Teaching data analysis for building performance simulation – Series: Building simulation and calculation tools in teaching

Hopfe, Christina Johanna; McLeod, Robert Scot; Gustin, Matej; Brembilla, Eleonora; McElroy, Lori

DOI
10.1002/bapi.202200027

Publication date
2022

Document Version
Final published version

Published in
Bauphysik

Citation (APA)
Hopfe, C. J., McLeod, R. S., Gustin, M., Brembilla, E., & McElroy, L. (2022). Teaching data analysis for building performance simulation – Series: Building simulation and calculation tools in teaching. *Bauphysik*, *44*(5), 285-290. https://doi.org/10.1002/bapi.202200027

DOI: 10.1002/bapi.202200027

Christina Johanna Hopfe, Robert Scot McLeod, Matej Gustin, Eleonora Brembilla, Lori McElroy

Teaching data analysis for building performance simulation

Series: Building simulation and calculation tools in teaching

The Erasmus+ project, entitled 'Digital Erasmus – a roadmap to using building performance simulation to achieve resilient design' (DesRes), seeks to transform the learning experience of students in built environment disciplines using a continuous digital learning cycle. Three universities play a part in this project: Graz University of Technology (TU Graz), Delft University of Technology (TU Delft) and the University of Strathclyde (UoS), each developing and delivering a module to complete the learning experience in building simulation. This paper describes the aims and learning objectives associated with the workshops taught at TU Graz as part of the module dedicated to energy monitoring. These workshops tackle the complexity of working with large data sets, which commonly arise from energy monitoring research. In particular, the workshops aim to provide a practical understanding of how to identify, handle, reshape, clean up and evaluate important summary statistics from incomplete data sets. These are fundamental skills in building simulation where model validation and calibration are increasingly commonplace.

Datenanalyse für die Gebäudesimulation in der Lehre – Serie: Gebäudesimulation und Berechnungstools in der Lehre
Das Erasmus+ Projekt mit dem Titel „Digital Erasmus – a roadmap to using building performance simulation to achieve resilient design" (DesRes) zielt darauf ab, die Lernerfahrungen von Studierenden in Fachbereichen der gebauten Umwelt durch einen kontinuierlichen digitalen Lernzyklus zu verändern. Drei Universitäten sind an diesem Projekt beteiligt: die Technische Universität Graz (TU Graz), die Technische Universität Delft (TU Delft) und die University of Strathclyde (UoS), Glasgow, die jeweils ein Modul entwickeln und anbieten, um die Lernerfahrung in der Gebäudesimulation zu vervollständigen. In diesem Beitrag werden die Ziele und Lerninhalte der Workshops beschrieben, die an der TU Graz als Teil des Moduls zum Energiemonitoring angeboten werden. Diese Workshops befassen sich mit der Komplexität der Arbeit mit großen Datenmengen, die üblicherweise in der Forschung zum Energieverbrauch anfallen. Insbesondere zielen die Workshops darauf ab, ein praktisches Verständnis dafür zu vermitteln, wie man wichtige zusammenfassende Statistiken aus unvollständigen Datensätzen identifiziert, handhabt, umformt, bereinigt und auswertet. Dies sind grundlegende Fähigkeiten in der Gebäudesimulation, wo die Validierung und Kalibrierung von Modellen immer wichtiger werden.

Keywords Building performance simulation BPS; Building Information Modeling BIM; learning cycle; energy monitoring; large data sets; sustainable construction; research and teaching

Stichworte Building Information Modeling BIM; Building Performance Tools BPS; Gebäudesimulation; große Datenmengen; nachhaltiges Bauen; Forschung und Lehre; Lernzyklus

1 Introduction

Building performance simulation (BPS) is traditionally taught in a classroom setting with access to computer labs, where students learn how to use the software with direct support from the teaching staff (Beausoleil-Morrison and Hopfe, [1], [2]). The DesRes project modules targetstudents who are enrolled in MSc programmes that include courses on building performance, building physics, but who are unable to participate in live classroom tutorials (e.g. due to Covid 19) and/or who would like to expand their learning expierence at a different university. DesRes proposes a new pedagogical methodology that addresses the challenges of digital teaching but at the same time provides a platform for students to acquire both the theory and practical skills needed to further their career in BPS (see Fig. 1).

