From plans to programs: A holistic toolchain for building data applications

Gerrit Bode, Florian Stinner, Marc Baranski, Erik Brümmendorf, Xiaoye Cai, Alexander Kümpel, Markus Schraven, Thomas Schreiber, Phillip Stoffel, Thomas Storek and Dirk Müller

RWTH Aachen University, E.ON Energy Research Center, Institute for Energy Efficient Buildings and Indoor Climate, Mathienesstraße 10, 52072 Aachen, Germany
E-mail: gbode@eonrc.rwth-aachen.de

Abstract. The rise of extensive monitoring systems and the availability of low-cost sensors as well as affordable computing power has led to the development of various big data and simulation model applications in the building sector. Nevertheless, many of these promising approaches face a common hindrance for the widespread application. In case of the big data applications, training data is often limited. Much the same, simulation models often lack required input data and require extensive work for calibration. Standard practices are often preferred to innovative approaches because construction and commissioning businesses are highly cost-sensitive. Therefore, we identified the need for a holistic approach for the combined use of machine learning and simulation techniques. In this paper, we present a toolchain to generate models and data needed for the application of innovative building automation and control tools. Using the data available during the construction process, machine-learning algorithms are employed to determine the type and location of data points in devices. From the relations of data points, the system architecture is derived and simulation models are generated automatically. Using these models, the data needed for the training of big data machine-learning algorithms can be generated. We describe the toolchain, already existing components and discuss the possible future implementation.

1. Introduction
In the past decades, buildings have been under investigation for their substantial share in worldwide energy consumption. Recent developments like digitalisation, new building codes, energy efficiency laws and the decrease of sensor prices have led to an increase in the data collected in buildings [1]. Lower costs for storage and computation power have incited the use of this data for artificial intelligence and simulation model development. Additional information can be imported from sources like building information modelling (BIM), computer-aided design (CAD) and other construction plans [2]. Applications for the developed models and algorithms include demand-side management [3], model-predictive control [4], fault detection and diagnosis [5] and control generation [6]. The data is also used in facility management and applications including predictive maintenance.

Nevertheless, while delivering promising results in the research applications, most of these techniques have not found wide-spread use in the building sector of today. For existing buildings, the work to implement new control and monitoring approaches is often far more costly than the possible energy savings. BIM, for example, is still only used in a minority of new building
In this paper, we propose the process depicted in Fig. 1: The raw meta data of the system is analyzed using artificial intelligence and natural language processing. The existing data descriptors are mapped to a standardized semantic. With this data set, the relationship between the data points can be taken into account to create and calibrate simulation models of the building systems. With these models, several promising techniques of building automation can be implemented automatically.

2. Automated analysis and model generation

To appropriately control or assess a system, a system description must be available. Additionally, if the control or assessment should be created automatically, the system description must be in a format that is understandable by computer algorithms. Hence, the first step is to translate the existing description of the building into a unified semantic that can afterwards be used by any algorithm. In the following, we present the translation and the subsequent creation of the simulation models from the semantic model.

2.1. Unified building semantics

Although often neglected in the scope of research activities, structuring of data is a crucial step in order to provide information in a machine-readable way and make use of the advantages of modern machine learning algorithms. Descriptions of the operation are often only available as prose and their interpretation is only possible through expert knowledge. There are many approaches to the description of buildings or building energy systems, and a comparison is often only possible for certain use cases. In particular, the description of sensor data is a big challenge [9]. Building Information Modelling (IFC 4) aims to prepare the data of buildings for the planning and construction process. However, this information is different from the information needed for advanced control, fault detection and analysis methods. Therefore, we developed a
method to describe data points comprehensively.

From the investigation of 40 schemas for describing data points in a building, we developed the BUDO Schema. It has the ability to map all categories, e.g. device type or location, of the schemas examined. This enables users to properly describe all data points in their components with little effort. A comprehensive description of the schema is available in [10].

There are various other approaches for normalization of sensors in air handling units [11]. However, these require time-series data, and it is difficult to generate simulation models from this normalized data, as it was not designed for this purpose.

To mitigate this, we presented AIKIDO, an algorithm that converts existing data point names into the BUDO Schema with little manual effort [12]. Only the data point names are used here. For this purpose, we integrated a graphical user interface, in which users can upload the data point names. Subsequently, these can be subdivided into their components, which can also be specified more precisely (e.g. whether it is a boiler). We use this algorithm in the presented tool chain, since it requires less input and can also be used by technicians without computer science knowledge. Fig. 2 shows the results of AIKIDO in five different buildings.

2.2. Automated model generation
In order to create a simulation model from existing meta-data, the metadata is converted into a topology model. This model can be used equally for control, machine learning approaches and the creation of simulation models. We use the Brick Schema [14] for this purpose, as it offers the advantage that the connections between technical equipment that the BUDO Schema does not include can be mapped. The process is described in more details in [13].

