[Prof. Dr.-Ing. habil. Christoph van Treeck, Aachen]

English

Latent space learning of energy consumption and indoor environmental quality data in context of building technology and construction informatics

Deutsch

Modellierung von Gebäudeenergieverbrauch- und Raumklimadaten in einem latenten Merkmalsraum im Kontext Gebäudeenergieeffizienz

Project Description

1 Starting Point

1.1 State of the art and preliminary work

This project's objective is to provide a contribution to the data-driven modeling using data related to occupant behavior (OB) in buildings, building control, and building operation. The objectives are based on existing research on data-driven modeling to ensure energy efficiency in buildings. In that regard, deep learning (DL) methods for energy consumption modeling and OB modeling have demonstrated a remarkable progress in recent years. As a continuing research question related to DL modeling, model applicability to different settings is explored by applying the techniques of domain adaptation and transfer learning. Finally, novel research directions such as the generation of synthetic data and physics informed data-driven modeling are of particular importance and constitute a reasonable step towards a full understanding of the latent space and semantics of building operation and energy consumption data.

Here, we refer to latent space learning as *"learning representations of the data that make it easier to extract useful information when building classifiers or other predictors"* [Bengio2013, p.1798]. Other The representation of the data considers posterior probabilities, that capture either the variables with high explanatory power or predictive power about the modeled output [Bengio2013]. The recent progress from several research directions is relevant for acquiring novel insights into the latent space modeling and semantics behind indoor environmental quality (IEQ) and building energy data. These are based on the recent progress on DL based modeling using data related to building monitoring, latent space learning, domain adaptation and transfer learning. Additionally, the synthetic building data generation is the objective that is aimed to be comprehensively fulfilled by understanding the representation behind the abstract, high-dimensional building monitoring data.

As a basis for latent space learning, **deep learning methods** have been widely researched in the context of energy efficient buildings. DL models are neural networks with multiple hidden layers [Goodfellow2016]. Utilizing their capability to approximate any kind of physical system, these models have been extensively used for building automation applications [Qian2020; Zhang2020; Yan2020], building energy consumption [Azadeh2008; Gonzalez2005; Kim2019; Nizami1995], and OB modeling [Markovic2018; Markovic2019; Markovic2021]. [Ben2004] used regression neural networks to predict hourly cooling load for three office buildings, while [Azadeh2008; Gonzalez2005; Kim2019; Nizami1995] predicted electric energy consumption using feedforward neural networks. After achieving accuracy with a multi-layer perceptron without transformed variables for estimating energy demand, [Pino2017] concluded that the neural networks are more suitable for combined heating and cooling energy consumption modeling, when compared to statistical regression models.

In order to optimize the DL model development, the training with small data sets has raised attention of the scientific community. [Qian2020] used neural networks to predict heating, ventilation, and air conditioning (HVAC) loads by training the model using sparse data. The results showed that utilizing only one month or one week of monitoring data could lead to an accuracy decrease of 6 % and 20 %, respectively. [Zhang2020] modeled building energy consumption data by using a machine learning framework consisting of a deep belief network and extreme learning machines. The proposed method could predict the energy consumption data with up to 10 % improvement in mean absolute error (MAE), compared to the support vector regression used as a benchmark. [Yan2020] implemented Generative Adversarial Networks (GANs) to generate faulty training samples for the automatic Fault Detection and Diagnosis (FDD) of chillers with imbalanced data sets. Without the implemented model, the classification accuracy of the original FDD approach could hardly reach 90 %, implying that generative modeling might also be a promising paradigm for DL modeling in the context of building energy modeling.

[Koschwitz2018] compared the accuracy of a Nonlinear Autoregressive Exogenous Recurrent Neural Network (RNN) with Support Vector Machines in order to predict the thermal loads of non-residential buildings. The results showed that the neural network outperformed the alternative machine learning method, presumably due the NN's capability to successfully learn the feature representation. [Sendra2020] implemented Long Short Term Memory Networks (LSTM) for the prediction of the next day energy consumption of a residential building in Madrid, based on the previous day of measurements. The model with the highest accuracy could predict the target variables with a test Pearson correlation coefficient of 0,797 and a normalized root mean squared error (NRMSE) of 0,13.

Latent space learning is closely related to DL research, since feature representation learning is included into DL. The models for latent space learning for energy consumption and OB modeling have mostly been researched with a methodological focus on the autoencoder neural networks. Autoencoders turn the input data into a compressed representation by the encoder, while the decoder learns to reconstruct the original data starting from the last encoded state [Goodfellow2016]. These models have been widely applied for FDD of building automation systems [Fan2018; Liu2020; Araya2016; Legrand2018; Benitez2020], considering their acknowledged capability to detect and reconstruct anomalous data patterns. [Fan2018] explored different autoencoder architectures for the anomaly detection of building monitoring data. The use of 1D convolutional autoencoders could both capture the intrinsic characteristic of the energy data and preserve their temporal information. [Liu2020] implemented autoencoder neural networks for the anomaly detection of vertical plant wall systems. This approach outperformed other machine leaning-based methods for both contextual and point anomaly detection. [Araya2016] successfully applied autoencoders for the detection of anomaly behavior in energy consumption data. [Legrand2018] discovered that RNN autoencoders can have better generalization capabilities than other configurations, when exploited for the anomaly detection of multiple houses' sensors. Finally, [Benitez2020] proposed variational autoencoders for the FDD of a subway system, outperforming conventional approaches.

Transfer learning and **domain adaptation** refer to the situation where what has been learned in one setting is exploited to improve generalization in another setting [Goodfellow2016]. The problem statement and the distinction between the domain adaptation, transfer learning, and model generalization in case of OB and energy consumption modeling is provided by [Markovic2021]. In that context, the domain adaptation refers to the situation where the model developed for a particular task in one domain is exploited to improve the performance for modeling the same task in a different domain. Transfer learning can be defined as the improvement of learning in a new task through the transfer of knowledge from a related task that has already been learned [Torrey2010].

As pointed out by [Markovic2020], the research on domain adaptation for OB and energy consumption data in buildings is still limited [Markovic2018; Zhang2019; Arief-Ang2018]. [Arief-Ang2018] applied semi-supervised domain adaptation to apply the model for occupancy count in significantly different building settings. Their results showed that the use of domain adaptation led to a slight improvement in the occupancy count estimation. [Zhang2019] applied domain adaptation to model occupancy counting for different rooms that were located within the same building. They relied on weight scaling for the target domain by conducting additional training iterations. The results pointed out a significant improvement in the occupancy count after the application of the domain adaptation. [Markovic2018] implemented a DL-based window opening model using data from a small group of offices in a commercial building. The proposed method was then adapted to other buildings, by running further iterations on the target domain data. The results indicated that the implemented algorithm was more accurate than other building-wise calibrated models. [Zhang2019] developed an occupancy model based on RNNs on a specific data set and applied a domain adaptation technique to transfer the acquired knowledge to a building located on a different continent. The results led to almost the same performance of a model trained entirely on target domain data. In summary, the aforementioned research shows that the domain adaptation can lead to more data-efficient model training, yet, the optimal approaches to conduct domain adaptations are still not sufficiently explored.