BPS and its application in enhancing the resilience of the built environment is by nature a digital activity. Paradoxically however, it is commonly taught in a traditional didactic manner, with students receiving direct instruction whilst experiencing little or no peer-to-peer interaction or experiential learning. Unsurprisingly pedagogic theory, including specific studies focused on the teaching of BPS in higher education have highlighted how such approaches have delivered exceptionally poor learning outcomes [1], [2] . This means that there is an urgent need to evolve new and innovative teaching practices, that better engage students, whilst yielding more robust learning outcomes (Hopfe et al. [3], McLeod and Hopfe [13]). In the midst of the ongoing global pandemic, with greater than ever demand for online learning, there has never been a more appropriate moment to make this change. The pedagogic model which DesRes advocates is rooted in the constructivist theory that learning occurs most profoundly when learners are actively involved in a process of meaning and knowledge construction. This means that DesRes applies a pedagogical methodology that addresses the challenges of digital teaching but at the same time provides a plat-

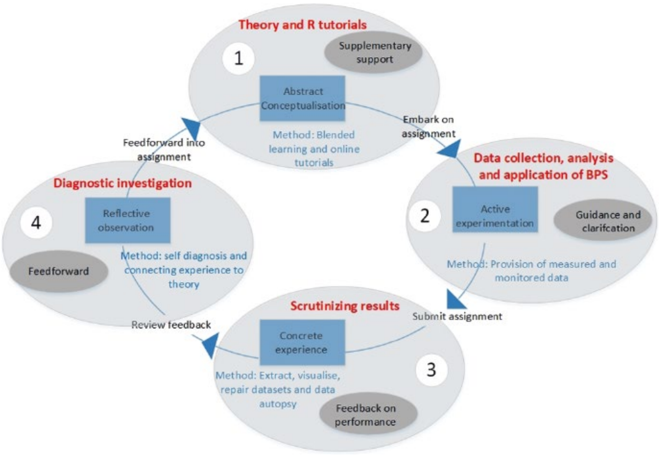

Fig. 1 Framework continuous learning cycle, adapted after Kolb [10].
Rahmenprogramm zum kontinuierlichen Lernzyklus, angepasst nach Kolb [10].

form for students which facilitates the acquisition of skills needed for a successful career in BPS (whether that be in academia, applied research or consultancy) (see Fig. 1) [4].

2 Module and workshops

In line with the overarching theme of the DesRes project and its objectives [4] a new module was conceived to challenge students to work in an interdisciplinary manner. In this module entitled “Energy Monitoring and the Effects of Indoor Climate”, students augment their existing knowledge in the field of building physics and energy-efficient construction with topics dedicated to energy monitoring and the impacts of indoor climate. Students get to understand metrological concepts, and how the important thermal, hygric and energetic properties of the building envelope are commonly measured and recorded. They learn about the effect of the building’s services and envelope on the indoor environmental performance and its interaction with user behaviour. This understanding of the interconnected of socio-technical design factors allows them to better evaluate and interpret data that is introduced and analysed as part of this module. The theoretical teaching uses the monitored operational data from buildings to demonstrate the high potential for optimization of the built infrastructure. Students also learn how the indoor climate in buildings can be assessed in connection with the building envelope and building services (such as ventilation systems) using methods of thermal comfort assessment. Measurement methods are introduced, to demonstrate how the room climate in real-buildings can be determined and evaluated experimentally, which can in turn be used to generate data sets to validate and/or calibrate white-box simulation models or to derive grey and black box machine learning models (Gustin et al., 2018 [5]; Gustin et al., 2019 [6]; Ahmed, 2020 [7]) or model predictive control functions.

Because building physics and BPS researchers and consultants are dealing with increasingly larger amounts of data, there is an imperative for students to learn to use the most appropriate tools and techniques needed to extract value from big data sets. Understanding the various statistical and visualisation methods that will allow them to examine and exploit large data sets is therefore fundamental to this goal (Table 1). The use of an open-source (i.e. free) programming language such as Python or R allows extremely fast manipulations of large amounts of data, creating automated and reproducible scripts and allowing the implementation of more complex and advanced statistical and visualisation methods (Gebhart, 2016 [8]; Joubert, 2019 [9]).

