climate scenarios in a short amount of time, a workflow has been developed that replaces physical-based building-specific simulations with a prognosis model based on a neural network. This workflow has been applied to the city of Berlin. To enhance the energy demand prognosis of the neural network, users are provided with the opportunity to adjust building parameters in an iterative process.

Simulations as training data

To carry out simulations with the software SimStadt (Coors, et al. 2021) for most of Berlin's buildings at building level, several building-specific datasets such as the cubature, the age and the type of use were compiled first. These attributes serve as key to merge further information from large statistical libraries for each individual building in the building-specific simulations: The 3D building models in CityGML format and the usage type of the building, which can be identified through ALKIS function numbers (Arbeitsgemeinschaft der Vermessungsverwaltungen der Länder der Bundesrepublik Deutschland (AdV) 2018), were merged. This information is provided by the State of Berlin through its Geoportal (Senatsverwaltung für Stadtentwicklung, Bauen und Wohnen Berlin 2024). Some of the 3D building models contain parts of buildings such as underground garages or driveways. These were removed from the building data sets with specially designed filters to keep only data that is most relevant to heating consumption. In addition to the type of use, each individual building model was enriched with information on the age of the building. The building ages from a 1992/93 building age map and a monument map were used for this purpose. For all buildings whose building age was not covered by these maps, the mean building age of the city block from German census data was used for simplification. In many other cities in Germany, there is building-specific information on building ages, which means that applying the workflow to other areas may even eliminate the need for time-consuming evaluations of building age maps and data.

Statistical data from IWU surveys were also used to describe the physical parameters of the envelope surfaces. One survey (IWU Institut für Wohnen und Umwelt 2016) was carried out for residential buildings and published in 2011. Here, the residential buildings were divided into five different residential building types and up to nine different construction age classes per residential building type. A second survey (IWU Institut für Wohnen und Umwelt 2021) on non-residential buildings was recently completed and published in 2021. The non-residential buildings were divided into eleven different usage types, each containing three different building age classes. The data surveys of residential and non-residential buildings refer to buildings throughout Germany and, although only statistical data is provided, the data is currently the best source for physical parameters as it is classifying non-residential buildings into 33 categories and was deducted from data of up to 800 buildings per category.

The data collection and categorization were carried out in different ways for residential and non-residential buildings. For residential buildings, SimStadt provides a library in which the refurbishment categories were unrefurbished, so-called "original" buildings; data collected by IWU from moderately refurbished buildings, labelled “medium” by SimStadt; as well as the data from advanced refurbished buildings which were labelled “advanced” by SimStadt, see Figure 1. There were also refurbishment scenarios where a building either only had a refurbished roof, refurbished windows, or walls. For all other parts of the building, it was assumed that they were still in their original or unrenovated condition.

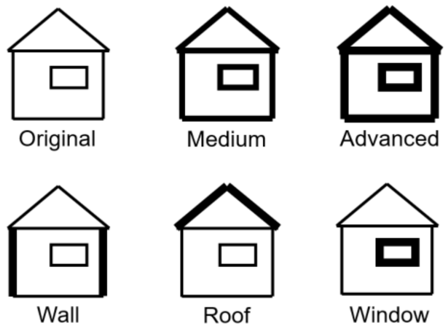

Figure 1 The six refurbishment scenarios used for building simulations in SimStadt with the highlighted refurbished components.

These refurbishment labels for residential buildings were then also used for the non-residential buildings. However, the data collected by IWU for the non-residential buildings was sorted differently in their survey: For every building category the mean average value and the standard deviations were determined. For simplification the category “medium” refurbishment was used for the mean average value. Furthermore, deviation indicating a worse refurbishment state was used for the “original” refurbishment category. The deviation indicating a better refurbishment state was used for the “advanced” category. In the example of a heat transfer coefficient this would be the lower value. For parameters such as the average storey height, the mean average value was used, as it is assumed that a refurbishment will not change the storey height in most cases.

In addition to the library of physical parameters, a usage library was created with 24 different building usage types. These 24 types of usage are made up of 22 different non-residential and only 2 residential building use profiles. The profiles for residential buildings are very similar, with only a distinction being made between apartment buildings and smaller houses, e.g. single-family houses. For each usage profile, data from usage profiles from DIN 18599 (Deutsche Institut für Normung e.V. 2018) were used as a basis. The internal gains as well as occupancy conditions were derived from the DIN 18599 profiles. Furthermore, for every building category there is an indirectly heated area ratio set so that an assumption is made for every building. Storeys that are underground are counted as cellars and those that are not high enough for living purposes are thus considered unheated areas. Unfortunately, no assumptions could be made regarding the zoning of non-residential buildings, as the interior of the buildings is completely unknown due to the available LoD2 3D geometries. All buildings are therefore treated as single-zone models. The assignment of buildings to a usage profile in the library was done via the ALKIS building function number of each building.