Related to the domain adaptation, transfer learning has been primarily explored for modeling the time-series of energy consumption data and thermal comfort data [Gao2020; Gao2021; Natarajan2019]. A remarkable spike of interest in novel methods such as latent space learning and transfer learning could also be observed in the field of thermal comfort research, which is so far mostly based on analytical methods and descriptive statistics. [Gao2021] proposed transductive transfer learning using TabularGAN models to model the thermal votes over different locations. [Natarajan2019] proposed a transfer learning approach to model the thermal comfort based on measured skin temperatures. They obtained promising results on a limited data sample of five participants.

Physics informed machine learning has been widely explored in different engineering areas in which the general system principles can be described using physical models, yet the volume of data that need to be proceeded exceeds the limit of conventional data analytic methods. Among others, the physics informed machine learning has been widely explored in computational fluid dynamics (CFD) [Wang2017; Wu2018], applied CFD in the context of climate modeling [Howland2019], geoscience modeling [Karimpouli2020] as well as non-linear dynamic systems [Raissi2017] and heat transfer [Zobeiry2021].

The physics informed data-driven modeling has recently gained significant attention in the building energy modeling community. In one of the pioneering works on this topic, [Drgona2020] proposed a constrained RNN to model buildings' thermal dynamics. The encoded underlying graph structure is inspired by the physical building components. The results pointed out that the model can be generalized well to different buildings and could represent the buildings' thermal dynamics including the indoor climate conditions with high validity. Their initial results demonstrated the model's accuracy and data-efficiency, which confirmed the large potential of the proposed modeling paradigm. However, the results presented in the latter publication were obtained only from a single building and the model predictive control (MPC) for a single actuator.

Therefore, a more comprehensive research on the applicability of physics informed modeling of different building and systems is required.

In summary, the existing methods pointed out that the combination of physical white-box approaches and advanced data-driven models is a suitable approach for describing a large variety of physical systems, while the identical methods for reconstructing the systems dynamics are a meaningful, yet rarely explored next step in research on this modeling paradigm.

Model development using synthetic data has been widely used throughout different scientific disciplines. The synthetic data have proved to be particularly useful in cases where the collection of a sufficient amount of historical data is cost-intensive or infeasible due to the non-proportional technical complexity related to data collection. The synthetic data generation methods can be classified by white-box approaches and data-driven data generation.

In the field of energy efficient buildings, building performance simulation (BPS) is one of the fundamental approaches to generate synthetic data using white-box models. The use of BPS belongs to the well-researched established techniques for assessing building energy needs, optimizing the indoor thermal environment during the design stage as well as a support tool for selecting appropriate building technologies in design stage. To conclude, BPS as the main white-box technique for generating synthetic data is undoubtedly a suitable methodology for optimizing building design and energy performance assessment. Nevertheless, BPS has a limited potential for supporting the real-time building control. In order to overcome the limitations of the conventional BPS approaches, [Hong2020] and [Li2021] proposed a novel framework for generating synthetic smart building data from buildings. In their recent work, the authors generated the baseline smart meter data using BPS and applied the data post-processing by adding artificial noise, missing data and further anomalies that are observable in real monitoring data. In summary, they successfully mimicked the smart meter data that could be further used to identify the energy savings and assess the energy performance, even in case of rare events or unusual operation. However, in order to unlock the full potential of the use of synthetic data in building operation, the data need to be synthesized at high data resolution, the stochastic factors driven by occupants should be reconstructed and the fine signal dynamics need to be reconstructed. On these places, the white box data synthetization is brought to its limits and the data-driven strategies show promising success potential.

The data-driven approaches for synthetic data generation are widely researched in the fields of computer vision and natural language processing. Related to computer vision, the synthetic data obtained from virtual reality proved suitable for training object trackers for autonomous vehicles [Alhaija2018], for 3D object reconstruction [Richardson2016] as well as for the semantic segmentation [Saleh2018; Sankaranarayanan2018]. Here, modeling using synthetic data followed the conventional pipeline as in case of modeling with real-data; with an additional step of synthetizing artificial training samples in a fashion suitable for the modeling domain in question. However, the data-driven techniques are still rarely applied in the field of IEQ and energy consumption modeling, which motivates further research in this direction.

Author's preliminary work

Machine learning methods have been in the applicant's research focus for energy consumption modeling [Koschwitz2018; Markovic2021; Liguori2021], computer vision for thermal image processing [Cao2017; Metzmacher2017], OB modeling [Markovic2018, Markovic2021] as well as IEQ modeling [Liguori2021]. Throughout the past research, the problems related to domain adaptation and time-series modeling were addressed, which was crucial for the research question formulation that will be tackled in the scope of this project.

Machine learning methods for OB modeling and energy consumption have been a focus of research of the applicant. Window states modeling and plug loads modeling using neural networks was extensively investigated in the scope of the applicant's preliminary work [Markovic2018; Markovic2019; Markovic2020; Markovic2021]. [Markovic2018] proposed one of the earliest applications of DL for modeling window opening behavior in buildings. The results showed that generic DL modeling leads to significantly improved performance, when compared to established statistical modeling approaches. [Markovic2019] developed a stacked multi-layer perception for modeling the window states prediction using time-series of IEQ and weather data from the short-term past. The results pointed out that the information rich part sequence regarding the future actions is between 30 and 60 minutes of past measurements. Additionally, the model could learn in end-to-end manner the information that is explainable using expert domain knowledge, which also poses one of the rare works on explainable machine learning for OB modeling.

Availability of the existing data for conducting the planned research

The prerequisite to conduct the research on data-driven methods for energy consumption in buildings and IEQ modeling is the availability of historical monitoring data. In this particular case, the modeling will be done using data from multiple data sources, such as 1) IEQ monitoring, 2) smart meter data, 3) building-wise energy consumption, and 4) OB data. Due to the applicant's preliminary work and related past projects, some suitable data sets are available. Among others, monitoring data with high resolution collected in around 80 offices in a building in Aachen between 2014 and 2018 and monitoring data collected at KfW Ostarkade [Wagner2005] are available for the research on the proposed models.

Furthermore, advantage will be taken of open source monitoring data. This is of particular importance for obtaining agnostic and large scale verified findings regarding the goodness of different modeling paradigms. Therefore, the following data sources will be considered:

1) Smart meter data and energy consumption data from 1,448 buildings that were released as a part of ASHRAE great energy predictor data competition[1]
2) ASHRAE Global occupant data base - open source publication of OB data from 42 buildings across different locations.

Related completed and running projects of the applicant

The aim of the project "Engineering-based generic modeling of occupant behavior for energy efficient buildings"[2] was to gain novel insights regarding time-series based modeling of OB in buildings. Among the main outcomes, this project provided a significant contribution towards standardizing occupant behavior in buildings [Carlucci2020]. Furthermore, the project results contributed to establishing the machine learning methods as a conventional modeling approach for OB modeling in buildings [Markovic2020; Carlucci2020; Markovic2021; Liguori2021].