While the formal lectures of the module developed at TU Graz focus mainly on understanding the measurement and modelling techniques used in the fields of energy and indoor air quality, they also include an introduction to data bases, data management and the statistical techniques used to clean and structure data. In order to put these skills into practice the three workshops of the module have the objective of providing an overview and introduction to a frequently used open-source tool (i.e. R programming) that is widely used in academia and industry for data science applications. R is introduced in a very practical way in these workshops with the aim of identifying some common pitfalls in working with big data and to show how data analysis and descriptive statistics can be used to leverage meaningful insights from large data sets. In addition to these general techniques, the workshops also provide specific experience in extracting and preparing bespoke weather datasets for building simulation (McLeod et al, 2012 [11]). The workshops are structured in three parts:

1. *Introduction to R*;
2. *Cleaning, structuring, and exporting climate data files from big datasets in R*; and
3. *Analysing big datasets in R*.

The building simulation related learning objectives (LOs) are described in Table 2.

Workshop 1

Prior knowledge or experience in using R is not a prerequisite for this module therefore the first workshop starts at the very beginning and the students get an introduction to programming in RStudio. During this session the students familiarise themselves with the use of R and how to perform basic calculations and define values, data frames and functions. The first workshop is not assessed; however, the students are encouraged to complete a series of exercises at the end of the workshop and to continue experimenting with what they have learnt.

Workshop 2

In the second workshop, the students learn to explore and analyse a large-scale monitoring data set. Learning to

Tab. 1 Aim and objectives of workshops (WS)
Zielsetzungen der Workshops (WS)

WS No.	Aim and objectives
1	*Introduction to R* aims to provide the students with the basic knowledge that is required to complete the subsequent workshops and to start using R independently for some basic operations. 1. Provide an overview of R and its strengths; 2. Understand the R Studio Graphical User Interface (i. e. the most popular and widely used GUI for R); 3. Learn how to perform basic calculations and define values, data frames and functions; and 4. Learn how dates and times work in R, how can they be converted to the R format and how to round them.
2	*Working with large data sets* aims to teach the process of cleaning and exporting climate data from big datasets in R, using a large energy monitoring dataset. 1. Highlight common issues that researchers have to deal with each time they adopt a new dataset; 2. Learn how to load, reshape, clean and extract the data; and 3. Learn how to export the cleaned weather data into a file suitable for BPS analysis (e.g. as a csv file).
3	*Statistical analyses (R) of data sets* aims to demonstrate the advanced statistical and visualisation capabilities of R and provide practical experience in analysing big datasets. The workshop builds up on the previous two workshops and focuses the analyses on the cleaned weather dataset (extracted from the *REFIT Smart Home dataset*) that was compiled for the first assignment (following workshop 2). 1. Understand how to explore big datasets and get some summary statistics from the data; 2. Learn how to detect and remove unwanted outliers; and 3. Learn how to use different types of plots to analyse the dataset and gain useful insights from data.

work efficiently with large data sets is becoming a necessity for aspiring BPS researchers. Spreadsheet programs (e. g. MS Excel) may be relatively easy to use, however, they have many limitations that quickly become evident when one attempts to open and analyse a large data set. To begin with the students are provided with a large open-source dataset. They are first given time to familiarise themselves with the dataset which is organised in four different files:

– *csv* file: contains all the time series data organised by sensor id and date and time of the observations;
– *xml* file: contains the ids and detailed information of all sensors, their location, the room and the building;
– *xsd* file: describes the structure and content of the *xml* file; and
– *txt* file: provides a description and basic information of the monitoring study and the dataset.

Students then learn how to extract the desired sensor data from the *csv* file. Once they understand this process (due to the limited amount of time available during the workshop) the students are then provided with a customised pre-built function, that specifies several arguments, in order to automate the above steps. Use of this function allows the students to extract all of the required weather variables.