The building simulations were carried out with weather data for the past 13 years. Past years were simulated as these can be compared with real consumption values. Additionally, weather data was also generated via Meteonorm (Meteotest 2020) for the next eight decades, i.e. 2030, 2040 etc. up to 2100. The three RCP climate scenarios 2.6, 4.5 and 8.5 (IPCC 2014) per decade were used for the simulations.

To initiate these simulations, a Python script was written that automatically simulates six refurbishment scenarios and a total of 37 weather scenarios for 408,000 buildings of Berlin using SimStadt. After the simulation, the data was processed and stored automatically in a PostGIS database.

The simulations can be repeated for other cities and municipalities in Germany. The libraries created can be used, only the 3D building models would need to be enriched with the information mentioned at the beginning and the climate data for the specific locations used in the simulation.

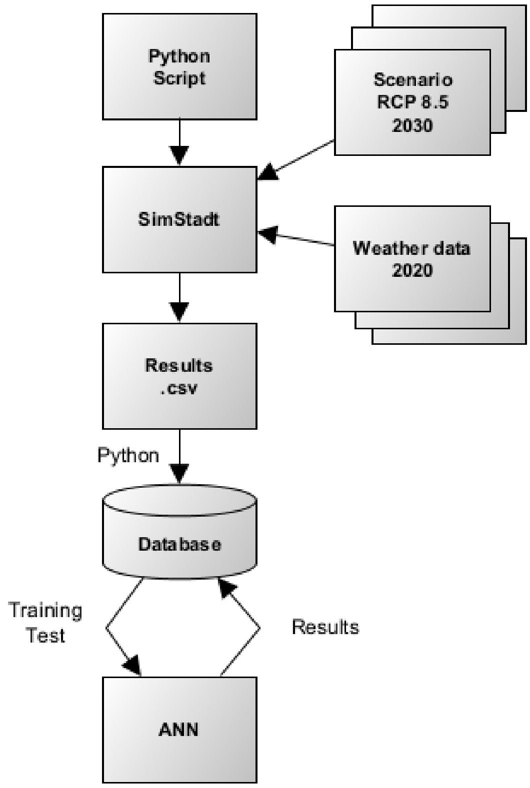

Figure 2 Training strategy for artificial neural networks (ANN) with synthetic training data derived from simulations.

Neural network as a surrogate model

In order to determine the heating and domestic hot water demands for further buildings, the extent to which ANNs can serve as surrogate models for traditional simulations is being investigated. ANN are already used in various domains to efficiently substitute complex and time-consuming simulations (Sun und Wang 2019), (Donnelly, Daneshkhah und Abolfathi 2023). Due to the enormous amount of data generated by the simulations, ANNs can achieve good forecast quality and provide quick and effective predictions for heating and hot water demands for additional buildings.

Trials with the data have shown that the prediction accuracy is higher when two different ANNs are trained for domestic hot water (DHW) and heating demands (HD). For this purpose, another Python script has been written, which accesses the simulation inputs and outputs in the database and chooses the relevant inputs and outputs for the training of the respective ANN. For the prediction of domestic hot water (DHW) far less input parameters are relevant than for the heating demand (HD). Table 1 shows the different parameters significant for the training of the prediction models.

Table 1: Input features for the distinct ANNs

Input features for ANNs	**HD**	**DHW**
Year of construction	x	x
ALKIS code	x	x
Primary and secondary usage zone areas	x	x
Heated area	x	x
Building type	x	x
Footprint area	x	
Window area	x	
Total wall above ground outwall/ shared wall/roof area	x	
Heat transfer coefficients for walls/ windows/ground/ roof/ceiling	x	
Gross volume and heated volume	x	
Storey number	x	
Average storey height	x	
Basement and/or attic heating	x	
Basement ceiling height above ground	x	
Surface area to volume ratio	x	
Climate year and scenario	x	

First, the data is prepared for the ANNs, and an exploratory data analysis is conducted. The script converts several non-numeric columns into a "category" data type. It also includes functions for displaying histograms and checking the normal distribution of the data. It analyses both numeric and categorical columns to provide insights into the distribution of the data.