The aim of the project "ValMoNuI - Validation and Modeling of Occupants' Actions"[3] was to study OB with respect to interactions between people and the building envelope in passively cooled buildings as well as human-machine interactions in the scope of building controls. In the scope of this project, initial proof of concept for the machine learning based OB modeling was presented

[1] https://www.kaggle.com/c/ashrae-energy-prediction
[2] https://gepris.dfg.de/gepris/projekt/418297274?language=en
[3] https://www.e3d.rwth-aachen.de/go/id/jmfz

[Markovic2017; Markovic2018]. Among the project deliverables, the novel machine learning driven OB models were developed and they were prototypical proofs of applicability in building automation systems were conducted [Markovic2020; Aleksy2020].

Finally, the applicant is involved in the project "DataFEE - Data mining, machine learning, feedback, and forward energy efficiency through user centric building systems"[4]. The objective of this project is to develop a human centered building automation system by making the technological advances in machine learning modeling, digital twins for building operation, advanced human centered building control and the design of the feedforward system for the persuasive user experience (UX) modeling.

1.2 Project-related publications

1.2.1 Articles published by outlets with scientific quality assurance, book publications, and works accepted for publication but not yet published.

1) Markovic, R., Grintal, E., Wölki, D., Frisch, J., & van Treeck, C. (2018). Window opening model using deep learning methods. *Building and Environment*, *145*, 319-329.
2) Markovic, R., Azar, E., Annaqeeb, M. K., Frisch, J., & van Treeck, C. (2021). Day-ahead prediction of plug-in loads using a long short-term memory neural network. *Energy and Buildings*, *234*, 110667.
3) Markovic, R., Frisch, J., & van Treeck, C. (2019). Learning short-term past as predictor of window opening-related human behavior in commercial buildings. *Energy and Buildings*, *185*, 1-11.
4) Liguori, A., Markovic, R., Dam, T. T. H., Frisch, J., van Treeck, C., & Causone, F. (2021). Indoor environment data time-series reconstruction using autoencoder neural networks. *Building and Environment*, *191*, 107623.
5) Koschwitz, D., J. Frisch, and C. van Treeck (2018). Data-driven heating and cooling load predictions for non-residential buildings based on support vector machine regression and NARX Recurrent Neural Network: A comparative study on district scale. *Energy 165*, 134-142.
6) Liguori, A., Markovic, R., Frisch, J., Wagner, A., Causone, F., van Treeck, C. (2021) A gap-filling method for room temperature data based on autoencoder neural networks. *IBPSA Building Simulation.*
7) Liguori, A., Yang, S., Markovic, R., Dam, T., Wagner, A., Frisch, J., van Treeck, C. (2021) Prediction of HVAC loads at different spatial resolutions and buildings using deep learning models. *IBPSA Building Simulation.*
8) Carlucci, S., De Simone, M., Firth, S. K., Kjærgaard, M. B., Markovic, R., Rahaman, M. S., ... & van Treeck, C. (2020). Modeling occupant behavior in buildings. *Building and Environment*, *174*, 106768.
9) Markovic, R., Wolf, S., Cao, J., Spinnräker, E., Wölki, D., Frisch, J., & van Treeck, C. (2017). Comparison of different classification algorithms for the detection of user's interaction with windows in office buildings. *Energy Procedia,* 122, 337-342
10) Markovic, R., Grintal, E., Nouri, A., Frisch, J., & van Treeck, C. (2019). Right on Time–Exploring Suitable Time Discretization for Occupant Behavior Co-Simulation. *IBPSA Building Simulation.*

1.2.2 Other publications, both peer-reviewed and non-peer-reviewed
n. A.

1.2.3 Patents
n. A.

1.2.3.1 Pending
n. A.

1.2.3.2 Issued
n. A.

2 Objectives and work programme

The project objectives will be addressed in the scope of work packages that are organized around the following research questions related to latent space learning: How to develop methods for representation learning that are successful in data reconstruction? How to apply these methods

[4] https://www.e3d.rwth-aachen.de/cms/E3D/Forschung/Projekte/~enekp/Datefee/?lidx=1

in different target domains and furthermore for different modeling objectives? Finally, how could the latent space be fully reconstructed so that realistic, fully synthetic data can be generated? In the context of the last question, the data generation is conducted in a data-driven fashion and using physics informed machine learning. In sum, the project is organized in six work packages with nine PM each, and a detailed Gantt is presented in Table 1.

1. Data-driven reconstruction of energy related data (9 PM)
2. Domain adaptation with adversarial learning (9 PM)
3. Physics informed reconstruction of energy related data (9 PM)
4. Transfer learning paradigm for energy related data (9 PM)
5. Data-driven generation of synthetic data (9 PM)
6. Physics-based data-driven generation of synthetic data (9 PM)

	Year 1												Year 2												Year 3		
	1	2	3	4	5	6	7	8	9	10	11	12	1	2	3	4	5	6	7	8	9	10	11	12	1	2	3
WP1: Data-driven reconstruction of energy related data	2	2	2	1	1	1																					
WP2: Domain adaptation with adversarial learning										2	2	2	1	1	1												
WP3: Physics informed reconstruction of energy related data													1	1	1	2	2	2									
WP4: Transfer learning paradigm for energy related data				1	1	1	2	2	2																		
WP5: Data-driven generation of synthetic data																			2	2	2	1	1	1			
WP6: Physics-based data-driven generation of synthetic data																						1	1	1	2	2	2

Total PM: 54

Table 1: Overview of the work packages and timeline.

2.1 Anticipated total duration of the project

The planed project duration is 27 months.

2.2 Objectives

The main objective of this project is to develop novel machine learning methods for latent space learning of time-series data related to the building monitoring, IEQ, and energy consumption. As shown in the following work programme, the latent space learning might indeed accelerate the transfer of knowledge across different domains. This is an important yet rarely explored step towards generic and large-scale modeling in data-scarcity scenarios. In order to obtain novel findings towards this research objective, problems of different complexities will be sequentially tackled and a set of promising modeling paradigms will be applied. An overview of the research objectives is presented in Figure 1. The proposed research builds up on the existing research on big data modeling and DL. The latent space learning is in the research focus in order to obtain computationally efficient embeddings of significant information from building monitoring data. Furthermore, the domain adaptation and transfer learning are applied to model data from various domains. Finally, using the gained knowledge, the realistic synthetic data will be generated.