The simulation models created can be coupled with real time series data and thus reproduce the real behavior of the plant very accurately. Figure 4 shows the results from the auto-generated simulation model and the time-series data of the original device.

3. Model application
In the following section, we present two examples on how the auto-generated models can be used to facilitate building control and future research.

3.1. Model-based control algorithms
Model-based control algorithms are considered to be very promising for achieving the goal of energy-efficient operation [15].
Figure 4. A comparison of the results of the simulation model vs. with the original time series data. [13]

The models can be used for mathematical optimization or could be coupled with weather prediction services to account for future changes in ambient conditions. However, many of those algorithms suffer from over-simplification or inconceivably high modelling effort [16]. Moreover, building automation systems must be sufficiently robust to operate autonomously for long periods in between maintenance. Model-based systems thus require continuous recalibration and adoption to account for changes in the system behavior such as wear and tear of control valves.

Thus, an important part of the toolchain is automated model generation and the ability of control algorithms to work with many smaller, distributed models instead of one large model. This facilitates the exchange of models in case of changes in the controlled system and supports the calibration and data-driven model generation. Therefore, we consider distributed model predictive control (DMPC) approaches particularly advantageous [17]. There are many studies showing that complex systems can be decomposed and, still, a near-global-optimum solutions can be achieved.

In practice, the use of advanced control algorithms such as (D)MPC in modern building automation and control systems is not yet very widespread. Main obstacles are the lack of locally available computing power and reservations of the building operators against the comparatively complicated and opaque algorithms. Especially for model predictive control, a high local computing power may be required, which the fewest state-of-the-art building automation controllers can provide.

One way of structuring the control algorithms is the use of defined operation modes [18]. The idea is to use a formalized description of the local control strategy, e.g. Petri-nets. The definition of transition conditions is a structured approach to initiate different operation modes like heating, cooling or a backup mode. Thus, it is possible to define boundaries in which a MPC algorithm can operate while ensuring a save building operation. The building operators keep the control of their system while the algorithm can improve the operation energy demand. Further, it is possible to run the algorithms in the cloud and use modern IoT protocols for communication. In further research will investigate how the Petri-net can be generated from the automatically generated physical models.
3.2. Fault detection and diagnosis

Faults can account for up to 30% of the energy consumption of a building [19]. As building systems become more complex, the process of finding faults needs to be automated to detect a faulty system behavior and to react in a timely manner. Approaches for fault detection system are usually divided into qualitative, quantitative model-based approaches and data-driven approaches [20].

Data-driven approaches like the application of artificial intelligence to the problem have gained popularity in recent years due to the expected availability of data, the reduced cost for computational effort and rising availability of user friendly implementations. They yield good results in experimental applications, and have resulted in widespread application in industries where products are produced in high quantities. Nevertheless, they have not yet permeated the building sector [21]. Quantitative model-based approaches are hindered by the need for physical modelling, with simple models limited in their general applicability and complex models limited by the large amount of work and information needed to create the model. Qualitative model-based methods, with their best known representative of rule-based expert systems, provide a pragmatic solution to fault detection that requires extensive, high-quality expert knowledge [20]. Data-driven approaches, on the other hand, are limited by the need for an extensive data set for training, which often does not exists in the required detail [22].

With the process of automated model generation we propose, we can approach this problem threefold: Firstly, the generated model can be used directly in model-based fault detection methods. Secondly, the model can be extended using purpose-built fault models to generate an extensive database of operation conditions. Thirdly, the model can be used to simulate faulty states and derive behavior information and performance indicators to assess the state of the system. We can additionally use the model to assess the impact of faults on the system and energy consumption and create a ranking of faults with regard to their severity, cross validated by comparison to a qualitative approach to fault detection.

In future work, we will focus on the second option, as it is most promising due to the possibility to use uncertainty functions, e.g. Monte-Carlo-Simulations, for the input parameters and then use the abstraction capabilities, e.g. the ability of the model to find the relevant reoccurring patterns among data, of the data-driven algorithms to reduce the impact of modelling and input deviations.

4. Conclusion and Outlook

In this paper, we presented steps to achieve the automatic application of new control technologies for the existing building stock by using data to unify the semantics and auto-generate the simulation models. Moreover, we reviewed their possible applications. The proposed holistic approach includes more than one field of modern building automation research. Using the BUDO schema throughout the process, we are able to develop algorithms compatible with each other. With the AIKIDO algorithm as starting point, we are able to quickly apply the algorithms to new energy systems with different data point naming schemes without the need for time-consuming manual mapping processes. All developed standards and practices are made public to enable other developers to fully integrate their work at any place in the process chain. The result of the automatic process is a model that can be used for a variety of building automation tasks.