After the weather variables have been extracted at the desired time-step resolution (e. g. hourly) and period (i. e. for a specified calendar year), they have to be combined into one single data frame. Whereas this appears to be a straightforward task, the data cannot be simply bound together (as recorded in the original dataset). This is because it is possible that certain variables are missing some data, with the risk that a column of values could be bound to incorrect dates and times. For the purpose of correctly binding columns of data together, students are first introduced to the various types of joins (see Fig. 2) and are then shown how they can perform a left join to bind a column of data onto a control date and time data frame. This process ensures that all the variables are joined at the correct dates and time, regardless of whether some data is missing. Once the com-

Tab. 2 Workshop and learning objectives (LO) for the Active Experimentation (AE) mode of learning at TU Graz
Workshops (WS) und Lernziele (LO) für das Aktive Experimentieren (AE) an der TU Graz

WS No.	Building Performance Simulation related learning objectives (LO)
1	(a) Learn how to create bespoke R scripts in order to define, shape and index data frames; (b) Create new functions to manipulate existing variables and define new ones;
2	(c) Gain practical experience in extracting useful information from a big dataset (d), Appreciate the complexity of large data sets and understand how to identify, handle, reshape and clean incomplete data sets; (e) Learn how to export a cleaned climate data file;
3	(f) Understand various statistical methods to rapidly examine, and summarise key environmental, climate and energy data; (g) Learn how to clean data, identify and remove unwanted outliers; and how to use different types of plots to leverage maximum insights from the data; (h) Appreciate different types of plots to get insight from data

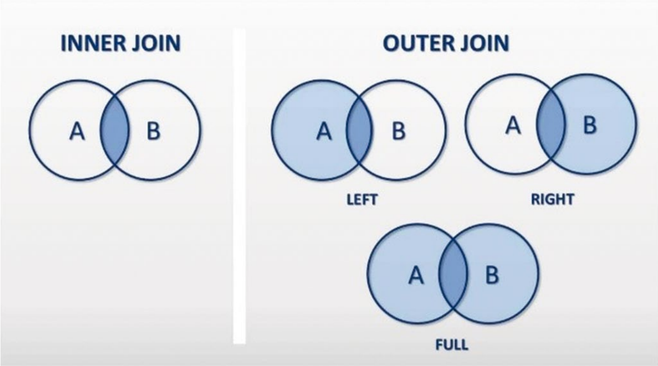

Fig. 2 Various types of join (Tecniche Hosting, 2019 [12]).
Verschiedene Arten von Datenmengen (Tecniche Hosting, 2019 [12]).

plete weather file has been compiled and inspected in this way, students learn how to export and save the weather data as a *csv* file.

Workshop 3

In order to make sense of large data sets and facilitate their analysis, the use of appropriate statistics and the creation of various plots is introduced as way of leveraging meaningful insights. The third workshop teaches students some basic descriptive statistics, data analysis and how to exploit various types of plots to thoroughly examine the data. Creating useful plots and communicating the findings is a necessary skill for every researcher and a prerequisite for most reports and publications. The workshop introduces the students to the skills needed to check the plausibility of monitored data, clean the data and analyse the subsequent datasets.

Students are provided with some pre-built functions to assist them with the task of exploring and visualising the data in a simplified way. To begin with, students learn to summarize and export the data set using different statistical quartiles. Then they focus on identifying potential outliers in the data. The limitations of using statistical approaches (e. g. simply removing data that is 1.5 times below/above the interquartile range) as a means to remove outliers is examined in the context of dealing with intermittent data (e. g. wind speed). The importance of researching and critically assessing whether the 'outlier' values of a given weather variables are realistic for the specific regional and national climate before removing them is carefully addressed. The students are then provided with a pre-built function which allows them to summarise the counts of data below, within and above two specified (i. e. critically determined) thresholds to more realistically identifying outliers. This process teaches them the importance of quality assurance in data management by allowing them to specify the counts and frequency of the outliers and determine when these occur (e. g. for the wind speed data). Students learn to plot the time series data (x-y plots) for any number of variables (with single-plot or multi-plot visualisation) at the specified time resolution (x-axis) for the desired analysis period (from/to the desired date and time). This skill allows them to visualise and explore the data, before and after the removal of outliers (Fig. 3), at different resolutions and over different periods of time.