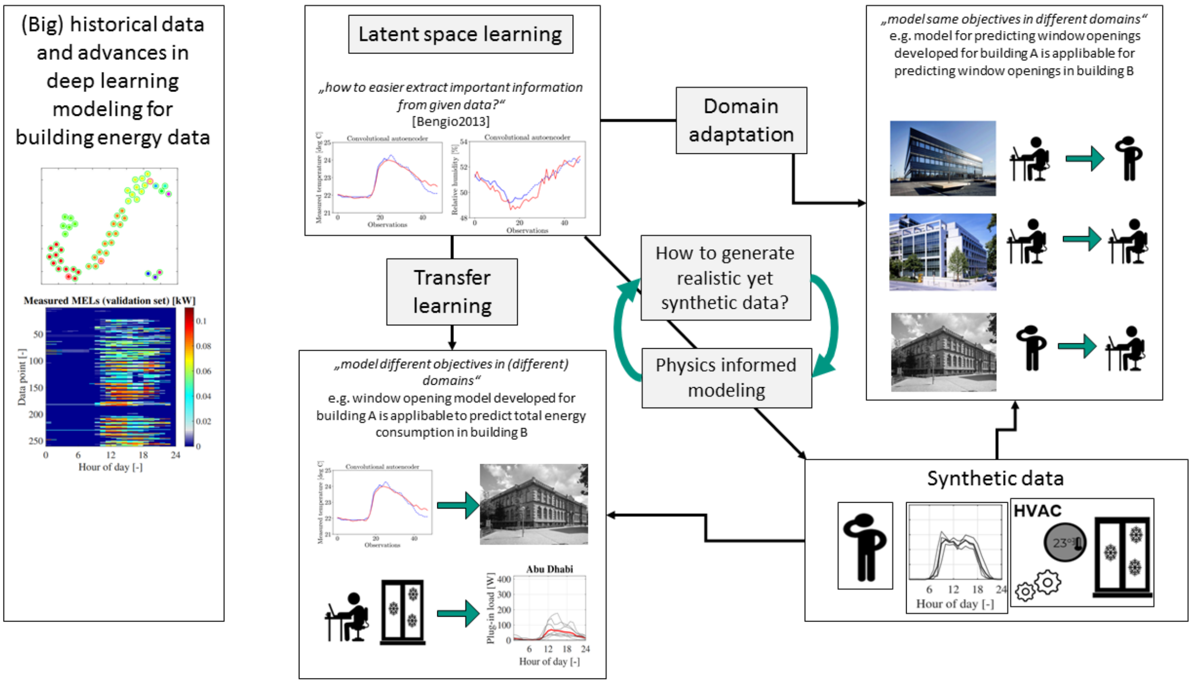

Figure 1: Research objective and overview of the proposed research activities.

Firstly, the project aims to propose latent space learning methods for reconstructing the missing part(s) of time-series from agnostic data sources. This problem was in part addressed successfully in the scope of preliminary work, while the goal of this particular activity is to propose a methodology that is applicable for widely versatile time-series obtained from various target domains. Up to date, there are only few empirical results and limited theoretical discussions on the factors that drive the machine learning performance for reconstructing data obtained from different building domains. Secondly, an adversarial method for the domain adaptation in building domains will be explored. The underlying motivation is the empirically proven suitability of the latter methods for comparable problems in natural language processing (NLP) and computer vision. Thirdly, the common knowledge between domains and learning different domains will be defined and a computationally efficient method for learning each domain's particularities will be explored. Here, the methodological focus is put on transfer learning methods.

Finally, when the efficient reconstruction of general knowledge has been achieved and the methodologies for domain specific information processing have been developed, it will be aimed to apply the knowledge to produce fully synthetic reliable data for different target domains. Conclusively, the synthetic data modeling is aimed to be improved by incorporating the physics-based modeling as a useful complement to data-driven methods.

2.3 Work programme including proposed research methods

Work package WP-A: Data-driven reconstruction of energy related data

Building data sets are often characterized by missing values that must be handled before further processing and analysis. On the one hand, the presence of unprocessed large and frequent data gaps reduces the size of the available data set, hindering further energy and thermal comfort analysis. On the other hand, using data-driven models in the post-processing phase, the use of simplified imputation methods may lead to inaccurate and biased evaluations.

In order to solve the problem of missing data in building automation, accurate and resource-efficient methods to impute the data gaps have to be implemented. This work package aims to achieve this objective by relying on generative models, such as autoencoder neural networks and GANs. The suitability of these models for reconstructing different levels of building monitoring data gaps will be tested separately and for a variety of temporal and spatial discretizations. Here, the validation of the proposed methods will be performed by applying random noise to the datasets and by quantifying the error between the reconstructed gaps and the real data (see [Liguori2021] for additional information). For that purpose, the used datasets will be single energy data streams collected in commercial buildings from previous projects (e.g. DFG - TR 892/4-1). Additional datasets might be shared with some of the researchers that agreed to cooperate on this project. Furthermore, in order to facilitate the inclusion in the building automation schemes, special attention will be given to the optimization of computational and memory requirements of the proposed methods.

The research steps may be summarized as follows:
- Development of autoencoder neural networks for the reconstruction of missing energy data time-series
- Development of GANs for the reconstruction of missing energy data time-series
- Model optimization and assessment of model limitations with respect to different levels of data gaps (e.g. hourly, weekly, monthly), varied temporal resolution and spatial resolution.
- Comparison to commonly adopted and newly proposed imputation approaches (e.g. copy-paste imputation)

Work package WP-B: Domain adaptation with adversarial learning

Data-driven methods, in particular DL approaches, usually require an extensive amount of training data and computational resources to be deployed effectively. This might hinder their practical integration into resource-constrained and real-time applications, such as MPC. In these scenarios, the reconstruction of monitoring data with time-consuming and memory-intensive DL models might be unfeasible and counterproductive.

In this regard, the dependence of DL models on large data sets might be reduced by exploiting the prior knowledge encoded in domains where data are easier or cheaper to collect. In order to accomplish this goal, the methodological focus of this package will be on domain adaptation with adversarial autoencoder neural networks. Here, two autoencoders will be jointly trained to reconstruct the time-series data from the respective novice (small training data set) and expert (large training data set) domains. In particular, the novice and expert domains will be single energy data streams with a high correlation coefficient, based on the results from work package WP-D. Meanwhile, an adversarial network [Goodfellow2014] will be trained to discriminate between the two latent space representations of the previous models. The adversarial network will therefore act as a regularization agent that matches the latent space representation of the novice domain to the expert domain. The accuracy and computational performance will be benchmarked to the models developed in the previous subtask. In particular, model performance will be evaluated and compared by re-training from scratch all the models on the same limited datasets.

The research steps may be summarized as follows:
- Selection of the novice and expert datasets, based on the results from work package WP-D
- Development of adversarial autoencoder neural networks for the reconstruction of missing energy data time-series through domain adaptation

- Model optimization and assessment of model limitations with respect to different levels of data gaps (e.g. missing hours, weeks, or months), varied temporal resolution and spatial resolution.
- Quantification of model's scalability and model development runtime analysis when compared to the autoencoder neural networks and GANs developed in the previous subtask and retrained on few training samples

Work package WP-C: Physics informed reconstruction of energy related data

As discussed in the previous sections, pure black-box models, such as standard neural networks, tend to rely heavily on the quality and size of the training data. This is generally due to the lack of initial knowledge regarding the underlying system. In case of physical systems, there is a variety of physical laws that pure black box models ignore at the beginning of the learning process. Encoding such knowledge from scratch results in an increasing amount of information needed by the algorithm.