Moreover, a correlation map is created to provide insights into data relationships to choose the most relevant features to feed into the network models. However, more methods, such as regularization need to be explored as the analysis of correlation between categorical features cannot be assessed by computing correlation coefficients. Moreover, a method for splitting the data into training and test sets and for detecting and handling outliers in the dataset are provided. Together, these functions ensure that the data is ready for model fitting and further analysis.

In the Python script two artificial neural network models are created using Keras (Google LLC 2024), a high-level API for the implementation of neural networks. It enables model creation with customizable architecture and optimization settings. The script provides functions to build and train the neural network as well as functions for visualizing the training history and making predictions on unseen test data. These functions are crucial for the development and fine-tuning of neural network models for various prediction tasks.

When implementing the neural network model, various hyperparameters concerning the model's architecture and training need to be optimized. A grid search was performed using 5-fold cross-validation with GridSearchCV from scikit-learn (scikit-learn 2024) to optimize hyperparameters such as batch size, epochs, neurons in the hidden layers, and learning rate. The optimal parameters found were a learning rate of 0.001 and a batch size of 2048. The train/validation/test split was 0.7/0.15/0.15.

A sequential model with two fully connected (dense) hidden layers was implemented, each with 128 neurons, which is effective for regression problems. The activation functions used were ReLU (rectified linear unit) for the hidden layers and a linear function for the output layer.

The training objective was to predict hot water demand and heating demand accurately. The Mean Squared Error (MSE) was chosen as the loss function, minimized using the Adam optimizer. Training was terminated early if the validation loss plateaued, with a maximum of 256 epochs and early stopping set at 16 epochs. The model achieved the following performance metrics:

- Loss: Train 0.0350, Validation 0.0464, Test 0.0323
- Mean Absolute Error (MAE): Train 0.0385, Validation 0.0323, Test 0.0294
- Mean Absolute Percentage Error (MAPE): 11.96%

Figure 3 is a scatter plot illustrating the relationship between simulated and predicted yearly heating demands. The blue dots, which represent data points, are primarily clustered close to a straight line, indicating a strong correlation and suggesting that the model's predictions are highly accurate. However, there are a few outliers further away from this ideal fit line, highlighting areas where the model's predictions deviate from actual values. This indicates some level of uncertainty and inaccuracy in the model's predictions, which must be critically evaluated to improve the overall quality and reliability of the results.

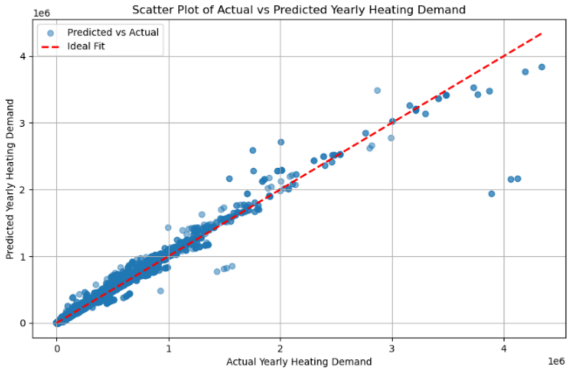

Figure 3 Scatter Plot of Actual (Simulated) vs Predicted Yearly Heating Demand

The training of the model takes about 0.0002% of the time of the simulations. Once the surrogate model is trained it can generate results in 0.000003% compared to the simulations. The Mean Absolute Percentage Error (MAPE) of 12% is not an outstanding prediction accuracy but it is considered good when compared with other calculation methods used for existing buildings.

The trained model can be saved and reloaded for further training or predictions on new datasets. For new data, input parameters need to be collected or calculated, with the outputs predicted by the model. The 3D geometry models, enriched with information from ALKIS, along with the gathered heat transfer coefficients and geometry parameters, support this process. Predictions for new data are stored georeferenced in the database.

Website for new data

Predicted heating and hot water demands will be published through a dedicated web application. All interested stakeholders will be able to access building specific prognosis data as well as underlying input data. Furthermore, users will have the ability to get data not only for single buildings but also use selection tools to extract data for the entirety of the buildings they are interested in. This will be useful e.g. for project developers who need to get an overview of the consumption in a development area. In the final implementation of the proposed web application users will even be able to submit their own building and consumption data for the already mentioned iterative refinement of the artificial neural networks.