Another example taught to the students is how to plot a 9-coloured heatmap for the entire year. They learn that if the heatmap is designed properly (e. g. using multiple colour gradients and organised by the year, month, day and hour), it is possible to visualise the range of a variable at an hourly resolution (i. e. one coloured cell for each hour) across an entire year or multiple years. Applying this approach, for example, to the outdoor air temperature, it is possible to identify the hottest (e. g. above 30 °C) and coldest (e. g. below 0 °C) periods with ease and with high accuracy (see Fig. 4).

Assignments and discussion

Following the second workshop, students have to extract additional weather variables and join all the weather variables into a new data frame. Weather files for simulations (e. g. *epw* – Energy Plus Weather file) typically require the computation of additional variables such as the dew point temperature. To learn how to compute new variables in R, students are tasked to implement equations to calculate these variables (from the measured data) and then extract them as a new weather file.

Following the third workshop, students have to generate tables and plots to analyse the data and answer 10 different questions based on a detailed analysis of the extracted weather variables. These include for example looking at potential outliers, frequency of exceedance and other base summary statistics (e. g. minimum, mean, maximum values etc.) during specific periods. There are different ways to answer the questions, so it is up to the students to decide which method(s) they want to adopt to answer the questions. In the assessment emphasis is placed on the importance on the rational and description provided for the adopted method. Herein, the diagnostic investigation following the continuous learning cycle adapted after Kolb, 2014 [10] (Fig. 1) drives the learning experience by placing the students at its centre. At this (post-graduate) level students need to understand the importance of critically examining their results in order to identify and diagnose sources of input error. This implies that they develop a degree of scepticism about their results, and develop the resources and skills necessary to diagnose, repair, and reduce errors. This is why we introduced the concept of the diagnostic investigation (see e.g. Beausoleil and Hopfe, 2015 [14]) as part of a working session in which the students, scaffolded by prompting and support where needed (from the instructors), collectively examine the results of the exercise and dissect the output files to identify potential sources of disagreement. This approach is used as a vehicle to help students develop critical thinking skills thereby building upon the active experimentation mode

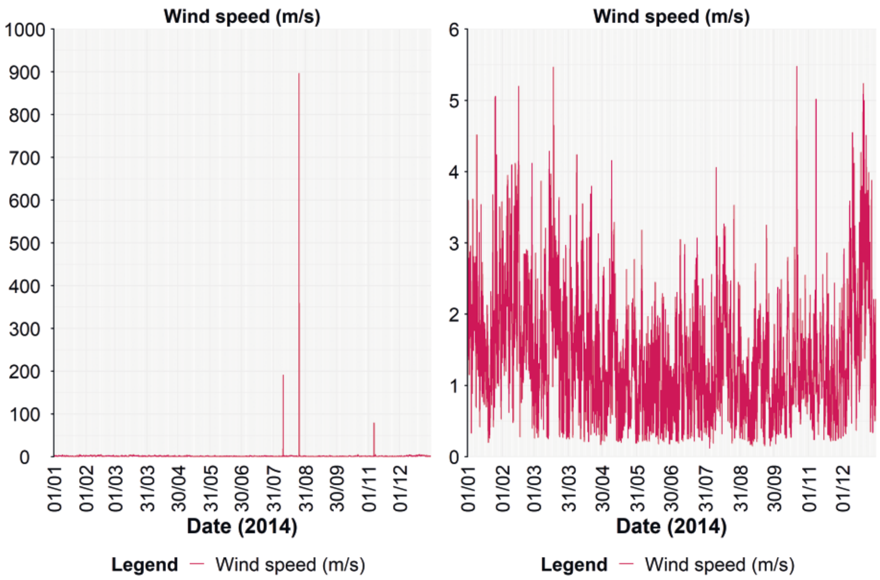

Fig. 3 Wind speed before (left) and after (right) the removal of the outliers.
Windgeschwindigkeit vor (links) und nach (rechts) der Entfernung der Ausreißer.