The aim of this package is to explore the possibility to use DL and physics-based modeling together, in order to handle the problem of missing data in building automation. For that purpose, different solutions should be researched. On the one hand, physics-informed neural networks (PINN) as proposed in [Raissi2017] exploit the recent progresses in automatic differentiation [Baydin2018] to solve the nonlinear partial differential equations that describe any generic physical system. Therefore, they represent a promising solution for the reconstruction of physical data such as environmental and energy time-series. On the other hand, the physical laws that describe the same environmental and energy data might be also used as a regularization agent for the neural network. This would keep the latent space representation of the model within meaningful boundaries, similarly to the adversarial training of sub-task B. Based on the previous consideration, in the first part of this work package, the state-space representation of a generic building's thermal dynamics will be implemented. This will consist of a system of partially observable partial differential equations (PDEs) based on the energy and mass conservation equations for the room air and walls of a multi-zone building [Yao2013]. In the next step, the loss function module of the existing autoencoder neural network, already implemented in the scope of WP-A, will be combined with a physics-based loss function. Therefore, the loss function of the newly proposed autoencoder-based PINN model will consist of one regression term and one physics-based term. The first component will define the mismatch between the reconstructed gaps and the real data; and it will therefore consider the stochastic influences such as occupant behavior. The last component will guide the network's outputs by means of the previously implemented state-space representation. As for WP-A, the validation of the proposed methods will be performed by applying random noise to the datasets and by quantifying the error between the reconstructed gaps and the real data (see [Liguori2021] for additional information). For that purpose, the used datasets will be single energy data streams collected in commercial buildings from previous projects (e.g. DFG - TR 892/4-1).

The research steps may be summarized as follows:
- Implementation of the state-space representation of a multi-zone building
- Inclusion of the physics-based loss function in the existing generative models
- Models training and validation using limited data samples
- Testing of autoencoder-based PINNs for the reconstruction of missing energy data for building automation

Work package WP-D: Transfer learning paradigm for energy related data

As discussed in section 2, there is a large number of target functions to be addressed in building energy consumption modeling and indoor environment modeling. In summary, each target point results commonly in a different target domain, and these domains are scaling up to a large number of buildings. As a result, either a distinct model should be developed for each target domain in question, or a general law that is valid for different domains should be identified and the pre-learned information should be shared over different domains. In this regard, the transfer learning may be a suitable task to solve the latter problem.

In order to address this problem, this work package focuses on the optimal problem formulation for the transfer learning using various domain sources, by following closely the approach proposed in [Wang2011]. In particular, in order to align this study with work package WP-B, the domain sources will be single energy data streams collected in commercial buildings from previous projects (e.g. DFG - TR 892/4-1). Additional datasets might be acquired from researchers that agreed to cooperate on this project.
The aim of this work package is to explore a modeling paradigm that can lead to developing data-efficient models for different target domains. Firstly, the similarity between different domains will be explored by learning a semantic graphical representation. Finally, the correlation between the various domains will be quantified by computing the similarities between the observed data patterns. The knowledge derived from this study will be used to select the suitable domains to be used in work package WP-B. Only the domains with high correlation factors will be selected.

The research steps may be summarized as follows:

- Approximation and alignment of the time series data streams from various domains with the hidden states of the source systems by data segmentation
- Clustering of the approximated time-series data streams for each of the analyzed domains
- Implementation of different Hidden Markov Models (HMMs) and learning of their parameters on each of the clustered time-series data streams
- Computation of the pattern correlations between the previously implemented HMMs
- Selection of the domains with higher correlation coefficients

Work package WP-E: Data-driven generation of synthetic data

As presented in section 1, the availability of large-scale historical data from buildings is a commonly observed bottleneck for the creation of novel machine learning models that are incorporated in building control or energy systems. Additionally, the processing of high granularity data is connected with data privacy risks which is a particular problem in case of detailed monitoring data from room-based controls. Nonetheless, sufficient and high-quality data is necessary for machine learning model creation.

In order to circumvent the significant data security threads and to minimize the data collection costs, we propose the creation of fully synthetic data that will be used for model training and validation. The synthetic data will be the product of “cloning” the existing indoor air temperature data, collected in an office building in Aachen, Germany [Fütterer2014]. Methodologically, the focus will be put on methods for latent space learning, such as GANs [Goodfellow2014] and autoencoder neural networks.
The overview of the required research steps may be summarized as follows:

- Development of GANs neural networks for the generation of the monitoring data
- Investigation of the suitability of the autoencoder neural networks for synthetic data generation
- Research on the optimal model formulation

- Model evaluation of the above listed modeling paradigms against the real monitoring data. Here, the difference between the synthetic and monitored test data will be quantified in terms of Root Mean Squared Error (RMSE) and coefficient of determination (R2)

A particular importance will be given to the definition of the suitable annotation protocol. The novelty of the synthetic data-driven data generation will be supported by the established research on analytical white-box data generation models, such as the BPS in the scope of WP-F.

Work package WP-F: Physics-based data-driven generation of synthetic data

The observed variables in form of monitoring data in buildings are result of complex, high-dimensional influences. This high-dimensional underlying system is impacted by the building's physics and the stochastic influences, such as OB. In order to realistically represent such a system, around 1,000 dimensions should be considered in case of a single zone building thermal model, while the whole building model including the HVAC system may result in up to 20,000 dimensions. Based on the state of the art, this physical system is commonly modeled using simplified RC models or some advanced solvers using BPS. Both of these approaches fulfill the purpose to represent the density distribution and the low-resolution trends in signals. However, even the well calibrated high-resolution BPS cannot reproduce the signal dynamics of data typically collected in the building automation. In order to incorporate the knowledge regarding the general system's dynamics and to overcome the limitations of the analytical models in reconstructing the high-resolution system's dynamics, we propose the physics-based DL driven data generation.

The basis for the data generation are the simulation results obtained from a BPS model, that will be further pre-processed using latent space learning models. In the first part of this work package, the BPS model will be implemented based on a two-person office located in Aachen, Germany, as presented in [Cali2015]. Here, the BPS model will be coupled with an occupancy simulator as presented in [Weber2020]. The aim will be to generate several synthetic days of indoor air temperature measurements based on stochastic occupant schedules. In particular, simulation will be performed for a group of similar rooms with different characteristics (e.g. different volumes), in order to obtain a general dynamic of the target variable. Therefore, no calibration with real data will be performed. In the second part, the models implemented in the scope of WP-E will be first pre-trained on the general simulated dataset and then adapted on the real data collected in one of the rooms of the same building. The goal is to generate a synthetic augmented copy of the target dataset, by reducing the collection of real training samples.
For that purpose, research steps will be structured as following:

- Implementation of the BPS model based on a two-person office located in Aachen, Germany
- Implementation of an occupancy simulator, based on [Chen2018]
- Simulation of the general indoor air temperature dataset
- Domain adaptation modeling using building simulation results and real collected data
- Quantification of model's scalability and model development runtime analysis when compared to the autoencoder neural networks and GANs developed in the previous subtask and retrained on few training samples

3 Bibliography concerning the state of the art, the research objectives, and the work programme

[Aleksy2020] Aleksy, M., & Bauer, P. (2020, April). Ensuring the Practical Applicability of Algorithms for User Behavior Modeling Through Integration into Building Automation Systems. In *International Conference on Advanced Information Networking and Applications* (pp. 715-721). Springer, Cham.