To achieve this a complete stack of web technologies needs to be implemented, see Figure 4. As a base a PostgreSQL (The PostgreSQL Global Development Group 2024) database management system with PostGIS (PostGIS PSC & OSGeo 2024) extension is setup. This can be achieved in several ways, one of which is to use a Docker Container of the 3DCityDB (Yao, et al. 2018) developed by TU Munich. While not being the core of the database system, the 3DCityDB and its dependencies correlate strongly with the requirements for this project. These requirements include a robust relational database management system with the capability to work with GIS based data. By starting with an instance of the 3DCityDB it is also immediately possible to import 3D geometries of entire cities into the database using the Importer/Exporter tool by the same institution.

The city of Berlin provides a 3D geometry model of the entire city in the CityGML format, that is freely available and is imported into the project database. Also imported were building input parameters from the feature list of the building database which originally comes from the ALKIS system and the mentioned libraries.

Within the project's GIS database all available building information is combined and an enriched CityGML model is exported. Only buildings and building parts that are considered to have heat energy consumption are exported.

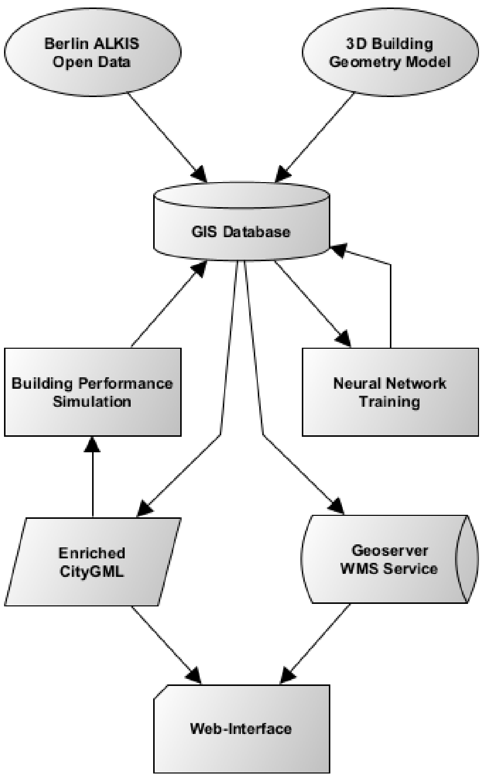

Figure 4 Data flow from source data to publication.

The same CityGML model is used for the afore mentioned simulation and training as well as for the visualization of 3D buildings within the web interface. For this the CityGML format is converted into Cesium 3D tiles, a proprietary data format that is supported by the freely available CesiumJS (Cesium GS, Inc. 2024) JavaScript library which is used as an engine for the map part of the web interface. While 3D models can be displayed it is also possible to get a 2D map. This map is hosted by Geoserver (Open Source Geospatial Foundation 2024) instance which serves a table of the central database with geometry information as a web map service (WMS). WMS can not only contain the 2D geometry information but also other attributes. Finally, a third way to access data is supplied with routes that are defined in the node.js backend of the described web application which can also query the central database and use predefined database functions to aggregate building information based on user requests and send them to the client's web browser for displaying.

The clients are running the presented web application by accessing the project website through a web browser, see Figure 5. When accessing they can view single building data by simply clicking individual building shapes on the 2D default view. When users choose to use one of several selection tools, they can select building groups by giving a point and a radius, by drawing a polyline with a specified distance or by drawing a polygon. The displayed data switches then to an aggregated summary of all buildings selected. Interested users can download the detailed data of all buildings by downloading a csv file.

The web interface gives stakeholders the opportunity to explore the results of our research for all buildings of Berlin. The predicted heat energy consumption is displayed together with the building energy efficiency class derived from these values. In addition, users can get aggregated results for city areas of their choice and use this data in their own projects.

Depending on how much knowledge the user has, they will be able to choose between an "expert mode" for people with domain knowledge and a simplified "citizen mode". Expert users are encouraged to also use provided APIs to get to the data more efficiently. It is possible to use the aforementioned URL routes to access data. The development of a Python API has been started, which will provide a more intuitive interface to the web API. The Python API will be published.