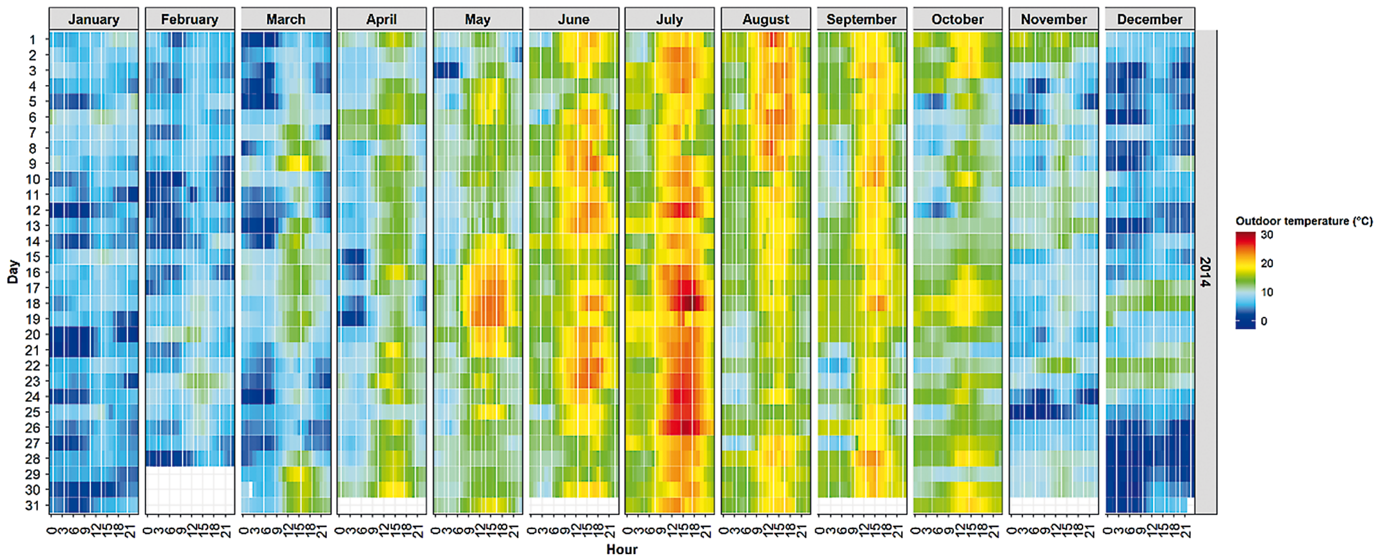

Fig. 4 Heatmap of the hourly outdoor air temperature for the entire year 2014.
Heatmap der stündlichen Außenlufttemperaturen für das gesamte Jahr 2014.

of learning and then relating the lessons learnt back to the theory acquired during the abstract conceptualisation mode.

Conclusion

This paper has described the practical workshop sessions which form part of the Erasmus+ DEsRes module on energy monitoring and provides a glimpse into the pedagogic framework used. This post-graduate taught module provides a blended digital learning experience for teaching data analysis as a core skill for building performance simulation. In these workshop sessions the students learn to work with the type of large data sets which they will encounter when creating bespoke weather files for building performance simulation. During this stepwise process they learn how to identify potential faults in the underlying data and use a more critical approach to remove potential outliers. This is a particularly important skill in order to compile accurate and high-quality data sets (e. g. weather files) for building performance simulations. In addition, student acquire a good understanding of basic statistical and visualisation methods to analyse data, which are a prerequisite for the use of more advanced analytical methods to extract insights on the energy performance of buildings and the development of data-driven approaches to building performance analysis.

Acknowledgment

The Digital Erasmus project has been funded with the support of the Erasmus+ Programme (grant agreement no.: 2020-1-AT01-KA226-HE-092643) of the European Union. This publication reflects the views of its author(s) only and the European Commission cannot be held responsible for any use that may be made of the information contained herein.

References

[1] Beausoleil-Morrison, I.; Hopfe, C. J. (2016) *Teaching building performance simulation: ever done an autopsy?* The 3rd Building Simulation and Optimization 2016 (BSO2016), Newcastle University, Sept12–14th

[2] Beausoleil-Morrison, I.; Hopfe, C. J. (2016) *Developing and testing a new course for teaching the fundamentals of building performance simulation*, Proc. eSim 2016, Hamilton, Canada.

[3] Hopfe, C. J.; Soebarto, V.; Crawley, D.; Rawal, R. (2017) *Understanding the differences of integrating building performance simulation in the architectural education system,* in: Barnaby, C. S.; Wetter, M. [ed.] 15th Conference of IBPSA San Francisco, August 7–9, 2017, www.ibpsa.org.