[Alhaija2018] Alhaija, H. A., Mustikovela, S. K., Mescheder, L., Geiger, A., & Rother, C. (2018). Augmented reality meets computer vision: Efficient data generation for urban driving scenes. *International Journal of Computer Vision*, *126*(9), 961-972.

[Araya2016] Araya, D. B., Grolinger, K., ElYamany, H. F., Capretz, M. A., & Bitsuamlak, G. (2016, July). Collective contextual anomaly detection framework for smart buildings. In *2016 IEEE international joint conference on neural networks (IJCNN),* pp. 511-518.

[Arief-Ang2018] Arief-Ang, I. B., Salim, F. D., & Hamilton, M. (2017, November). DA-HOC: Semi-supervised domain adaptation for room occupancy prediction using CO2 sensor data. In *Proceedings of the 4th ACM International Conference on Systems for Energy-Efficient Built Environments,* pp. 1-10.

[Azadeh2008] Azadeh, A., S. Ghaderi, and S. Sohrabkhani (2008). Annual electricity consumption forecasting by neural network in high energy consuming industrial sectors. *Energy Conversion and Management 49*(8), 2272-2278.

[Baydin2018] Baydin, A. G., Pearlmutter, B. A., Radul, A. A., & Siskind, J. M. (2018). Automatic differentiation in machine learning: A survey. *Journal of Marchine Learning Research*, *18*, 1-43.

[Ben2004] Ben-Nakhi, A. E. and M. A. Mahmoud (2004). Cooling load prediction for buildings using general regression neural networks. *Energy Conversion and Management 45*(13-14), 2127-2141.

[Benitez2020] Loy-Benitez, J., Li, Q., Nam, K., & Yoo, C. (2020). Sustainable subway indoor air quality monitoring and fault-tolerant ventilation control using a sparse autoencoder-driven sensor self-validation. *Sustainable Cities and Society*, *52*, 101847.

[Bengio2013] Bengio, Y., Courville, A., & Vincent, P. (2013). Representation learning: A review and new perspectives. *IEEE transactions on pattern analysis and machine intelligence*, *35*(8), 1798-1828.

[Cali2015] Calì, D., Matthes, P., Huchtemann, K., Streblow, R., & Müller, D. (2015). CO2 based occupancy detection algorithm: Experimental analysis and validation for office and residential buildings. *Building and Environment*, *86*, 39-49.

[Cao2017] Cao, J., Metzmacher, H., O'Donnell, J., Frisch, J., Bazjanac, V., Kobbelt, L., & van Treeck, C. (2017). Facade geometry generation from low-resolution aerial photographs for building energy modeling. *Building and Environment*, *123*, 601-624.

[Chen2018] Chen, Y., Hong, T., & Luo, X. (2018, February). An agent-based stochastic Occupancy Simulator. In *Building Simulation, 11*(1), pp. 37-49).

[Drgona2020] Drgona, J., Tuor, A. R., Chandan, V., & Vrabie, D. L. (2020). *Physics-constrained Deep Recurrent Neural Models of Building Thermal Dynamics* (No. PNNL-SA-156966). Pacific Northwest National Lab (PNNL), Richland, WA. NeurIPS.

[Fan2018] Fan, C., Xiao, F., Zhao, Y., & Wang, J. (2018). Analytical investigation of autoencoder-based methods for unsupervised anomaly detection in building energy data. *Applied energy*, *211*, 1123-1135.

[Fütterer2014] Fütterer, J. & A. Constantin (2014). Energy concept for the E. ON ERC main building. *E.ON Energy Research Center Series 4*(9). https://publications.rwth-aachen.de/record/443118

[Gao2021] Gao, N., Shao, W., Rahaman, M. S., Zhai, J., David, K., & Salim, F. D. (2021). Transfer learning for thermal comfort prediction in multiple cities. *Building and Environment*, *195*, 107725.

[Gao2020] Gao, N., Xue, H., Shao, W., Zhao, S., Qin, K. K., Prabowo, A., ... & Salim, F. D. (2020). Generative adversarial networks for spatio-temporal data: A survey. *arXiv preprint arXiv:2008.08903*.

[Gonzalez2005] Gonzalez, P. A. and J. M. Zamarreno (2005). Prediction of hourly energy consumption in buildings based on a feedback articial neural network. *Energy and buildings 37 (6),* 595-601.

[Goodfellow2014] Goodfellow, I., Pouget-Abadie, J., Mirza, M., Xu, B., Warde-Farley, D., Ozair, S., ... & Bengio, Y. (2014). Generative adversarial nets. *Advances in Neural Information Processing Systems*, *27*.

[Goodfellow2016] Goodfellow, I., Bengio, Y., & Courville, A. (2016). Deep Learning MIT Press (2016).

[Hong2020] Hong, T., Macumber, D., Li, H., Fleming, K., & Wang, Z. (2020, December). Generation and representation of synthetic smart meter data. In *Building Simulation* (Vol. 13, pp. 1205-1220). Tsinghua University Press.

[Howland2019] Howland, M. F., & Dabiri, J. O. (2019). Wind farm modeling with interpretable physics-informed machine learning. *Energies*, *12*(14), 2716.

[Karimpouli2020] Karimpouli, S., & Tahmasebi, P. (2020). Physics informed machine learning: Seismic wave equation. *Geoscience Frontiers*, *11*(6), 1993-2001.

[Kim2019] Kim, J.-Y. and S.-B. Cho (2019). Electric energy consumption prediction by deep learning with state explainable autoencoder. *Energies 12*(4), 739.

[Koschwitz2018] Koschwitz, D., J. Frisch, and C. Van Treeck (2018). Data-driven heating and cooling load predictions for non-residential buildings based on support vector machine regression and narx recurrent neural network: A comparative study on district scale. Energy 165, 134-142.

[Legrand2018] Legrand, A., Niepceron, B., Cournier, A., & Trannois, H. (2018, December). Study of autoencoder neural networks for anomaly detection in connected buildings. In *2018 IEEE Global Conference on Internet of Things (GCIoT)* (pp. 1-5). IEEE.

[Li2021] Li, H., Wang, Z., & Hong, T. (2021). A synthetic building operation dataset. *Scientific Data*, *8*(1), 1-13.

[Liguori2021] Liguori, A., Markovic, R., Dam, T. T. H., Frisch, J., van Treeck, C., & Causone, F. (2021). Indoor environment data time-series reconstruction using autoencoder neural networks. *Building and Environment*, *191*, 107623.