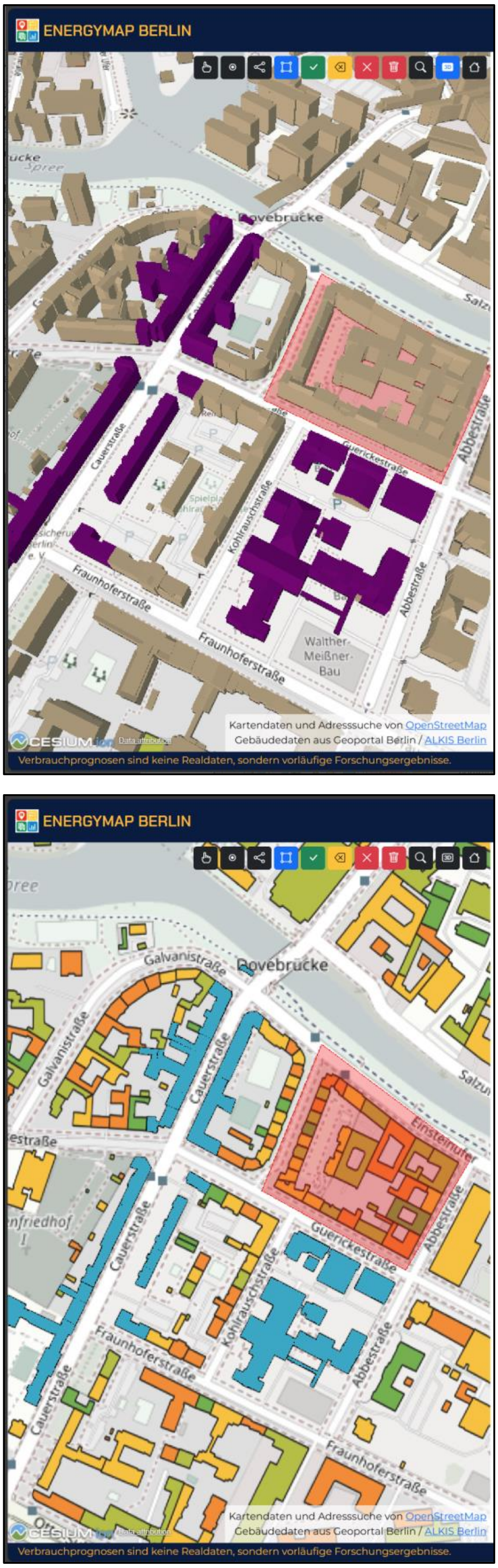

Figure 5 Web-interface including 3D building view.

Conclusion

In conclusion, the integration of 3D geometries, open data, and artificial intelligence (AI) has enabled significant progress in predicting heating and hot water demands for buildings in Berlin. By developing sophisticated physics and usage libraries, coupled with extensive simulations and the implementation of artificial neural networks (ANNs), this study has established a robust framework for building-specific estimations.

The iterative process undertaken facilitates faster predictions and establishes a scalable methodology applicable to city-wide or even nation-wide heating planning. Incorporating user-centric features, such as a web interface for adjustments and corrections based on actual demands, ensures continuous improvement and adaptability to diverse climate scenarios and refurbishment plans.

However, while the utilization of ANNs as surrogate models offers a high-speed alternative to traditional simulations, it does not achieve the highest accuracy, with a Mean Absolute Percentage Error (MAPE) of about 12%. This indicates that while the ANNs provide quick results, there is room for improvement in terms of accuracy. Future research should explore methods to enhance the precision of these models, such as by splitting the data into different building types or incorporating additional relevant features.

The comprehensive Python scripts developed for data preparation, model creation, and evaluation contribute to the transparency and reproducibility of the methodology. The implementation of a dedicated web application further democratizes access to prognosis data, empowering stakeholders to explore and utilize building-specific information for informed decision-making. With features allowing for selection tools, data extraction, and APIs for expert users, the platform promotes collaboration and engagement among various stakeholders involved in city planning and development.

Overall, this study emphasizes the importance of leveraging advanced technologies and open data initiatives to address challenges in communal heating planning. While the current approach provides a solid foundation, ongoing efforts to refine and enhance model accuracy will be crucial for maximizing the benefits of AI in energy optimization and climate mitigation. As cities worldwide strive towards decarbonization and resilience, the methodologies and tools developed in this research serve as valuable assets in shaping future strategies for sustainable and efficient energy practices in urban environments.

Acknowledgement

The project outcomes were only made possible through the comprehensive data acquisition efforts conducted by the Institut für Wohnen und Umwelt, in conjunction with the utilization of the sophisticated simulation tool, SimStadt, developed by the HFT Stuttgart. The statistical data as well as the simulation tool played a pivotal role in the successful attainment of the objectives.

This research is funded by the German Federal Ministry of Economic Affairs and Climate Action (FKZ 03EN3051A). We would like to express our sincere gratitude for their support.

References

Arbeitsgemeinschaft der Vermessungsverwaltungen der Länder der Bundesrepublik Deutschland (AdV). „Dokumentation zur Modellierung der Geoinformationen des amtlichen Vermessungswesens - ALKIS-Objektartenkatalog DLKM.“ ALKIS-Katalogwerk, 2018.