[4] Hopfe, C. J.; Brembilla, E.; McElroy, L. B.; McLeod, R. S.; Gustin, M.; Costola, D. (2022). *Digital Erasmus – a pan-European approach to teaching building performance and resilient design*, Proceedings of CLIMA 2022- 14th REHVA HVAC World Congress. Available at: https://doi.org/10.34641/clima.2022.129.

[5] Gustin, M.; McLeod, R.; Lomas, K. (2018) *Forecasting indoor temperatures during heatwaves using time series models*, Building and Environment 143, pp. 727–739.

[6] Gustin, M.; McLeod, R.; Lomas, K. (2019) *Can semi-parametric additive models outperform linear models when occupant interventions are incorporated for the prediction of overheating in dwellings?* Energy and Buildings Vol. 193, pp. 250–266.

[7] Ahmed, A.; McLeod, R.; Gustin, M. (2020) *Forecasting underheating in dwellings to detect excess winter mortality risks using time series models*, Applied Energy, Vol. 286, 15 March 2021, 116517.

[8] Gebhart, K. (2016) *Understanding R programming over Excel for Data Analysis*, Available at: https://www.gapintelligence.com/blog/understanding-r-programming-over-excel-for-data-analysis/ [Accessed: 26 July 2022].

[9] Joubert, S. (2019). *R vs Excel: What's the Difference?* Northeastern University. Available at: https://www.northeastern.edu/graduate/blog/r-vs-excel/ [Accessed: 26 July 2022].

[10] Kolb, D. (2014) *Experiential Learning: Experience as the Source of Learning and Development*, 2nd edition, London: Pearson FT Press,.

[11] McLeod, R.; Hopfe, C.; Rezgui, Y. (2012). *A proposed method for generating high resolution current and future climate data for Passivhaus design*. Energy and Buildings, Vol. 55, pp. 481–493, https://doi.org/10.1016/j.enbuild.2012.08.045

[12] Tecniche Hosting (2019). SQL Outer Join. Available at: https://www.ionos.it/digitalguide/hosting/tecniche-hosting/outer-join-sql/ [Accessed: 27 July 2022].

[13] McLeod, R. S.; Hopfe, C. J. (2021) *Experiential learning in Building Physics – the icebox challenge*, J. Build. Phys. Vol. 45, No. 3, pp. 391–401.

[14] Beausoleil-Morrison, I.; Hopfe, C. J. (2015) *Teaching building performance simulation through a continuous learning cycle* Proc. Building Simulation 2015, pp. 2757–2764, Hyderabad, India.

Authors

Univ.-Prof. Dipl.-Ing. Dr. Christina Johanna Hopfe (corresponding author)
c.j.hopfe@tugraz.at
Technische Universität Graz
Institut für Bauphysik, Gebäudetechnik und Hochbau
Lessingstr. 25/III
8010 Graz/A

Assoc.Prof. FHEA FIMechE M.Sc. Ph.D CEng. Robert Scot McLeod
mcleod@tugraz.at
Technische Universität Graz
Institut für Bauphysik, Gebäudetechnik und Hochbau
Lessingstr. 25/III
8010 Graz/A

Dott. Mag. Ph.D Dott. Matej Gustin
m.gustin@tugraz.at
Technische Universität Graz
Institut für Bauphysik, Gebäudetechnik und Hochbau
Lessingstr. 25/III
8010 Graz/A

Dr.-Ing. Eleonora Brembilla
E.Brembilla@tudelft.nl
Delft University of Technology (TU Delft)
Architectural Engineering and Technology
Building Physics and Services
Building 8
Julianalaan 134
2628 BL Delft/NL

Prof. Lori McElroy
lori.mcelroy@strath.ac.uk
University of Strathclyde
Architecture
James Weir Building
75 Montrose Street
Glasgow G1 1XJ
Scotland/UK

How to Cite this Paper

Hopfe, C.; McLeod, R.; Gustin, M.; Brembilla, E.; McElroy, L. (2022) *Teaching data analysis for building performance simulation – Series: Building simulation and calculation tools in teaching.* Bauphysik 44, H. 5, S. 285–290. https://doi.org/10.1002/bapi.202200027