[Liu2020] Liu, Y., Pang, Z., Karlsson, M., & Gong, S. (2020). Anomaly detection based on machine learning in IoT-based vertical plant wall for indoor climate control. *Building and Environment*, *183*, 107212.

[Markovic2021] Markovic, R., Azar, E., Annaqeeb, M. K., Frisch, J., & van Treeck, C. (2021). Day-ahead prediction of plug-in loads using a long short-term memory neural network. *Energy and Buildings*, *234*, 110667.

[Markovic2020] Markovic, R. (2020). *Generic occupant behavior modeling for* commercial buildings (Doctoral dissertation, Doctoral Thesis. RWTH Aachen University).

[Markovic2019] Markovic, R., E. Grintal, A. Nouri, J. Frisch, and C. van Treeck (2019, Sep). Right on Time: Exploring Suitable Time Discretization for Occupant Behavior Co-Simulation. In *Proceedings of Building Simulation 2019, 16th Conference of IBPSA*: Rome, Italy, September 2019.

[Markovic2018] Markovic, R., Grintal, E., Wölki, D., Frisch, J., & van Treeck, C. (2018). Window opening model using deep learning methods. *Building and Environment*, *145*, 319-329.

[Metzmacher2017] Metzmacher, H., Wölki, D., Schmidt, C., Frisch, J., & van Treeck, C. (2017, August). Real-time assessment of human thermal comfort using image recognition in conjunction with a detailed numerical human model. *In 15th International Building Simulation Conference* (pp. 691-700).

[Natarajan2019] Natarajan, A., & Laftchiev, E. (2019). A Transfer Active Learning Framework to Predict Thermal Comfort. *International Journal of Prognostics and Health Management*, *10*, 13.

[Nizami1995] Nizami, S. J. and A. Z. Al-Garni (1995). Forecasting electric energy consumption using neural networks. *Energy Policy 23*(12), 1097-1104.

[Pino2017] Pino-Mejas, R., A. Perez-Fargallo, C. Rubio-Bellido, and J. A. Pulido-Arcas (2017). Comparison of linear regression and articial neural networks models to predict heating and cooling energy demand, energy consumption and CO2 emissions. *Energy 118*, 24-36.

[Raissi2017] Raissi, M., Perdikaris, P., & Karniadakis, G. E. (2017). Physics informed deep learning (part i): Data-driven solutions of nonlinear partial differential equations. *arXiv preprint arXiv:1711.10561*.

[Richardson2016] Richardson, E., Sela, M., & Kimmel, R. (2016, October). 3D face reconstruction by learning from synthetic data. In *2016 fourth international conference on 3D vision (3DV)* (pp. 460-469). IEEE.

[Saleh2018] Saleh, F. S., Aliakbarian, M. S., Salzmann, M., Petersson, L., & Alvarez, J. M. (2018). Effective use of synthetic data for urban scene semantic segmentation. In *Proceedings of the European Conference on Computer Vision (ECCV)* (pp. 84-100).

[Sankaranarayanan2018] Sankaranarayanan, S., Balaji, Y., Jain, A., Lim, S. N., & Chellappa, R. (2018). Learning from synthetic data: Addressing domain shift for semantic segmentation. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition* (pp. 3752-3761).

[Sendra2020] Sendra-Arranz, R. and Gutiérre, A. (2020). A long short-term memory artificial neural network to predict daily HVAC consumption in buildings. *Energy and Buildings 216*, 109952.

[Torrey2010] L. Torrey, J. Shavlik, Transfer learning, in *Handbook of research on machine learning applications and trends: algorithms, methods, and techniques*, 2010, pp. 242–264. IGI global

[Wagner2005] Wagner, A., Kleber, M., & Rohlffs, K. (2005). A bank saves money: The KfW Ostarkade building project at Frankfurt/Main; Eine Bank spart-Ergebnisse und Erfahrungen aus dem SolarBau-Projekt KfW-Ostarkade in Frankfurt.

[Wang2011] Wang, P., Wang, H., & Wang, W. (2011, June). Finding semantics in time series. In *Proceedings of the 2011 ACM SIGMOD International Conference on Management of data* (pp. 385-396).

[Wang2017] Wang, J. X., Wu, J. L., & Xiao, H. (2017). Physics-informed machine learning approach for reconstructing Reynolds stress modeling discrepancies based on DNS data. *Physical Review Fluids*, *2*(3), 034603.

[Weber2020] Weber, M., Doblander, C., & Mandl, P. (2020, November). Detecting building occupancy with synthetic environmental data. In *Proceedings of the 7th ACM International Conference on Systems for Energy-Efficient Buildings, Cities, and Transportation* (pp. 324-325).

[Wu2018] Wu, J. L., Xiao, H., & Paterson, E. (2018). Physics-informed machine learning approach for augmenting turbulence models: A comprehensive framework. *Physical Review Fluids*, *3*(7), 074602.

[Yao2013] Yao, Y., Yang, K., Huang, M., & Wang, L. (2013). A state-space model for dynamic response of indoor air temperature and humidity. *Building and Environment, 64*, 26-37.

[Yan2020] Yan, K., Huang, J., Shen, W., & Ji, Z. (2020). Unsupervised learning for fault detection and diagnosis of air handling units. *Energy and Buildings*, *210*, 109689.

[Zhang2020] Zhang, G., Tian, C., Li, C., Zhang, J. J., & Zuo, W. (2020). Accurate forecasting of building energy consumption via a novel ensembled deep learning method considering the cyclic feature. *Energy*, *201*, 117531.

[Zhang2019] Zhang, T., & Ardakanian, O. (2019, April). A domain adaptation technique for fine-grained occupancy estimation in commercial buildings. In *Proceedings of the International Conference on Internet of Things Design and Implementation* (pp. 148-159).

[Zobeiry2021] Zobeiry, N., & Humfeld, K. D. (2021). A physics-informed machine learning approach for solving heat transfer equation in advanced manufacturing and engineering applications. *Engineering Applications of Artificial Intelligence*, *101*, 104232.

4 Relevance of sex, gender and/or diversity

The fundamental objective of the proposed methodologies is to develop methods for processing the IEQ and energy consumption data from agnostic buildings. In that regard, the models will be applicable to data from diverse and heterogenic populations. This will be ensured by conducting the modeling using the open source data obtained from different locations.

5 Supplementary information on the research context

5.1 Ethical and/or legal aspects of the project

5.1.1 General ethical aspects

There are no ethical concerns with regard to the planned research activities. In the scope of the project, no personal data in sense of the general data protection regulation (GDPR) will be processed. The (virtual) experiments will be conducted using the publicly available anonymized historical building monitoring data. No experiments with participants are planned in the scope of the project.

5.1.2 Descriptions of proposed investigations involving experiments on humans or human materials

Not applicable.

5.1.3 Descriptions of proposed investigations involving experiments on animals

Not applicable.