In future work on MPC, we will investigate how Petri-nets can be used to structure control algorithms and how they can be generated from the auto-generated model.

Future work in the area of fault detection will focus on using the abstraction capability of data driven approaches to mitigate the uncertainty of the automatically generated models to improve the quality of the predictions. We are developing a fault model library to be compatible with the generated model to fully automate the process of fault simulation for any given building
energy system. Judging from our experience with the single components of the presented toolchain, we consider potential of the toolchain an important step towards the application of advanced energy-efficiency measures in buildings.

Acknowledgments
We gratefully acknowledge the financial support provided by the BMWi (Federal Ministry for Economic Affairs and Energy), promotional references 03SBE006A and 03ET1657A. We would like to thank our project partners aedifion GmbH, AutomationOne GmbH and werkkraft GmbH for their support.

References
[1] Stimmel C L 2015 Big data analytics strategies for the smart grid (Boca Raton, FL: CRC Press) ISBN 9781482218282
[2] Tang S, Shelden D R, Eastman C M, Pishdad-Bozorgi P and Gao X 2019 Automation in Construction 101 127–139 ISSN 09265805
[3] Bode G, Behrendt S, Füttner J and Müller D 2017 Energy Procedia 122 997–1002 ISSN 18766102
[4] Baranski M, Meyer L, Füttner J and Mueller D Comparative study of simulation-assisted approaches for distributed model predictive control in building energy systems ECOS 2018 - Proceedings of the 31st International Conference on Efficiency, Cost, Optimisation, Simulation and Environmental Impact of Energy Systems
[5] Cheung H and Braun J E 2015 Development of Fault Models for Hybrid Fault Detection and Diagnostics Algorithm
[6] Schumacher F and Fay A 2014 Control Engineering Practice 33 84-93 ISSN 09670661
[7] Reiß und Hommerich Bericht zum Thema Building Information Modeling (BIM): Bundesweite Befragung
[8] Veichtlbauer A, Pfeifferberger T and Schrittesser U 2012 Generic Control Architecture for Heterogeneous Building Automation Applications 6th International Conference on Sensor Technologies and Applications (SensorComm 2012)
[9] Bhattacharya A, Ploennigs J and Culler D 2015 Short Paper: Analyzing Metadata Schemas for Buildings: The Good, the Bad, and the Ugly Proceedings of the 2nd ACM International Conference on Embedded Systems for Energy-Efficient Built Environments (2821674: ACM) pp 33–34
[10] Stinner F, Kornas A, Baranski M and Müller D 2018 Structuring building monitoring and automation system data The REHVA European HVAC Journal - August 2018 REHVA Journal ed REHVA pp 10–15
[11] Gao J and Bergés M 2018 Advanced Engineering Informatics 37 14–30 ISSN 14740346
[12] Stinner F, Neiller-Deiters P, Baranski M and Müller D 2019 AIKIDO: Structuring data point identifiers of technical building equipment by machine learning CISBAT 2019
[13] Stinner F, Yang Y, Schreiber T, Bode G, Baranski M and Müller D 2019 Generating generic data sets for machine learning applications in building services using standardized time series data ISARC - 36th International Symposium on Automation and Robotics in Construction (Banff, AB, Canada)
[14] Balaji B, Bhattacharya A, Fierro G, Gao J, Gluck J, Hong D, Johansen A, Koh J, Ploennigs J, Agarwal Y, Bergés M, Culler D, Gupta R K, Kjærgaard M B, Srivastava M and Whitehouse K 2018 Applied Energy 226 1273–1292 ISSN 03062619
[15] Afram A and Janabi-Sharifi F 2014 Building and Environment 72 343–355 ISSN 0360-1323
[16] Privara S, Cigler J, Vaňa Z, Oldewurtel F, Sagerschnig C and Žáčková E 2013 Energy and Buildings 56 8–22 ISSN 03787788
[17] Christofides P D, Scattolini R, de la Pena, David Munoz and Liu J 2013 Computers & Chemical Engineering 51 21–41 ISSN 0098-1354
[18] Schild T, Futterer J and Muller D 2017 A mode-based implementation framework for advanced control methods in building automation systems with Petri-Nets 2017 IEEE International Conference on Advanced Intelligent Mechatronics (AIM) (IEEE) pp 213–218 ISBN 978-1-5090-5998-0
[19] Yu Y, Woradchujumroon D and Yu D 2014 Energy and Buildings 82 550–562 ISSN 03787788
[20] Katipamula S and Brambley M 2005 HVAC&R Research 11 3–25 ISSN 1078-9609
[21] Turner W, Staino A and Basu B 2017 Energy and Buildings 151 1–17 ISSN 03787788
[22] Burak Gunay H, Shen W and Newsham G 2019 Automation in Construction 97 96–109 ISSN 09265805