Behörde für Stadtentwicklung und Wohnen, Stadt Hamburg. *Wärmekataster Handbuch.* Stadt Hamburg, 2019.

Bundesregierung. *Gesetz für die Wärmeplanung und zur Dekarbonisierung der Wärmenetze (Wärmeplanungsgesetz - WPG).* 12 20, 2023. https://www.gesetze-im-internet.de/wpg/WPG.pdf (accessed 02 08, 2024).

Cesium GS, Inc. *CesiumJS - 3D geospatial visualization for the web.* 2024. https://cesium.com/platform/cesiumjs/.

Coors, Volker, et al. *EnEff:Stadt- SimStadt 2.0 - Schlussbericht.* Hochschule für Technik Stuttgart, 2021.

Deutsche Institut für Normung e.V. "DIN V 18599-10 | 2018-09." *Energetische Bewertung von Gebäuden - Berechnung des Nutz-, End- und Primärenergiebedarfs für Heizung, Kühlung, Lüftung, Trinkwarmwasser und Beleuchtung - Teil 10: Nutzungsrandbedingungen, Klimadaten.* 09 2018.

Dochev, I., H. Seller, and I. Peters. *Aggregating and visualizing urban heat demand using graph theory. A case study from Hamburg, Germany.* preprint version, 2019.

Donnelly, James, Alireza Daneshkhah, and Soroush Abolfathi. "A Physics-informed Neural Network Surrogate Model for Tidal Simulations." *5th International Conference on Uncertainty Quantification in Computational Sciences and Engineering.* 2023. 836-844.

Google LLC. *Keras.* 2024. https://keras.io/.

IPCC. *Representative Concentration Pathways.* 2014. https://sedac.ciesin.columbia.edu/ddc/ar5_scenario_process/RCPs.html (accessed 03 11, 2024).

IWU Institut für Wohnen und Umwelt. *Datenerhebung Wohngebäudebestand 2016.* 2016. https://wohngebaeudedaten2016.iwu.de/ (accessed 03 11, 2024).

—. *ENOB DataNWG: Forschungsdatenbank.* 2021. https://www.datanwg.de/forschungsdatenbank/ (accessed 03 11, 2024).

Kahlenborn, Walter, et al. *Climate Impact and Risk Assessment 2021 for Germany - Summary.* Umwelt Bundesamt, 2021.

Meteotest. *Meteonorm Version 8.* 2020. https://meteonorm.meteotest.ch/meteonorm-version-8 (accessed 03 11, 2024).

Nytsch-Geusen, Christoph, Werner Kaul, Philipp Wehage, and Michael Färber. *Nachhaltige und energieeffiziente Bestandsentwicklung von Stadtquartieren – Entwicklung eines Analyse- und Planungstools zur energetischen Bewertung und Sanierung von Stadtquartieren (Open eQuarter).* Stuttgart: Fraunhofer IRB Verlag, 2016.

Open Source Geospatial Foundation. *GeoServer is an open source server for sharing geospatial data.* 2024. https://geoserver.org/.

Paardekooper, Susana, et al. *Heat Roadmap Europe 4.* Aalborg University Denmark, 2018.

PostGIS PSC & OSGeo. *PostGIS.* 2024. https://postgis.net/.

scikit-learn. *scikit-learn Machine Learning in Python.* 2024. https://scikit-learn.org/stable/.

Senatsverwaltung für Stadtentwicklung, Bauen und Wohnen Berlin. *Geoportal Berlin / ALKIS Berlin (Amtliches Liegenschaftskatasterinformationssystem).* 2024. https://fbinter.stadt-berlin.de/fb/wfs/data/senstadt/s_wfs_alkis?REQUEST=GetCapabilities&SERVICE=wfs.

Sun, Gang, and Shuyue Wang. "A review of the artificial neural network surrogate modeling in aerodynamic design." *Proceedings of the Institution of Mechanical Engineers, Part G: Journal of Aerospace EngineeringVolume 233, Issue 16*, 2019: 5863-5872.

The PostgreSQL Global Development Group. *PostgreSQL: The World's Most Advanced Open Source Relational Database.* 2024. https://www.postgresql.org/.

Yao, Zhihang, et al. "3DCityDB - a 3D geodatabase solution for the management, analysis, and visualization of semantic 3D city models based on CityGML." *Open Geospatial Data, Software and Standards*, 2018.