5.1.4 Descriptions of projects involving genetic resources (or associated traditional knowledge) from a foreign country

Not applicable.

5.1.5 Descriptions of investigations involving dual use research of concern, foreign trade regulations

Not applicable.

5.2 Data handling

The research data, developed models, and software will be hosted within the institute's server infrastructure. The input historical data will consist primarily of publicly available data sets. The developed models will be hosted in git repositories. Scientifically relevant final models will be released as open source software via git repositories. Furthermore, no personal data will be collected or processed as a part of research data during the planned research activities.

5.3 Other information

Not applicable.

6 People/collaborations/funding

6.1 Employment status information

Since 2012, Dr.-Ing. habil. Christoph van Treeck is a full University Professor (W3) employed for lifetime. He is head of the Institute of Energy Efficiency and Sustainable Building E3D at RWTH Aachen University.

6.2 First-time proposal data

Not applicable.

6.3 Composition of the project group

The Institute of Energy Efficiency and Sustainable Building E3D employs an interdisciplinary team that consists of 56 scientific- and non-scientific staff (as of 15.03.2022). Beside the head of the institute, there are currently 23 scientific research assistants. In addition, there are eight technical and administrative employees and around 25 student assistants. All team members that will be involved in the proposed project have perennial collaboration and publication-vita with the project proposer. Such a well-established and experienced team is crucial for carrying out the project. In particular, for the proposed project, the following team members have a vital advisory role:

1. Dr.-Ing. Jérôme Frisch (Akademischer Oberrat, Funding: State of NRW), field of research: Computational engineering, parallel computing, numerical modelling
2. Antonio Liguori, M.Sc., Research Assistant
3. Christoph Emunds, M.Sc., Research Assistant (M.Sc. in Computer Science)
4. Dr. phil. Marc Syndicus (Group leader, Funding: State of NRW), field of research: Thermal comfort, occupant behavior, thermal ergonomics

6.4 Researchers in Germany with whom you have agreed to cooperate on this project

Not applicable.

6.5 Researchers abroad with whom you have agreed to cooperate on this project

During the course of previous projects (see Section 1.1 “Related completed and running projects of the applicant”), the Institute of Energy Efficiency and Sustainable Building E3D has established several scientific collaborations with researchers from abroad. These collaborations have been beneficial in terms of data sharing, research visits and joint publications (see Section 1.2.1). Therefore, the scope of the collaboration with external researchers is to promote joint studies that might be beneficial for the research objectives described in the previous work packages.

In particular, the Institute has scientific collaborations with the following researchers from abroad: Prof. Bing Dong, Syracuse University, USA; Dr. Tianzhen Hong, LBNL, USA; Prof. Elie Azar, Khalifa University, United Arab Emirates; Jan Drgona, PhD, PNNL, USA; Salvatore Carlucci, The Cyprus Institute, Cyprus; Prof. Francesco Causone, Politecnico di Milano, Italy; Prof. Clayton Miller, National University of Singapore, Singapore.

6.6 Researchers with whom you have collaborated scientifically within the past three years

Prof. Bing Dong, Syracuse University, USA; Bjarne Olesen, Danish Technical University, Denmark; Christoph Nytsch-Geusen, UDK Berlin; Dirk Müller, RWTH Aachen; Prof. Elie Azar, Khalifa University, United Arab Emirates; Erhard Mayer, Fraunhofer-Institute for Building Physics; Francesco Causone, Politecnico di Milano; Italy; Gunnar Grün, University of Stuttgart and Fraunhofer Institute for Building Physics; Jan Drgona, PhD, PNNL, USA; Karsten Voss, Wuppertal University; Liam O‘Brien, Caleton University; Prof. Marcel Schweiker, RWTH Aachen; Prof. Markus Reischl, Karlsruhe Institute of Technology; Rajan Rawal, CEPT University India; Prof. Ralf Mikut, Karlsruhe Institute of Technology; Runa Hellwig, Aalborg University; Salvatore Carlucci, The Cyprus Institute; Dr. Tianzhen Hong, LBNL, USA; Vladimir Bazjanac, Civil and Environmental Engineering, Stanford University.

6.7 Project-relevant cooperation with commercial enterprises

Not applicable.

6.8 Project-relevant participation in commercial enterprises

Not applicable.

6.9 Scientific equipment

The institute is equipped with a PC with a GPU unit (Nvidia GeForce GTX 1080) suitable for conducting parts of the proposed computations. For the second researcher, equipment is applied for in section 7.1.2.1. Furthermore, the university's computational infrastructure (RWTH Cluster) for running computation-intensive experiments can be accessed.

6.10 Other submissions

The funding application for this project has not been submitted to any other funding agency. In case on any change of status, including the funding proposal with third parties, the German Research Foundation (DFG) will be immediately informed.

7 Requested modules/funds

7.1 Basic Module

7.1.1 Funding for Staff

1) Two PhD candidates with a background in Engineering, Information Science or related technological fields of study will work on the project. In order to achieve the goals of the work packages, these persons are required for the full project duration (27 months; € 5.975 per month for two candidates = total € 322.650,00).
2) Two Student Assistants with Bachelor Degree (10 hrs./week = € 668,07 per month for two student assistants = total € 36.075,78). The student assistants will support the research work by performing minor, clearly defines tasks including (partial) code development, organizational support of model training and cluster computing, or literature research.

7.1.2 Direct Project Costs

7.1.2.1 Equipment up to € 10,000, Software and Consumables

The requested funds cover the costs of a new dedicated development server of € 5.000 necessary for performing the computations described in Section 2 as well as the pre-cluster model-testing.

7.1.2.2 Travel Expenses

The requested funds include the attendances at national and international conferences (one international conference per year) as well as participation in relevant continuing education measures such as workshops and summer schools. A total of € 7.500 will be requested for the above-mentioned scientists over the funded period. Potential additional costs will be covered by the institute's budget.

7.1.2.3 Visiting Researchers (excluding Mercator Fellows)

Visiting researchers are currently not involved. In case traveling and quarantine regulations due to Covid are further derestricted, funding opportunities will be sought.

7.1.2.4 Expenses for Laboratory Animals

Not applicable.

7.1.2.5 Other Costs

Not applicable.

7.1.2.6 Project-related Publication Expenses

The requested funds include a gold open access publication in the journal „Energy and Buildings“, which is one of the most important academic journals in the related field. A total of € 4.100 (taxes included) are requested for one open access publication in the aforementioned journal.

7.1.3 Instrumentation

7.1.3.1 Equipment exceeding € 10,000

Not applicable.

7.1.3.2 Major Instrumentation exceeding € 50,000

Not applicable.

7.2 Module Temporary Position for Principal Investigator

Not applicable.

7.3 Module Replacements

Not applicable.

7.4 Module Temporary Substitute for Clinicians

Not applicable.

7.5 Module Mercator Fellows

Not applicable.

7.6 Module Project-Specific Workshops

Not applicable.

7.7 Module Public Relations

Not applicable.

7.8 Module Standard Allowance for Gender Equality Measures

Not applicable.